1.Introduction

Welcome to the E-learning module on Statistics. In this session we are going to cover the concept of Scatter plot and Pie Chart.

At the end of the this session you will be able to :

- Explain Scatterplot and the types of correlations in Scatterplot
- Appreciate the use of Scatterplot in real-life situations
- Explain Pie Chart and the construction of a Pie Chart
- Explain the examples of a Pie Chart
- List the limitations of a Pie charts

In a scatter plot, the correlation between variables is shown in terms of a collection of points at the Cartesian coordinates of the variables which constitute the set of data. Scatterplot are also called Scatter graphs, scatter charts, and scatter diagrams.

Very often we come across the need to understand the correlation between two variable data points. For example, we may want to understand the relationship between size of diamonds and their price?

Or what is the correlation to educational qualification and income of people?

Scatterplot are used to investigate the possible relationship between two variables that both relate to the same "event”.

Data is shown as a collection of points marked on the Cartesian coordinates of the two variables. Here is an example on how they will look:

A scatter plot is a type of diagram to display values for two variables for a set of data, using Cartesian coordinates.

The data is displayed as a collection of visual points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

Let us take an example. Suppose we have to analyze the correlation between Height and Weight in people.

Figure 1

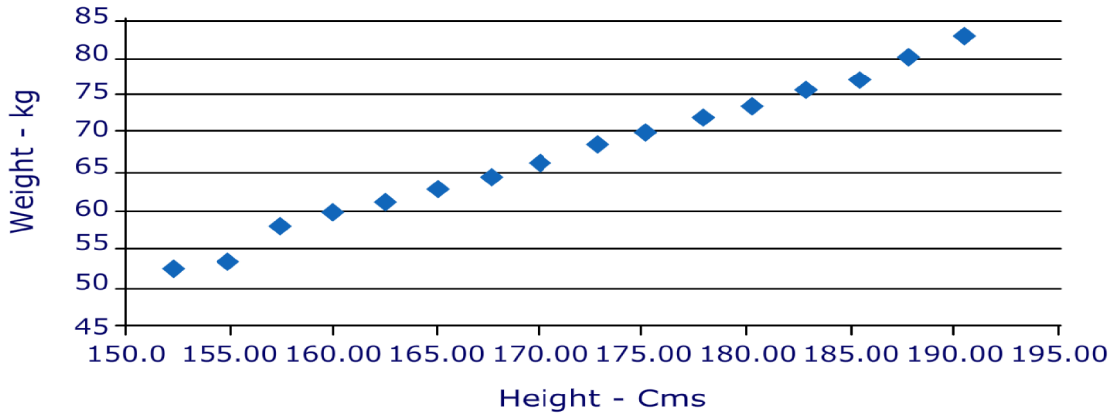
Healthy Men	
Height (cms)	Weight(kg)
152.30	52.6
154.80	53.55
157.40	58.3
159.90	59.65
162.40	61.2
165.00	63.05
167.50	64.45
170.00	66.25
172.60	68.3
175.10	70.1
177.70	71.9
180.20	73.7
182.70	75.75
185.30	77
187.80	80.55
190.40	82.85

First let us look at the data for healthy people (in this case, Indian man). Here, Height in cms is given, like 152.30, 154.80, 157.40, 159.90 and so on. Next column shows the weight in kg. like 52.6, 53.55, 58.3, 59.65 and so on.

Let us build a plotter chart with height in the X-axis and Weight in the Y-Axis. To make the chart easier to understand, we will use minimum values of 150.00 cm in the X-axis and 45 kg in the Y-axis.

The resulting plotter chart will look like this:

Figure 2



2. Types of Correlations

Correlations: Now we will discuss about correlation.

Correlation is a single number that describes the degree of relationship between two variables.

Let us take the same example of analyzing the height and weight of people.

Here, it clearly shows that pattern of dots slopes from left bottom corner to right top corner. We can draw an almost straight line connecting the points, in the direction of the slope. This is called a positive correlation. In our example, this means that weight increases proportionally to height.

