Presentation of data by frequency trebles, diagrams and graphs partition values

1. Introduction

Welcome to the series on E-Learning module in Statistics and Statistical methods. In this module, we are going to cover a basic yet important concept called Frequency Distribution. Frequency distribution forms a base for many statistical analysis methods.

At the end of this session, you will be able to:
• Prepare the Frequency Distribution (FD) table
• Discuss the steps involved in preparing a Frequency Distribution table
• Explain the usage of diagrams and graphs in representing a data

Let us have a recap on what statistical analysis is all about. What we understand by frequency distribution? And the steps involved in preparing the frequency distribution table.

Statistical analysis is a mechanism, which transforms a disorganized meaningless data into organized, structured, meaningful, and useful data, which can be used effectively in planning and development activities. Frequency distribution helps in transforming the raw data into organized and important data for statistical analysis and inferences.

What do you understand by Frequency Distribution?
Frequency Distribution is a process of arranging data into sequences and groups or separating them into different parts according to certain chosen common characteristics.

It shows a summarized grouping of data and presents the data in a form that clearly describes a thematic representation of the data based on the selected characteristics.

For example:
• A Consumer goods company would like to know the sales performance of a certain product (say TV or Refrigerator or Washing machine) in a region for the current year

• Central Government would like to know how different states are performing on the agricultural yield or milk production
An academic institution would like to know the academic performance of their students in various exams

• A company would like to know how it has performed year on year on parameters like revenue, profitability vis-à-vis its competitors

Steps involved in preparing a frequency distribution table are Arraying, condensing, classification, grouping, and representing data.

2. Various Problems to Understand the Preparation of a Frequency Distribution Table

Let us now work out various problems to understand the preparation of a frequency distribution table.

(Refer to Figure 1)

Problem 1:
The following data is the number of years that 50 workers have worked for a small factory. Construct a frequency distribution starting from 0 and under 2 and having equal width in each class.

As a first step, we need to arrange the data in ascending order. The least value in the data is 0.4 and the largest value is 15.0. Once the data is arranged, let us now condense the data by grouping into various class intervals.

(Refer to Figure 2)

As already mentioned in the problem, the data needs to be grouped starting with 0 as the lower limit and 2 as the upper limit for the first class interval, two as the lower limit and 4 as the upper limit and so on up to 14 to 16 as the maximum value is 15. It is also mentioned that we need to have equal class width. Keeping these points in mind we have the data grouped as 0 to 2, 2 to 4, 4 to 6, 6 to 8, 8 to 10, 10 to 12, 12 to 14 and 14 to 16.

(Refer to Figure 3)

Now, to condense the data we will use the tally marks and mark each variable into the group it belongs. The variable 1.4 comes into the group 0-2. Similarly, all the data points are grouped into the table using the tally marks. When we total the tally marks we get the frequencies, which explains the number of data points in each group. Here, in this case, we have 8 data points in the first class interval 0-2. Similarly, we have 11 data points in the group 2-4; thus, we have taken the total of all the data points under various groups. Therefore, the frequency distribution table is ready.

(Refer to Figure 4)

Construct a cumulative frequency distribution table from the following grouped frequency distribution table.

(Refer to Figure 5)

We need to transform the grouped frequency distribution into cumulative frequency distribution table. For this as a first step, we will convert the class intervals into less than or more than class intervals. If it is less than class interval we consider the upper limit and while preparing the more than class interval we

will consider the lower limit.

Now, we will prepare the table by considering the less than class interval. Therefore, the first class interval is less than 10, the next class intervals will be less than 20, less than 30, less than 40 and less than 50.

(Refer to Figure 6)

The next step is calculating the cumulative frequency. The first frequency is taken as it is and the subsequent frequencies are added to the preceding frequencies. Here, the first frequency is 7 and the cumulative frequency is 7. In the second class we have the frequency 10 plus the preceding frequency 7, which makes the cumulative frequency 17. The third class we have the frequency 18 plus the total of the preceding class 17, which is equal to 35. Then, we have 35 plus 9, which is equal to 44, and the last class we have 44 plus 6 is equal to 50. Thus, the cumulative frequency distribution table is ready.

(Refer to Figure 7)

In a survey, it was found that 64 families bought milk in the following quantities (litres) in a particular month. Using sturges rule, convert the above data into a frequency distribution by inclusive method.

(Refer to Figure 8)

Sturges formula is used to calculate the number of classes of a data. The Sturges formula is as follows:
$n=1+3.322 \log 10N$
where,
n=Number of classes (rounded to the next whole number)
N=The total number of observation

1+3.322*log10(64)
1+3.322 (1.806) =6.9995532 rounded of to 7
Therefore, we will have 7 classes of the data. The lowest value of the data is 5 and the highest value of the data is 39 so, we will have class intervals with width as 5 that is 39 minus 5 divided by 7 is equal to 4.8.

