RESEARCH METHODOLOGY AND IPR (SHS04)

Unit-II

SURVEY OF LITERATURE

After defining a research problem, the researcher has to do literature survey connected with the problem. It is a collection of research publications, books and other documents related to the defined problem. It is very essential to know whether the defined problem has already been solved, status of the problem and other related details one can survey

1. The journal which publishes abstract of papers published in various journals.
2. Review articles related to the topic chosen.
3. Journals which publishes research articles
4. Advanced level books on chosen topic.
5. Proceeding of conference.
6. Internet
7. Reprint collections available with the supervisor and near by experts working on the chosen topic.

IMPORTANCE OF LITERATURE SURVEY

Once the research problem is defined, the next logical step will be survey of background literature. This is essential whether the proposed problem deals only with theoretical or with experimental aspects or both. All available literature uncerning the problem at hand must necessarily be surveyed and examined. The researcher must be well conversant with concerned theories in the field, reports and records and all other literature.

The literature survey is mainly to find out what material and other data are available for operational purposes. If somebody has done experimental study or simulation, he/she might have mentioned the problem and difficulties failed by them in their study, which may be useful for the current study. Knowing what data are available aften serves to narrow down the problem itself.

Literature survey would also help a researcher to know if there are any gaps in the theories, or whether the existing theories applicable to the problem under study. All these will enable the researcher to further farm the existing.

SOURCES OF INFORMATION

The sources of information can be classified as

1.PUBLIC SOURCES

I.Central government departments [defence, energy, science and technology etc.,]

II.State and local government [highways, pollution control board etc.,]

III. Libraries

IV. Universities

V. Internet

2. PRIVATE SOURECES

I. Nonprofit organizations and services [professional societies, trade unociations, membership organization]

II. Profit – oriented organization [manufacturers, vendors, catalogues, test data etc, consultants]

III. Induvial [direct conversation or correspondence, personal friends, faculty]

Depending upon the nature of the problem under taken for the research study, the researcher has to seek the needed information from different sources. However, the major sources of information nowadays are library, there is a hierarchy of information sources as given below

- Technical dictionaries
- Encyclopaedias
- Handbooks
- Indexing and abstract services
- Technical and professional journals
- Translations
- Technical reports
- Books
- Patents

The various sources from which information is gathered can also be grouped into

1. PRIMARY SOURCES

2. SECONDARY SOURCES

Primary sources:

These sources provide original information. They contain raw information e.g., data gather out of an experimental study

- Papers of original authors
- First published documentation of new information
- Government documents, patents, photographs, audio recordings, video recordings, films, speeches, published books etc.,

Secondary sources:

These are the generalizations, analysis, interpretation and of primary sources.

e.g., Documentary about the historical events journal's articles, dictionaries, books that interpret analyse, political commentary, biographies.

ASSESMENT OF QUALITY OF JOURNALS AND ARTICLES

The quality of a journals can be assessed by using "IMAPCT FACTOR". The impact factor of a journal is calculated by dividing the references cited in one year by the number of citable articles published in the same journal over the previous 2 years. This ratio is published annually in Thomson scientific journal citation reports (JCR). There are many databases or perish.

The scientific journals citation report (CJR)impact factors are all based on data from journals indexed in web of science, but SCOPUS uses a measure called 'h-index", the "h-index" is an index that attempts to measure both the scientific productivity and the apparent impact of scientists. This is based on the set of the scientists most cited papers and the number of citations that they have received in other people's publication.

[the calculation of h-index can be explained as follows. A scientist has index 'h' if 'h' of his papers has at least h citations each and other [N-h] papers have almost h citation each.]

Science citation index (SCI) is a citation index originally produced by the institute for scientific information (ISI) in 1960 and now owned by Thomson reuters. It covers more than 6500 journals across more than 150 disciplines.

Quality of its journal is judged by the impact factor of the journal, the quality of an article by the number of citations, h-index etc.,

INFORMATION THROUGH INTERNET [WEB AS A SOURCES]

From the past one decade or so the internet. Internet became an important source of knowledge and an effective medium for research. For researchers, it is providing a range of new opportunities for collecting information, networking, conducting research, collecting data and disseminating research results.

Electronic mail, e journal, online submission of articles to journals, online focus groups, online video conference and online questionary or some of the latest tools opened up by the internet.

The most recent and rapidly growing components of the internet is the world wide web [www].

