ISBN: 978-93-94198-49-4

S. No	Author's Profile	
1.	**Dr. Diwakar Chaudhary** **Assistant Professor** **Mangalmay Institute of Management and Technology,** **Greater Noida, UP.** **Email: diwakarchaudhary2@gmail.com**	
2.	**Dr. Rahul Ratnakar Mahamuni,** **Assistant Professor,** **Department of Conservation of Biodiversity,** **Gopinathrao Munde National Institute of Rural Development and Research** **Maharashtra.** **Email: rahulrmahamuni@gmail.com**	
3.	**Dr. J. Manjunathan,** **Assistant Professor,** **Department of Biotechnology,** **School of Life science,** **Vels Institute of Science Technology and Advanced Studies, Pallavaram,** **Chennai, Tamil Nadu.** **Email: jmanjunathan@gmail.com**	
4.	**Vinitha S** **Assistant Professor** **Department of Psychology** **Marudhar kesari Jain college for women,** **Vaniyambadi, Tamilnadu** **Email: vinithas@mkjc.in**	

Published and Printed By

SALIHA PUBLICATIONS

ISBN: 978-93-94198-49-4

HANDBOOK OF RESEARCH METHODOLOGY

Dr. Diwakar Chaudhary
Dr. Rahul Ratnakar Mahamuni
Dr. J. Manjunathan
Ms. Vinitha S

First Edition: June, 2024.

ISBN: 978-93-94198-49-4

Place of Publication: Vaniyambadi, Tirupattur District, Tamil Nadu, India.

HANDBOOK OF RESEARCH METHODOLOGY

Authors

Dr. Diwakar Chaudhary
Dr. Rahul Ratnakar Mahamuni
Dr. J. Manjunathan
Ms. Vinitha S

Published and Printed By

SALIHA PUBLICATIONS

Published and Printed By

SALIHA PUBLICATIONS

Vaniyambadi, Tirupattur District, Tamil Nadu, INDIA

https://salihapublications.wordpress.com/

E. mail: salihapublications2016@gmail.com

AUTHORS PROFILE

They have sufficient teaching and research experience with expertise in HANDBOOK OF RESEARCH METHODOLOGY. They have written this book for the benefits of undergraduate and postgraduate students.

S. No	Author's Profile	
1.	Dr. Diwakar Chaudhary Assistant Professor Mangalmay Institute of Management and Technology, Greater Noida, UP. Email: diwakarchaudhary2@gmail.com	
2.	Dr. Rahul Ratnakar Mahamuni, Assistant Professor, Department of Conservation of Biodiversity, Gopinathrao Munde National Institute of Rural Development and Research Maharashtra. Email: rahulrmahamuni@gmail.com	
3.	Dr. J. Manjunathan, Assistant Professor, Department of Biotechnology, School of Life science, Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu. Email: jmanjunathan@gmail.com	
4.	Vinitha S Assistant Professor Department of Psychology Marudhar kesari Jain college for women, Vaniyambadi, Tamilnadu Email: vinithas@mkjc.in	

PREFACE

Research methodology is a structured and scientific approach used to collect, analyze, and interpret quantitative or qualitative data to answer research questions or test hypotheses. A research methodology is like a plan for carrying out research and helps keep researchers on track by limiting the scope of the research. Several aspects must be considered before selecting an appropriate research methodology, such as research limitations and ethical concerns that may affect your research.

The research methodology section in a scientific paper describes the different methodological choices made, such as the data collection and analysis methods, and why these choices were selected. The reasons should explain why the methods chosen are the most appropriate to answer the research question. A good research methodology also helps ensure the reliability and validity of the research findings. There are three types of research methodology—quantitative, qualitative, and mixed-method, which can be chosen based on the research objectives.

We welcome comments by readers of **"HANDBOOK**

OF RESEARCH METHODOLOGY” for ways to improve the book and to increase its value. Such suggestions will be seriously considered in the preparation of subsequent editions.

 ISBN: 978-93-94198-49-4

CHAPTER- I

BASICS OF RESEARCH

 ISBN: 978-93-94198-49-4

Meaning of research

Research is the process of discovering new information in order to answer a question. The word research consists of the parts re (again) and search (find), which suggest that we are engaging in an activity to investigate an issue again or that we are seeking new knowledge about something. If a Front Office Executive wants to become a great sales expert in addition to being a host, he or she must learn about the hotel's facilities, timetables, and important elements of the products and services offered.

Definitions of research

Research is manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art (Kothari, C.R.)

Another definition of research is given by Creswell who states that ―Research is a process of steps used to collect and analyse information to increase our understanding of a topic or issue‖. It consists of three steps: Pose a question, collect data to answer the question, and present an answer to the question.

 ISBN: 978-93-94198-49-4

NATURE OF RESEARCH

Research, as explained earlier, is systematic and critical investigation of phenomena. It identifies the variables, collects and analyses data on such variables to find answers to certain crucial questions. These answers contribute further to increase human knowledge. Orderliness is the hallmark of research. Research has to have an organic unity. This becomes essential if the knowledge which accrues from research is to be verified; for, it must be verifiable by anybody who takes the trouble to do so. In fact, research is considered to be a formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of investigating, usually resulting in some sort of formal record of procedures and results or conclusions.

There are two main categories of research methods: qualitative research methods and quantitative research methods. Quantitative research methods involve using numbers to measure data. Researchers can use statistical analysis to find connections and meaning in the data.

In order to obtain results when testing hypotheses and evaluating data, researchers employ a variety of techniques.

 ISBN: 978-93-94198-49-4

Analysts and researchers employ quantitative and qualitative research methodologies for data interpretation, such as focus groups, interviews, and surveys. This fosters a deeper comprehension of the data. A research professional can acquire fruitful outcomes from their investigations by employing the most appropriate research method.

Characteristics of Research

The major characteristics of any research are; objectivity, precision, design and verifiability. Let us look at these attributes more closely:

Objectivity: Ideally, research is beyond the subjective bias of the researcher. The researcher makes deliberate efforts to eliminate personal preference resisting the temptation to seek only such data that supports his/her hypothesis. The emphasis is on testing, rather than proving the hypothesis. The researcher is willing to suspend personal judgement and permit the data and logic to lead independently to a sound conclusion.

Objectivity is achieved through standardisation of research instruments, choosing appropriate research design and analytical tools and ensuring dependability of data.

 ISBN: 978-93-94198-49-4

Precision: Precision in scientific research is achieved through the uses of statistical methods and techniques. As such, research conclusions convey the exact meaning to the reader, e.g., measures of central tendency, variability, correlation, regression etc. are the most precise expression in quantitative research which explains or represents the truth. Precise language describes the study accurately so that the study may be replicated or the results correctly used.

Design: In scientific research, the researcher has to have a much specified design of carrying out the investigation. This will imply that any scientific inquiry will, in general, undergo the following steps: – defining of the problem, – statement of the hypothesis, – collection and analysis of data testing and confirmation or rejection of hypothesis, and – reporting of results. Only if the research has been carried out by using a specified process, it can be replicated for verification.

Verifiability: This is an important characteristic of every research. Research methods and findings presented to the professional community for other researchers to analyse, confirm or reject them. Research is a social enterprise and its information is open for public scrutiny. This characteristic of research, i.e., verifiability, is related to the criteria of objectivity and precision. Only through further investigation or replication of studies can the results of a single study be confirmed or revised. Through this process, a body of new knowledge is developed and new questions identified

Importance of research

Research unfolds many unexplored things that stimulate working capability and develops the understanding and decision-making of learners. Without research, one has to believe in his or her intuition or imagination which can't be accurate. That is why research is important for scaling up performance and better results. Here are 5 reasons why research is important for stimulating better results and lending success to the people.

Research helps to achieve your goal: Many companies have achieved success by conducting research about their products and services through consumer satisfaction exercises. Through online surveys, companies offer feedback about their products and services to know customer satisfaction and it provides them with an opportunity to tailor their businesses and improve their services.

Research reveals new ideas and facts: Research sparks new ideas and facts; enables the researcher to draw novel conclusions about his research topic. The more deeply one researches a particular subject better he may get into the field and properly help position his or her ideas. Research stimulates different ideas and helps build discerning and analytical concepts.

Research develops understanding and decision-making:

It is the most valuable tool to understand the complexities of a problem, disapprove of lies, uphold truth and build on to create knowledge that is reliable and authentic. Conducting research develops a better understanding and enhances decision-making capabilities. Researchers analyse the details of a project and help take the right and well-informed decisions.

Research helps understand what's not working:

Research provides you with an opportunity to detect and investigate many new things that are not working for your project. Research helps you find a new concept or some amazing information about any particular topic, it also unfolds many antagonist elements not suitable for your project.

Research not only helps explore what is beneficial and working for your project, but it also explains different aspects that are invalid and not working for your venture.

Research builds credibility: Any argument or statement based on imagination or intuition affects the reputation and credibility of the subject and researcher. Research on the subjects' scales up credibility and weight to it and critics wouldn't find holes and would appreciate the research project. It also helps boost the confidence of the researcher. Research builds a solid foundation when novel ideas are studied, experimented and explained to the learners and organizations for accelerating the success rate to gain credibility for the research project.

Objective of research

The main purpose of research is to discover answers to the meaningful questions through scientific procedures and systematic attempt.

The main objectives of research are:

1. To gain familiarity or to achieve new insights into a phenomenon.
2. To describe the accurate characteristics of a particular individual, situation or group.

 ISBN: 978-93-94198-49-4

3. To determine the frequency with which something occurs or with which it is associated with other things.
4. To test a hypothesis of a causal relationship between variables.

What is a research problem?

It's a statement that outlines the specific problem, challenge, or question that you try to address or answer. It should essentially address what is the problem and why is it important.

Research problem definition: A research problem is defined as an area of concern that requires a meaningful understanding of a specific topic, a condition, a contradiction, or a difficulty. So, what is research problem? A research problem means finding answers to questions or strengthening existing findings to bridge the knowledge gap to solve problems.

What is the purpose of a Research problem statement?

A problem statement in research seeks to achieve the following:

1. Introduce the importance of the topic in the research proposal.

 ISBN: 978-93-94198-49-4

2. Position the problem in an appropriate context.
3. Provide a framework to analyse and report results.

Selection of Research Problem

There are some suggestions for the graduate students and researchers which are drawn from the different areas of education, social sciences as well as psychology.

There are two factors in the selection of topic **external and personal.** External criteria involve how the topic is important for the field, availability of both data and data collection methods and the administration is cooperative or not. Personal Criteria means researcher own interest, time and cost. Criteria for selection of research problem depend on the following characteristics.

Personal Inclination. The chief motivation in the way of selecting research problem is the personal inclination of the researcher. If a researcher has personal interest in the topic, he would select that problem for his research work.

Resources Availability. During the selection, a researcher will see to the resources available. If these resources like money, time, accommodation and transport are available to the selection place, then the selection of the problem is easy.

Relative Importance. The importance and the problem also play a vital role in the selection of research problem. If the problem is relatively important, then the researcher tends towards the selection of the problem.

Researcher Knowledge. The researcher knowledge should play a vital role in the selection of the research problem. The wisdom and experience of an investigator is required for well collection of the research data. He can bitterly select a problem.

Practicality: Practicality is also responsible for the selection. The practical usefulness of the problem is the main motivation for a researcher to attend it.

Time-lines of the Problem. Some problems take little time for its solution while others take more time. So, it depends on the time in which we have to complete his research work.

Data Availability. If the desired data is available to the researcher, then the problem would be selected.

Urgency. Urgency is a pinpoint in the way of the selection of research problem. Urgent problem must be given priority because the immediate solution can benefit the people.

Feasibility. Feasibility is also an important factor for the selection of the research problem. The researcher qualification, training and experience should match the problem.

Area Culture. The culture of the area for which a researcher conducts his research is also responsible for the selection of research problem.

Characteristic of Research Problem

Any research is a difficult task to achieve and research needs to do a great effort. Selection of research topic is the first step to success.