After arranging the data and using the sturges rule we will get the following:

(Refer to Figure 9)

So the first class interval for this data will start with the least value 5. Add the width 5 which is equal to 9. As the requirement for the data is an inclusive series, the lower limit of the next class interval and the upper limit of the previous class interval will not be the same. Therefore, we will have the next class interval as 5-9, 10-14, 15-19, 20-24, 25-29, 30-34 and 35-39.

(Refer to Figure 10)

In the next step, we prepare the frequency distribution by taking each of the data points and marking it against the variable group using the tally bars. Now, we get the frequencies of the data as 7, 10, 13, 18, 8, 5 and 3. Thus, the frequency distribution table is ready.

(Refer to Figure 11)

3. Discuss the Diagrams and Graphs

A market survey collected responses from 50 people regarding acceptability of a new product. The scores of the respondents on appropriate scale are as follows. Prepare a frequency table.

(Refer to Figure 12)

As a first step for preparing the frequency distribution table, we will arrange the given data in ascending order. We get the following table.

(Refer to Figure 13)

Now we need to group the data into various class intervals, which can be done using the sturgers formula. In practice, the number of classes is determined keeping in mind the requirement given in the problem. Classes should be chosen in such a way that they cover all the data to be categorized.

Here, we have the lowest value of the data as 8 and the highest value as 48, taking a width of 5 we will get around 9 class intervals. The class intervals will be 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40 40-45 and 45-50. Now, let us use the tally mark to group the data points in the respective class intervals and we get the frequency as 2, 2, 1, 2, 9, 3, 10, 16 and 5. Thus, the frequency distribution table is ready.

(Refer to Figure 14)

Let us discuss the Diagrams and Graphs.
Data once collected and summarized are presented in a systematic manner using statistical techniques. However, we need to keep in mind that the common person uses most of the statistical data and an easy representation of the same is needed for easy understanding of the concepts. The best way for such a representation is done by using diagrams and graphs.

Various sources like magazines, newspaper, journals use diagrams and graphs to display results pictorially. This kind of representation needs a lot of artistic talent and imagination on the part of the person preparing the diagrams and graphs.

Let us now work out few problems to understand the preparation of diagrams and graphs.
Average wages of some firms are given below. Represent this by a simple bar diagram.

(Refer to Figure 15)

The firms A, B, C, D, E and F are taken on the X-axis and the average wages of these firms are taken on Y-axis. One small division in Y-axis is equal to 100/10 is equal to 10. On X-axis, equal width is given to each variable and distance between these bars is constant.

(Refer to Figure 16)

A survey was conducted in a major city with a group of 300 people to give their preferences for the best metro to live. The data obtained are as follows. Summarize the data above and construct a pie chart a bar graph and a percentage frequency distribution.

(Refer to Figure 17)

o represent the data by circular diagram, we express each city as percentage of aggregate.
Mumbai is equal to 120 divided by 300 into 100 is equal to 40.
Similarly, Delhi is equal to 80 divided by 300 into 100 is equal to 26.67, Chennai is equal to 60 divided by 300 into 100 is equal to 20 and Kolkatta is equal to 40 divided by 300 into 100 equal to 13.33.

Now 1 percent corresponds to 3.6 degrees, so the angles at the centre of the corresponding metros are Mumbai is equal to 144 degrees (40X3.6), Delhi is equal to 96 degree (26.67X3.6), Chennai is equal to 72 degree (20X3.6) and Kolkatta is equal to 48 degree (13.33X3.6)
Note: Addition of all percentage should be = 100 and addition of all angles should be 360 degree.

(Refer to Figure 18)

After all the calculation, we represent the data in a pie chart format and simple bar diagram.

(Refer to Figure 19)

4. Examples

Following is the population of few countries. Present this data using pictogram.

(Refer to Figure 20)

In this type of problems we take a suitable sign so as to depict or make an effect of the desired element. Therefore, here we are picking up the sign for population.

Let one person is equal to 10 crores then, we represent the above data as follows:

The population of India is 90 crores, so we have represented it with 9 pictures of a person. Similar representation is done for Pakistan, China, USA and Russia.

Present the following data on a graph paper.

(Refer to Figure 21)

In this case, we will take the time on the X-axis at equal distance and plot the production on the Y-axis. Scale on y-axis: 1 division is equal to 20 divided by 10 is equal to 2000 metric tonnes.

