# UNIT 13 : CLASSIFICATION AND TABULATION OF DATA

## UNIT STRUCTURE


13.1 Learning Objectives

13.2 Introduction

13.3 Objectives of Classification

13.4 Types

13.5 Continuous and Discrete Variable

13.6 Frequency Distribution

13.6.1 Preparation of Frequency Distribution

13.6.2 Construction of Frequency Distribution Table

13.6.3 Cumulative Frequency Distribution

13.7 Let Us Sum Up

13.8 Further Reading

13.9 Answers to Check Your Progress

13.9 Model Questions


## 13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

understand the meaning of classification of data

- discuss Its types and objectives
- explain the meaning of frequency distribution
- explain What is cumulative frequency distribution
- discuss the way prepare a frequency distribution table

## 13.2 INTRODUCTION

In any type of research data collection is an important part. But the data collected by the investigator is raw data. These collected data need to be classified for making them suitable for statistical research. The technique of arranging data in different homogenous groups is known as classification.

**L.R. Connor defines classification in the following way —**

Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities and gives expression to the unity of attributes that may subsist amongst a diversity of individuals. In this unit we shall discuss about objectives and types of Classification.

## 13.3 OBJECTIVES OF CLASSIFICATION

- **To make the data simple and brief:** Classification aims at making the complex data simple and to be presented in brief so that they are easily understandable by a lay man.
- **To clearly bring out similarities and dissimilarities:** Classification facilitates to find out similarities and dissimilarities in a set of data, such as the difference between males and females, etc.
- **To facilitate comparison:** Classification of data is particularly important in order to compare two sets of data. For example, suppose we want to compare the number of students studying higher education in Assam and Meghalaya. Classifying the data into technical and non-technical, and further into Arts, Science, Commerce and Medical and Engineering, etc. will make the comparison easier.
- **Scientific and critical arrangements:** With the help of classification, data can be presented scientifically. For example, the number of students in a class can be classified on the basis of their weights or marks obtained.
- **To prepare the basis of tabulation:** Classification prepares the basis for tabulation and statistical analysis of the data. Unclassified data cannot be presented in tables.

## 13.4 TYPES OF CLASSIFICATION

There are four types of classification of data. They are –

- **Geographical Classification:** When the raw data collected by the investigator is classified according to location we call it geographical

or spatial classification. For example, suppose one collects data about the world female population. That data is first classified into countries, and then each country's population is classified into states and then subsequently into districts.

- **Chronological Classification:** If data is classified with respect to different periods of time, then this classification is known as chronological classification. For example, in order to show population of India from 1951 to 2016, we write population of the country for different years starting from 1951.
- **Qualitative Classification:** When the collected data is classified in terms of certain attributes such as caste, creed, religion, sex, marital status, etc., i.e., which cannot be quantified, it is known as qualitative classification. Qualitative classification may be of two types –
  - **Simple Classification :** When we classify the data into two classes of a single attribute, we call it simple classification. For example, suppose we classify the population of a village into males and females. This is simple classification.
  - **Complex classification :** When classification of data is done according to more than one attributes, it is called complex classification. For example, take the case of classifying the population of a country. The total population may be first classified into male and female. Then each category may be classified according to age and so on.
- **Quantitative classification:** Quantitative classification includes classification of data that can be measured, such as, income, expenditure, revenue, import, export, etc. For example, if we have to classify the import and export of all the Asian countries, then that classification will be in numerics and will be termed as quantitative classification.

In order to know about the quantitative classification in detail, it is important to know the continuous and discrete variable first.

## 13.5 CONTINUOUS AND DISCRETE RANDOM VARIABLE

- **Continuous Variable :** Those variables which can take all the possible values, integral as well as fraction, in a given specified range are termed as continuous variable. For example, the height of students in a class lies between 4 feet and 5 feet. So, one may have 4.2 feet, another may have 4.9 feet, etc. the heights may take any values between 4 and 5. Just like height, weights, time, distance, etc. are the examples of continuous variable.
- **Discrete Variable:** Variable which can take only exact value and cannot take fractional value is called a discrete variable. For example, the number of students in a class, number of cars, etc. is examples of discrete variable.