1. Research topic must be very clear and easy to understand. It should not distract people.
2. If a topic is well defined is the only way to successful research. The topic should not create doubt and double impression.
3. Easy language is a key to success. Use technical words, if necessary, otherwise focus of simplicity.
4. Research title should be according to the rules of titling. There are different rules of titling, a researcher must aware before writing a research title.
5. While selecting a research topic current importance of a researcher should also be considered. Topic

should not be obsolete and it should have great importance in the current day.

Steps to formulate a research problem

Here are the five basic steps to formulate a research problem:

1. **Identify the broad research area:** Begin your research by identifying a broad research area based on your interest, specialty, profession, expertise, and knowledge. This area must possess some kind of significance regarding your knowledge interest and specialty. For example, a researcher studying sports education can select areas like football, soccer, hockey, and baseball. These are the broader areas that can be further subdivided into various research topics to figure out marketing strategies.
2. **Divide the broad area into sub-areas:** After you choose a broad area to study, drill down to a specific topic that is manageable and researchable. To do this, break down the broad area into sub-areas and choose a specific topic. For example, if your broad area is soccer, it can be further divided into the following subcategories:

1. Profile of soccer players
2. Profile of soccer clubs
3. Level of soccer clubs
4. Impact of the club on the city
5. Revenue generating areas
6. Sponsors of the soccer clubs

3. **Choose a sub-area:** It is not possible to study all the sub-areas due to time and money constraints. Thus, choose one sub-area of interest and one that is manageable and feasible for you. The area you select must have some research significance and must be significant to your research knowledge.
4. **Formulate research questions:** After you choose a specific sub-area, think about the areas you must explore and research about. Start noting down important questions that you deem important for the research study. Many questions may arise but narrow down and choose the most important and impactful questions. The length of the research depends on the number of questions you formulate. Choose the questions, depending on the expected length of your research.

5. **Set research objectives:** You must draw a plan about the objectives of the research that you need to explore. The objectives of the research study help to identify the research questions. There is a difference between the research question and the research objective. The difference is the way they are written. Research questions generally consist of an interrogative tone. On the other hand, the research objectives are aim-oriented. They include terms like to examine, to investigate, to explore, and to find out.

What is a research objective?

Research objectives describe what your research project intends to accomplish. They should guide every step of the research process, including how you collect data, build your argument, and develop your conclusions.

Your research objectives may evolve slightly as your research progresses, but they should always line up with the research carried out and the actual content of your paper.

Research aims

A distinction is often made between research objectives and research aims.

A research aim typically refers to a broad statement indicating the general purpose of your research project. It should appear at the end of your problem statement, before your research objectives.

Your research objectives are more specific than your research aim and indicate the particular focus and approach of your project. Though you will only have one research aim, you will likely have several research objectives.

Example: Research aim

To examine contributory factors to muscle retention in a group of elderly people.

Example: Research objectives

To assess the relationship between sedentary habits and muscle atrophy among the participants.

To determine the impact of dietary factors, particularly protein consumption, on the muscular health of the participants.

To determine the effect of physical activity on the participants' muscular health.

Objectives - Research problem is the goal, and the objectives define strategies and implementation steps to attain this goal.

 ISBN: 978-93-94198-49-4

Thus, it highlights **how** you plan to address your research problem.

Meaning of scientific literature

Scientific literature is the principal medium for communicating the results of scientific research and, as such, represents the permanent record of the collective achievements of the scientific community over time.

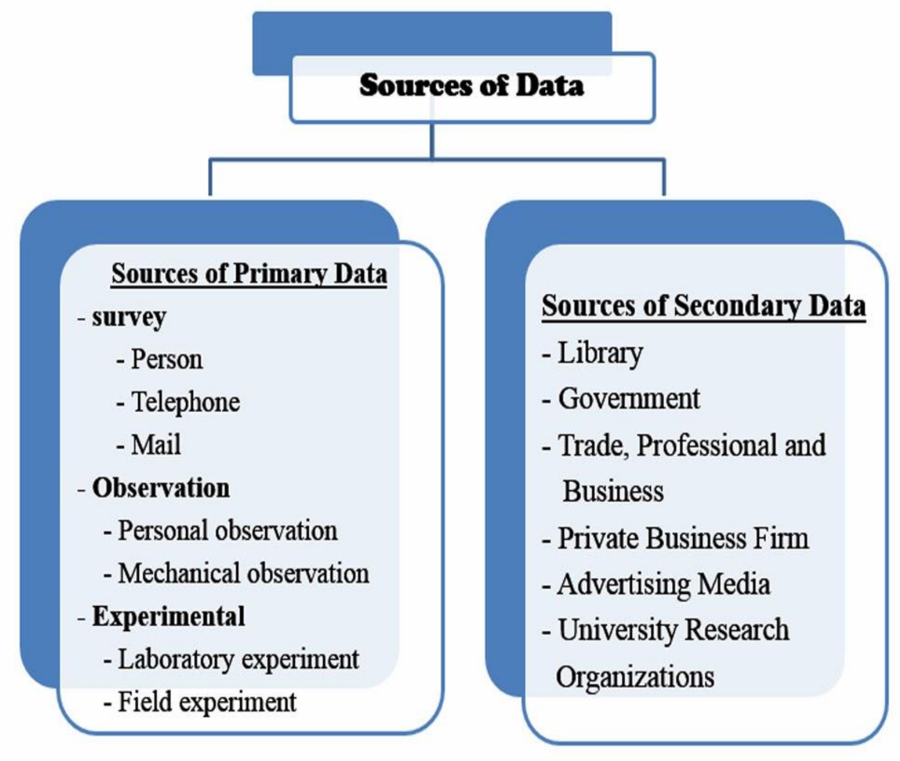
A literature review for scientific research is a survey of scientific books, scholarly articles, and any other systematic scientific sources pertinent to a particular issue, area of study, or theory, in order to provide a description, summary, and evaluation of a concept, school of thought, or ideas relevant to the research question under investigation. In addition, the literature study familiarises the author with the breadth of their field's knowledge. When included in the paper, it demonstrates to the reader the depth of the author's comprehension and knowledge of the subject.

The scientifically relevant literature in a certain discipline includes, among other things, past studies, established schools of thought, scholarly papers, and scientific periodicals. Literature reviews vary from discipline to discipline. In the hard sciences, the literature consists primarily of facts, and a review may consist merely of a synopsis of the essential sources. In the field of soft sciences, the survey gives an overview and synthesis of several schools of thought and their interrelationship. A summary or overview is a concise account of the informative highlights from major sources, whereas synthesis is the restructuring or reorganisation of the material to inform the dissertation's plan for researching the research problem.

Data collection

The data collection technique is different for different types of research design. There are predominantly two types of data: (i) the primary data and (ii) the secondary data. Primary data is one a researcher collects for a specific purpose of investigating the research problem at hand. Secondary data are ones that have not been collected for the immediate study at hand but for purposes other than the problem at hand.

 ISBN: 978-93-94198-49-4



Primary sources provide raw information and first-hand evidence. Examples include interview transcripts, statistical data, and works of art. A primary source gives you direct access to the subject of your research.

Secondary sources provide second-hand information and commentary from other researchers. Examples include journal articles, reviews, and academic books. A secondary source describes, interprets, or synthesizes primary sources.

Citation indexes

Cite can be defined in numerous ways, but in the context of academic writing, it is the acknowledgment of another person's intellectual property.

Citation indexes allow researchers to trace the impact of an article upon later publications. Besides including the bibliographic information about an article (author, article title, journal title, date, etc.), citation indexes also provide each article's references or bibliography (the list of sources cited).

It is essential to keep in mind that citations are required for all types of information sources. Academic journals, internet, books, newspapers, and magazines are among the most frequently used source types. However, there are numerous alternative options, such as movies, audio recordings, lectures, presentations, personal interviews, surveys, and photographs. Regardless of the sort of source used, a citation is required to provide credit to the original author.

Scientific Style and Format presents three systems for referring to references (also known as citations) within the text of a journal article, book, or other scientific publication: 1) citation–sequence; 2) name–year; and 3) citation–name. These abbreviated references are called in-text references. They refer to a list of references at the end of the document.

The system of in-text references that you use will determine the order of references at the end of your document. These end references have essentially the same format in all three systems, except for the placement of the date of publication in the name–year system.

In scientific writing, we don't cite references by using footnotes. Instead, citations including author's last name and year are inserted directly into the text, like this (Kusnick, 1997). If there is more than one author, the citation might look like this (Kusnick and Waterstraat, 1996) or like this (Kusnick and others, 1995). If you use a reference more than once, you just insert the citation again (Kusnick, 1997).

If you reference several authors in the same sentence, you can list them alphabetically within the same citation (Kusnick, 1997; Waterstraat, 1996). If an author has more than one citation in a single year, label them a, b,c, etc. after the year (Kusnick, 1997a).

If you use someone else's stuff for a whole paragraph, cite them once in the first sentence and again at the end of the paragraph. If you include a direct quote, put it in quotes and include the page number in the citation (Kusnick, 1997a, p. 37). Be sure to insert a citation anytime you reference factual material or someone else's opinion. It is better to have too many citations than too few (which can be plagiarism).

The most frequently used citation style used in science is **APA (American Psychological Association) style or Vancouver**, created by the IEEE (Institute of Electrical and Electronics Engineers).

Typical APA style papers have four main sections:

- Title Page
- Abstract
- Main Body
- References

Formatting References

The reference list is alphabetical, and should be either in APA bibliographic style, or in a style used by a scholarly scientific journal (e.g., Geological Society of America style). The style you choose is less important than being consistent. A typical scientific style might look like this:

Journal article:

Kusnick, Judith E, 1993, Review of Revitalizing Undergraduate Science: Why Some Things Work and Most Don't by Sheila Tobias, *Thought and Action*, v. 9, no. 1, p. 145-146.

American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.). https://doi.org/10.1037/0000165-000

Patent

A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem. To get a patent, technical information about the invention must be disclosed to the public in a patent application.

According to the World Intellectual Property Organization (WIPO), patent literature is the collective term for relevant data containing research and development results that have been applied or confirmed to be a discovery, utility model, or industrial design, as

 ISBN: 978-93-94198-49-4

well as published or unpublished documents of materials that protect the rights of the inventor and patentee.

Patent literature often refers to the official publications of each country's patent office, such as the patent specification, patent gazette, patent abstract, patent index, and patent classification table, among others. It should be emphasised that the patent specification in the new application phase may consist of either disclosed literature or unpublished data (still kept confidential).

Various types of patents:

There are three types of patents: invention, utility model, and industrial design.

Invention—refers to the novel technical scheme proposed for a product, a method, or its enhancement, such as the method or procedure for manufacturing the product, the material formula, the medicine formula, etc.

Utility model—refers to the new applicable technical system proposed for the form, construction, or combination thereof of a product, commonly referred to in practise as "little invention."

 ISBN: 978-93-94198-49-4

Industrial design—refers to a new design that is created based on shape, pattern, or their combination, as well as the combination of colour, shape, and pattern, and is suitable for industrial use. This section emphasises the "look" of the packing case, packaging bag, and packaging box.

CHAPTER–II

MIXTURES, SUBSTANCES AND TECHNIQUES

 ISBN: 978-93-94198-49-4

Introduction

There are various kinds of substances present in nature. There are three states or phases of matter - solid, liquid and gas. Matter can exist in different states, for example water can exist in solid state in ice, liquid state or gas state in steam. Temperature and pressure are the two important factors that can change the states of matter. The matter can be classified based on the physical features (solid, liquid and gas) and on the chemical features (mixtures and pure substances).

Pure substances can be divided into **elements** - such as sodium, chlorine, hydrogen or oxygen; and **compounds** – such as sodium chloride and water. Two or more elements chemically bond to form a **molecule.** Compounds can be broken to smaller substances, but elements cannot be broken down to smaller ones because they are made up of atoms containing electrons, protons and neutrons.

A **mixture** consists of two or more substances that are physically mixed, but not chemically bonded.

Most of the substances present around us are not in the pure form. They are present in mixtures of two or more forms. Several kinds of separation techniques are used to separate different substances from the mixture. These separation techniques are useful in removing all the unwanted materials and obtain the desired substances in a pure form. In simple terms, we can say that separation is a process of segregation, where unwanted and impure substances are removed from essential substances. For example: removal of stones and husk from rice, wheat and other food products.