In case the pattern of dots slopes in the opposite direction (from right top to left bottom) it means that, as one variable raises, the other decreases. This is called a negative correlation.

For example, as the age of a car increases, its market price reduces.

If no clear trend is visible, then the variables can be called uncorrelated.

Depending upon the general direction of the slope, we can also determine the correlation as “strong” or “weak”. A straight line indicates strong correlation whereas a jagged line but steadily in same direction indicates “weak” correlation.

Let us now plot a scatter chart using data for People (Indian men) affected with a particular disease in order to compare trends using Scatterplot.

This data shows the height and weight of affected men.

Height in centimeters like 152.30, 154.80, 157.40 and so on.

Similarly weight like 68.38 kg., 64.26 kg., 64.13 kg. and so on.

Figure 3

Healthy Men	
Height (cms)	Weight(kg)
152.30	68.38
154.80	64.26
157.40	64.13
159.90	76.35
162.40	80.78
165.00	76.92
167.50	82.50
170.00	86.13
172.60	89.47
175.10	95.34
177.70	87.72
180.20	84.76
182.70	87.11
185.30	87.78
187.80	96.66
190.40	101.08

This Scatter Graph shows that pattern of dots is neither consistently rising nor reducing consistently. There is no clear visible trend and hence the variables (height and weight) have become uncorrelated.

Let us now plot a graph with both the data i.e., of healthy men and of affected men.

In this table height is given in cms.

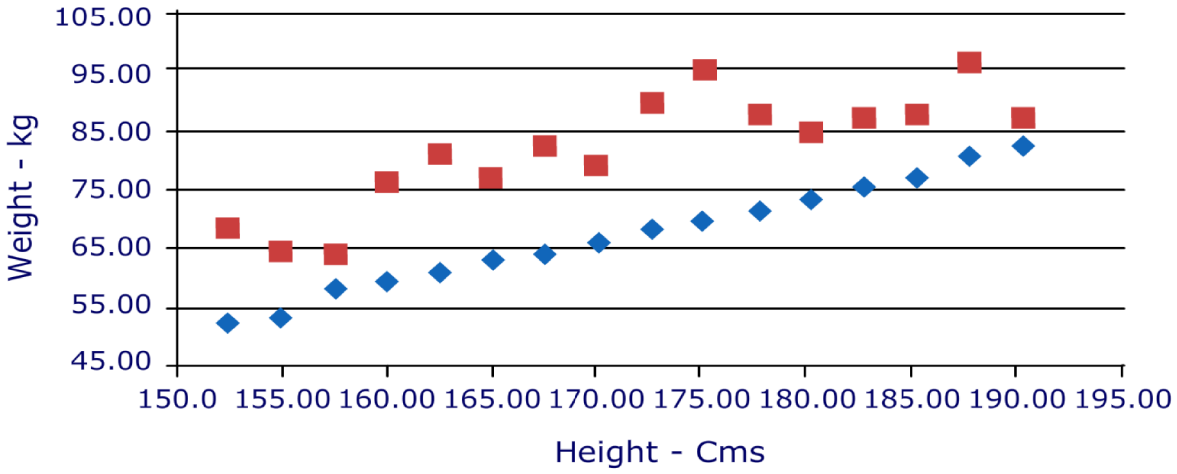
Figure 4

	Comparison	
Height (cms)	Weight(kg) of healthy men	Weight(kg) of affected men
152.30	52.6	68.38
154.80	53.55	64.26
157.40	58.3	64.13
159.90	59.65	76.352
162.40	61.2	80.782
165.00	63.05	76.921
167.50	64.45	82.496
170.00	66.25	79.5
172.60	68.3	89.473
175.10	70.1	95.336
177.70	71.9	87.718
180.20	73.7	84.755
182.70	75.75	87.1125
185.30	77	87.78
187.80	80.55	96.66
190.40	82.85	86.9925

Whereas comparison between weight of healthy men and weight of affected men is shown.