## 13.6 FREQUENCY DISTRIBUTION

A frequency distribution is a comprehensive way to classify raw data of a quantitative variable. It shows how different values of a variable are distributed in different classes along with their corresponding class frequencies. A frequency distribution table consists of class interval and frequencies. Suppose a class of 50 students obtains marks in economics between 0 and 50. This information may be presented in a frequency distribution table in the following way:

| Class interval | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| Frequency | 1 | 10 | 12 | 18 | 9 |

In the above example, against each class interval different frequencies are plotted. The class intervals or class limits are the two ends of a class the lower value is called the lower limit and the upper value is called the upper limit. For the class interval 10-20, 10 is the lower limit and 20 is the upper limit.

The difference between the upper limit and lower limit is called the length or height of the class. In the above example, the height is (30 – 20)= 10.

The middle value of a class is called mid value or mid-point of the class. It is also known as class mark and is calculated in the following way:
Mid Value = (upper limit + lower limit)/2

### 13.6.1 Preparation of Frequency Distribution

- **NUMBER OF CLASSES:**

It is important to decide as to how many classes we need. A frequency distribution with too many classes or too few classes may not reveal the true set of the data, and it will also be difficult to interpret such frequency distribution. Although there is no hard and fast rule, yet generally the number of classes generally range between 5 and 13.

Before we determine the class interval, we need to find the range of the data, which is the difference between the maximum number and minimum number, i.e., **range = (max-min).**

- **Size of the class**:

After we determine the range and number of classes, we need to decide upon the height of the class interval using the formula ***h=range/no. of classes***

Given the range of the variable, number of classes can be easily determined if we know the class intervals. Similarly, class interval can be known if the number of classes is known. So, these decisions are interlinked with one another.

- **Determining the class limit**:

The value of the upper class limit of a class is obtained by adding the class interval with the value of the lower limit of that class. For example, in the class limit 30 – 40, 30 is the lower class limit and by adding 10, which is the class interval, to 30, we get 40, i.e., 30+10=40. In order to determine the lower class limit of the first class, one has to choose the minimum value of the frequency.

However, before determining the class limit, one must see that the classes are mutually exclusive. This means that no data value can fall into two different classes. Also, the classes must be all inclusive or exhaustive. This means that all data values must be included.

Let us consider the following example.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57 | 44 | 86 | 75 | 00 | 18 | 45 | 14 | 04 | 64 | 72 | 51 | 69 | 34 | 22 | 83 | 70 |
| 20 | 57 | 28 | 100 | 56 | 50 | 47 | 10 | 34 | 61 | 64 | 80 | 46 | 22 | 10 | 84 | 50 |
| 47 | 73 | 42 | 33 | 48 | 65 | 10 | 34 | 66 | 53 | 75 | 90 | 58 | 46 | 38 | 69 | |

In the above example, 0 is the minimum value so the lower class limit of the first class will be 0. Now adding 10, the upper class limit of the first class will be 10. The same process will be followed for the other classes as well.

Table 13.1

| Class interval | Frequency |
|---|---|
| 0 - 10 | 2 |
| 10 – 20 | 5 |
| 20 – 30 | 4 |
| 30 – 40 | 5 |
| 40 - 50 | 8 |
| 50 – 60 | 8 |
| 60 – 70 | 7 |
| 70 – 80 | 5 |
| 80 – 90 | 4 |
| 90 - 100 | 2 |

In the above table, upper class limit of the first class = lower class limit of the second class = 10. This is true for other classes as well. Here classification is done by exclusive method, where upper limit of one class coincides with the lower limit of the next corresponding class. So, 40-50 and 50-60 are two consecutive classes. Now the question is in which of the two classes will 50 fall. To answer this we need to know the exclusive method and the inclusive method.

- **EXCLUSIVE METHOD :**

When the classes are so fixed that upper limit of one class is the lower limit of the next class, it is called exclusive method. So, the continuity of the data is maintained. This method is more useful in case of continuous variable. Under this method, the value which is equal to the lower limit of a

class is included in that very class. On the other hand, the value which is equal to the upper class limit of a class is not included in that very class. So, in case of class 30-40, 30 is included in the class 30-40, but 40 is included in the class 40-50.

➢ **INCLUSIVE METHOD :**

In this method, both the lower limit and upper limit of a class are included in that very class. So, in this case, the upper limit of one class is not equal to the lower limit of its next class. Consider the following example.

