Special Issue of First International Conference on Innovations in Engineering Sciences (ICIES 2020)

Introduction on Data Analysis and Graphical Representation of Data

T. Sukeerthi[1], K. Sukanya[2]

[1]T. Sukeerthi, Academic Consultant, Department of Statistics, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India Pincode:517501.
e-mail:tptsukeerthi2309@gmail.com
[2]K. Sukanya, Academic Consultant, Department of Statistics, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India Pincode:517501.
[2]sukanya.sudha27@gmail.com

Abstract

In everyday life, most of the members using spreadsheets for maintenance of their professional or research works. A small amount of statisticians or researchers using the Microsoft excel to do statistical computations and mathematical functions which are in-built. The implementation of this paper leads to guide the researchers or statisticians based on techniques used by professional data management teams to save their time and get better knowledge about data analysing. By doing themselves, they get correct and accurate results on their collection of data. Through Microsoft Excel, researchers would collect data, summarizes data, interpret data and perform specific techniques that can suits their work. The main aim of the study is to enrich their knowledge on data analysis and graphical representation in computational method.

Keywords: *Data analyzing, Data classification, MS Excel, graphical representation*

1. Introduction

Data can be defined as a systematic record of a particular quantity. It is the different values of that quantity represented together in a set. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. The source of data is also an important factor.

Fig.1. Data

1.1. Types of Data

Data Types are an important concept of statistics, which needs to be understood, apply statistical measurements to your data and therefore to correctly conclude certain assumptions about it.

Fig.2. Types of Data

Table.1. Data Types Table

S.no.	Data Types		Description	Example
1.	**Categorical data**	Nominal	It represent discrete units and label variables.	Gender, race
		Ordinal	Variables with an ordered series.	Blood group, performance
2.	**Numerical data**	Discrete	This type of data **can be counted**. It gives information that can be characterized.	No. of parts damaged in shipment.
		Continous	This data is information that **can be measured.**	Length, Size , width etc.
		Interval	Interval values represent **ordered units that have the same difference.**	Years, temperature and personality tests.
		Ratio	Ratio values are **the same as interval values, with the difference that they do have an absolute zero.**	Speed, salary, word count.

1.2 Data Collection

In Statistical studies, data collection is the first step. Data are collected by an investigator from personal experimental studies is called **primary data**. In this method, the investigator would interact directly with the respondent and gather an information. When data is obtained from some secondary source such as journals, magazines, reports etc., then it is called **secondary data**. (Sarma, K.V.S (2010)) [1].

1.3 Analyzing Data

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. This purpose is to extract useful information from data and taking the decision based upon the data analysis. There are several types of Data Analysis techniques that exist based on business and technology.

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. This purpose is to extract useful information from data and taking the decision based upon the data analysis. There are several types of Data Analysis techniques that exist based on business and technology.

Table.2. Types of Data Analysis Table

S.no	Types of Data Analysis	Explanation
1.	Text Analysis	Through this analysis, we can discover a model in large data sets using databases or data mining tools.
2.	Statistical Analysis	Data Collection, Analysis, interpretation, presentation, and

		modelling of data can be done through this statistical method. There are two categories - Descriptive Analysis and Inferential Analysis. (Anderson, T.W (1983)) [5].
3.	Diagnostic Analysis	This Analysis is useful to identify behaviour patterns of data.
4.	Predictive Analysis	This analysis makes calculation about future outcomes based on past or present data. It can accurate or estimate values based on our detailed information.
5.	Prescriptive Analysis	This analysis is last phase of business analytics. It predicts the questions, "what is likely to happen?"

1.4 Classification of Data

Connor defined classification as: "the process of arranging things 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". The classification is also called categorization of data and it can be broadly classified in to four types as follows:

Table.3. Types of Classification Table

S.no	Types of Classification	Description	Examples
1.	Qualitative	It describes and explains the data.	gender, colour, literacy, beauty
2.	Quantitative	It measures a basis of variable	height, weight etc
3.	Chronological	It measures on the basis of time	Weekly, monthly and yearly
4.	Geographical	It is based on geographical locations.	Village ,city and country

1.5 Tabulation of Data

The tabulation is also known as frequency distribution of variable. The main objective of tabulation is to reduce the data and to make comparison easy. In the tabular form of data, the required interpretation is easily reachable. The preparation of table depends upon size and nature of data. The data collected from field and recorded in the way it was collected would be raw data. Monthly reports of trade and commerce, food stocks and hospital statistics are some of the examples. These tabulated data may need further 'finishing' before interpreting them.
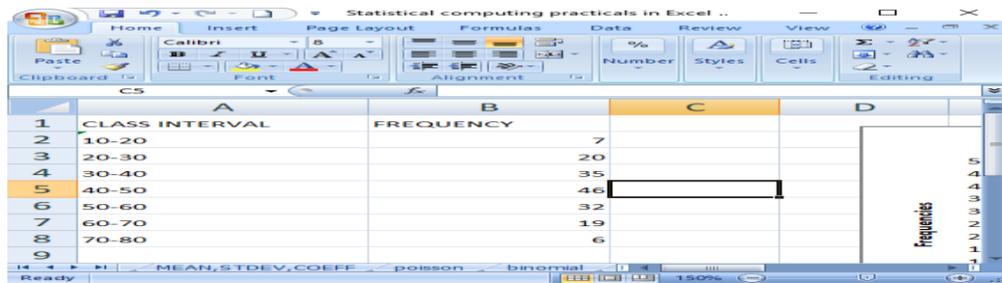
2. MS EXCEL:

