



ePathshala



ePathshala
पाठशाला

A Gateway to all Post Graduate Courses

An MHRD Project under its National Mission on Education through ICT (NME-ICT)

Subject: LAW

Production of Courseware

e-Content for Post Graduate Courses



Paper : Research Methodology

Module : Data Analysis



ज्ञान-विज्ञान विमुक्तये

A Gate



Role	Name	Affiliation
Principal Investigator	Prof. (Dr.) Ranbir Singh	Vice Chancellor, National Law University, Delhi
Co-Principal Investigator	Prof. (Dr.) G.S. Bajpai	Registrar, National Law University Delhi
Paper Coordinator	Prof. (Dr.) G.S. Bajpai	Registrar, National Law University Delhi
Content Writer/Author	Prof. (Dr.) G.S. Bajpai Ms Deepika Prakash	Registrar, National Law University Delhi National Law University Delhi
Content Reviewer	Prof. V.K.Srivastva	Department of Anthropology , University of Delhi

DESCRIPTION OF MODULE

Items	Description of Module
Subject Name	Law
Paper Name	Research Methodology
Module Name/Title	Data Analysis
Module Id	XIII
Objectives	To study the concept and method of analyzing data in a research
Key words	Data processing, tabulation, graphical representation, analysis , statistics, statistical software, interpretation

LEARNING OUTCOME

This module will elaborate on the meaning and utility of data analysis. It will provide a brief understanding of data processing, analysis and interpretation in the research process. The major focus of the module is to guide data analysis -how to plan, collect and manage the data so collected in a quantitative research for a meaningful research outcome.

1. OVERVIEW OF THE STEPS IN RESEARCH

In every research, following are the general steps involved:

- Defining problem
- Reviewing the available literature
- Formulation of hypothesis or research questions
- Creating a research design
- Collection of data with the help of various research tools
- Processing of the data collected
- Analysis and interpretation of the data
- Report writing



The present module seeks to understand how to handle the data which has been collected in the research process in order to come up with some concrete findings in a scientific and systematic manner. The data analysis in this module refers to data collected only in a quantitative study. In such a study numerical data which has been gathered by the researcher presents quantities and variables which have been collected using tools such as structured observations, questionnaire and various tests.

2. MEANING OF DATA ANALYSIS

In any research, the step of analysis of the data is one of the most crucial tasks requiring proficient knowledge to handle the data collected as per the pre decided research design of the project.

Analysis of data is defined by **Prof Wilkinson and Bhandarkar** as-

A number of closely related operations that are performed with the purpose of summarizing the collected data and organizing these in such a manner that they will yield answers to the research questions or suggest hypothesis or questions if no such questions or hypothesis had initiated the study.

According to **Goode , Barr and Scales**,

analysis is a process which enters into research in one form or another form the very beginning...It may be fair to say that research consists in general of two larger steps –the gathering of data, but no amount of analysis can validly extract from the data factors which are not present.

In his book on research methodology, **C. R. Kothari** explains that the term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. He quotes **G.B.Giles** to further elaborate the concept as “in the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions”

Hence, whether it is a qualitative or quantitative research even if the data is sufficient and valid, it will not serve any purpose unless it is carefully processed and scientifically analyzed and interpreted.

3. DIFFERENCE BETWEEN DATA ANALYSIS, PROCESSING AND INTERPRETATION

The general understanding is that data analysis and processing are one and the same. However a number of researchers and authors are of the opinion that both of them are two very distinct steps in the research process where data processing leads to data analysis. Lets us understand the difference between the two in more detail.

Prof. John Gauing is of the opinion that processing of data refers to concentrating, recasting and dealing

with the data so that they are as responsive to analysis, while analysis of data refers to seeing the data in the light of hypothesis of research questions and the prevailing



theories and drawing conclusions that are as amenable to theory formation as possible.¹

According to **Francis Rummel**, “the analysis and interpretation of data involve the objective material in the possession of the researcher and his subjective reaction and desires to derive from the data the inherent meaning in their relation to the problem. To avoid making conclusions or interpretations from insufficient or invalid data, the final analysis must be anticipated in detail when plans are being made for collecting information.