This scatter plot helps us to compare the trends easily.

Figure 5



The key finding here is that, people affected with the disease seem to be gaining weight in general (the dots for affected people are consistently above healthy people of same height). But the exact gain is not necessarily proportionate to their height (the dots for affected people are not sloping in any order). Weight is proportionate to height in case of healthy men; their weight is determined by their height.

But weight is not determined by height in case of affected men.

Now we will see some of the Key Consideration while building Scatterplot:

As you have already seen, two variables are used to build a scatter plot.

1. These variables should be related to the same event or context (otherwise the scatter plot becomes meaningless).

   There are a few other points to keep in mind, while building Scatterplot.

2. If one of the variables is systemically incremented and/or decremented that is called the independent variable (or the control parameter) and is typically plotted on the horizontal X-axis.

3. The other variable is the measured or dependent variable and is customarily plotted along the vertical Y-axis.

4. Multiple values of the depended variable can be plotted.

In the previous example, height is the independent variable whose confidence intervals have been determined in advance. Weight is the dependent (measured) variable.

The minimum or maximum or incremental values on both X-axis and Y-axis should be chosen such a way to avoid high-density clutter of points in the graph.

More often a scatter diagram is used to prove or disprove cause-and-effect relationships.

Some of the uses of Scatterplot are:

- Scatterplot are typically used to represent the correlation between variables so as to analyze the trends and causes. Through positive or negative correlation, it can also show that one variable is a surrogate of another

However it is very important to remember that a scatter plot pattern by itself does NOT prove that one variable causes the other (a third variable could be the cause).

Scatterplot can also help in determining the extreme values ("outliers") or boundary limits for the given variable combination

Scatterplot are used for a variety of purposes in our daily lives. Some examples are:

- Market surveys: For example: finding out the correlation between income and buying preferences
- Healthcare: For example: finding the correlation between age and diseases
- Economics: For example: impact of oil prices on consumer spending
- Behavioral analysis: For example: determining the relationship between age and divorces
- Environmental studies: For example: studying migration patterns of animals based on specific weather conditions

Scatterplot with very high number of reliable data points collected over long periods can predict future trends accurately. They are used for accurate forecasting and planning, by governments as well as business organizations.

3.Pie Chart and the construction of a Pie Chart

Let us now understand another one-dimensional diagrammatically representation of data using a circle i.e. Pie Chart concept.

Pie Chart is named for its resemblance to a pie, which has been sliced.

The pie chart is perhaps the most widely used statistical chart in the business world and the mass media. However, it has been criticized, and some even recommend avoiding it, pointing out in particular that it is difficult to compare different sections of a given pie chart, or to compare data across different pie charts.

Pie charts can be an effective way of displaying information in some cases, in particular if the intent is to compare the size of a slice with the whole pie, rather than comparing the slices among them.

Pie charts work particularly well when the slices represent 25 to 50% of the data, but in general, other plots such as the bar chart or the dot plot, or non-graphical methods such as tables, may be more adapted for representing certain information.

A pie chart is also called as a circle graph. It is a circular chart divided into sectors, illustrating proportion.

In a pie chart, the arc length of each sector and consequently its central angle and area, is proportional to the quantity it represents.

As explained, a pie chart is a circular chart divided into sectors, illustrating proportion. Let us now understand how a pie chart looks and how it can be constructed, through an example.

Let us look into the data in this table, which is based on the Government of India Survey during the year 1999-2000 about the Production of Food Grains in million tonnes over the last 50 years.

Figure 6

Year	Production of Food Grains (Million Tonnes)
1950-51	50.8
1960-61	82.0
1970-71	108.4
1080-81	129.6
1990-91	176.4

The data shows the Production of Food Grains in million tonnes from 1950-51 for every decade thereafter.

If we have to represent this data in a pie chart or a circular graph, we need to convert the Production units, into central angles and construct the pie.