The popularity of www comes from the fact that it makes distribution and accessing digital information simple and inexpensive.

To search the world wide web, we need a search engine usually refers to a web search engine. There are other kind of search engines such as enterprise such engines which searches on intranets, personal search engines, mobile search engines etc., all search engines will not produce the same result for the specific enquiry.

The researcher should have knowledge of many URLS devoted to engineering topics such as:

- NASA technical information service -http://tech reports.Nasa.gov/egi-bin/NTRs
- National technical information service – http://ntis.gov

Some of the scientific research information available on internet is

- Titles and other relevant information of articles published in various journals
- Preprit of papers submitted by the researchers in certain websites
- Information about scientific meetings
- Contact details of other researches
- Database of refence materials etc.,

EFFECTIVE LITERATURE STUDIES APPROACHES

Effective literature studies approach in research methods involve a specialized focus on legal, ethical and research methodologies.

Some of the approaches of literature are

1.UNDERSTANDING THE LEGAL FRAME WORK:

Begin by studying fundamental legal frame works related to intellectual property rights. This includes patents, copyrights, trademarks, and trade secrets understand now these rights are granted, enforced, and protected.

2.DEFINE THE SCOPE:

Clearly define the scope of your literature study. Determine specific aspects of RM and IPR you want to explore, such as research funding, project management, patents laws, copyright issues, etc.

3.IDENTIFY KEY CONCEPTS AND KEYWORDS:

Develop a list of key concepts and keywords related to your research. Include terms specific to RM (e.g. patents, copyrights, trademarks).

4.SEARCH STRATEGY:

Design a comprehensive search strategy using Boolean Operators (AND, OR to combine and refine your key words)

5.USE RELEVANT DATA BASE:

Utilize academic database, institutional repositories, and specialized platforms for RM and IPR Research. Examples includes PubMed, IEE Explore, science direct, Google scholar, and legal databases like west law and Lexis Nexis.

6.REVIEW KEY JOURNALS AND CONFERENCES:

Identify reputable journals and conferences in the fields of RM and IPR. Focus on publications that are known for high quality research in these areas.

7.INCORPORATE GREY LITERATURE:

Include grey literature such as reports, thesis, conference proceedings, and working papers.
This can provide valuable insights and perspectives not found in traditional academic journals.

8.SYSTEMATIC REVIEW METHODOLOGY:

Consider using systematic review methodologies, especially if your goal is to synthesize existing knowledge comprehensively. This involves a structured and transparent approach to literature selection and analysis.

9.CITATION ANALYSIS:

Analyse citations in relevant papers to identify seminal works and influential authors. This can help you understand the evolution of ideas and key contribution to the field.

10.STAY UPDATED:

Setup alerts and notifications for newly published articles and updates in RM and IPR. Subscribe to relevant journals and newsletters.

PLAGIARISM:

"The process or practice of using another person's ideas or work and pretending that is your own".

"Presenting work or idea from another source as your own with or without consent of the original other by incorporating it into your work without full acknowledgement".

Plagiarism means using someone else work without giving them proper credit . in academic writing, plagiarizing involve using words, ideas or information from a source without citing it correctly.

Examples:

- Copying parts of a textword for word, without quotation marks.
- Turning in someone else work as your own.
- Copying large pieces of a text from a source without citing that source etc.

Types of plagiarism:

Plagiarism comes in various forms, each differing in intent, severity and technique. Some of the common types are.

1. **Direct plagiarism**: Copying someone else work word for word without giving credit or using quotation marks.

2. **Self plagiarism**: reusing your own previously published work without proper citation or permission. This is often happens when a person submits the some paper or project for different classes or publications.

3. **Mosaic plagiarism:(also known as patchwriting):** Borrowing phrases or pieces text from different sources without proper attribution blending them with one's own writing. The structure the original text might to be changed slightly ,but the language is still essentially same.

4. **Paraphasing plagiarism:** Re-writing someone else's ideas or text in your own words without citing the original source

5. **Accidental plagiarism:** Unintentionally failing to cite sources properly due to carelessness.
6. **Source-based plagiarism:** Misrepresenting or fabricating a source. This includes citing incorrect or non-existent source or proving details about a source.
7. **Global plagiarism:** Submitting an entire work(like an essay or research paper) that is copied from another source such a website or purchasing the another source such a website or purchasing it from a 3rd party and presenting an one's own.
8. **Retweet plagiarism:** Using an author's entire argument or idea in the same sequence or structure but rephrased without citing the original source.

