



SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.



An Autonomous Institution

COURSE NAME : 19CST203 - DATA ANALYTICS

II YEAR /IV SEMESTER

1.UNIVARIATE ANALYSIS

Univariate Analysis is a type of data visualization where we visualize only a single variable at a time. Univariate Analysis helps us to analyze the distribution of the variable present in the data so that we can perform further analysis

Univariate Analysis is a type of data visualization where we visualize only a single variable at a time. Univariate Analysis helps us to analyze the distribution of the variable present in the data so that we can perform further analysis

DATA VISUALIZATION:

Data visualization is the process of representing data using visual elements like charts, graphs, etc. that helps in deriving meaningful insights from the data. It is aimed at revealing the information behind the data and further aids the viewer in seeing the structure in the data.

Need for visualizing data :

- Understand the trends and patterns of data
- Analyze the frequency and other such characteristics of data
- Know the distribution of the variables in the data.
- Visualize the relationship that may exist between different variables

METHODS:

- Frequency Distribution Tables.
- Bar Charts.
- Histograms.
- Frequency Polygons.
- Pie Charts.

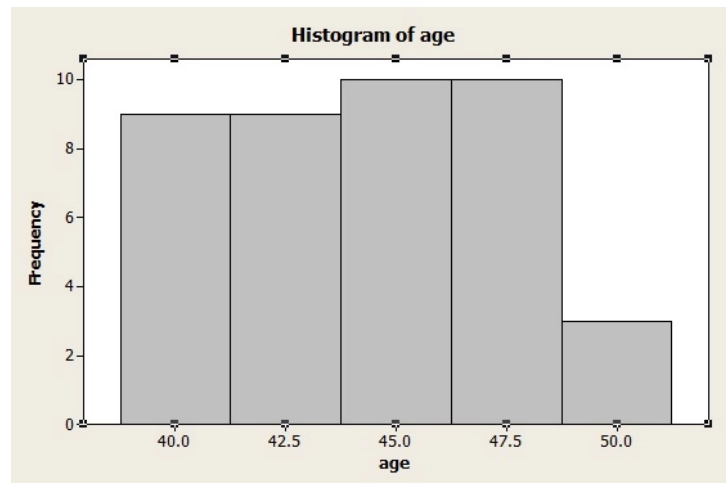
Frequency Distribution Tables

- *Frequency* tells you **how often something happened**. The frequency of an observation tells you the number of times the observation occurs in the data. For example, in the following list of numbers, the frequency of the number 9 is 5 (because it occurs 5 times):
- 1, 2, 3, 4, 6, 9, 9, 8, 5, 1, 1, 9, 9, 0, 6, 9.

- Tables can show either categorical variables (sometimes called qualitative variables) or quantitative variables (sometimes called numeric variables). You can think of categorical variables as categories (like eye color or brand of dog food) and quantitative variables as numbers.

Bar Charts.

A **bar chart** is a graph with rectangular bars. The graph usually compares different categories. Although the graphs can be plotted vertically (bars standing up) or horizontally (bars laying flat from left to right), the most usual type of bar graph is vertical.



The horizontal (x) axis represents the categories; The vertical (y) axis represents a value for those categories. In the graph below, the values are percentages.

Histograms.

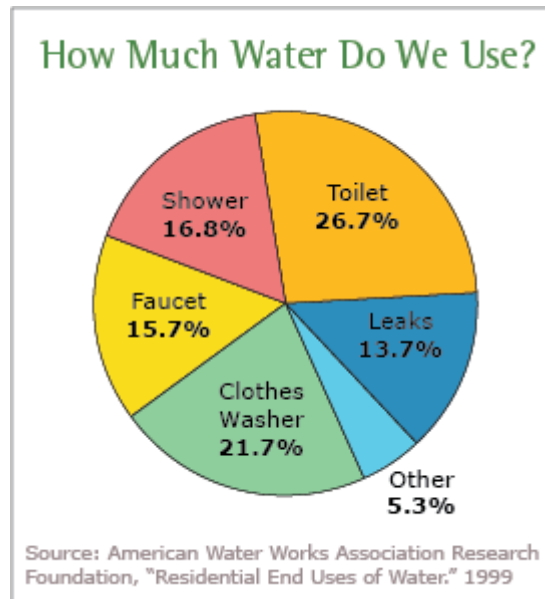
Histograms are similar to bar charts; they are a way to display counts of data. A bar graph charts actual counts against categories; The height of the bar indicates the number of items in that category. A histogram displays the same categorical variables in “bins”.

