**Presentation of data**

“The device of gathering data and result by using statistical concepts is called presentation of data.”

Here we discuss some of the methods which are used for gathering of data, accommodation of data which are classification, tabulation and graphical methods.

**Classification:**

“Classification is the sorting of data into homogenous classes.”

Or

“It is the statistical process through which we divide a data into classes or small groups according to their being a like or not.”

In classification the data are homogenous internally classes but non-homogenous externally to other classes.

**Types of classification:**

There are three types of classification:

1. **One way classification**:

“When the data are classified according to single characteristics, then classification is known as one way classification.”

Or

“If the data is classified according to one criteria only, it is called one way classification or simple classification.”

E.g: Human beings classified according to sex (male or female)

1. **Two- Way Classification:**

“When the classify the data by considering two characteristics then we are doing two way classification.”

Or

The data in two- way classification classified according to two criteria.

E.g

i. Population is classified according to sex and literacy.

ii. Population of world classified by region and sex.

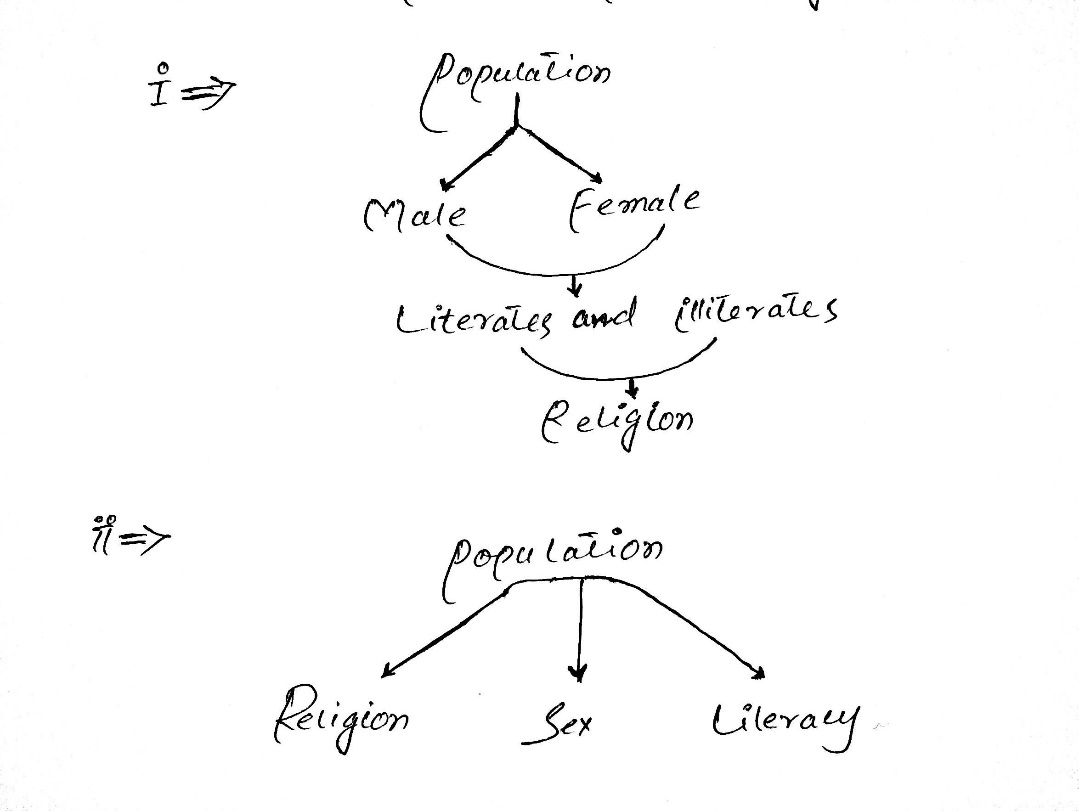
1. **Manifold Classification:**

“When we are considered more than two characteristics at a time to classify the data called manifold classification.”

Or

“If more than one attributers two, three ……, are classify simultaneously, the classification is said to be manifold classification”

E.g



**Aims of Classification:**

The aims of classification are:

1. To reduce the large set of data/
2. To display the point of similarity and dissimilarity.
3. Eliminate unnecessary detail.
4. To save mental strain.
5. To reflect the important aspects of data.
6. Prepare a ground for comparison.

**Basic principles of classification:**

In classification we classify a large set of data. While classifying the following points should be kept in mind:

1. The classes or categories should be mutually exclusive and no over-lap should exist between successive classes.
2. The classes or categories should include all the dat.
3. The conventional classification procedure should be adopted.
4. All the data adjusted into one or two classes.

**Distribution:**

“The arrangement of data according to the characteristics of variables is called distribution.”

**Spatial Or Geographical Distribution:**

“When the variables are expressed in terms of location we get a geographical distribution”.

Temporal arrangement of values refers to the time series.

**Tabulation:**

“Tabulation is the systematic presentation of data classified under suitable heads and sub-heads on the basis of rows and columns.”

Or

“The arrangement of data in rows and columns is called tabulation”.

It is the second method which is used for representing data or a group of data.

There are two types of tabulation on the basis of purpose:

1. General purpose table
2. Special purpose table
3. **General Purpose table:**

The general purpose tables are very large in size, extensive with waste coverage and constructed for reference purpose.

1. **Special Purpose table:**

It is very simpler in structure. It deals with one or two criteria of variables such tables are used to a system analyzing data.

When the classification corresponds to once, two or many criteria or characteristics, the tabulation is called single, double, or manifold tabulation respectively.