Consequences of plagiarism

Plagiarism can have serious consequences depending on the context, whether in academia, the work place, or creative fields. Here are some common consequences.

1. **Academic consequence:**

- Falling grade
- Disciplinary action
- Revocation of degrees
- Damage to academic reputation

2. **Professional consequences:**

- Job termination
- Damage to career and reputation
- Legal consequences

3. **Legal consequences:**

- Copyright infrightment
- Financial damages

4. **Reputational damages:**

- Loss of trust
- Public shame

5. **Loss of professional licenses:**

- In certain fields like journalism or law, plagiarism can lead to relocation of licenses or certifications.

6. **Hindrance to learning and development:**

- Plagiarism deprives individuals of the learning process, resulting in a lack of skills and knowledge that can hinder future academic or professional growth.

Research ethics:

It refers to the guidelines and principles that govern the conduct of research to ensure that it is conducted with integrity, respect and responsibility.

Key principles of research ethics:

1. **Respect the persons:-**

- *Informed consent*: - Researches must enure that participants voluntarily agree to participate in the study with a full understanding of its purpose procedure, risks and benefits.
- *Autonomy:*- participants should have the right to make their own decisions about participating in research.

2. **Beneficence:-**

- *Maximizing benefits, minimizing harm:*-Researches should ensure that it potential benefits of its research outweigh the risk or harm to participants.

3. **Justice:-**

- *Fair treatment:*- The benefits and burdens of research should be distributed fairly among all groups.
- *Equitable selection of participants:*- Researcher's must ensure that participant recruitment is done fairly and withot bias.

4. **Confidentiality:-** The researcher must ensure that the participants data is handled confidentially.

5. **Integrity and transparency:**-Researcher's must avoid fabrication, plagiarism, or any form of misconduct.

All aspects of the research process from methodology to reporting should be clear and transparent to ensure trust worthiness.

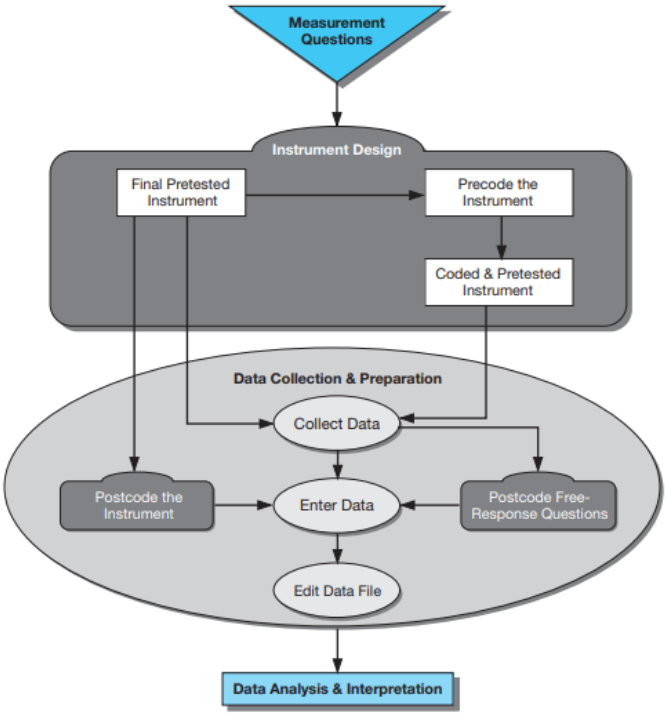
6. **Social responsibility:-** Researcher's should ensure that their work contribute positively to society.

7. **Accountability:**- Researcher's are accountable to their institutions, funding agencies, and the public.

Data – preparing, exploring, examining and displaying:

Data preparation :-

Data preparation includes editing , coding and data entry and is the activity that ensures the accuracy of the data and their conversion from raw form to reduced and classified forms that are more appropriate for analysis.



The above figure reflects the steps in preparing the data in the research process.