Let us look into the second table, which shows us the percentage of the Production of Food grains, in the year mentioned in the table, out of the total food grains produced over the 5 years data, shown in the table. The total of the food grains produced over the 5 years is 547.2 Million tonnes. Within this total, 82 million tonnes or about 15% of the total was produced in the year 1960-61 as shown in this table.

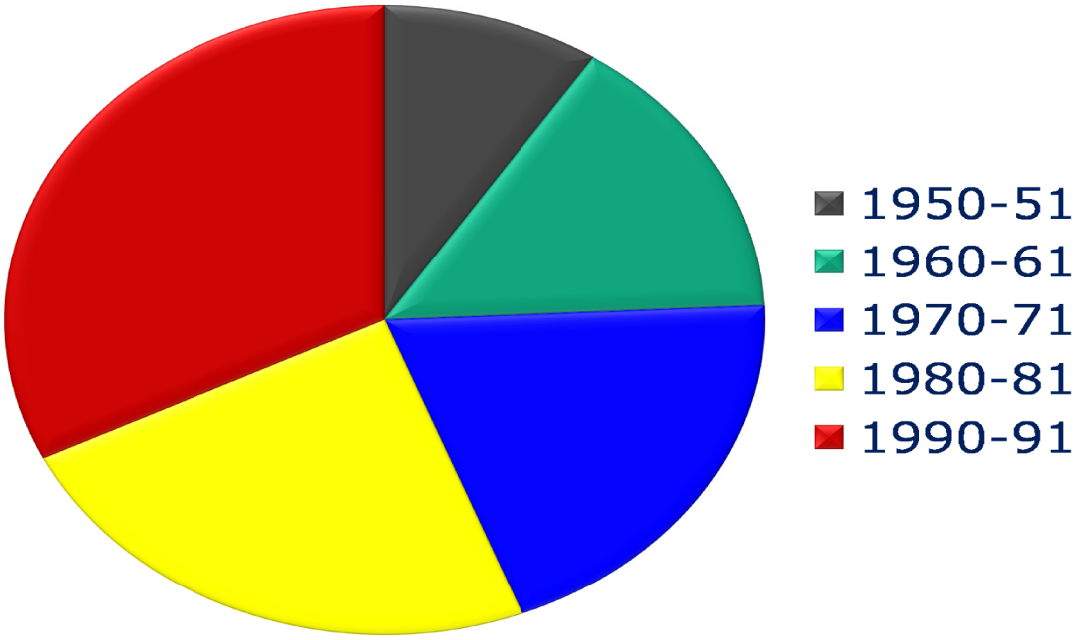
Now that we have the percentages of the Production of Food Grains for the five years, we will try to represent the same in the form of a Pie Chart. To do that, we will have to convert the percentages into 'central angles' to be able to construct the sectors of the circle, representing Production of grains in each of the 5 years.

Now, let us look into this table where, the last column data is computed by converting the percentages of the total share of 100%, into a proportionate 'central angle' of the total share of the circle's angle, which is 360 degrees. That is - the Central angle is calculated by the following formula – In 1950-51 if 9% of 100% is the Production of Food Grains, the same is converted to a central angle with the formula 9% multiplied 360 divided by 100%. This gives us a 'central angle' value of 32.4 degrees for the Year 1950-51. Similarly, we can calculate the 'central angle' for all the other 4 years. The total of the 'central angles' for the 5

years in the table should amount to 360 degrees. With this data, we can construct a circle with sectors broken down by the 'central angle' for each of the 5 years.

The pie chart representing the Production of Food Grains in Million Tonnes for 5 years looks like this. The sector for each year is shown in a different colour to distinguish from year to year. It gives a clear picture about the growth of the Food Grain production over the years.

Figure 7



The pie diagram is also known as an angular sector diagram, though in common usage the term pie chart or pie diagram is used.

We have seen how to represent data in a 'pie chart' format representing one category of data. We have also discussed how to calculate, convert the tabular data into the 'central angle' data in order to construct a 'pie chart'.