3.1 Data Processing

Once the data is collected, following steps are taken to process the data into more measurable and concise manner:

a. Editing

In the stage of editing all the raw data that is collected is checked for errors, omissions sometimes legibility and consistency as well. This ensure basic standard in the data collected and facilitate further processing.

b. Coding

Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also be exhaustive (i.e., there must be a class for every data item) and also that of mutual exclusively which means that a specific answer can be placed in one and only one cell in a given category set.² Coding can also be pre or post. Pre coding meaning codes being assigned while the questionnaire or interview schedule is being prepared. In the case of post coding, codes are assigned to the answers after they are collected.

c. Classification

Once the data is collected it is to be divided into homogeneous groups for further analysis on the basis of common characteristics.

d. Tabulation

Tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in columns and rows.

Tabulation is essential because of the following reasons-

1. It conserves space and reduces explanatory and descriptive statement to a minimum.
2. It facilitates the process of comparison.
3. It facilitates the summation of items and the detection of errors and omissions.
4. It provides the basis for various statistical computations.

¹ Dr. Y.K.Singh and Dr. R.B Bajpai, Research Methodology: Data Presentation, p 151, APH Publishing Corporation, ed 2012.

² C.R.Kothari, Research Methodology- Methods and Techniques, P 123, new age international limited publisher, 2nd ed.



Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, cost considerations, time pressures and the availability of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation if other factors are favorable and necessary facilities are available.³

Tabulation may be a very effective way of making legal research manageable, readable and understandable.

Types of table

There are generally two types of tables simple and complex. They are discussed following:

(i) Simple table/ frequency distribution

Under it, the different attribute are stated in the left hand column and the frequency or extend of occurrence of each of these classed are written in another column. In this three things are essential) the classes made must be mutually exclusive, b) the tabulation must have internal logic and order, and c) the class intervals must carefully and reasonably selected.⁴ Following is an illustration of the same.

Table 1- Univariate⁵

Age of the respondents	Frequency	Percentage
Below 10	14	10.8
11-20	18	13.8
21-30	22	16.9
31-40	42	32.3
41-50	26	20
Above 50	8	6.2
Total	130	100

In the above table the only variant is age.

(ii) Complex or cross table

³ Supra note 2.

⁴ Dr R. Kumar, *Methodology of social Science Research*, p223, Book Enclave, Jaipur, 2002.

⁵ R.Ahuja, *Research Methods*,



In a complex table, bi or multivariate are used. These have become more popular in the research representation in recent years. Following is an example of the same.

Table 2- Multivariate

Income (Rupees)	SEX				Total
	Male		Female		
	Rural	Urban	Rural	Urban	
Below 100	20	23	8	12	63
101-500	18	30	10	36	94
501-1000	10	28	5	21	64
Above 1000	5	15	2	14	36
Above 5000	2	10	0	8	20
Above10000	1	8	0	5	14

In the above table there are three variants i.e income, residence and sex are being studied and tabulated.

Preparation of a table

Following are certain guidelines to be kept in mind while preparing a table:

1. **Title of the table** - give suitable heading to each table which should be short and appropriate
2. **Sub headings and captions** - subheadings to different columns and rows must be given. Captions are given to the various classifications made like income, age, sex etc.
3. **Size of the column**- each column must have the correct size which make them look more attractive
4. **Arrangement of items in rows and columns** - items must be arranged in one order like alphabetically, chronologically etc.
5. **Totals** - the total for different columns must be different.
6. **Demarcation of columns** - If columns have been divided further into sub groups, they should be in an suitable order and sub headings
7. **Footnotes** - If there is anything special about the table or figures which need to be brought attention to, the same should be mentioned in a footnote.