Editing:- First step is to edit the raw data. Editing detects errors and omissions, corrects them when possible and certifies that maximum data quality standards are achieved. The editor's purpose is to guarantee that the data are

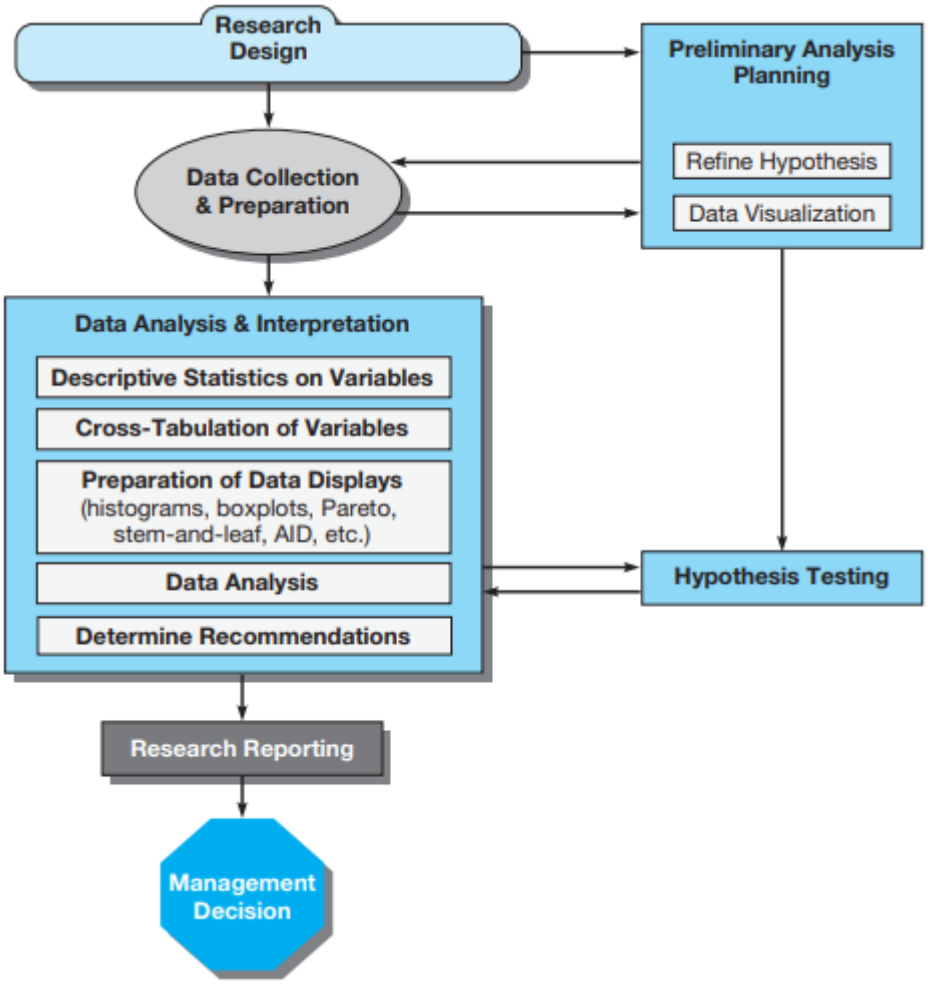
- Accurate
- Uniformly entered
- Complete
- Arranged to simplicity coding and tabulation.

Coding:- It involves assigning number or other symbols to answers so that the response can be grouped into a limited number of categories. In coding, categories are the partition of a data set of a given variable. [e.g., If the variable is gender, the partitions are male and female].

Categorizations of data sacrifices some data detail but it is necessary for an efficient analysis.

Data entry:- It converts information gathered secondary or primary methods to a medium for viewing and manipulation. Keyboarding is mainly used to create data file and store it in a minimal space on a variety of media. Researcher have profited from more efficient ways of speeding up research process, like barcoding and optical character and mark recognition.

Exploratory data analysis (EDA)



It provides a perspective and set of tools of search for clues and patterns in the data in addition to numerical summaries of location, spread and shape. EDA emphasizes on visual representation and graphical techniques to provide a complete and accurate impression of distributions and variable relationships.

Confirmatory data analysis is an analytical process guided by classical statical inference in its use of significance testing and confidence

EDA is like police defective searching for clues as it gathers evidence needed for confirmatory data analysis, uses classical statistics to test significance and confidence.

Confirmatory data analysis to an analytical process guided by classical statistical inference in its use of significance testing and confidence.

Exploratory data analysis is the first step in search for evidence, without which confirmatory analysis has nothing to evaluate.