Different methods of separation techniques are available for the separation of substances based on their physical and chemical properties. These kinds of separation techniques are used at home, factory, industry, laboratory, research and pharmaceutical companies.

What is a mixture?

A mixture is a combination of two or more substances, combined together. The components in the mixture may be linked to each other or present separately to form a blend of solution, suspension or colloid. They do not undergo any change and retain their individual properties. The individual substances preserve the physical or chemical identity even when it is present in the mixture. These physical or chemical properties can be made use to separate the individual components from the mixture.

For example,

- **Air** is a mixture of various gases like nitrogen, oxygen, carbon dioxide and other inert gases.
- **Seawater** is a mixture of various salts and water.
- **Crude oil** is a mixture of many organic hydrocarbons.

Types of mixtures

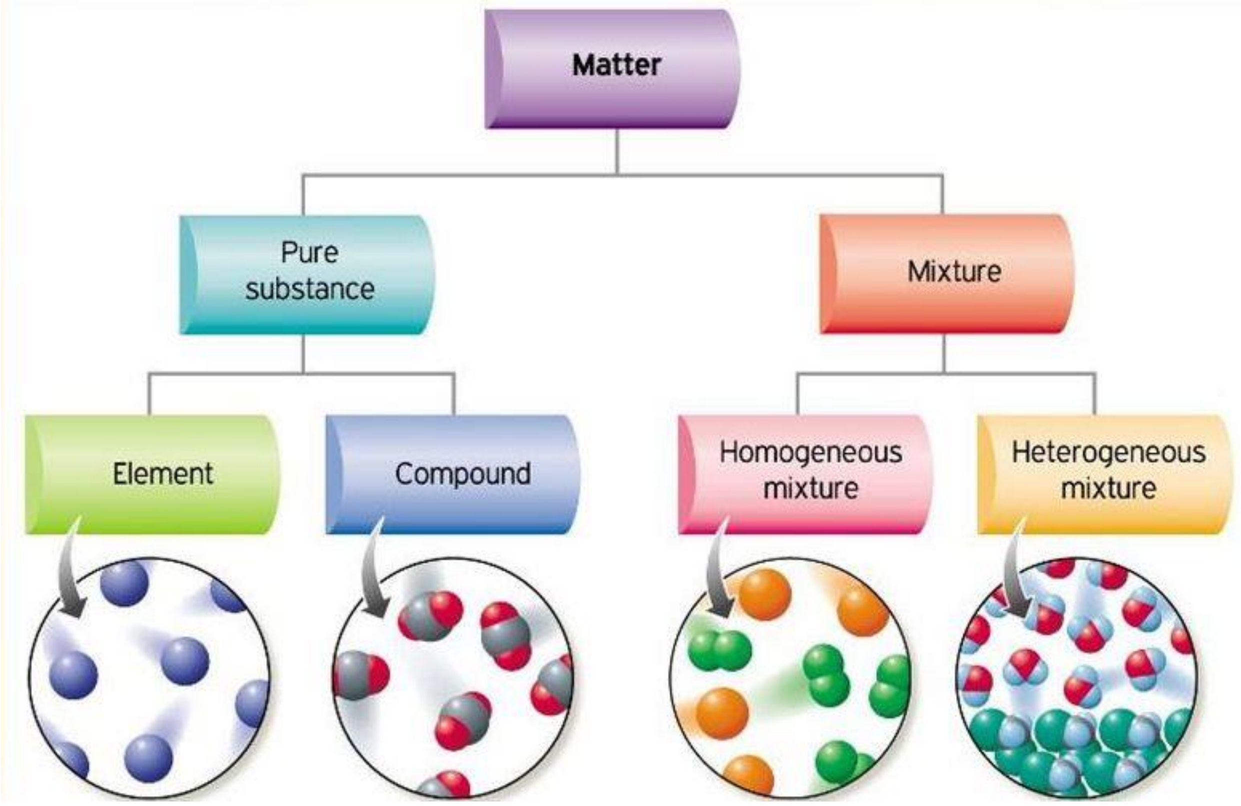


Fig. 1 Types of mixtures

We come across different kinds of mixtures in our daily life. Mixtures are available in different kinds and can be identified based on their physical and chemical properties. Let us discuss the properties of these mixtures. There are mainly two types of mixture: homogenous mixture and heterogenous mixture (fig 1).

 ISBN: 978-93-94198-49-4

Homogenous mixture

Homogenous mixture can be defined as the mixture of different substances or components with the same property and mixed uniformly without any boundary. In other words, we can say that a homogenous mixture has a uniform composition. The constituents in the mixture do not have physically distinct parts and is not visible. The individual components are difficult to separate. For example: mixture of sugar/salt in water.

Heterogenous mixture

Heterogenous mixture can be defined as the mixture of different substances or components with different property. It does not have a uniform composition and has visible boundaries between the components. The constituents in the mixture have physically distinct parts and are easily visible. The individual components are easily separated from each other. For example: mixture of stones and grains or mixture of oil and water.

There is a difference between compounds and mixture:

Compound:

A compound is a pure substance, chemically combined with two or more elements. The elements in the compound are present in a fixed ratio and cannot be altered. If the composition is altered, it changes the characteristic feature of that compound. Compounds are always uniform and homogenous in nature.

Mixture:

A mixture is an impure substance, containing two or more elements or substances that are not chemically combined. The elements in the mixture are not present in a fixed ratio and vary according to the mixture. The components in a mixture are combined physically without altering their properties. They do not undergo any chemical change and retain their individual properties. The mixture can either be homogenous or heterogenous. Examples of Mixtures are smog, cement, soil, blood, brass, sea water, bleach etc.

The components in the mixture can be separated by physical methods like purification, distillation, filtration, centrifugation, electrophoresis or chromatography methods.

The components in a mixture can be separated based on the properties like solubility, boiling point, melting point, density etc. For example, a mixture of water and alcohol can be separated based on the difference in their boiling point.

Physical properties of substances

A physical property of a substance describes the physical nature of the atoms, elements and compounds. All substances present in nature have different physical properties, which can be used to separate them from a mixture. Some of the physical properties are appearance, size, texture, solubility, melting point, boiling point and density.

The different components in a mixture can be separated based upon the following properties.

1. The original physical and chemical properties of individual substances in a mixture do not undergo any change.
2. The individual components in a mixture do not share any chemical bonding.
3. The individual components in a mixture are dissolved in various proportions.
4. The substances in a mixture exist together inspite of the absence of any force between them.
5. The substances in a mixture are either homogenous or heterogenous in nature.
6. The property of the mixture depends on the properties of individual components.
7. The substances in the mixture can be separated by physical methods.

Based on the particle size of the individual components, mixture can be classified into solution, colloid and suspension (fig 2).

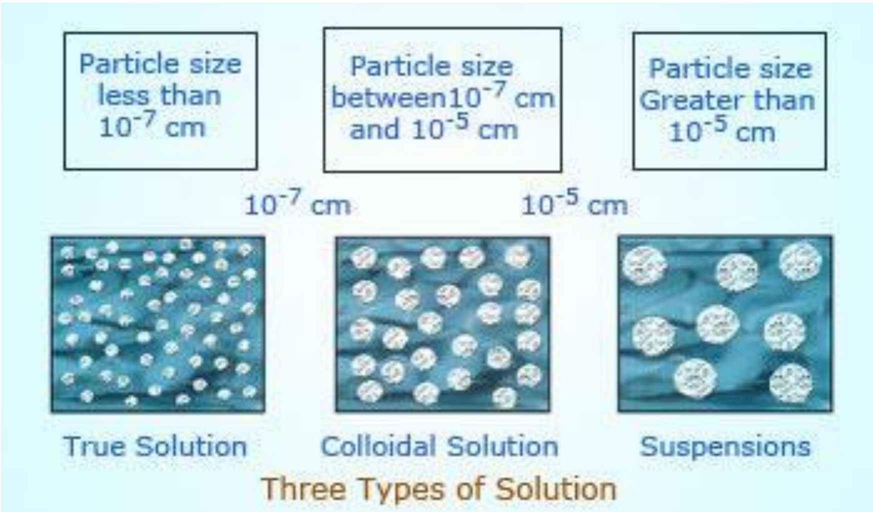


Fig. 2 Types of solution

Solution

A solution is a uniform homogenous mixture of tiny particles with a size of less than a nanometer that remains same throughout. The particles in a solution are not visible to the naked eye and cannot be separated easily with physical methods. Examples are a mixture of salt or sugar in water.

Colloids

A colloid is a mixture of particles that range between 1 nanometer and 1 micrometer and are dispersed in a solvent and are separated by centrifugation. Examples are aerosol, sols, emulsion, gel.

Suspension

A suspension is a heterogenous mixture of solid particles larger than 1000 nanometers spread throughout the liquid without dissolving in it. The particles in the suspension settles down on standing and can be separated by filtration or centrifugation. Examples are muddy water, sand particles in water.

Separation techniques

Separation techniques are methods used to separate one component from another in a mixture. Separation is a very important step in the purification process. The differences in the physical or chemical properties of the substances in a mixture are used as a basis of separation. The properties may include size, shape, mass, density, boiling or melting points etc. The purification of a compound depends upon the different separation techniques employed.

The separation of substances can be on a small scale or on a large scale. The small scale techniques are used in the laboratories or research labs for analytical purposes. Large scale separation techniques are used in the industry for large scale preparative purposes.

The separation techniques include filtration, evaporation, distillation, crystallization, sublimation, dialysis, centrifugation, chromatography and electrophoresis.

Preliminary separation techniques used for separating components in the mixture for preparative or analytical techniques include the following methods (fig 3):

Methods of separation of dry substances

1. Handpicking is the method by which the impure unwanted substances are separated from useful substances by hand. For example, we separate the impurities such as stones, husks and other unwanted substances from grains by handpicking.

2. Threshing is the method used to separate the grains from the dried husk by beating during harvesting. This is the process used by farmers for separating the grains from dried stalks by threshing by hand, animals or machine.

Fig. 3 Separation of substances

3. Winnowing is the method where light weightless components are separated from the heavy components with the help of wind or blowing. This is also a method used by farmers to separate the husk and chaff from grains.

4. Sieving is the method by which particles of different sizes are separated by passing through a sieve with different mesh size. By this method, the smaller particles are passed through the pores while the larger particles are retained by the sieve.

5. Magnetic separation is the method by which a substance can be separated based on its magnetic property.

 ISBN: 978-93-94198-49-4

Strong magnets are used to separate magnetic substances like iron, nickel, steel, oxides and cobalt from impure substances.

Methods of separation of wet substances

1. Evaporation

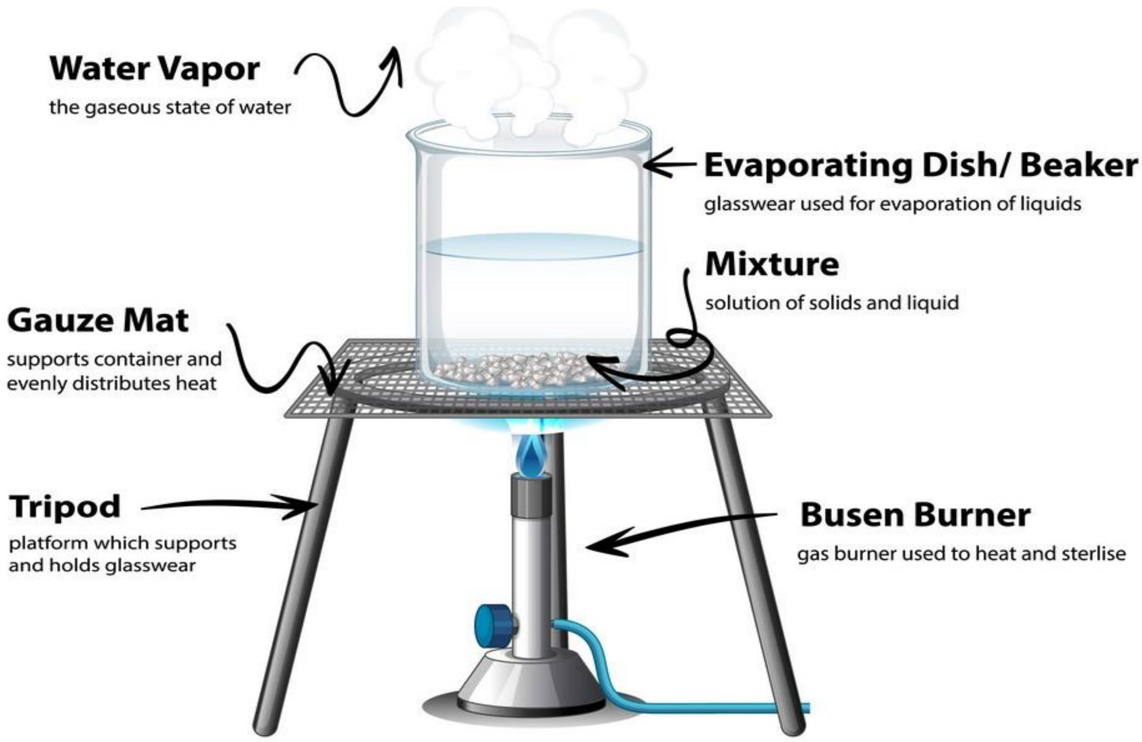


Fig. 4 Evaporation

Evaporation is the method by which a solute is separated from a solution by evaporating the solvent. Water is converted into water vapor by sunlight or heating to its boiling point. Salt is separated from seawater by evaporation (fig 4).