3.2 Data Interpretation

Once the data has been processed and analyzed, the final step required in the research process is interpretation of the data. The line between analysis and interpretation is very thin. Through interpretation one understands what the given research findings really mean and what is the underlying generalization which is manifested through the data collected. This can be descriptive or analytical or theoretical. The data is



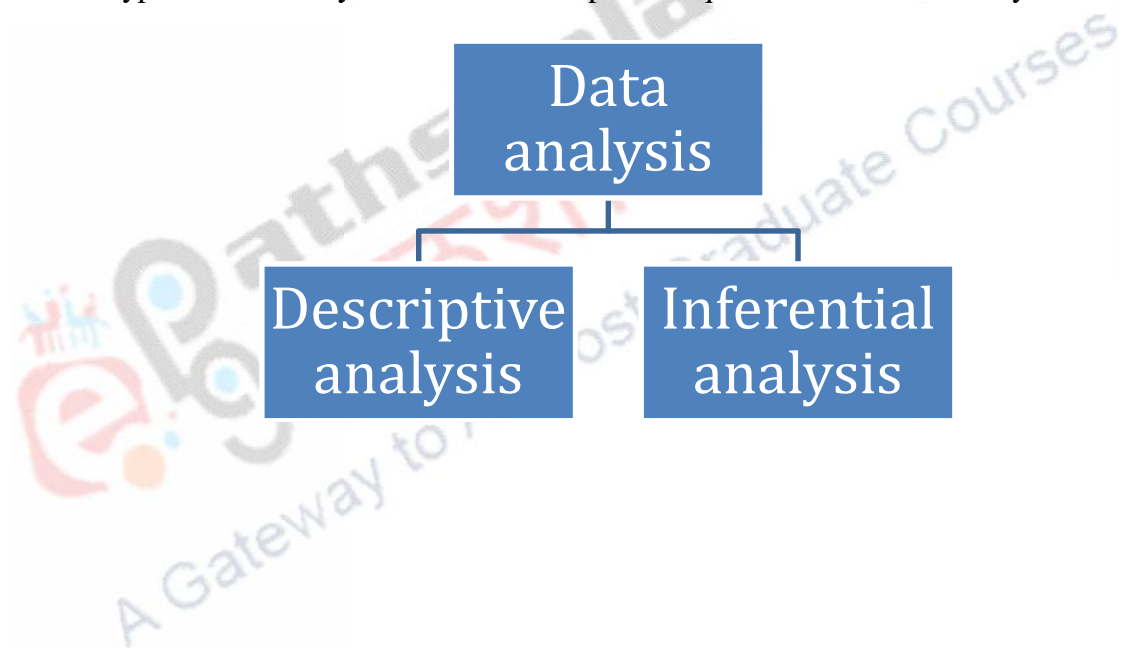
interpreted from the point of the research questions and hypothesis is tested. While interpretation is being done, generalizations are drawn. Thus, interpretation consists of conclusions that the researcher has reached after the data has been processed and analyzed.

It is interesting to mention that **Bloom's taxonomy** has laid down a structure on data presentation⁶:

1. Describe - Pen down the 'facts' observed/ heard after filtering the non relevant data.
2. Classify - Group the material based similarities, categorize, and make headings.
3. Interpret - identify important features and patterns in the light of the research questions or hypothesis and then represent them.

4. TYPES OF DATA ANALYSIS

Data analysis depends upon the nature of research that the researcher is undertaking. Types of data analysis vary depending upon whether the research is qualitative or quantitative in nature. In the present module, as earlier stated we will be studying various types of data analysis from the stand point of quantitative research only.



4.1 Descriptive analysis

According to **C Emory**, “*descriptive analysis* is largely the study of distribution of one variable. This study provides us with profiles of companies, work groups, persons and other subjects on any multiple characteristics such as size, composition, efficiency, preferences, etc.”⁷

Illustration: The researcher is collecting data from various law colleges in India to map the job preferences of the students in the final year of LL.B. In such a research job preferences like litigation, corporate, further studies, judiciary etc becomes the variable.