1. **Single Tabulation:**

“Table with one criteria of classification is called single tabulation.”

E.g: Students against weight.

1. **Double Tabulation:**

“Table with two criteria of classification is called double tabulation.”

e.g Gender and marital status.

Gender and height.

Gender and weight.

1. **Manifold:**

“Table with more than two criteria of classification is called manifold classification”.

E.g: Population of a country by age, by gender, by residence, by literacy etc.

**Frequency Distribution:**

Frequency distribution can be define as:

“The values/ data along with respective probability is called frequency distribution”.

Or

“The organization of data in a table showing the data dividing into different classes along with respective frequency/ probability of each class is called frequency distribution”.

**Class Frequency:**

“The numbers of observation falling in a particular class is called class frequency”

**Group data:**

“The data presented in a form of frequency distribution is called group data.”

**Ungroup Data:**

“The data is in original form is called un-group data.”

**Class limit:**

Class limit can be define as:

"The numbers / values of the variable which describe classes is called class limit”.

The larger number is the upper class limit.

**Class boundaries:**

Class boundary can be defined as:

“The precise number which separate one class from another.”

The selection of these numbers removes the difficulties.

**Class Mark or Mid-point:**

“The class mark can be calculated by adding two class limits and dividing by two is called class mark or mid-point.”

Class mark divides class into two equal parts.

**Class width or Class interval:**

Class width is equal to the difference between class boundaries.

Class width can also be calculated form the two successive lower class limits or the two successive class marks.

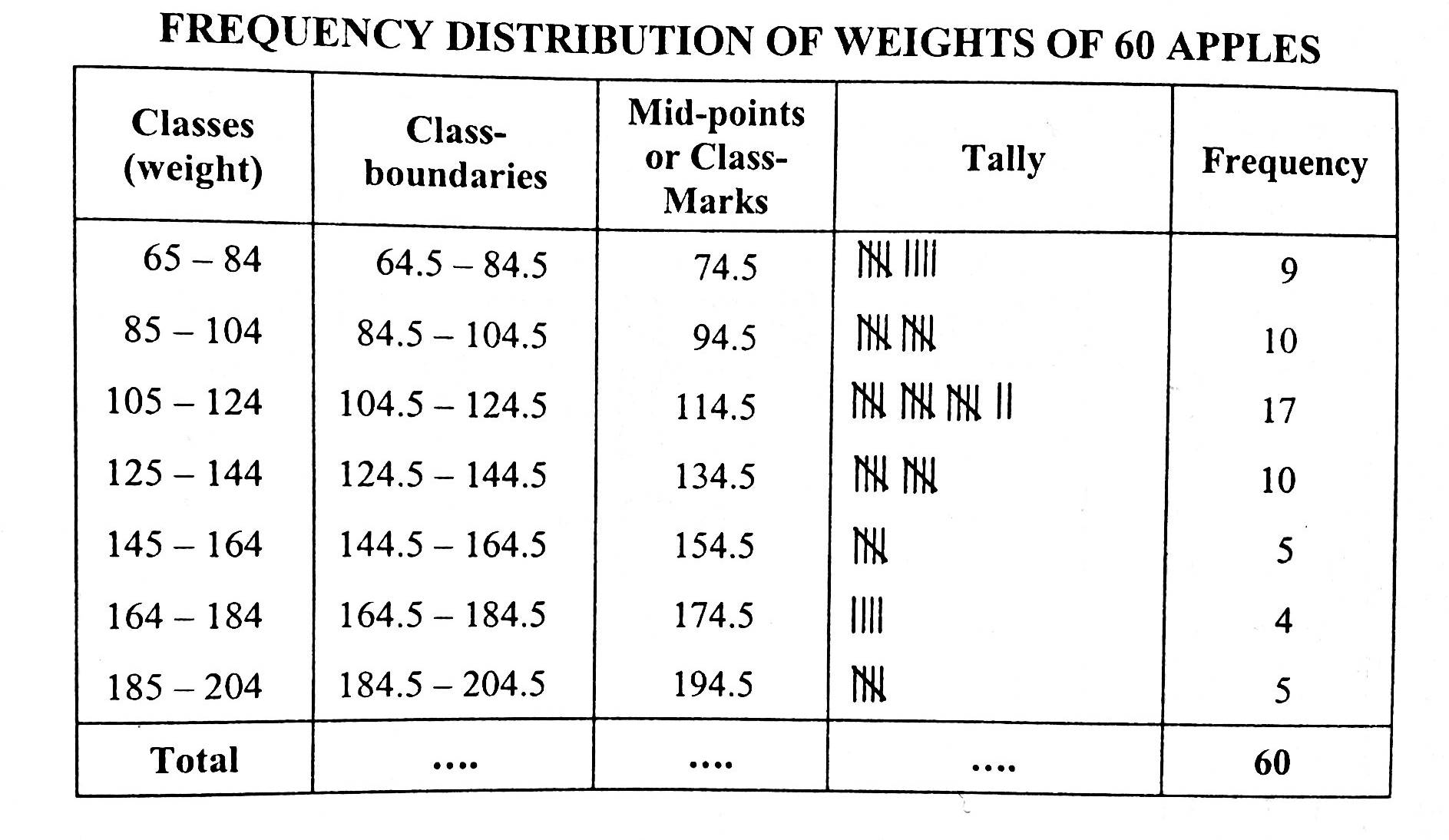
**Example:**

Make a grouped frequency distribution from the following data, relating to the weight recorded to the nearest grams of 60 apples picked out at random from a consignment.