4. Example 1 of Pie Chart

Figure 8

Items	Expenditure as percent of total
Food	50
Clothing	15
Housing	10
Fuel and Lighting	5
Education	10
Recreation	5
Miscellaneous	5
Total	100

Let us look into a real life example of a family expenditure in a given month. Let us review the data in the table. The table depicts Items of expenditure and the percent of the total expenditure for each of the item. For example, Item 'Food' constitutes 50% of the total expenses of the month for the family. Similarly, Item 'Education' constitutes 10% of the total expenses of the month for the family. The total of the percentage constitution of each of the Item is equal to 100%.

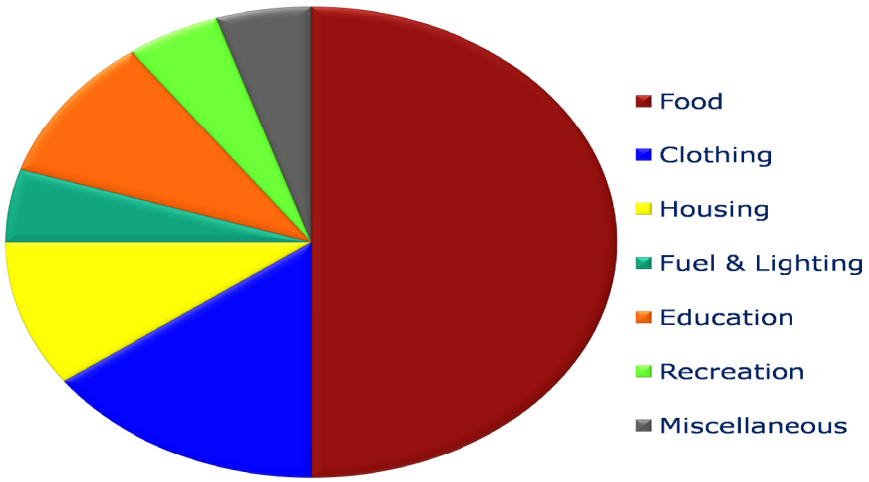
Figure 9

Items	Expenditure as percent of total	Central Angle
Food	50	"50/100*360=180 degrees
Clothing	15	"15/100*360=54 degrees
Housing	10	"10/100*360=36 degrees
Fuel and Lighting	5	"5/100*360=18 degrees
Education	10	"10/100*360=36 degrees
Recreation	5	"5/100*360=18 degrees
Miscellaneous	5	"5/100*360=18 degrees
Total	100	360 degrees

Now, we can calculate the 'central angle or the 'sector area' for these items to construct the pie chart. You can see the third column in the table depicting the same. For example, 'fuel and lighting' item has an expenditure of 5% of the total expenditure for the family. If this needs to be converted into a sector in a pie

chart, we should convert this as a percentage of the total pie, which is a percentage of 360 degrees. Therefore, for fuel and lighting, the angle will be 5/100 * 360, which is 18 degrees.

Figure 10



Similarly, the central angle can be calculated for all the expenditure items of the family. The total of all the central angle will be 360 degrees. Using these central angles, we can construct the pie chart for percentage expenditures of the family.