Techniques for displaying data

(i) A Frequency Table (Minimum Age for Social Networking)

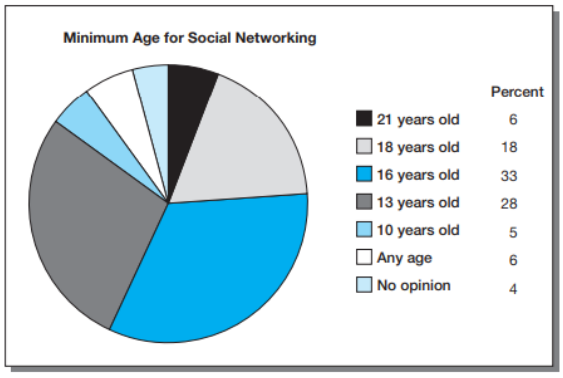
Value Label	Value	Frequency	Percent	Valid Percent	Cumulative Percent
21 years old	1	60	6	6	6
18 years old min	2	180	18	18	24
16 years old min	3	330	33	33	57
13 years old min	4	280	28	28	85
10 years old min	5	50	5	5	90
Any age	6	60	6	6	96
No opinion	7	40	4	4	100
		1,000	100	100	

Valid Cases 1,000; Missing Cases 0

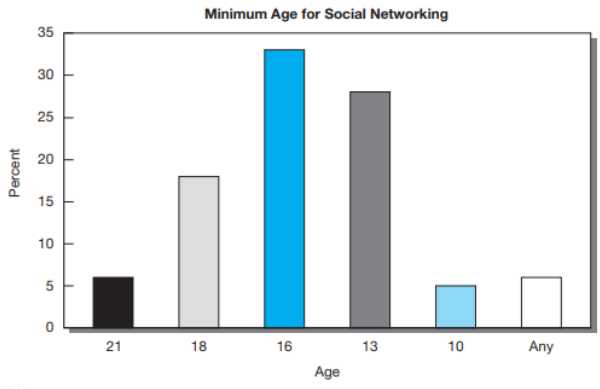
Frequency tables They array data from lowest to highest values with counts and percentages. They are most useful for inspecting the range of responses and their repeated occurrence. Ex: Figure shows a frequency table of the perceived minimum age for owning a social networking account.

(ii) **Bar charts and pie charts**

They are suitable for relative comparisons of nominal data. Ex: The same data of minimum age for social networking are presented in figure using a pie chart and a bar chart. The values and percentages are more readily understood in this graphic format.



Pie chart



Bar chart

(iii) **Histograms**

They are used with continuous variables when it is possible to group the variable's values into intervals. They are useful for (1) displaying all intervals in a distribution, even those without observed values, and (2) examining the shape of the distribution for skewness, kurtosis, and the modal pattern. Histograms can help answer the questions: Is there a single mode? Are subgroups identifiable when multiple modes are present? Are straggling data values (Outliers) detached from the central concentration?

Note: A histogram cannot be used for a nominal variable like minimum age for social networking (Figure) that has no order to its categories.

Ex: Figure(a) gives average annual purchases and figure(b) presents a histogram for same. The midpoint for each interval for the variable of interest (average annual purchases) is shown on the horizontal axis; the frequency or number of observations in each interval, on the vertical axis. The height of the bar corresponds with the frequency of observations in the interval.

Value	Frequency	Percent	Cumulative Percent	Value	Frequency	Percent	Cumulative Percent
54.9	1	2	2	75.6	1	2	54
55.4	1	2	4	76.4	1	2	56
55.6	1	2	6	77.5	1	2	58
56.4	1	2	8	78.9	1	2	60
56.8	1	2	10	80.9	1	2	62
56.9	1	2	12	82.2	1	2	64
57.8	1	2	14	82.5	1	2	66
58.1	1	2	16	86.4	1	2	68
58.2	1	2	18	88.3	1	2	70
58.3	1	2	20	102.5	1	2	72
58.5	1	2	22	104.1	1	2	74
59.9	2	4	26	110.4	1	2	76
61.5	1	2	28	111.9	1	2	78
62.6	1	2	30	118.6	1	2	80
64.8	1	2	32	123.8	1	2	82
66.0	2	4	36	131.2	1	2	84
66.3	1	2	38	140.9	1	2	86
67.6	1	2	40	146.2	1	2	88
69.1	1	2	42	153.2	1	2	90
69.2	1	2	44	163.2	1	2	92
70.5	1	2	46	166.7	1	2	94
72.7	1	2	48	183.2	1	2	96
72.9	1	2	50	206.9	1	2	98
73.5	1	2	52	218.2	1	2	100
				Total	50	100	