2. Filtration

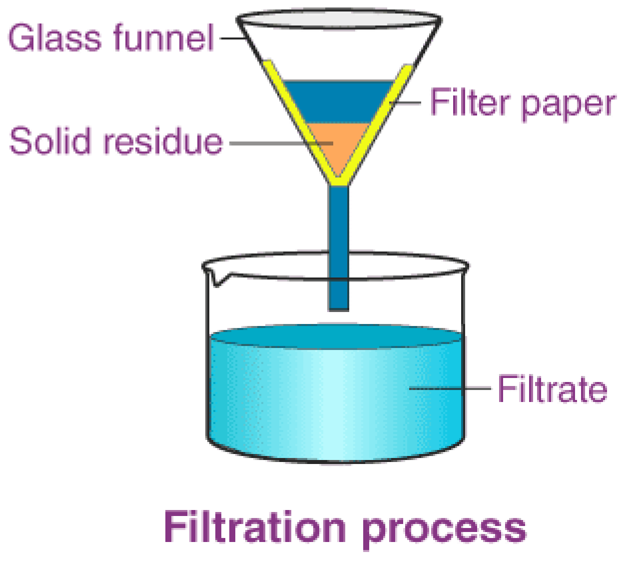


Fig. 5 Filtration

Filtration is the method by which insoluble solid is separated from liquid by filtration using filter paper. The insoluble solid will not pass through the filter paper allowing the liquid to be separated in a pure form (fig 5).

3. Sedimentation

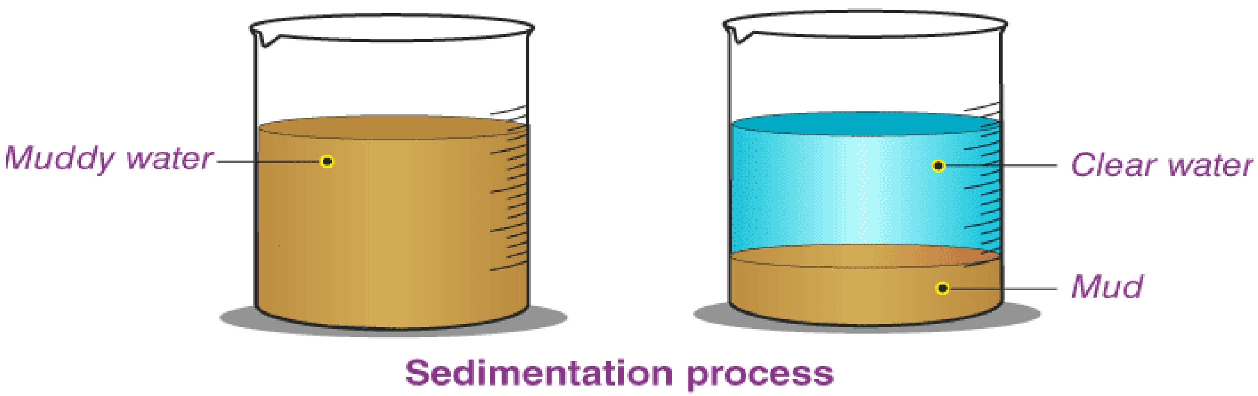


Fig. 6 Sedimentation

Sedimentation is the method by which solid insoluble particles are separated from liquid by allowing it to settle at the bottom. This is followed by decantation where the clear liquid is separated from the solids without disturbing the solids settled at the bottom (fig 6).

4. Distillation

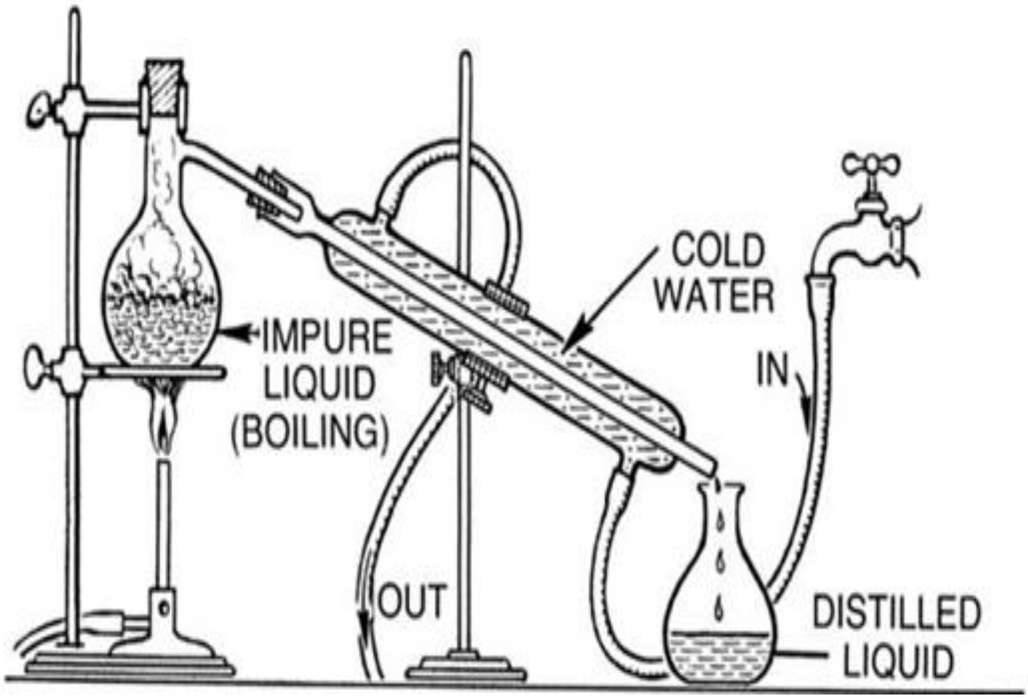


Fig. 7 Distillation

Distillation is a method by which two or more liquids are separated based on the boiling point where one liquid is vaporized, condensed and separated. The solution is heated to boiling, and the pure vapors are condensed by allowing it to pass through the condenser (fig 7).

 ISBN: 978-93-94198-49-4

5. Sublimation

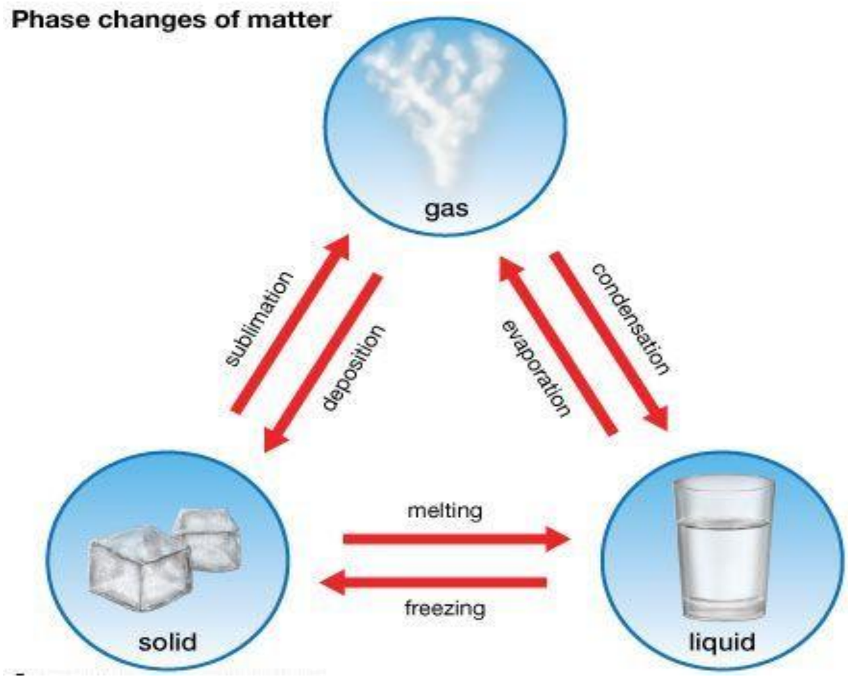


Fig. 8 Sublimation

Sublimation is the method by which volatile substances are separated from non-volatile substances. This is a method which converts the solid directly to vapor state. Substances like camphor, naphthalene, anthracene and ammonium chloride are sublime substances (fig 8).

6. Crystallization

Crystallization is the method that is used to separate solids from a solution by forming pure crystals. Salt is separated from seawater by crystallization (fig 9).

 ISBN: 978-93-94198-49-4

Fig. 9 Crystallization

7. Separation funnel

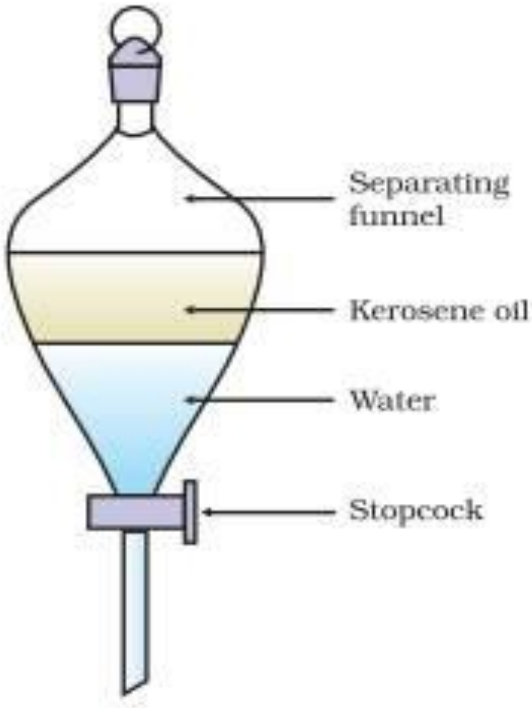


Fig. 10 Separation funnel

Separation funnel is the method by which a separating funnel is used to separate two liquids which do not mix with each other (fig 10).

What is the difference between Separation and Isolation techniques?

Separation is an important method to separate or purify one compound from another; whereas isolation is a technique used to separate or purify a compound from its original source.

Why do we have to separate compounds?

1. To separate a compound in a pure form.
2. To remove unwanted substances from the compound of interest.
3. To identify the nature of the compound that is separated.
4. To analyze the physical, chemical and biological properties of the compound.
5. The isolated compounds can be converted into pills or formulations that can be taken as medicine.
6. Isolation of phytochemicals are necessary for current research in medicine

 ISBN: 978-93-94198-49-4

Extraction

Extraction is a method of separation technique that is involved in the separation of biologically active compounds in a pure form from a plant or animal tissue using solvents. The solvent dissolves the compounds present in the source, depending upon the solubility of the compound. The solvent penetrates into the source and helps in the softening of the tissues, bringing the active compounds into solution (fig 11).