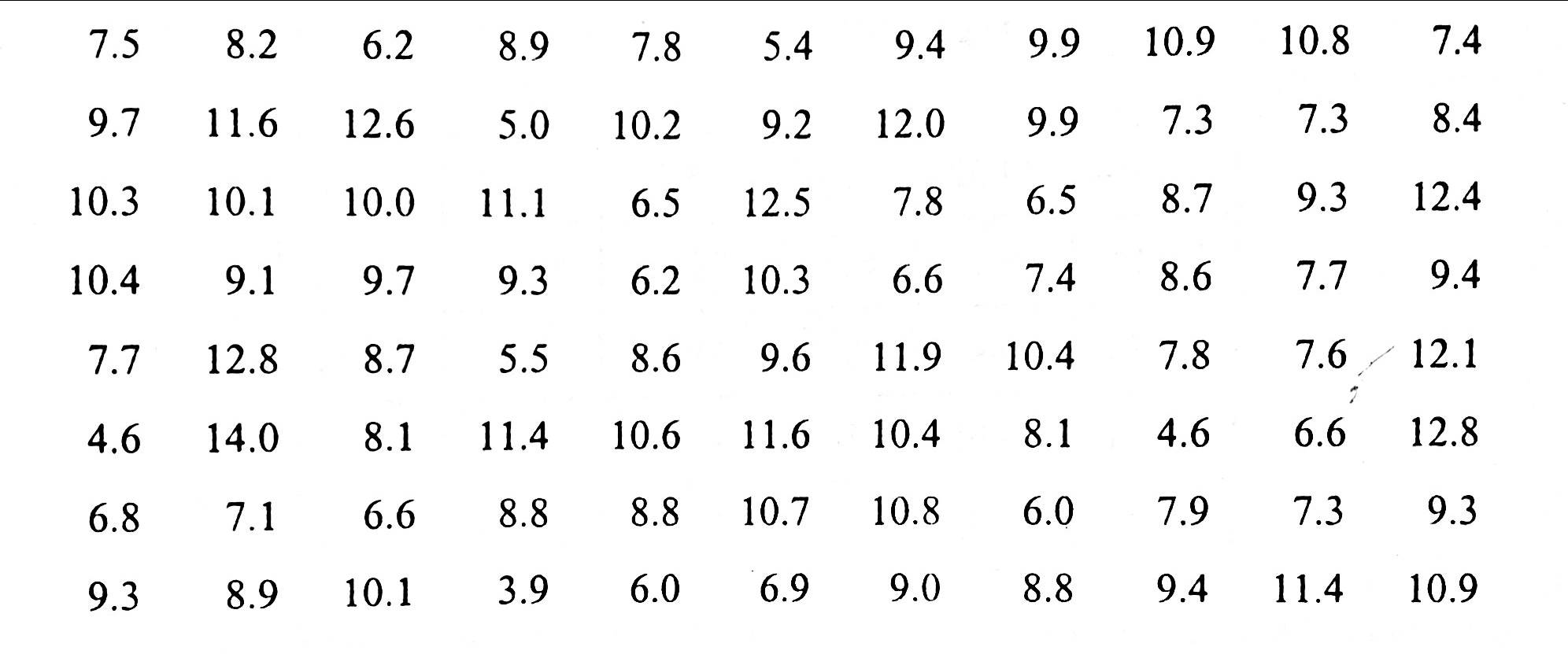
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 106 | 107 | 76 | 82 | 109 | 107 | 115 | 93 | 187 | 95 | 123 | 125 |
| 111 | 92 | 86 | 70 | 126 | 68 | 130 | 129 | 139 | 119 | 115 | 128 |
| 100 | 186 | 84 | 99 | 113 | 204 | 111 | 141 | 136 | 123 | 90 | 115 |
| 98 | 110 | 78 | 185 | 162 | 178 | 140 | 152 | 173 | 146 | 158 | 194 |
| 149 | 90 | 107 | 181 | 131 | 75 | 184 | 104 | 110 | 80 | 118 | 82 |

By scanning the data, we find that the largest weight is 204 grams and the smallest weight is 68 grams so that the range is 204-68 =136 grams.