Table 13.2

| Class interval (daily wage) | Frequency (no. of persons) |
|---|---|
| 100 – 199 | 50 |
| 200 – 299 | 30 |
| 300 – 399 | 40 |
| 400 – 499 | 80 |
| 500 – 599 | 120 |
| 600 - 699 | 100 |
| | Total = 420 |

Here no. of persons earning daily wage between 100 and 199 will be included in the class 100-199. If someone earns a daily wage of 200, that person will be included in the class 200-299.

➢ **Conversion of Inclusive Series in Exclusive Series**

For converting the inclusive series into exclusive series, following steps may be followed:

- Find the difference between upper limit of a class and lower limit of a class. In the above example, the difference is 1.
- Divide the difference by 2. S, ½= 0.5
- Add 0.5 to upper limit and subtract 0.5 from lower limit of each class. This kind of adjustment restores continuity of data.

The above example, when converted into exclusive series, will be as follows:

Table 13.3

| Class interval (daily wage) | Frequency (no. of persons) |
|---|---|
| 99.5 – 199.5 | 50 |
| 199.5 – 299.5 | 30 |
| 299.5 – 399.5 | 40 |
| 399.5– 499.5 | 80 |
| 499.5– 599.5 | 120 |
| 599.5 - 699.5 | 100 |
| | Total = 420 |

### 13.6.2 Construction of Frequency Distribution Table

Consider the following example of marks obtained by 50 students:

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 57 | 44 | 86 | 75 | 00 | 18 | 45 | 14 | 04 | 64 | 72 | 51 | 69 | 34 | 22 | 83 | 70 |
| 20 | 57 | 28 | 100 | 56 | 50 | 47 | 10 | 34 | 61 | 64 | 80 | 46 | 22 | 10 | 84 | 50 |
| 47 | 73 | 42 | 33 | 48 | 65 | 10 | 34 | 66 | 53 | 75 | 90 | 58 | 46 | 38 | 69 | |

The above information needs to be classified into a frequency distribution table, as shown below.

Table 13.4

| Classes | Observations | Tally marks | Frequency | Mid Value |
|---|---|---|---|---|
| 0 - 10 | 0,4 | Ⅱ | 2 | 5 |
| 10 – 20 | 18,14,10,10,10 | ⅠⅠⅠⅠ | 5 | 15 |
| 20 – 30 | 22,20,28,22 | ⅠⅠⅠⅠ | 4 | 25 |
| 30 – 40 | 34,34,33,34,38 | ⅠⅠⅠⅠ | 5 | 35 |
| 40 - 50 | 44,45,47,46,47,42,48,46 | ⅠⅠⅠⅠ ⅠⅠⅠ | 8 | 45 |
| 50 – 60 | 57,51,57,56,50,50,53,58 | ⅠⅠⅠⅠ ⅠⅠⅠ | 8 | 55 |
| 60 – 70 | 64,69,61,64,65,66,69 | ⅠⅠⅠⅠ Ⅱ | 7 | 65 |
| 70 – 80 | 75,72,70,73,75 | ⅠⅠⅠⅠ | 5 | 75 |
| 80 – 90 | 86,83,80,84, | ⅠⅠⅠⅠ | 4 | 85 |
| 90 - 100 | 100,90 | Ⅱ | 2 | 95 |

The above table has five columns – classes, observations, tally marks, frequency and mid value. It is evident from the above table that most of the observations are concentrated in classes 40-50 and 50-60. Their frequencies are 8. Out of 50 observations, 16 observations are in

these two classes. That is, 32% of the data lies in between 40 – 60. The other classes are sparsely populated.

### 13.6.3 Cumulative Frequency Distribution

From table 13.5, one can know as to how many students got marks between 40-50, or 50-60 and so on. But if one wants to know as to how many students got above 90 or less than 40, then that table cannot furnish the information. For that one has to calculate cumulative frequency. Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes according to a certain rule. It is written as c.f.