Figure(a): Average Annual Purchases of PrimeSell's Top 50 Customers

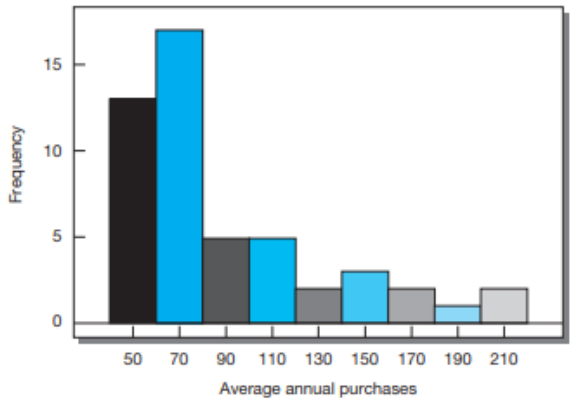


Figure (b): Histogram of PrimeSell's Top 50 Customers' Average Annual Purchases

(iv) Stem-and-leaf displays:

They are EDA techniques that provide visual representations of distributions. In contrast to histograms, which lose information by grouping data values into intervals, the stem-and-leaf presents actual data values that can be inspected directly. Adv1: EDA reveals the distribution of values within the interval and preserves their rank order for finding the median, quartiles, and other summary statistics. Adv2: It eases linking a specific observation back to the data file and to subject that produced it.

Adv3: Visualization: The range of values and both shape and spread are apparent at a glance. Patterns in the data—such as gaps where no values exist, areas where values are clustered, or outlying values that differ from the main body of the data—are easily observed. Ex: To develop a stem-and-leaf display for the data given in figure, the first digits of each data item are arranged to the left of a vertical line. Next, pass through the values in the order they were recorded and place the last digit, in unit position, for each item to the right of the vertical line (ignore the decimal values). The stem-and-leaf display is shown in figure.

5	4 5 5 6 6 6 7 8 8 8 8 9
6	1 2 4 6 6 7 9 9
7	0 2 2 3 5 6 7 8
8	0 2 2 6 8
9	
10	2 4
11	0 1 8
12	3
13	1
14	0 6
15	3
16	3 6
17	
18	3
19	
20	6
21	8

v. Pareto diagrams :

Pareto diagrams are named after a 19th century Italian economist and are used in quality management. To illustrate the 80/20 rule, where 80% of quality improvements can come from addressing 20% of the causes. In these diagrams, responses are sorted by importance, with bars arranged from tallest to shortest. Ex: Figure shows MindWriter customer complaints, highlighting that the top two issues accounted for 80% of perceptions of inadequate repair service.

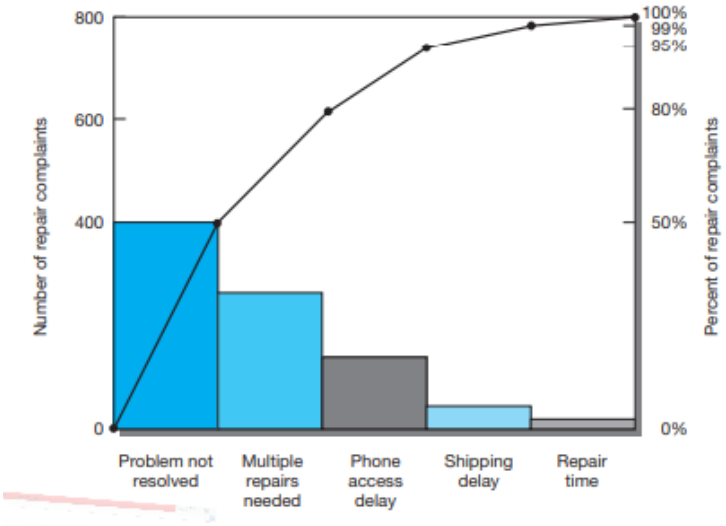


Figure: Pareto Diagram of MindWriter Repair Complaints

vi. Boxplots:

Boxplots convey a detailed picture of the distribution's location, spread, shape, tail length, and outliers. They use the five-number summary that consists of the median, the upper and lower quartiles, and the largest and smallest observations.