Microsoft Excel is a spreadsheet program included in the Microsoft Office group of their applications. In MS Excel, there are some features like calculation, graphing tools, pivot tables and a macro programming language called Visual Basic for Applications. It also offers a suite of statistical analysis functions and other tools that can be used to run descriptive statistics and to perform several different statistical tests used in business and research.[1-4].

2.1 Data Entry

Data entry in the worksheet is an easy process. There is no need to create a structure as same as in access table. We can enter different types of fields in rows and columns with headings.(Russel A. Stultz (1997)) [4-7].

Fig.3. Process of Data entry

2.2 Graphical Representation of Data In MS Excel:

The data can be displayed with the half of graph and diagram instead of classification of tabulation. There are many reasons to draw a graph. The most important reasons is that one simple graph says more than words. Before preceding the formal statistical calculation, usually suggested that the graphical representation to understand the data in an easy way. The graphs gives us visualization of data. (Tukey JW 1990) [6]. Most frequently used graphs are Bar chart, Pie chart, Histogram and Scatter diagram.

2.2.1 Bar Chart:

In bar chart the values for each category is shown in vertical or horizontal bar format whose length is proportional to the value. If the bars are in vertical direction, then it is called a column chart / bar chart. Bar chart is used to represent the values of categorical variables like gender, age groups, social categories, etc.

To construct a bar chart in excel, first enter data in a worksheet with headings in two columns with a row (or) two rows and a column. Select all cells with headings. Click on 'Insert tab' and then click on column chart.

Fig.4. Dialogue box of chart types
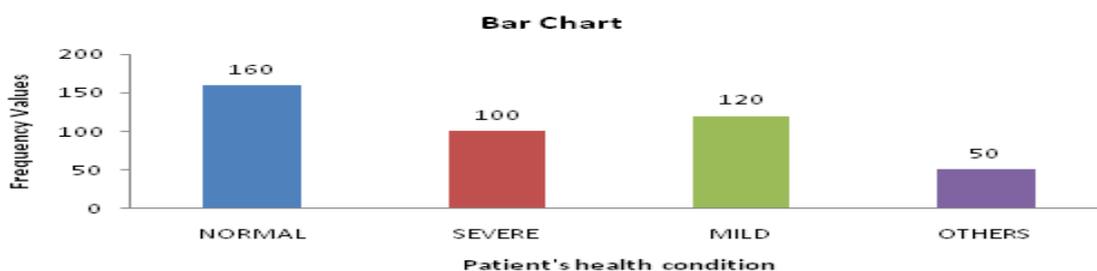
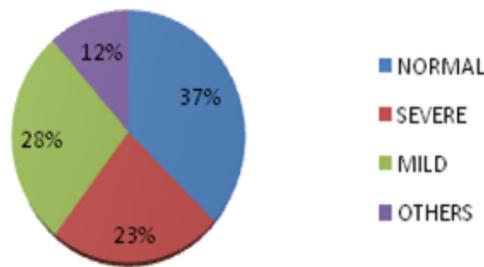**Fig.5. Bar Chart Diagram**

2.2.2 Pie Chart:

First enter the data in worksheet to construct a pie chart. Then select the data of the first series and choose pie chart in the insert tab of excel sheet. The pie chart will appears with different sectors as per data values.
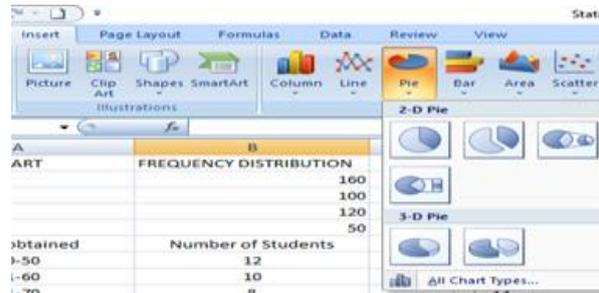
Fig.6. Dialogue box of Pie Diagram

Fig.7. Pie Chart on Patient's Health Condition.

A circle is divided into sectors having an area which is proportional to the frequencies or percentages of cases under various categories. A Pie chart is used to represent the distribution of categorical data as a percentage of the total values. This chart is popularly used to describe financial and economic data.

2.2.3 Histogram and Frequency Distribution in Excel:

It is a graphical presentation of frequency distribution in which variable characteristics are presented on the x-axis of the graph while the frequencies on y-axis. In histogram, each bins allows a continuation of values without a gap between the adjacent columns. The curve of histogram consists of a series of adjacent rectangles drawn for a grouped frequency.