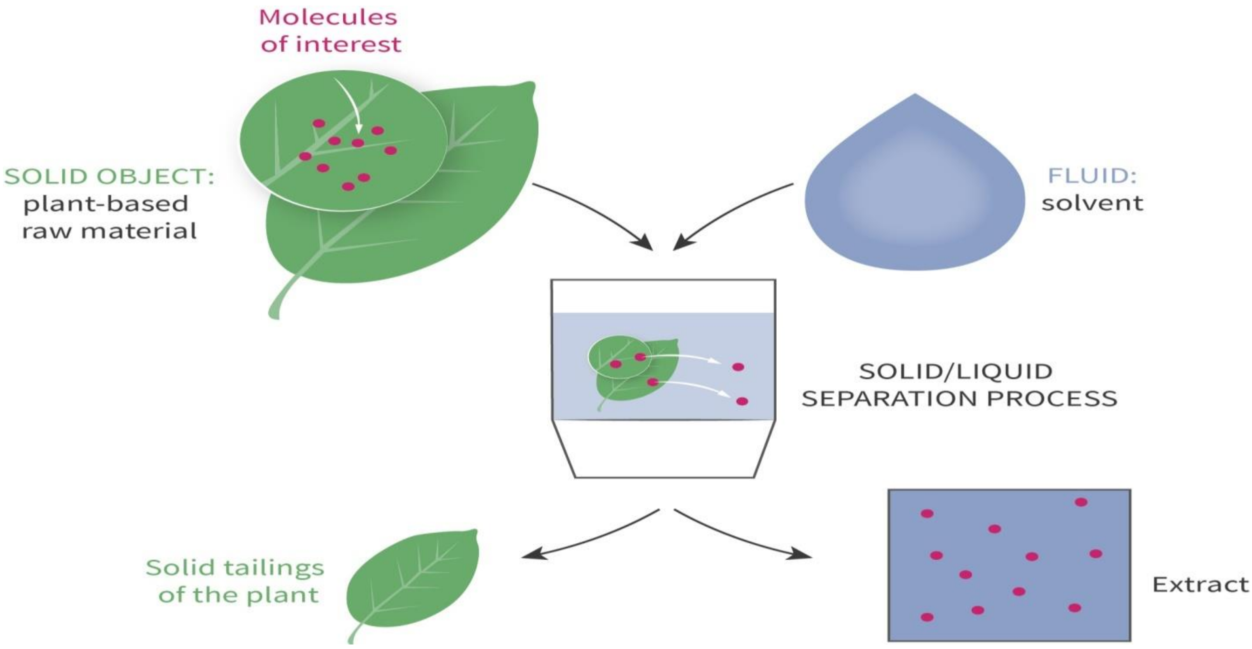


Fig. 11 Extraction of bioactive compounds from plants

Preliminary pre-extraction steps

The first step in the extraction procedure is to identify the source of the compound to be isolated, whether it is a plant source or animal source. The availability and the origin of the source have to be studied either from the literature or from the natural source.

Whole plants can be used as a source or the individual parts of the plants like leaves, stem, bark, root, flowers or fruits can be used as plant source. The different organs of an animal like brain, liver, pancreas or kidneys can be used as an animal source to isolate organelles or organic compounds like nucleus, mitochondria, ribosomes, endoplasmic reticulum, DNA, RNA, proteins etc. Let us study about the separation and isolation of plant materials.

Selection and collection of plants

Plants contain many bioactive compounds called phytochemicals like phenols, flavonoids, alkaloids, tannins, steroids, glycosides, oils that are present in various parts of the plants such as leaf, stem, root, seed, fruit or flower. The bioactive compounds can be isolated from the source by the following conventional and non-conventional (modern) methods (fig 12).

Conventional methods

Conventional methods of extraction include methods like maceration, percolation and decoction, in which a large amount of solvents are used with or without temperature for a long period of time.

Non-conventional methods

Non-conventional methods of extraction include methods like ultrasound assisted extraction, microwave assisted extraction, supercritical fluid extraction, pulse field extraction, enzyme assisted extraction methods and so on.

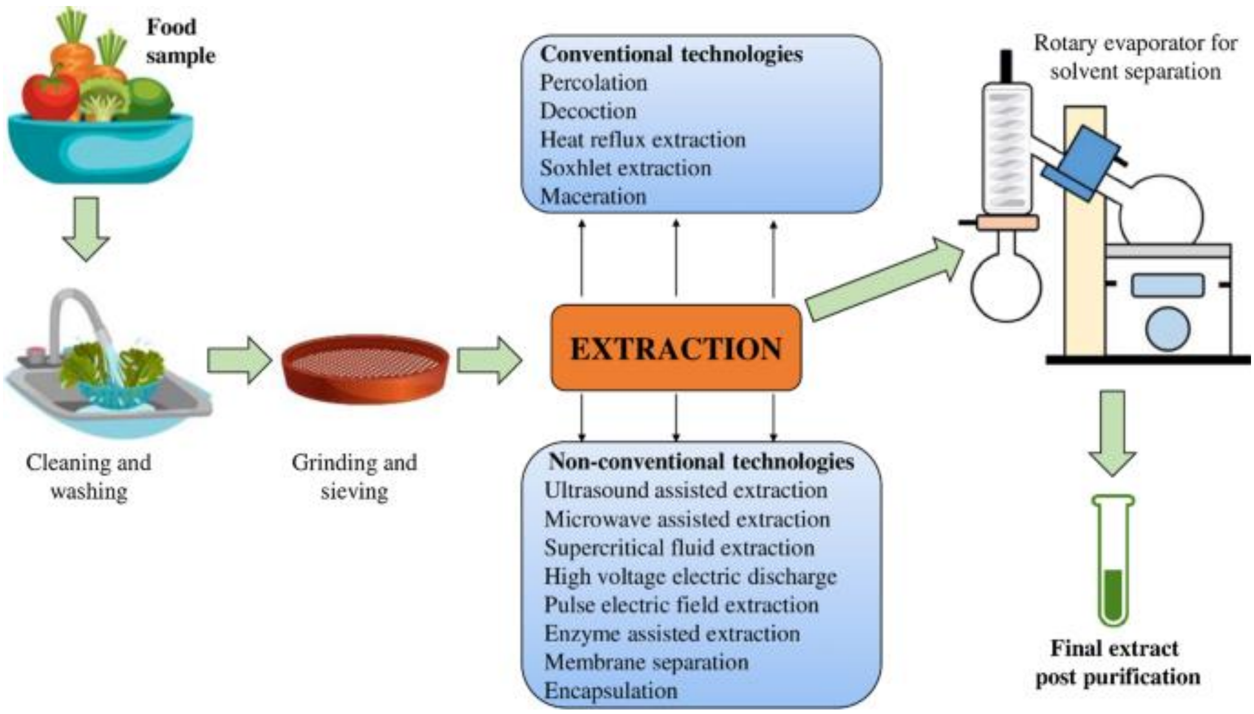


Fig. 12 Conventional and non-conventional extraction methods

Maceration (softening by liquid)

Maceration is a popular and simple technique used for the extraction of essential oils and active compounds from plant materials. The plant material is coarsely ground to increase the surface area and dissolved in a solvent in a closed vessel. The soaking of the plant material in a solvent for several hours leads to the softening of the material (fig 13).

Fig. 13 Extraction methods – decoction, infusion and maceration

Maceration ruptures the leaf structure, exposing the chemical constituents of the cells like polyphenols, lipids and amino acids. The connective tissue fibers are dissolved to release the bioactive components. Maceration is the process used in wine making where the phenolic materials like tannins and anthocyanins are leached from the grapes. This is a very old technique which requires long extraction time and large amount of solvent and low extraction efficiency as it is done in cold condition. This method can be favorable for the extraction of thermo labile substances.

Percolation (to penetrate)

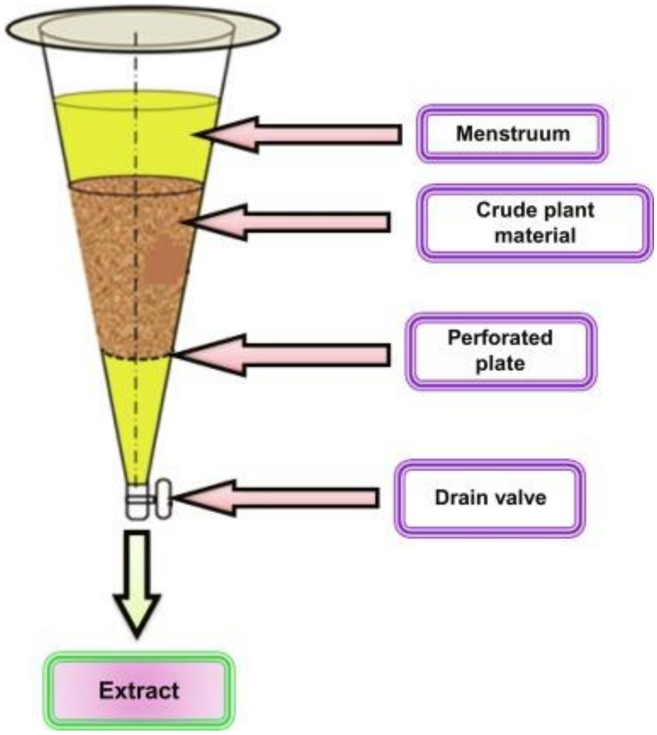


Fig. 14 Percolation technique

Percolation is the downward movement of water through soil due to gravity and capillary forces. Percolation is a conventional method for the extraction of phytochemicals from plants using solvents. The plant material is placed in a tank with excess of solvent for several hours. The solvent percolates into the plant material, leaching the bioactive compounds into the solvent. The solvent (called as menstruum) is further collected by filtration and the compounds separated from the solvent by distillation or evaporation. Percolation method is more efficient than maceration as it uses fresh solvent to replace the saturated solvent.

Decoction (to boil down)

Decoction is also a conventional method of extraction used from olden days to extract the plant material by boiling in a suitable solvent especially water (fig 13). This is a common method of preparation of home-made herbal medicine, mostly followed in India and China. The plant material is dried and cut into small pieces and stored in air tight containers for medicinal use. The preparation of coffee or tea decoction is a very good example. Decoction is the resultant liquid obtained after boiling the plant material. Water is generally used as a solvent, but sometimes aqueous solvents with ethanol will also be used. The plant material is boiled with the solvent to extract the bioactive compound for a few hours to reduce the solvent and to concentrate the extract in the form of Kashayam, used in traditional herbal medicine. This method is not suitable of extracting thermo labile substances. Water decoctions should be prepared fresh and consumed immediately. Alcohol decoctions can be stored for some time and used as tincture.