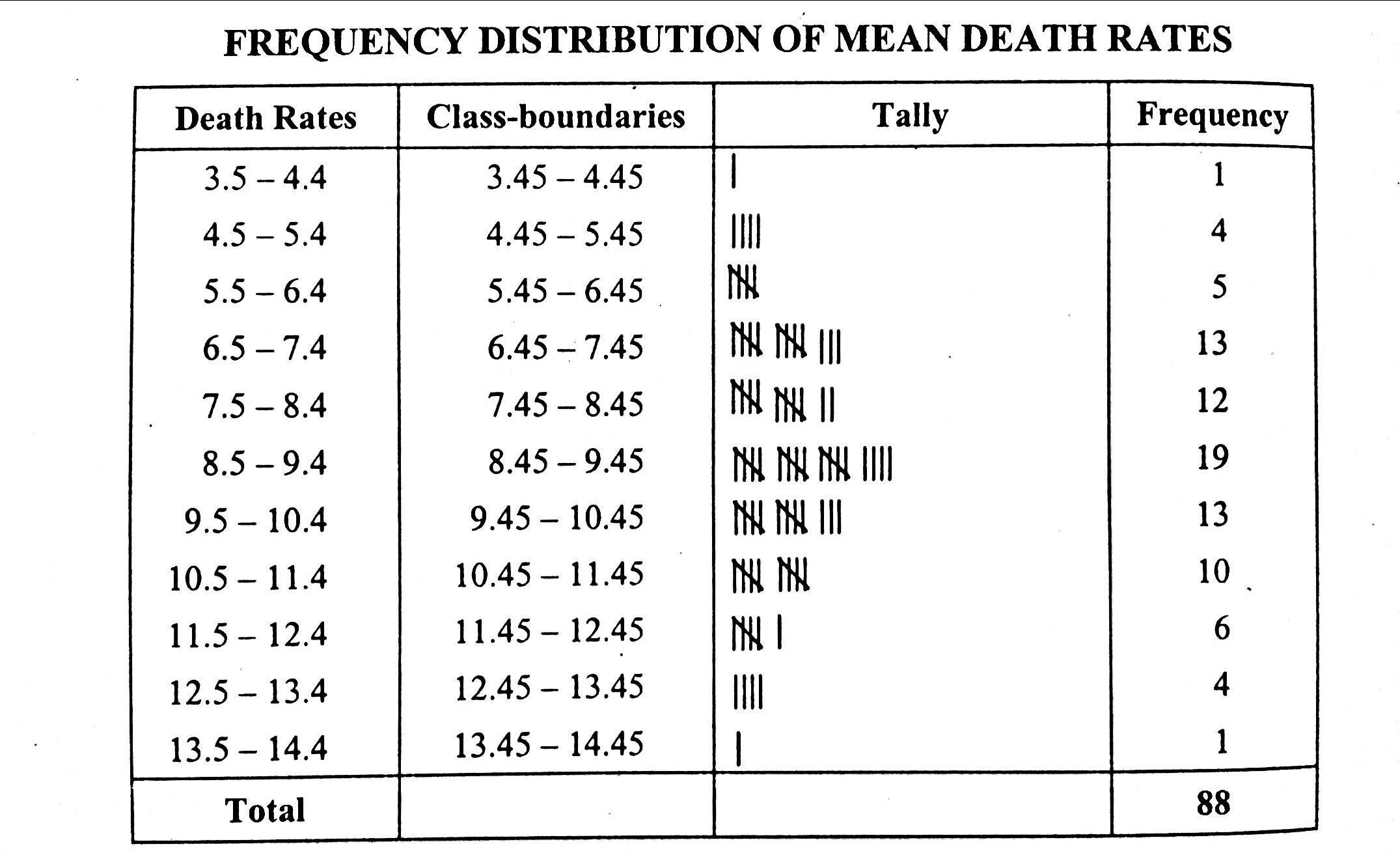
Suppose we decided to take 7 classes of equal size. Then size or width of the equal class interval would be 136/7 =19.47. But we take h=20 the next integral value higher than 19.47 to facilitate the numerical work.

 Let us decide to locate the lower limit of the lowest class at 65 with this choice, the class limit will be 68-84, 85-104, 105-124…., the class boundaries become 64.5-84.5, 84.5-104.5, 104.5-124.5,…., and the class marks are 74.5, 94.5, 114.5,….. the grouped frequency distribution is then constructed as follow:

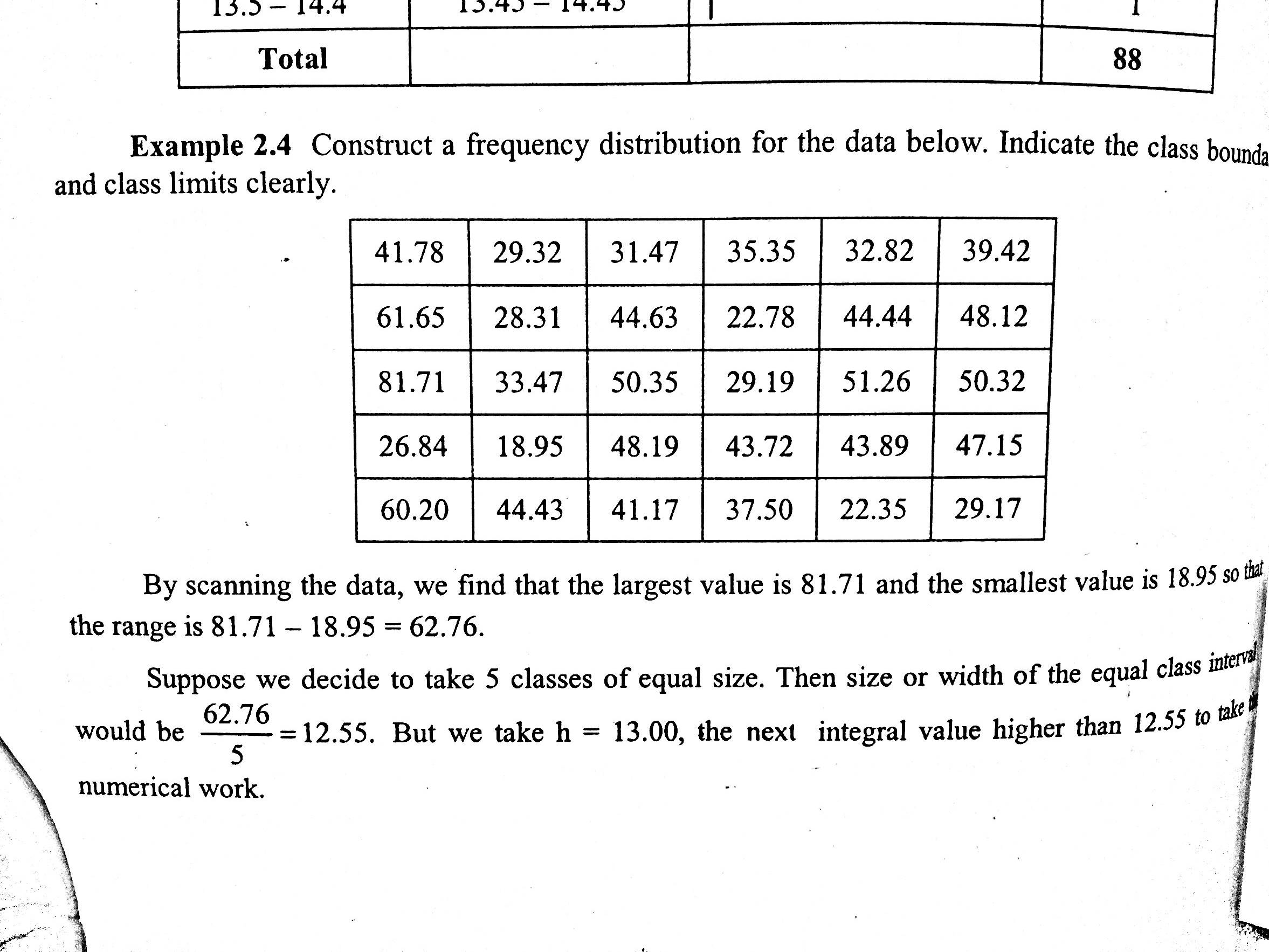
**Example:** Given below are the mean annual death rates per 1000 at ages 20-65 in each of 88 occupational groups. Construct a frequency distribution.