⁶ G. Guthrie, *Basic Research Methods an entry to social science research*, p 158, Sage publication , 3rd edition 2012

⁷ C. William Emory, *Business Research Methods*, p. 356.



Under its statistical tools like percentage and means are used and the data is then represented through a graph. The data analysis may be having one variable also known as one-dimensional analysis or two variables/ bivariate analysis or more than two variables also described as multivariate analysis.

4.2 Inferential analysis

Inferential analysis is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

Illustration:

The researcher is studying the access to justice system in India and his hypothesis beings that the India justice delivery system favors the haves and marginalizes the have not's. The data collected is from various stages in the delivery system like police station, courts of justice, litigants etc. Once the data is collected, proceeded then the researcher does inferential analysis to test the validity of the hypotheses.

5. GENERAL CHARACTERISTICS OF ANALYSIS OF THE DATA

1. The researcher should keep in mind that the analysis of data will vary depending upon the type of study i.e. qualitative or quantitative or mixed in nature.
2. The researcher should possess thorough knowledge of the area of research as well as the data collected by him which will help in the analysis of data.
3. The data to be analyzed and interpreted should:
 - a. Be reproducible,
 - b. Be readily disposed to quantitative treatment
 - c. Have significance for some systematic theory, and can serve as broad generalization.
4. The researcher should keep a clear set of hypothesis formulated at the very start of the research which will lead to clearer actions and better data collection as well as analysis.
5. In case the data collected is from vague clues rather than according to the specific hypothesis, in such cases the data are analyzed inductively or investigated during the process and not by means of any prescribed set of rules.
6. For a successful study, the task of analysis and interpretation should be designed before the data is actually collected.

6. STATISTICAL ANALYSIS OF DATA

Statistics is an important tool in the hands of a researcher for a good research. **Croxtton and Cowden**, two well known statisticians have introduced a simple, definition of statistics. In their words, "statistics may be defined as the science of collection, presenting and analysis and interpretation of numerical data."⁸

Statistics is not merely a device for collecting numerical data but also a means of sound techniques for their handling, analysis and drawing value inferences from them.

⁸ S Gupta, *Research Methodology and Statistical techniques*, p 200, Deep and Deep Publication, 2007.



When the data are collected, edited, classified, tabulated, it is analyzed and interpreted with the help of various statistical techniques and tools depending upon the nature of the investigation.

6.1 Uses of statistics

Statistics is useful in all fields of research and study. One of the greatest advantages of the use of statistics is that in a research with large data, it helps in reducing such data into a more manageable size for the purpose of analysis and interpretation. It also helps in comparing two or more series as well as draw inferences and conclusions of the research.

Illustration- The researcher is doing an impact analysis of the National Food Security Act, 2013 in the National Capital Territory. The universe of the researcher in such a case is Delhi, and the population is all the segments of people who are eligible for the food under the said Act. The tool of data collection chosen by the researcher is survey method. Once the data is collected, the size of the data would be big. Here, statistical tools would be of great assistance to the researcher to achieve his research objective.

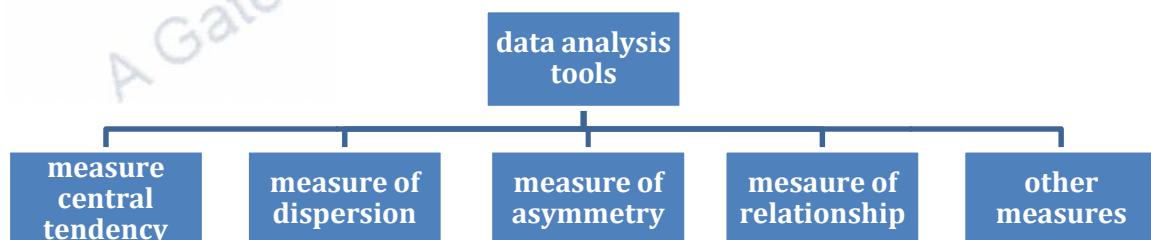
6.2 Limitations of statistics

Though statistical methods are of great value to a researcher, they carry with themselves certain limitations which must be kept in mind while deciding a tool of data analysis. They are:

1. Qualitative values like subjective perceptions, qualities and attributes are not considered under statistics. It only considers quantities. This by far is the greatest limitation of statistics.
2. Statistics studies and analysis group attributes rather than individual characteristics and values.
3. Statistical analysis is mostly based on average; hence the inferences drawn through them are only approximate and not exact like that of mathematics.
4. Statistics only help discover, analyze certain characteristics. It does not explain the picture. Hence, it only forms a part of the inference and interpretation.