(Refer to Figure 22)

Prepare Histogram for the following data:

(Refer to Figure 23)

As discussed in the module for histogram, we have two types of histogram namely:
Histogram with equal class intervals
Histogram with Unequal class intervals

Here, we are preparing a histogram for a data with equal class intervals.
Let us take a false base line for OX as the smallest term is 325. We take scales for both the axes. For X-axis, 1 division is equal to 25 the width of the class interval divided by 10 units is equal to 2.5 units and on Y-axis 1 division is equal to 10 divided by 10 is equal to one unit. With this scale we take the wages in the x-axis and the frequencies in the Y-axis. We have now plotted the frequencies against the upper limit of the class interval and the histogram is prepared.

(Refer to Figure 24)

Prepare a histogram for the following data:

(Refer to Figure 25)

Here we have observed that the class interval is unequal series, we will now convert this into equal interval series by adjusting the frequencies. Here the smallest class interval is of width 10 so let us convert all the intervals to this size. Here the two class intervals 30-60 and 60-80 are unequal and can be divided into two class intervals of 10 each. So the adjusted frequencies are frequency 18 divided by 3 is equal to 9 each and the other class interval 60-80 has 4 frequencies divided by 2 is equal to 2 each. Therefore, the new set of data will be as shown in the table.

(Refer to Figure 26)

Now, let us take the scales for the X-axis and Y-axis. We have 1 division is equal to 10/10 is equal to 1unit. Plot the data in the graph we will get a histogram as follows.

(Refer to Figure 27)

5. Construction of Frequency Polygon

Construct frequency polygon for following data:

(Refer to Figure 28)

In a frequency polygon graph, we can represent the data in three forms. The first form is that it can be done by just plotting the curve directly in the graph and join the points with straight lines.

The second form is that the points can also be joined with free hand curve to get a smooth curve, remove the ruggedness of the polygon, and present it in a good form or shape.

We can draw the frequency polygon with a histogram or without a histogram. The curve begins and ends in the base line. We have represented the data with histogram, frequency polygon and frequency curve as shown in the diagram.

(Refer to Figure 29)

Draw the less than and more than ogives for the following data:

(Refer to Figure 30)

In this case, we have two types of ogives namely less than ogive and more than ogive.
The less than ogives are got by adding the frequencies from top to bottom to get the cumulative frequencies and plotting against the upper limit of the class intervals and joining those lines or by free hand curve. The lower point is joined to the X-axis at the lower limit of the first class interval.
The more than ogives are frequencies cumulated from bottom to top, and such cumulated frequency is plotted against the lower limit of corresponding class intervals. The upper most point is joined to the Y-axis at the upper limit of the last class interval.

First let us change the above given discrete series to continuous intervals by adding the correction factor 0.5 to get the upper limit and minus the correction factor from the value to get the lower limit. So in this case, we will get the first class interval minus point 5 to point five, the second class interval is 0.5 to 1.5, the third class interval is 1.5 to 2.5. Similarly, we will calculate all the class intervals as seen in the table. Then, we will calculate the cumulative frequencies both the less than and more than by adding the

frequencies.

(Refer to Figure 31)

Plot the less than frequency and the more than frequency in a graph we will get both the ogive curves.

(Refer to Figure 32)

For the following data, draw less than and more than ogive.

(Refer to Figure 33)

The given data is already in the form of less than cumulative frequency. As we have to calculate the more than frequency, let us get the data in the normal data format.
So we will take the class intervals as 0-20, 20-40, 40-60, 60-80 and 80-100.
And the frequencies will be 41, 92 minus 41 is equal to 51, 156 minus 92 is equal to 64, 194 minus 156 is equal to 38 and 201 minus 194 is equal to 7.

Let us now calculate the less than and more than cumulative frequencies. The less than ogives are got by adding the frequencies from top to bottom to get the cumulative frequencies and plotting against the upper limit of the class intervals and joining those lines or by free hand curve. The lower point is joined to the X-axis at the lower limit of the first class interval.

The more than ogives are frequencies cumulated from bottom to top, and such cumulated frequency is plotted against the lower limit of corresponding class intervals. The upper most point is joined to with the Y-axis at the upper limit of the class of the last class interval.
So the new data set will look like the one in the table with both the less than frequencies and more than frequencies.

(Refer to Figure 34)

Now, when plotting both the values against the lower limit and the upper limit of the class interval we get the ogives.

(Refer to Figure 35)

Here's a summary of our learning in this session where we have understood:
- The preparation of the Frequency distribution (FD) table
- The steps involved in preparing a Frequency Distribution table
- The usage of diagrams and graphs in representing a data