A sum of the data shows that the largest value is 14.0 and the smallest value is 3.9, so that the range is 14.0 – 3.9 = 10.1

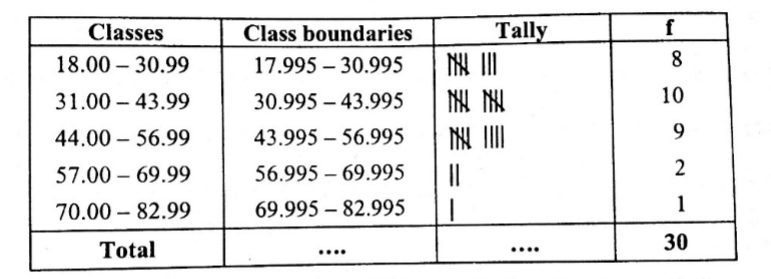
 As the data are recoded to one decimal place, we may therefore locate the lower limit of the first group at 3.5. let us choose a class interval of 1.0 then the class limits are spaced as 3.4 - 4.4, 4.5 - 5.4 , 5.5 - 6.4……, with this choice, the boundaries are 3.45 -4.45, 4.45 – 5.45, 5.45 – 6.45…… which do not coincide with the given values.

**Example:**

 Construct a frequency distribution for the data below. Indicate the class boundaries and class limit clearly:

By scanning the data, we find that the largest value is 81.71 and smallest value is 18.95. so that the range is 81.71-18.75 = 62.76

Suppose we decide to take 5 classes of equal size. Then size or width of the equal class interval would be 62.76/5 = 12.55. But we take h = 13.00, the 5 nest integral value higher than 12.55 to take the numerical work.

 As the data are recorded to two decimal places, we may locate the lower limit of the first group at 18.00 with this choice, the class limits will be 18.00- 30.99, 31.00 – 43.99 …. The class boundaries became 17.995 – 30.995, 30.995-43.995, the grouped frequency distribution is then constructed as follow.