Reflux extraction

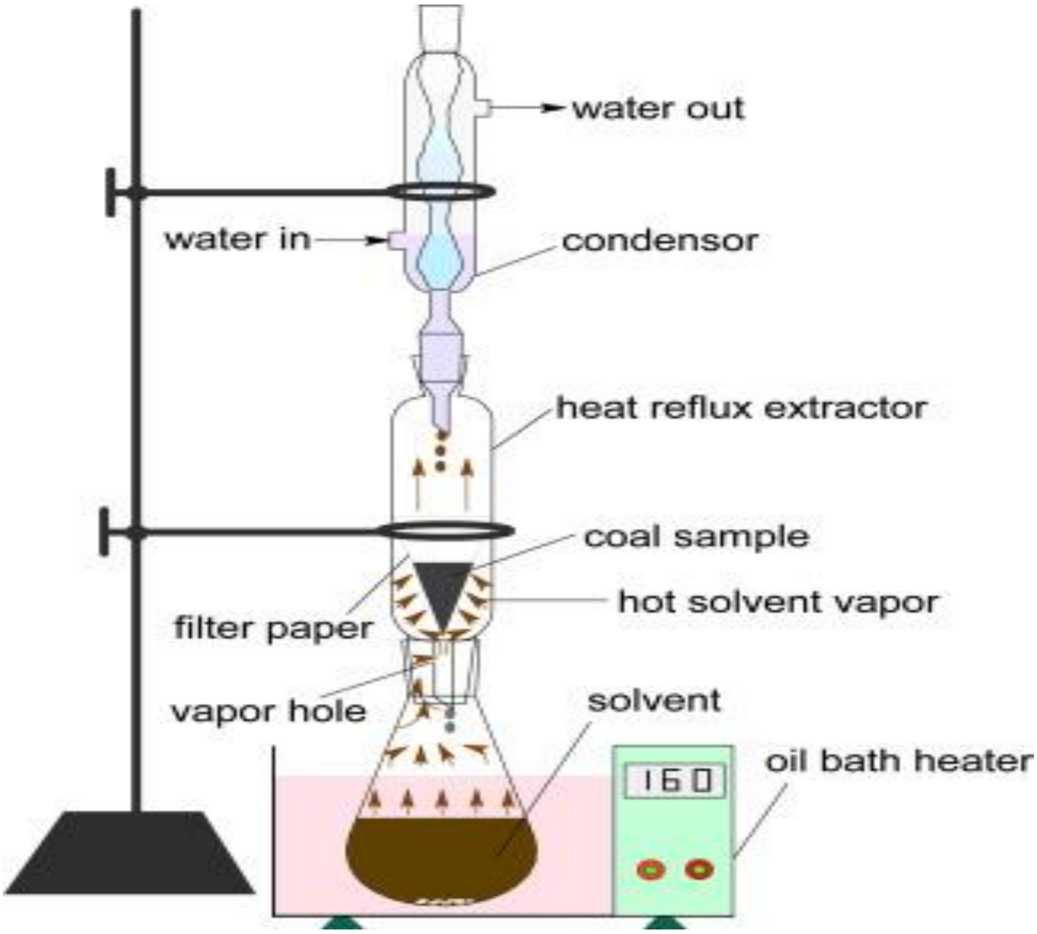


Fig. 15 Reflux extraction

Reflux extraction is another method of extraction that is more efficient than percolation and maceration. This method uses heat and requires less time and solvent than the previous methods. This method is used for the isolation and extraction of thermostable compounds. However, this method is not useful for the extraction of thermolabile compounds. Solvents like alcohol and other organic solvents can be used. Alcohol can be used in combination with aqueous solvents like water.

This reflux extraction method gives a better yield of separated compounds than percolation and maceration methods (fig 15).

Soxhlet extraction

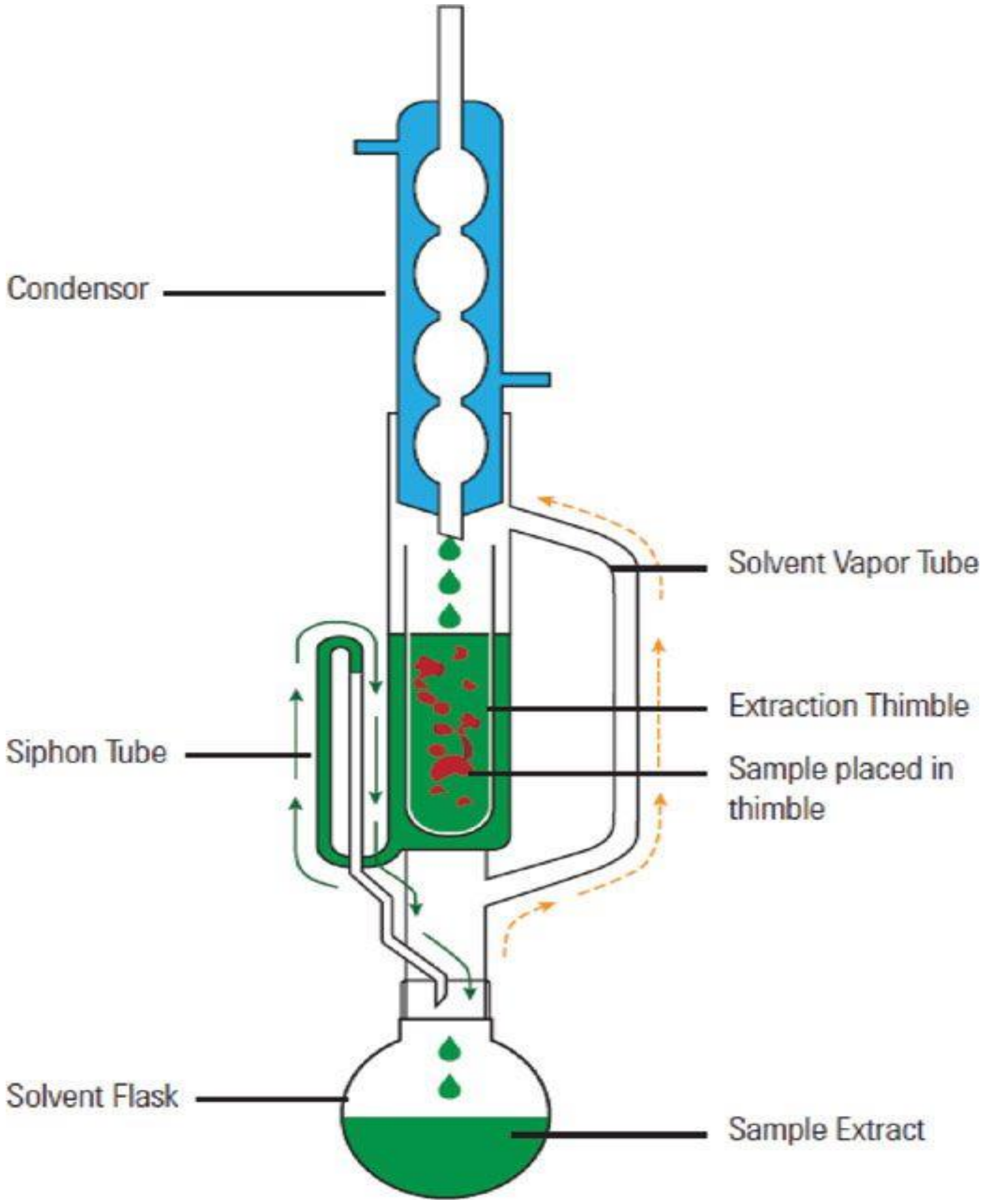


Fig. 16 Soxhlet extraction

Soxhlet extraction is a very efficient method when compared to percolation and maceration. This method is a combination of reflux extraction and percolation and utilizes fresh solvent. The extraction is done by heating the solvent and the compound is extracted by evaporation and condensation of the solvent and requires less time. The hot solvent is allowed to flow through the crude material several times to ensure the complete extraction of the bioactive compound in the solvent. The solvent can be recovered by distillation and the bioactive compound by evaporation.

- The soxhlet apparatus contains the following components:
- A round bottom flask containing the solvent
- Extraction chamber with thistle that contains the crude sample
- Heating mantle
- Condensor

The crude sample in the powdered form is placed in a filter paper or gauze cloth and packed inside the thistle. The solvent is kept in the round bottom flask for boiling in the heating mantle. The vapour of the solvent enters the condenser via a side tube. The vapour is condensed and falls on the sample inside and fills the thistle covering the sample kept in the cloth. The solvent percolates the crude sample and the bioactive compound gets extracted in the solvent and falls back into the round bottom flask (fig 16).

The process is repeated for several cycles until the compound is completely extracted and the solvent becomes colourless. The extracted solvent is further removed by distillation and the bioactive compound is isolated by evaporation and drying or lyophilization. This method is only suited for the isolation of thermostable substances, but not suitable for thermolabile substances as extraction requires temperature. The solvent used should be pure and have a constant boiling point.

Nowadays, modern extraction methods like supercritical fluid extraction (SFC), pressurized liquid extraction (PLE), microwave assisted extraction (MAE) are used for the extraction of natural products. These modern methods are more advantageous than the previous methods as they require less solvent, shorter extraction time and high selectivity.

Pressurized liquid extraction (PLE)

This method is otherwise called as accelerated solvent extraction or enhanced solvent extraction method. This is a high pressure solvent extraction method, which applies high pressure to keep the solvent in the liquid state above the boiling point. This results in high solubility and penetration into the matrix of the plant.

This method is suitable for the diffusion and extraction of lipids into the solvent. This method requires less time and gives high reproducibility. Many types of natural compounds like saponin, flavonoid and essential oils. Due to the use of high temperature, this method is not suitable for the extraction of themolabile compounds.

Supercritical fluid extraction (SFE)

In this method, supercritical liquids and gases such as carbon dioxide, nitrogen, methane, ethane, ethylene, nitrous oxide, sulfur dioxide, propane, propylene, ammonia and sulfur hexafluoride are used for the isolation of bioactive ingredients from natural sources. The plant material is extracted with gas instead of solvent under controlled temperature and pressure. The active compounds are dissolved in the gas at reduced temperature and pressure and will be transferred as a mass in the supercritical solvent (fig 17).

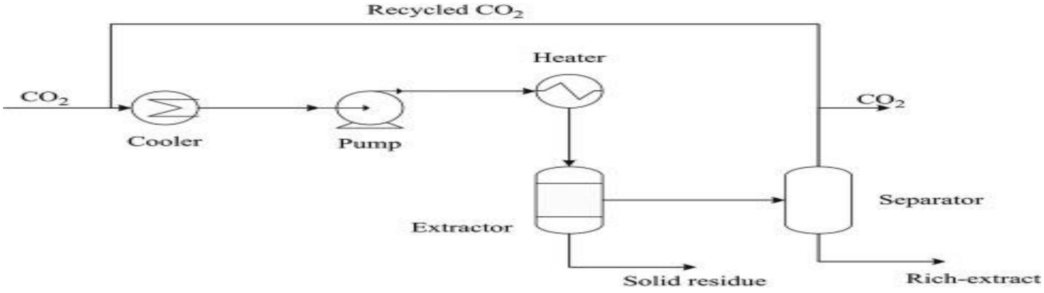


Fig. 17 Supercritical fluid extraction

 ISBN: 978-93-94198-49-4

However, pressure has a greater influence on the mass transfer of active compounds than temperature. As the pressure increases, more amounts of active ingredients will be transferred into the supercritical fluid. A wide variety of natural compounds can be separated by this method. Supercritical carbon dioxide (S-CO_2) is widely used to extract thermally labile compounds because of its low cost, non-toxicity and selectivity. The S-CO_2 is ideal for the extraction of thermally labile compounds because of its low polarity.

Ultrasound assisted extraction (UAE)

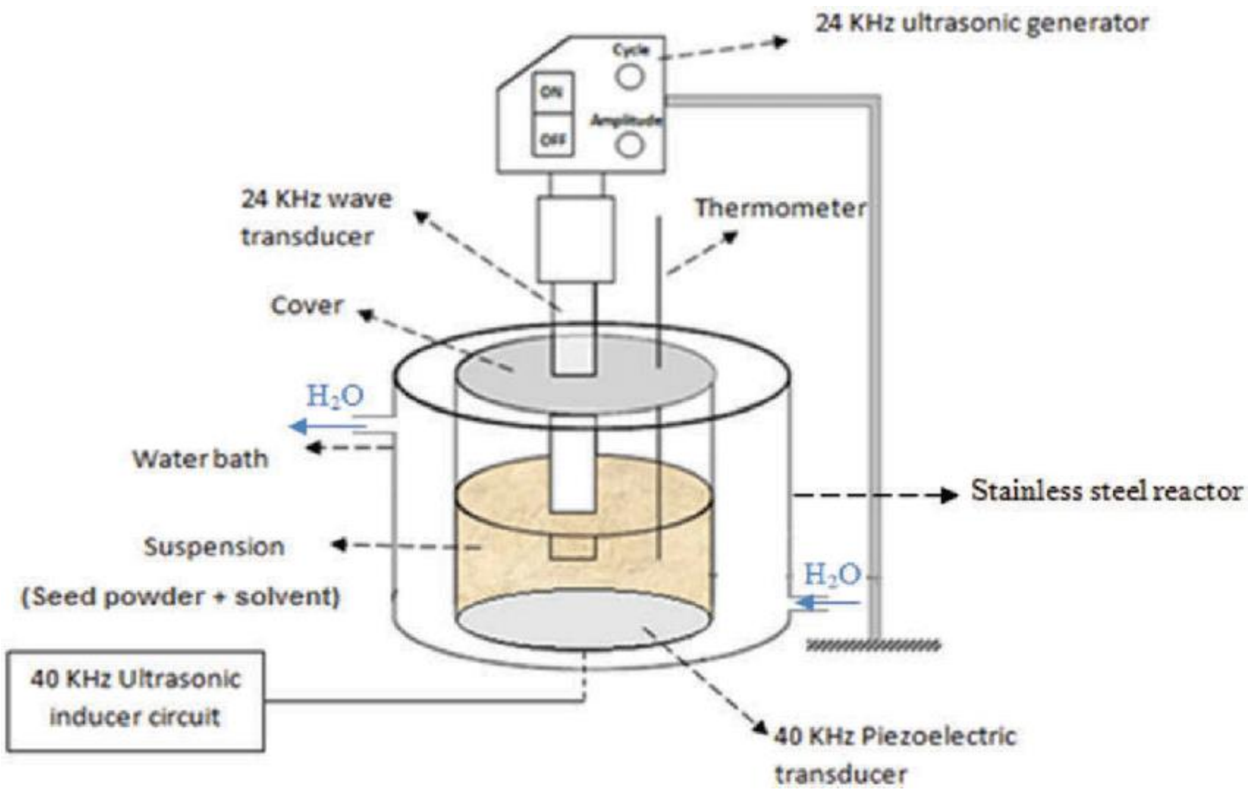


Fig. 18 Ultasound assisted extraction

This is a very advanced method of plant extraction otherwise called as ultrasonic extraction or sonication. This method used ultrasonic waves for the extraction of large amounts of bioactive compounds within a short period of time. The solvent penetrates into the plant matrix due to the disruption of cell walls due to the ultrasound energy. The extraction can be achieved at low temperatures and uses less solvent for the separation of thermally unstable compounds also (fig 18).