6.3 Tools of statistical analysis

There are various statistical tools which are available for the researcher's assistance.



1. Measure central tendency

The term central tendency connotes the average. The most common central tendency tools are average or mean, median, mode, geometric mean and harmonic mean.

2. Measure of dispersion

The measure of dispersion or variability is the most common corrective measure for the concept of average. The most common method of the same is standard deviation. Others are mean deviation and range.

3. Measure of asymmetry



The tools used under it are skewness and kurtosis. Skewness is a measure that refers to the extent of symmetry or asymmetry in a distribution. It is used to describe the shape of a distribution. Kurtosis is a measure that indicates the degree to which a curve of a frequency distribution is peaked or flat-topped.

4. Measure of relationship

Correlation and coefficient is commonly used to measure the relationship. It is mostly used for prediction. Higher the degree of correlation, greater the accuracy with which one can predict a score. Karl Pearson's coefficient of correlation is the frequently used measure in case of statistics of variables, whereas Yule's coefficient of association is used in case of statistics of attributes. Multiple correlation coefficient, partial correlation coefficient, regression analysis, etc., are other important measures often used by a researcher.⁹

5. Other measures

Index number and analysis of time series are some of the other tools of data analysis.

Index numbers are indicators which reflect the relative changes in the level of a certain phenomenon in any given period called the current period with respect to its values in some other period called the base period selected primarily for this comparison.

Illustration: Index number is used to compare the changes in the national income of India from independence (1947) to the year 2014.

Analysis of time series

A time series is an arrangement of statistical data in accordance with its time of occurrence. If the values of a phenomenon are observed at different periods of time, the values so obtained will show appreciable variations.

6.4 Statistical software packages

To assist the researcher in quantitative data analysis, there are various statistic softwares available for computerized statistical data analysis. Some of them are available in the open source/ public domain i.e. free of cost while others are paid and purchased softwares. They are of great help when analyzing large quantities of data. The two most commonly used softwares are **SAS** (Statistical Analysis System) and **SPSS** (Statistical Package for Social Sciences).

7. ANALYSIS WHEN HYPOTHESIS EXISTS

When specific hypothesis has been set down, then the major part of analysis involves getting the appropriate combinations of data and reading them so as to verify or falsify the hypothesis. A hypothesis which is tested for possible rejection is known as 'null hypotheses. Null hypothesis is very much useful in testing the significant difference between assumed and observed values.

8. PRECAUTIONS IN ANALYSIS AND INTERPERTATION OF DATA

Following are some of the common precautions to be kept in mind while analyzing and interpreting the data:

1. Comprehensive knowledge and proper perspective

⁹ Supra note no 2.



The researcher while analyzing and interpreting the data must have thorough knowledge of the research from a wider perspective rather than analyzing the immediate element of the problem.

2. Take into account all pertinent elements

The researcher must keep all relevant factors/elements into consideration while analyzing and interpreting the data. Failure to do so will make the generalizations drawn inaccurate.

3. Limitations of the study

The researcher must mention all the limitations in the study like non-representation in sampling, bias in the data, inadequacy in the design, inaccurate statistical analysis etc .

4. Proper evaluation of data

Suitable interpretation of data lies on proper evaluation of facts. The researcher must interpret and analyze the data thoroughly himself for better results.