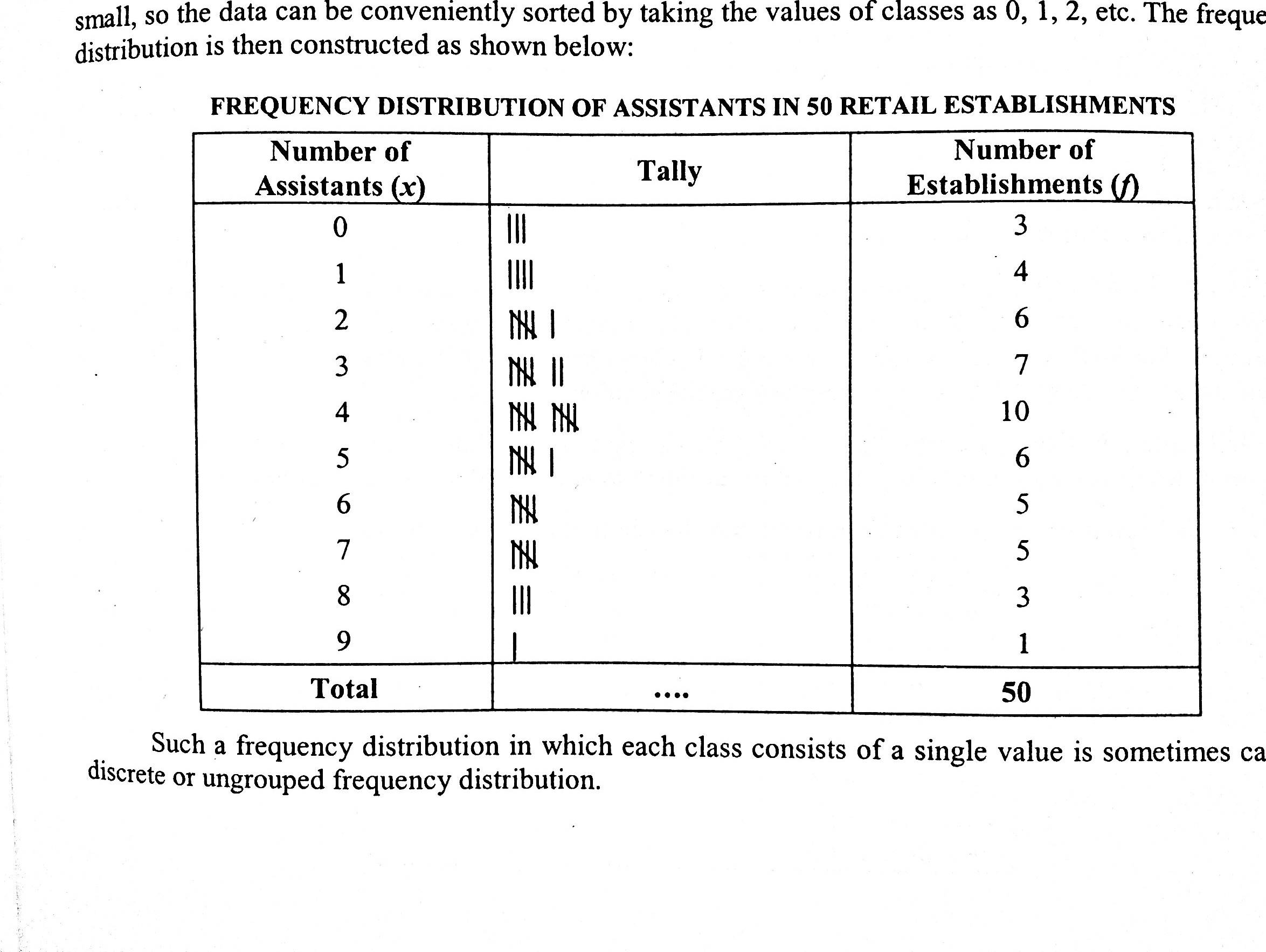
**Example:**

A survey of 50 retail establishments had assistants, excluding proprietors, as follow:

2,3,9,0,4,4,1,5,4,8,5,3,6,6,0,2,2,7,6,4,8,4,3,3,1,0,8,7,5,1,3,4,2,4,7,5,2,6,3,1,7,5,4,6,4,2,5,3,4.

Arrange the values as a frequency distribution.

By scanning the data, we find that the number of assistants is a discrete variable and the range is small, so the data can be conveniently sorted by taking the values of classes as 0,1,2 etc. the frequency, distribution is then constructed as shown below:

Frequency distribution of assistants in 50 Retail Establishment.

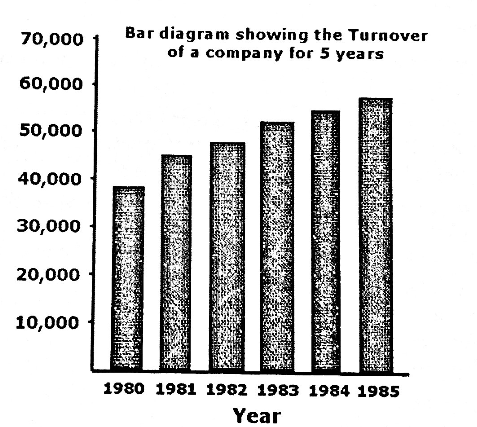
Such a frequency distribution in which each class consists of single value is sometimes called a discrete or ungrouped frequency distribution.

**Example:**

Draw a simple bar diagram to represent the turnover of company for 6 years.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Years: | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 |
| Turn Over: | 38,000 | 45,000 | 48,000 | 52,000 | 55,000 | 58,000 |

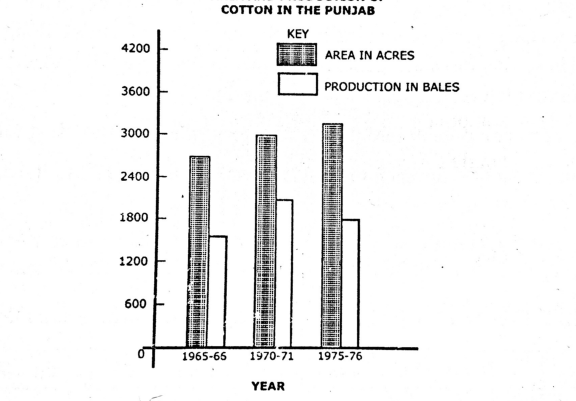
The bar chart is drawn below:



**Example**:

Draw multiple bar charts to show the area and production of cotton in the Punjab from the following data:

|  |  |  |
| --- | --- | --- |
| Year | Area (000 acres) | Production (000 Bales) |
| 1965 - 66 | 2866 | 1588 |
| 1970 – 71 | 3233 | 2229 |
| 1975 – 76 | 3420 | 1937 |

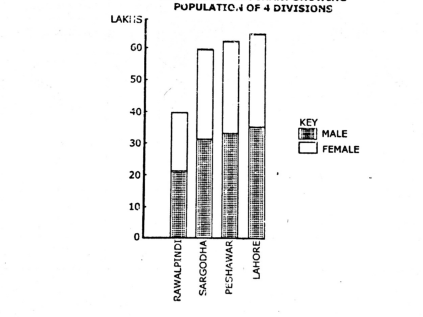
The Multiple bar charts are drawn below:

**Example:**

Draw a component bar chart for the following data:

Population in lakhs

|  |  |  |  |
| --- | --- | --- | --- |
| Division | Both Genders | Male | Female |
| Peshawar | 64 | 33 | 31 |
| Rawalpindi | 40 | 21 | 19 |
| Sargodha | 60 | 32 | 28 |
| Lahore | 65 | 35 | 30 |

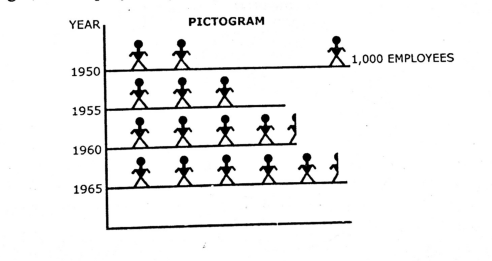
The Appropriate component bar chart after arranging the population figure in ascending order is drawn below:

**Example:**

The following table shows the number of employees in a certain textile mills, represent the data by means of a pictogram:

|  |  |
| --- | --- |
| Year | No. Of Employees |
| 1950 | 2,004 |
| 1955 | 2,990 |
| 1960 | 4,240 |
| 1965 | 5,380 |

Representing 1000 employees by one picture, the pictogram is down below.