Microwave assisted extraction (MAE)

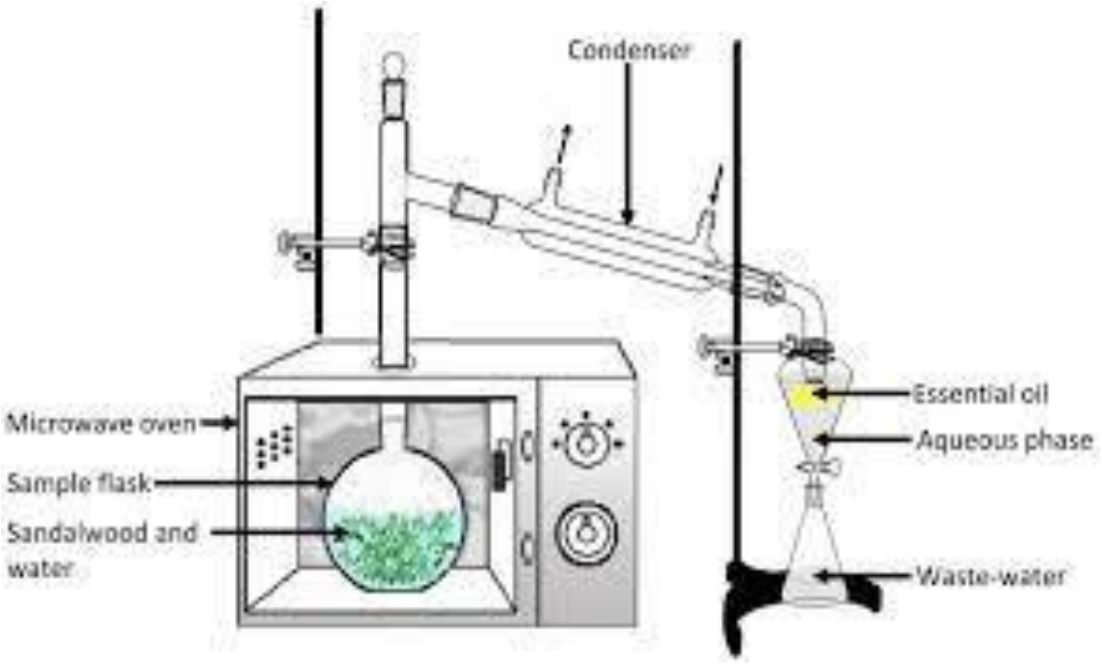


Fig.19 Microwave assisted extraction

This method uses microwave energy to separate the active ingredients from the plant source. Microwave possesses both electric and magnetic field that are perpendicular to each other. The electric field generates heat by interacting with

 ISBN: 978-93-94198-49-4

polar compounds by rotation and ionic conduction, so that the sample heats up simultaneously and extracts the bioactive compound. Heat disrupts the weak hydrogen bonds due to the dipole rotation of molecules and the migration of dissolved molecules increases the penetration of solvent into the plant matrix and extracts the active ingredients into the solvent. This method increases the extraction yield by preventing the thermal degradation of compounds (fig 19). There are two types of extraction methods – solvent free extraction for volatile compounds and solvent extraction for non-volatile compounds.

Pulsed electric field extraction

Pulse electric field extraction is a non-thermal method and suitable for the extraction of thermolabile compounds. This method of separation takes place in the field of electricity; it increases the yield and decreases the extraction time. The extraction depends on the field strength, energy input, pulse number and temperature.

Enzyme assisted extraction

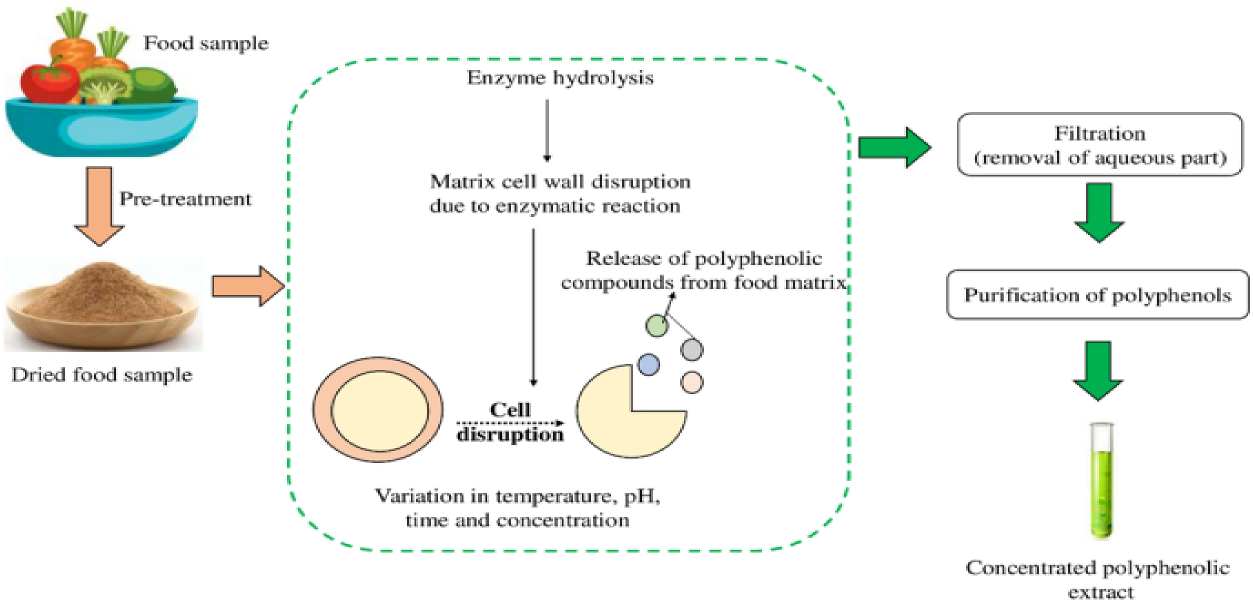


Fig. 20 Enzyme assisted extraction

This method uses the hydrolytic action of enzymes on the cell wall, cell membrane to release the bioactive compounds from the plants. Cellulose, α-amylase, pectinase, glucose oxidase are the enzymes commonly employed in this process. Enzymes enhance the extraction efficiency of the compound. This method overcomes the denaturation of bioactive compounds when high temperatures are used.

Hydro distillation and steam distillation

The hydro distillation and steam distillation method is used for the extraction of volatile oil. Hydro distillation is a better for the extraction of primary and secondary essential oils than steam distillation. But this method has some disadvantage, where the yield of extraction is less with increased time. Molecular distillation method is a more advanced technique where the molecules can be separated by distillation under vacuum at a temperature below the boiling point. Thermosensitive and high molecular weight compounds are separated by this method.

Chemistry of working with hazardous materials

Many laboratories are required in the schools, colleges, scientific and research institutions. These laboratories use many different kinds of chemicals involving acids, bases, corrosive, toxic chemicals. Certain other laboratories also use many explosive and radioactive materials for their research purposes. The importance of safety and preventive measures has to be emphasized and followed in all these laboratories for the safety of people handling the chemicals. According to the Occupational Safety and Health Act in 1970, certain precautions have to be followed to protect the safety and health of employees, teachers, students or researchers.

Though the safety measures have been followed strictly in industries, it has been adopted very slowly in educational institutions. Science programs involving teaching and practice are followed in all educational institutions as a part of the academics. These science programs have some potential dangers during the use of chemicals and laboratory reagents.

Most dangers can be avoided by following safety measures and good laboratory practices.

Safety and health should be an important part of planning, implementation and practice of science program, wherever it may be. Each person involved in the program should have an obligation and responsibility in following the safety measures. Teachers and instructors play an important role in imparting and instructing the students about safety measures to be followed in a laboratory. Each and every stage of the laboratory experiment should be planned, instructed and supervised carefully. When proper instructions are given by the instructors before the experiment, the students can develop the habit of assessing the risk and dangers associated with working with chemicals, toxic, poisonous or hazardous materials.

Everyone working in the laboratory should wear personal protection equipments (PPE) like gloves, coat, mask to demonstrate safety behavior and promote the culture of practicing safety.

The following aspects should be monitored periodically to avoid any risk in the laboratory:

Instruments and Equipments in the laboratory

- The instruments and equipments in the laboratory should be checked regularly to ensure its proper usage and to check any defect in the working.
- Proper quality check should be conducted and the date of quality check should be recorded with signatures of authorities.
- The instruments and equipments should be properly labeled to identify the instrument. Also, if there are any hazardous or defective instruments, it should be mentioned and labeled. Care should be taken not to use any defective instruments.
- Records should be maintained for the usage of the instruments; laboratory training of staff and students also should be recorded.
- Educate students on the location and proper usage of instruments and equipments.
- Emergency safety equipments, first aid kits and fire extinguishers should be available for the proper usage of the people working in the laboratory.
- If there are any unwanted incidents or hazard, it should be immediately reported to the concerned person; and if there is an emergency situation it should be taken

 ISBN: 978-93-94198-49-4

seriously and call the attention of the police or hospital. The unwanted incident can be fire, chemical spill, leak of toxic or hazardous chemicals, explosion etc.

Chemicals

- Proper labeling of all the chemicals should be done and entered in the stock register. Chemicals should be stored in locked cupboards and the usage should be recorded. The toxic and hazardous chemicals should be separated from the normal chemicals and should be labeled properly. The acids should be kept in fumigation chamber with proper ventilation.
- Stock register and log books should be maintained properly and the students should be instructed to make a note on the usage of chemicals and instruments.
- The used chemicals and reagents should be disposed properly without causing any danger or damage to the people or property.
- Ensure that all safety rules are practiced by students and researchers at all times. Avoid eating, drinking and smoking inside the laboratory. Never allow anyone to take the chemicals out of the laboratory.

- Protective gloves should be worn at all times when handling the chemicals.
- The reagents prepared from the chemicals should be labeled with name, concentration and date and initial of the person who prepared.
- Spatula, spoon, pipettes, measuring cylinder and burettes should be used for the transfer of chemicals and reagents.
- Flammable liquids and chemicals should not be heated directly, use a hot water bath for heating.
- Never add water directly to acid, instead acid should be slowly added into water little by little.
- The chemicals should never be touched, smelled, tasted or inhaled.
- Chemicals and reagents should be placed in a tray, bucket or appropriate containers during transport.

According to the periodic table, there are 112 elements present, and are arranged according to the atomic number horizontally. The horizontal rows are called *periods.* The vertical rows are called *groups or families.* Elements within a specific group have the same chemical properties as they have the same number of electrons in the outer most shell. The groups are divided into three categories: metals, non-

 ISBN: 978-93-94198-49-4

metals and metalloids. The elements lower in the periodic table are solids and those elements higher in the table are gases.