9. DIAGRAMMATIC REPRESENTATION

A very convenient and appealing method of data representation is by using various forms of diagrams. They in a very meaningful way highlight the salient features of the data which makes them easy to understand. Following are examples of some of the diagrammatic representations that may be employed in the research report. It may be noted that all the diagrams are fictitious and made only for illustrative purpose here:

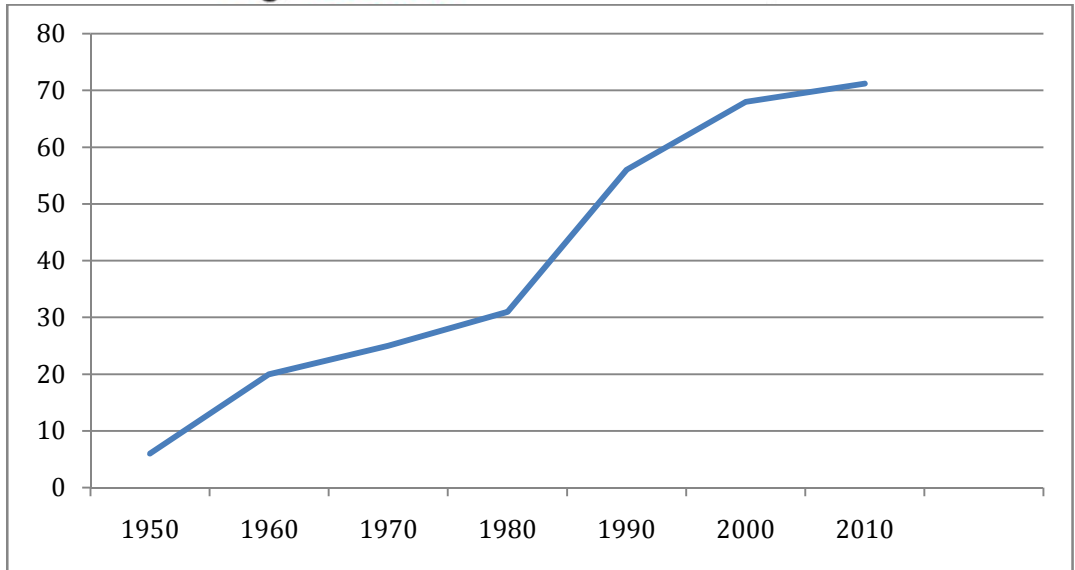
a) Graph

In a graph there are two axis the X and Y axis. X axis is horizontal and the Y axis is vertical intersecting the X axis. The point where intersection occurs is the place of origin. The independent variables are scaled on the X axis and the dependent one on the Y axis.

Following is an illustration of the same.

In the graph the growth of female literacy in India since independence has been shown. The X axis has the years while the Y axis has the rate of growth of women literacy in India.