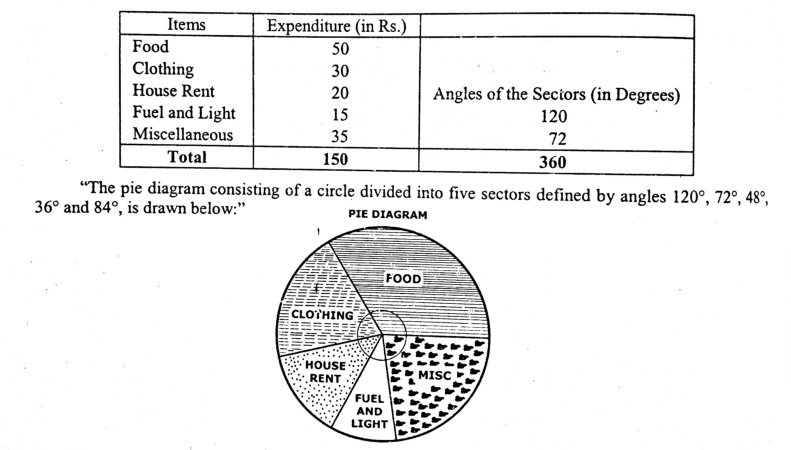


**Example:**

Represent the total expenditure and expenditure on various items of a family by a pie diagram.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Items | Food | Clothing | House Rent | Fuel and light | Misc. |
| Expenditure (in Rs.) | 50 | 30 | 20 | 15 | 35 |

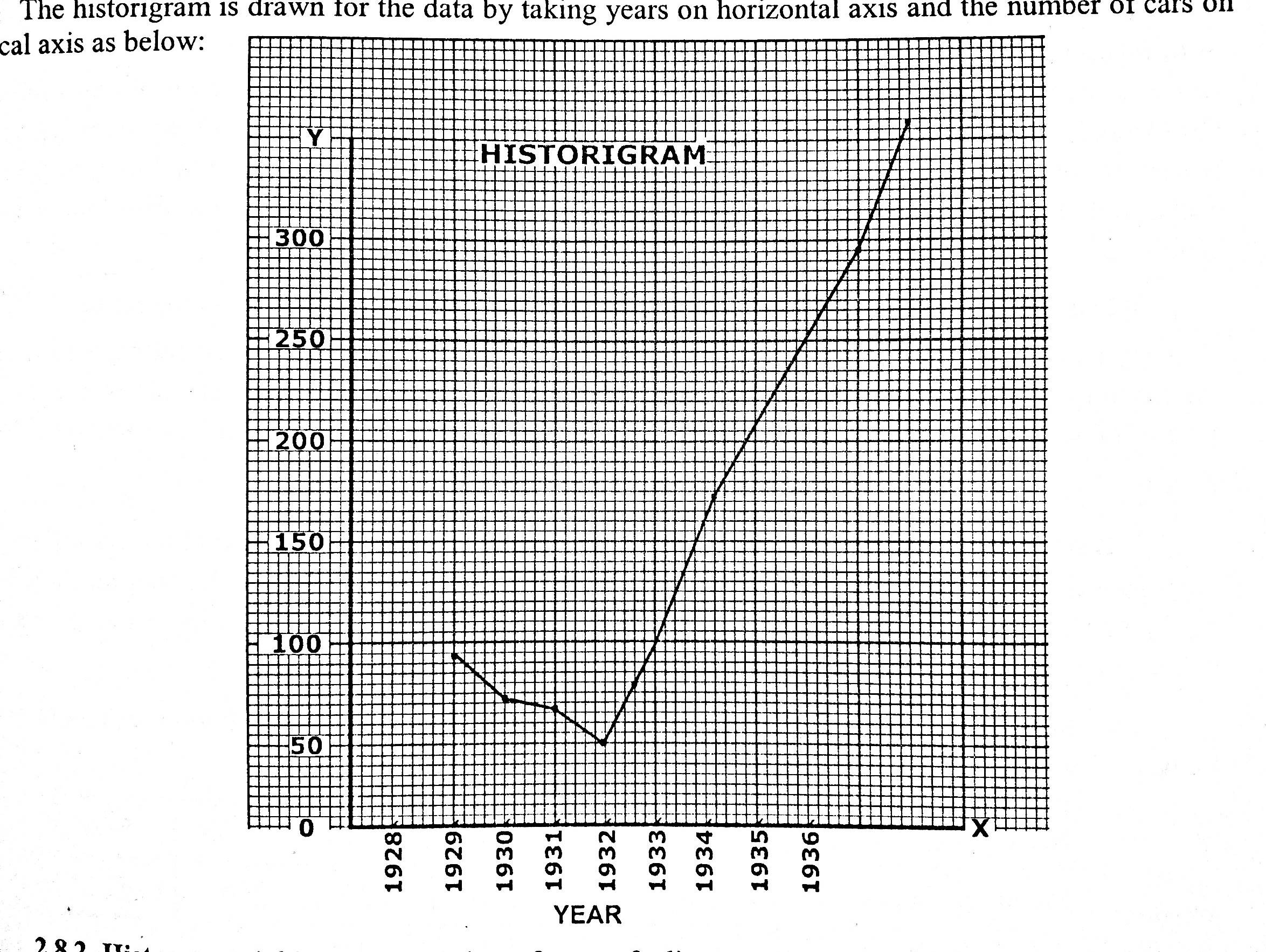
The Corresponding angles needed to Draw the chart are computed below.