Cumulative frequency distributions are of two types – 'less than' and 'more than'**.** In a 'less than' frequency distribution, the frequencies of each class are added successively from top to bottom. It indicates the cumulative number of observations less than or equal to the class frequency to which it relates. In a 'more than' cumulative frequency distribution, the frequencies of each class are added successively from bottom to top. It indicates the cumulative number of observations more than or equal to the class frequency to which it relates. This is shown in the following table:

| Classes | Frequency | Cumulative frequency | |
|---|---|---|---|
| | | Less than c.f. | More than c.f. |
| 0 – 10 | 2 | 2 | 2+48=50 |
| 10 – 20 | 5 | 2+5=7 | 5+43=48 |
| 20 – 30 | 4 | 7+4=11 | 4+39=43 |
| 30 – 40 | 5 | 11+5=16 | 5+34=39 |
| 40 – 50 | 8 | 16+8=24 | 8+26=34 |
| 50 – 60 | 8 | 24+8=32 | 8+18=26 |
| 60 – 70 | 7 | 32+7=39 | 7+11=18 |
| 70 – 80 | 5 | 39+5=44 | 5+6=11 |
| 80 – 90 | 4 | 44+4=48 | 2+4=6 |
| 90 – 100 | 2 | 48+2=50 | 2 |

The above table is presented is presented below in a more simplified way –

| Less than | | More than | |
|---|---|---|---|
| Marks | No. of students | Marks | No. of students |
| Less than 10 | 2 | 0 and more than 0 | 2+48=50 |
| Less than 20 | 2+5=7 | More than 10 | 5+43=48 |
| Less than 30 | 7+4=11 | More than 20 | 4+39=43 |
| Less than 40 | 11+5=16 | More than 30 | 5+34=39 |
| Less than 50 | 16+8=24 | More than 40 | 8+26=34 |
| Less than 60 | 24+8=32 | More than 50 | 8+18=26 |
| Less than 70 | 32+7=39 | More than 60 | 7+11=18 |
| Less than 80 | 39+5=44 | More than 70 | 5+6=11 |
| Less than 90 | 44+4=48 | More than 80 | 2+4=6 |
| Less than 100 | 48+2=50 | More than 90 | 2 |

➢ **Purpose/Use/Necessity of Cumulative Frequency Distribution**

- ❍ Cumulative frequencies are used to know as to how many values of the variables are larger than a particular value or less than a particular value.
- ❍ Cumulative frequencies are used in determining mean, median and mode.
- ❍ Cumulative frequencies are also used while drawing Lorenz curve.
- ❍ We also require cumulative frequencies to represent some given data by ogive.



## 13.7 LET US SUM UP

- The technique of arranging data in different homogenous groups is known as classification.
- There are four types of classification of data – geographical, chronological, qualitative and quantitative.
- Those variables which can take all the possible values, integral as well as fraction, in a given specified range are termed as continuous variable.

- Variable which can take only exact value and cannot take fractional value is called a discrete variable.
- A frequency distribution is a comprehensive way to classify raw data of a quantitative variable.
- A frequency distribution table consists of class interval and frequencies.
- Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes according to a certain rule.



**CHECK YOUR PROGRESS**

Q1 : What do you mean by classification?·

.....

.....

Q2 : What do you mean by frequency distribution?·

.....

.....

.....

.....

Q3: What is cumulative frequency distribution?

.....

.....



## 13.8 FURTHER READING

1) Gupta, S.C. (2009). ‘ Fundamentals of Statistics, Himalayan Publishing House, Mumbai, Sixth Edition Reprint.

2) Gupta, S.P. (2010). ‘Statistical Methods’, Sultan Chand & Sons, New Delhi.

## 13.9 ANSWERS TO CHECK YOUR PROGRESS

**Ans to Q No 1 :** The technique of arranging data in different heomogenous groups is known as classification.

**Ans to Q No 2 :** A frequency distribution is a comprehensive way to classify raw data of a quantitative variable.

**Ans to Q No 3 :** Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes accending to a certain rule.



## 13.10 MODEL QUESTIONS

**A) Short Questions** (Answer each question in about 150 words)

Q1 : Define classification.

Q2 : What do you mean by frequency distribution?

Q3 : What do you mean by cumulative frequency distribution?

Q4 : Differentiate between discrete and continuous variable.

Q5 : Differentiate between qualitative and quantitative classification of data.

Q6 : What are the uses of cumulative frequency distribution?

**B) Long Questions** (Answer each question in about 300-500 words)

Q1 : Discuss the objectives of classification of data.

Q2 : Discuss the different types of classification of data.

Q3 : Discuss the steps involved in the preparation of frequency distribution.

Q4 : Discuss the steps involved in the conversion of Inclusive Series in Exclusive Series.

*** ***** ***