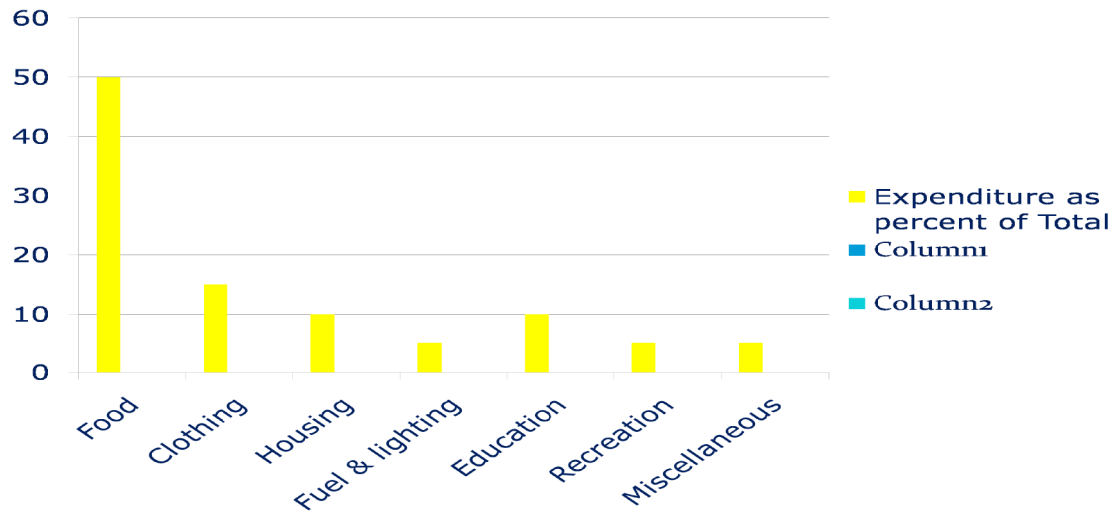
The pie chart for the family expenditure in a given month looks like this. We can notice the following in the pie chart:

- Different colours of the sectors in the pie indicate different expenditure item as shown in the legend

The sector angle percentage is shown within the sector, and that represents the percentage of expenditure of that item in relation to the total expenditure of 100 percent. We can notice that as the items or sectors get smaller, it becomes difficult for a user to comprehend the data. In addition, the number of such small items makes it challenging to read and interpret.

Now, let us contrast the pie chart you just now learnt with a bar chart for the same data.

Figure 11



As you can see, the bar chart has more clarity and can be easily interpreted than a pie chart. However, the depiction is a lot nicer in a pie chart.

While there are pros and cons for each of these type of charts but Bar chart or Bar diagram holds the edge in most situations.

5. Example 2 of Pie Chart

Let us take another example of a practical application of a pie chart in the field of cricket. A pie chart is used in television coverage of cricket to show the viewers the percentage of runs scored by a batsman in different areas of the cricket field. Here a circle is drawn to represent a cricket field and this is divided into various parts to depict areas of the cricket field like the offside, legside, behind the wicket, down the ground etc. If a batsman has scored 100 runs these hundred runs would have been scored by hitting the ball to various parts of the ground. The percentage of runs scored in different areas will be represented in a pie chart. Let us say 30 runs or 30 per cent of total runs are scored square on the leg side , 20 runs or 20 per cent of the runs are scored down the ground on the offside etc. This pie chart is commonly called the wagon wheel by cricket commentators. This type of chart can also be used to depict a teams score in an innings. This will clearly show a viewer which area of the ground, a team has scored its runs.

Pie charts are also very popular in business presentations. They are used to show market shares of competing products or also market share of a number of products of a single organization. Other commercial information like sales figures, profits etc, can also be represented in a attractive manner by pie charts. Mass media like TV Channels and Newspapers use pie charts to depict detailed data as it can be presented in a manner that is pleasing to the eye. Financial data of an organization can also be represented in pie charts showing elements of expenditure where the entire pie shows total expenditure or individual items of income where the pie represents total income.

Now, we can discuss the limitations of Pie chart.

- They are not as effective as bar diagrams for accurate reading and interpretation
- This limitation is even more profound, when there is large number of components in the overall pie or the differences among the components are too small
- It is confusing to differentiate the relative values of several small sectors having more or less the same size in the pie

Although, pie diagrams are used frequently, it is usually inferior to a bar diagram most times.
Here’s a summary of our learning in this session:

- Explain Scatterplot and the types of correlations in Scatterplot
- Appreciate the use of Scatterplot in real-life situations
- Explain Pie Chart and the construction of a pie chart
- Explain the examples of a pie chart
- List the limitations of a Pie charts