**Example:**

The Following tables given the number of cars produced in Germany during the year 1929 – 1936. Draw a suitable graph i.e Histogram of the series.

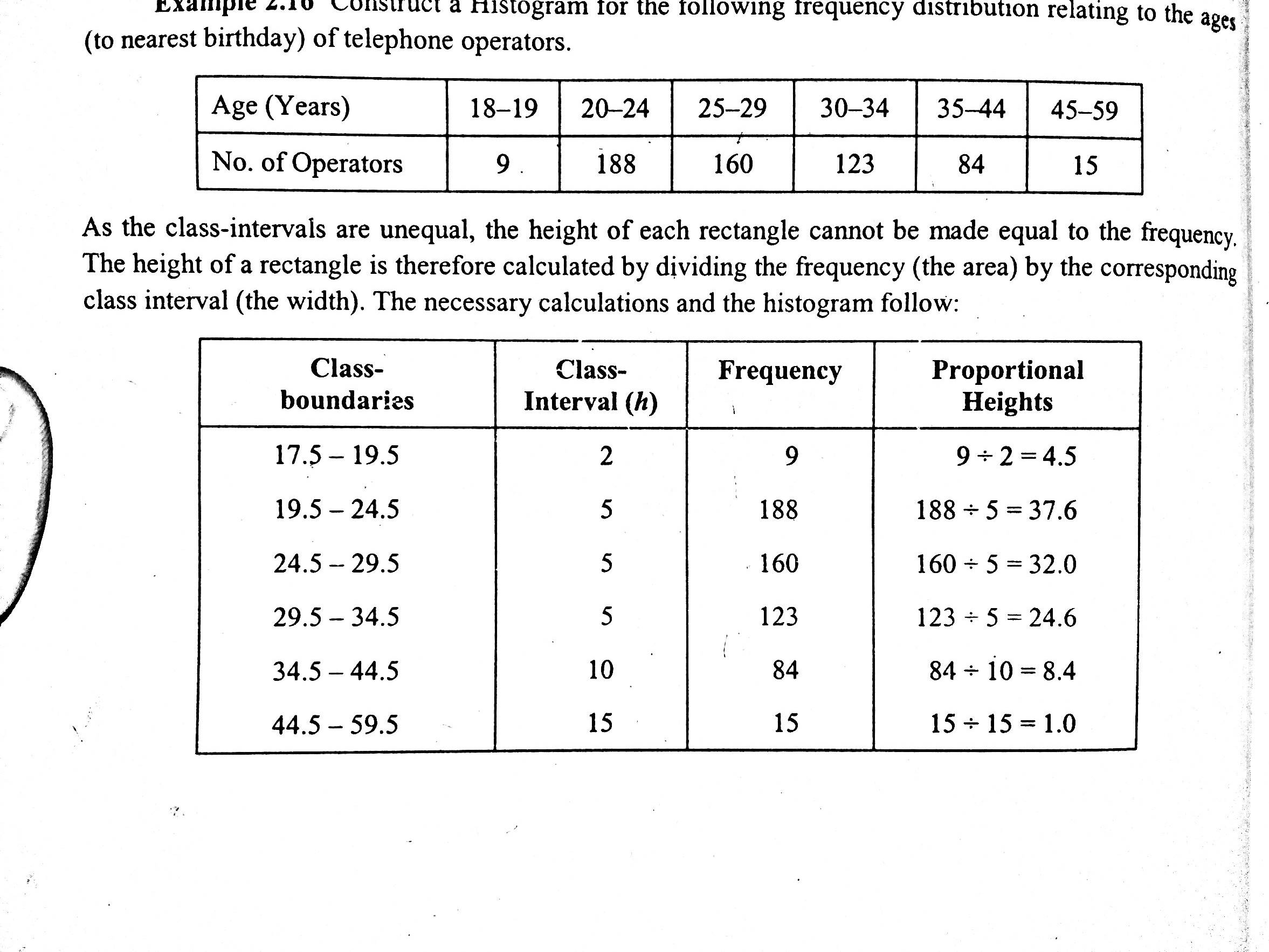
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 1929 | 1930 | 1931 | 1932 | 1933 | 1934 | 1935 | 1936 |
| Cars | 98 | 74 | 68 | 50 | 99 | 172 | 245 | 302 |

 The histogram is drawn for the data by taking years on horizontal axis and the number of cars on vertical axis as;

**Example:**

Construct a histogram for the following frequency distribution relating to the ages (to nearest birthday) of telephone operator.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Age | 18 – 19 | 20 – 24 | 25 – 29 | 30 – 34 | 35 – 44 | 44 – 45 |
| NO of operation | 9 | 188 | 160 | 123 | 84 | 15 |

As the class – interval are unequal, the height of each rectangle cannot be made equal the frequency. The height of a rectangle is therefore calculated by dividing the frequency (the area) by the corresponding class interval (the width). The necessary calcuations and the histogram follow.