The basic ingredients of the plot are:

1. The rectangular plot that encompasses 50 % of the data values.
2. A center line marking the median.
3. The edges of the box, called hinges.

4. The "whiskers" that extend from the right and left hinges to the largest and smallest values. These values may be found within 1.5 times the interquartile range (IQR) from either edge of the box. These components and their relationships are shown in Figure.

Outliers, data points that exceed +1.5 the interquartile range, reflect unusual cases and are an important source of information for the study. Outliers that are entry mistakes should be corrected or removed during editing. Ex: In figure, multiple boxplots compare five sectors of PrimeSell's customers by their average annual purchases data.

The overall impression is one of potential problems for the analyst: unequal variances, skewness, and extreme outliers. Note the similarities of the profiles of finance and retailing in contrast to the high-tech and insurance sectors.

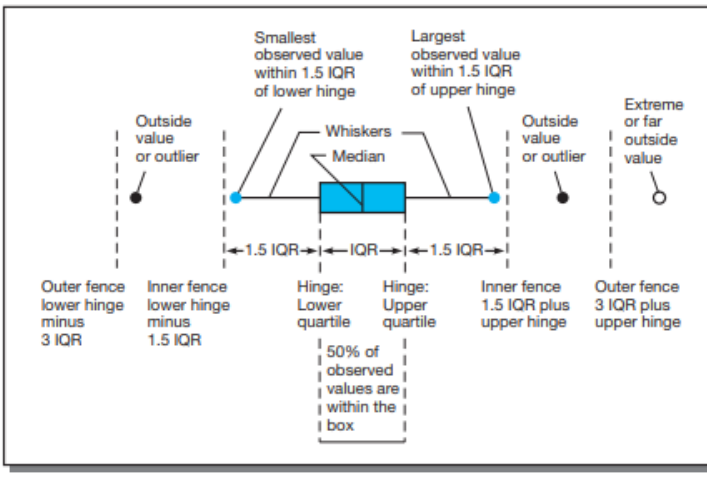


Figure: Boxplot Components

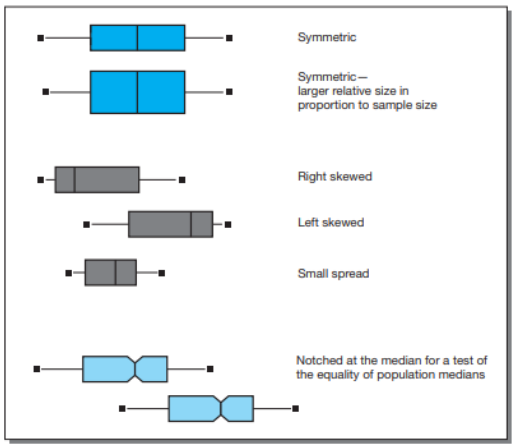


Figure: Diagnostics with Boxplots

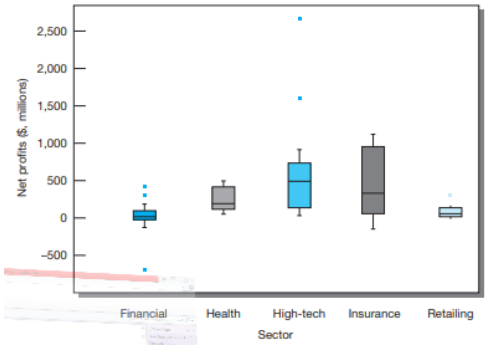


Figure: Boxplot Comparison of Customer Sectors

vii. Mapping:

Geographic Information System (GIS) software and coordinate measuring devices work by linking data sets to each other with at least one common data field (e.g., a household's street address). The GIS allows the researcher to connect target and classification variables from a survey to specific geographic-based databases like U.S. Census data, to develop a richer understanding of the sample's attitudes and behavior. The most common way to display such data is with a map. Colors and patterns denoting knowledge, attitude, behavior, or demographic data arrays are superimposed

over street, county, state, or country maps.

Cross-tabulation:

Cross-tabulation is used to examine relationships involving categorical variables. The tables used for this purpose consist of cells and marginals (row and column total). The cells may contain combinations of count, row, column, and total percentages. It serves as a framework for later statistical testing. Computer software for cross-classification analysis allows for efficient table-based data visualization and decision making by incorporating one or more control variables. An advanced variation on n-way tables is automatic interaction detection (AID).