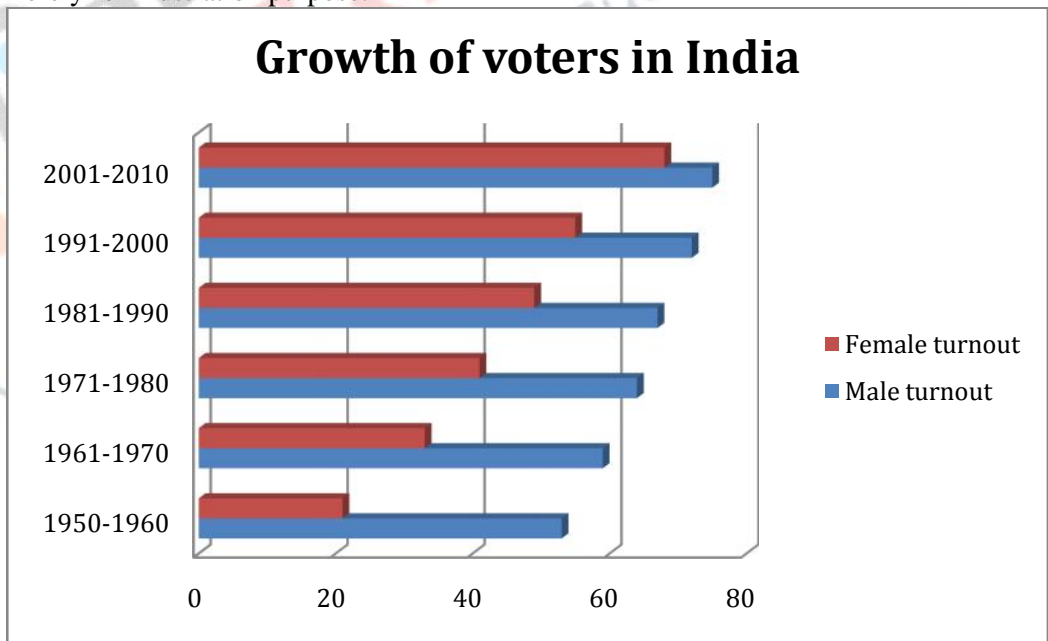
Graph



b) Bar diagram

The bar diagrams are drawn either vertically or horizontally. Each bar indicates the value of the variable.

Illustration- The following bar diagram shows by way of example what was the voters turn out till the year 2010 general election in the state of Delhi. The data is merely for illustration purpose.

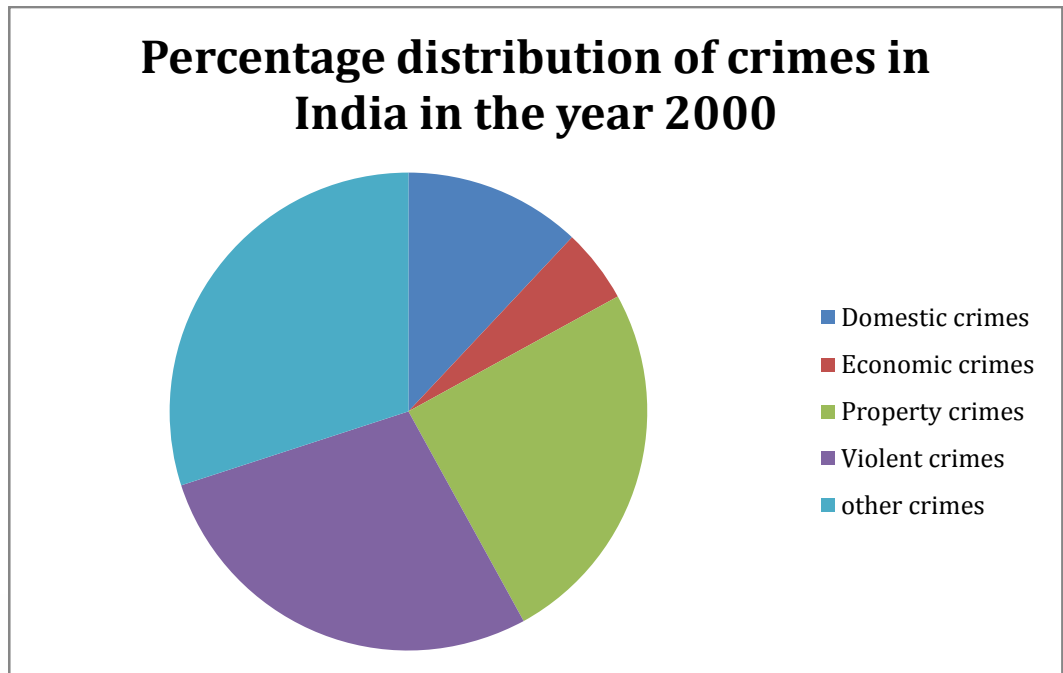


c) Pie chart

In a pie chart, the data is presented in the form of a circle with each category occupying a segment that is proportional according to the size of its data.



Following is an illustration of the same:



10. CONCLUSION

In the research process, data analysis is a very important and scientific step especially when the researcher is conducting a quantitative research. The researcher must understand the research area comprehensively and do the processing, analysis and finally interpretation with the help of various techniques and tools of analysis depending upon the nature, scope and aims of the research being conducted.