Frequency Polygons.

A frequency polygon is very similar to a histogram. In fact, they are almost identical except that frequency polygons can be used to compare sets of data or to display a cumulative frequency distribution. In addition, histograms tend to be rectangles while a frequency polygon resembles a line graph.

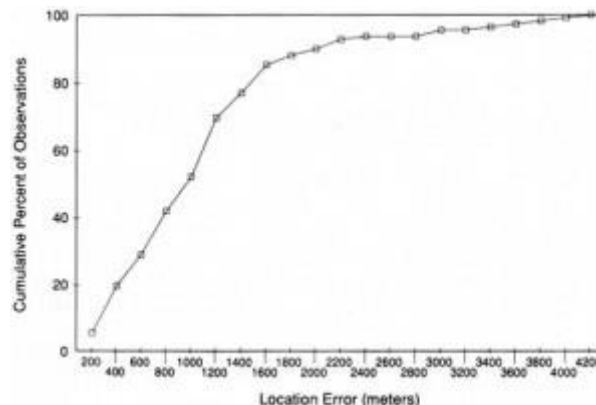
Pie Charts

A Pie Chart is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. In other words, **each slice of the pie is relative to the size of that category** in the group as a whole. The entire “pie” represents 100 percent of a whole, while the pie “slices” represent portions of the whole.



2.BIVARIATE ANALYSIS:

A Bivariate analysis will measure the correlations between the two variables. Multivariate analysis is a more complex form of statistical analysis technique and used when there are more than two variables in the data set. A doctor has collected data on cholesterol, blood



pressure, and weight.

TEMPERATURE(IN CELSIUS)	ICE CREAM SALES
20	2000
25	2500
35	5000
43	7800

Suppose the temperature and ice cream sales are the two variables of a bivariate data (figure 2). Here, the relationship is visible from the table that temperature and sales are directly proportional to each other and thus related because as the temperature increases, the sales also increase. Thus bivariate data analysis involves comparisons, relationships, causes and explanations. These variables are often plotted on X and Y axis on the graph for better understanding of data and one of these variables is independent while the other is dependent.

3. MULTIVARIATE DATA

When the data involves **three or more variables**, it is categorized under multivariate. Example of this type of data is suppose an advertiser wants to compare the popularity of four advertisements on a website, then their click rates could be measured for both men and women and relationships between variables can then be examined. It is similar to bivariate but contains more than one dependent variable. The ways to perform analysis on this data depends on the goals to be achieved. Some of the techniques are regression analysis, path analysis, factor analysis and multivariate analysis of variance (MANOVA). There are a lots of different tools, techniques and methods that can be used to conduct your analysis. You could use software libraries, visualization tools and statistic testing methods. However, this blog we will be compare Univariate, Bivariate and Multivariate analysis.

Univariate	Bivariate	Multivariate
It only summarize single variable at a time.	It only summarize two variables	It only summarize more than 2 variables.
It does not deal with causes and relationships.	It does deal with causes and relationships and analysis is done.	It does not deal with causes and relationships and analysis is done.

<p>It does not contain any dependent variable.</p>	<p>It does contain only one dependent variable.</p>	<p>It is similar to bivariate but it contains more than 2 variables.</p>
<p>The main purpose is to describe.</p>	<p>The main purpose is to explain.</p>	<p>The main purpose is to study the relationship among them.</p>
<p>The example of a univariate can be height.</p>	<p>The example of bivariate can be temperature and ice sales in summer vacation.</p>	<p>Example, Suppose an advertiser wants to compare the popularity of four advertisements on a website. Then their click rates could be measured for both men and women and relationships between variable can be examined</p>