A frequency distribution table in excel gives a detail note of how the data is spread out. It is an usual pair of frequency distribution table with a histogram. In order to make a frequency distribution table in Excel with a histogram. For this, data analysis tool pak must be installed. (Sarma, K.V.S (2010)) [1]. Enter the data into a worksheet into columns with headings, click on

"Data tab" ⟶ "Data Analysis"
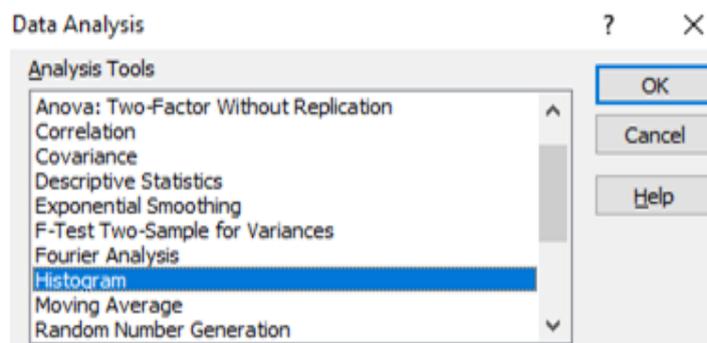
Fig.8. Dialogue box of Data Analysis

In this dialogue box click on histogram and then select a location where you want your output to appear in worksheet.
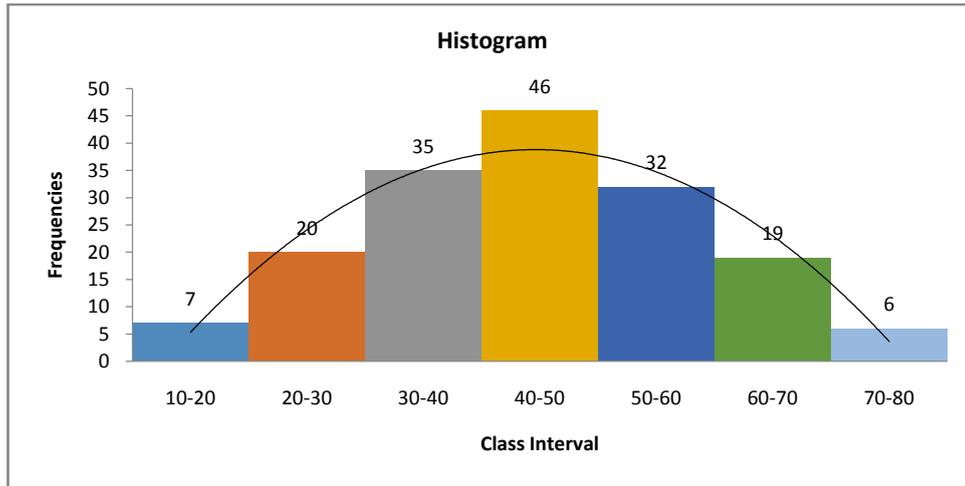

Fig.9. Histogram Chart and Frequency Polygon

2.2.4 Scatter Diagram:

Scatter diagram is the simplest method of studying relationship between variables. It shows direction of correlation between two variables. It is primary tool in correlation and regression studies. The graphical presentation in the form of dots. So, it is also called as dot diagram. It is easy to understand and easy to interpret the given data. Construction of this chart is similar as other charts. It can be created by choosing scatter diagram in chart options in excel sheet.
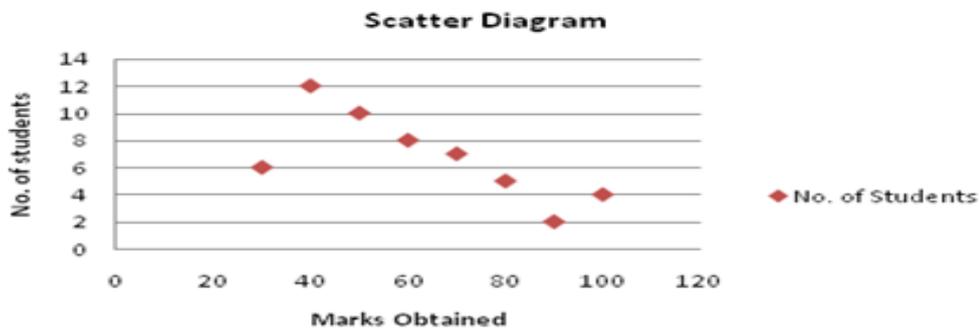
Fig.10. Scatter Diagram

References:

[1]. Sarma, K.V.S (2010), Statistics Made Simple, Do it Yourself on PC , Prentice Hall of India.

[2]. Microsoft Excel for Business Managers by Levin and Rubin.

[3]. Business Statistics using Excel by Davin and Pecker.

[4]. Russel A. Stultz (1997), Learn Microsoft Office 97, BPB Publications.

[5]. Anderson, T.W (1983), An introduction to Multivariate Statistical Analysis, Wiley, 2nd Edition.

[6]. Tukey JW. Data-based graphics: visual display in the decades to come. Stat Sci 1990; 5:327-39.

[7]. Rossman, A.J., Chance, B.L., & Locke, R. H. (2001) Workshop statistics: Discovery with data and Fathom. Emeryville, CA: Key Curriculum Publishing.