There are four significant hazardous *groups or families* that have similar chemical features that are very reactive. They can be classified as follows:

1. Group I – Alkali metals
2. Group II – Alkaline earths
3. Group VII – Halogens
4. Group VIII – Noble gases

Group I – The alkali metals

The alkali metals include lithium, sodium, potassium, rubidium, cesium and francium. These elements are highly reactive. When they come in contact with water, they produce flammable hydrogen gas and enough heat to ignite the hydrogen. Sodium and potassium are the most common alkali metals that are water-reactive (fig 21).

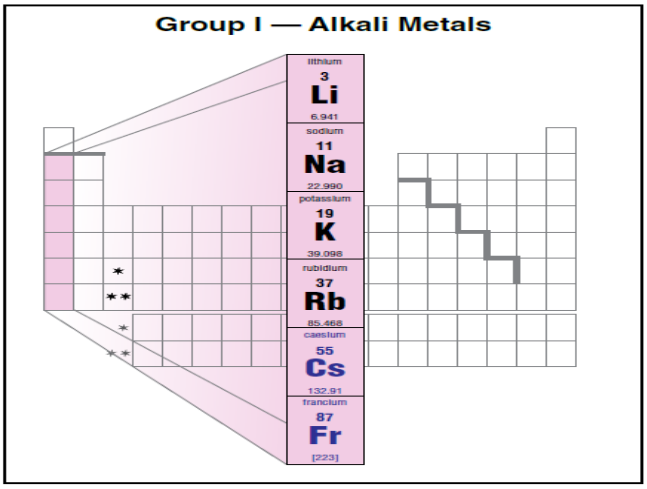


Fig. 21 Group I Alkali metals (Lithium, Sodium, Potassium, Rubidium, Caesium, Francium)

These metals are usually stored under kerosene or vacuum to prevent the exposure to air and water. Sodium and potassium combine with free oxygen and hydrogen atoms and forms sodium hydroxide (NaOH) and potassium hydroxide (KOH), which are both very corrosive. The reaction of sodium and potassium metals with water are very violent and generates enough heat to melt the metal, ignite the hydrogen gas and any other combustible near them. The fire caused by sodium and potassium should be put off with special extinguishing agents.

Group II – The Alkaline Earths

 ISBN: 978-93-94198-49-4

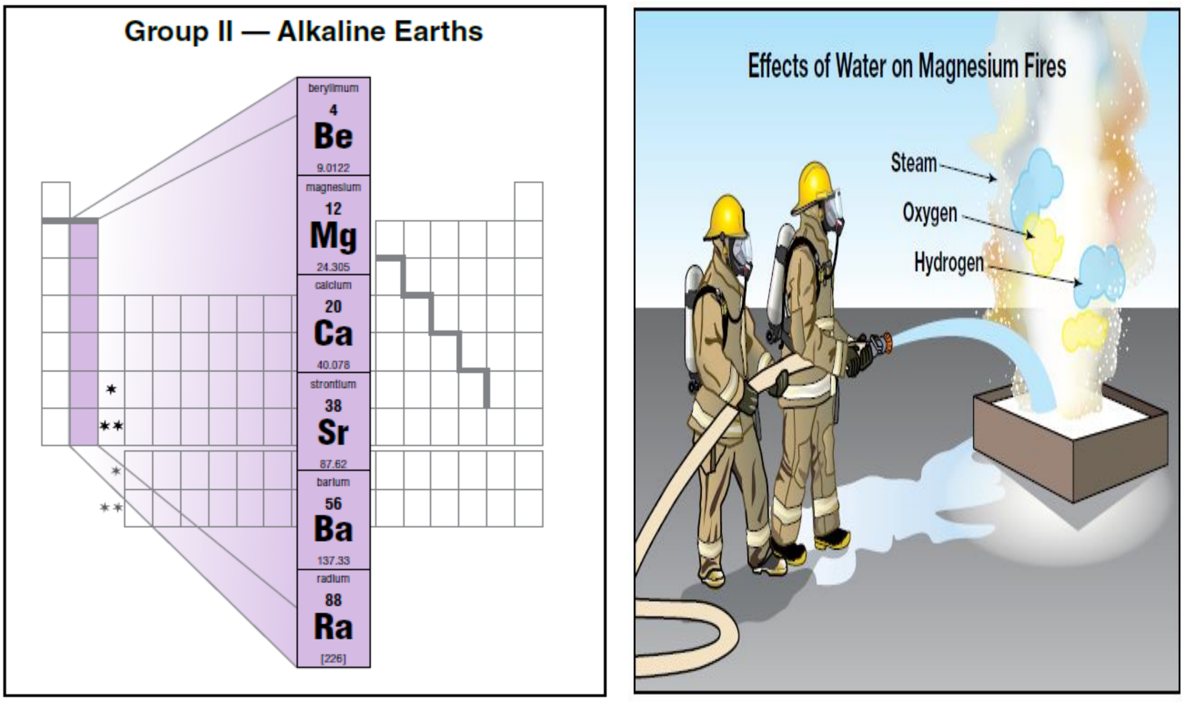


Fig. 22 Group II Alkaline earth

The alkaline metals include beryllium, magnesium calcium, strontium, barium and radium and are also water-reactive, but lesser than alkali metals (fig 22). Magnesium powder, dust, chips and shavings are more dangerous than solid magnesium, which can explode on contact with water. Magnesium burns very intensely with a bright white flame and cause retinal damage on staring. Water intensifies the flame as it is broken down to an oxidizer (oxygen) and fuel (hydrogen); and hence water extinguishers should not be used to put out the fire. Water molecules can be instantly converted to steam by the heat of the fire, resulting in steam explosion.

 ISBN: 978-93-94198-49-4

Group VII – The Halogens

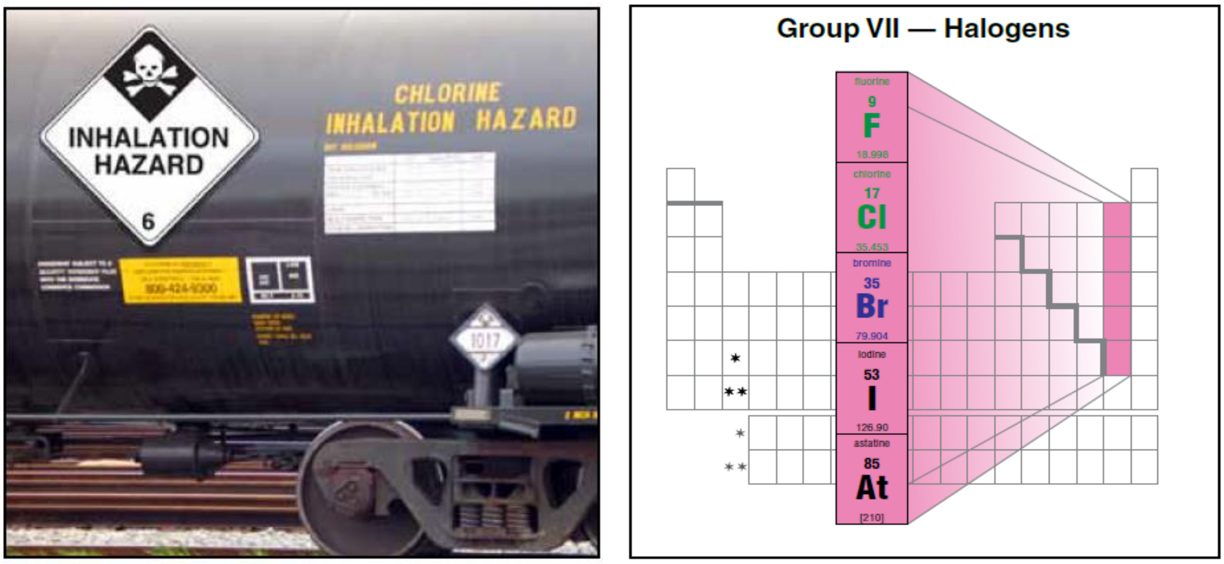


Fig. 23 Halogens

The halogens include fluorine, chlorine, bromine, iodine and astatine. Though these elements are non-flammable, they are highly reactive and toxic (fig 23). Chlorine is the most common halogen and has a strong odor. Chlorine reacts with moisture in the respiratory system and skin to form corrosive hydrochloric acid (HCl) and hypochlorous acid (HOCl), which is very toxic. Fluorine reacts with water to form hydrofluoric acid (HF) and hydrofluorous acid, which is more corrosive than chlorine. Both are very powerful oxidizers that can burn without the presence of oxygen. People handling these halogens should wear chemically protective clothing.

Group VIII – The Noble gases

 ISBN: 978-93-94198-49-4

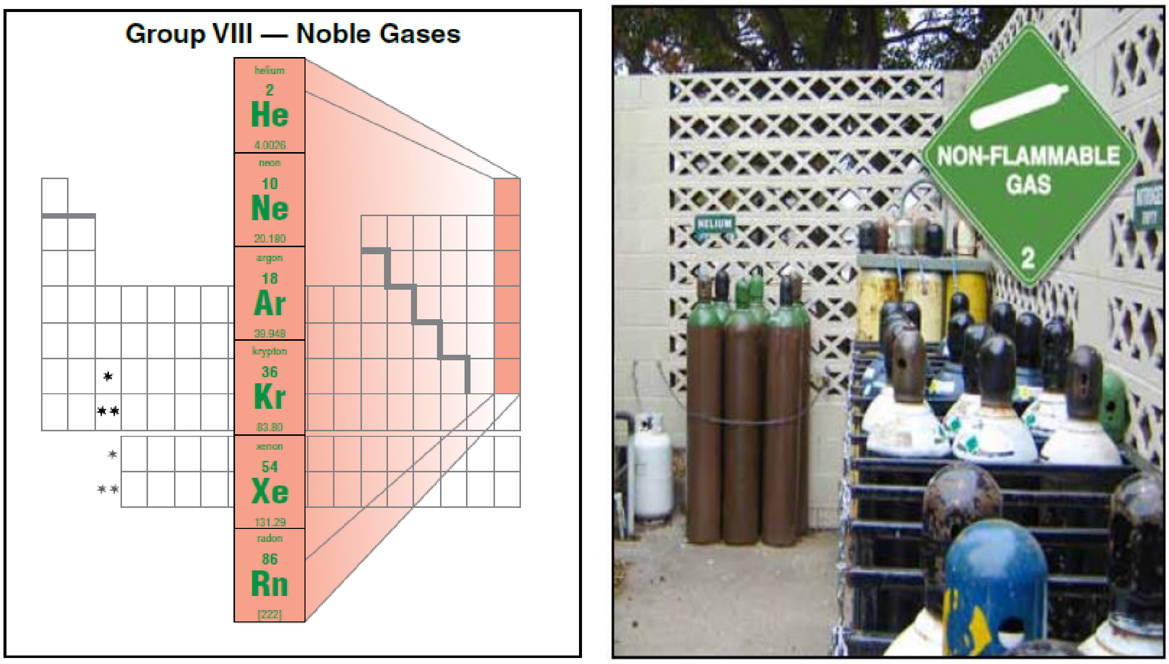


Fig. 24 Group VIII Noble gases (Helium, Neon, Argon, Krypton, Xenon, Radon)

The Nobel gases include helium, neon, argon, krypton, xenon and radon and are called as inert gases (fig 24). The inert gases are not toxic or flammable and will not react with other metals; but when these gases are released in an uncontrollable manner, the gas expands to fill the entire space, whether it is a room or a closed space and will displace the oxygen from the environment leading to decreased availability of oxygen and asphyxiation. These noble gases are transported as cryogenic liquid (-130° F) that can cause severe damage to people and property.

Reactions of chemicals

 ISBN: 978-93-94198-49-4

When two or more chemicals are mixed, or when chemicals are exposed to heat, fire, water or air; it will result in some kind of reactions. Some of the reactions are safe, while others are hazardous and lead to some unexpected or unwanted reactions viz. industrial fire, combustion of flammable chemicals. Reactions can be classified into exothermic and endothermic reactions. When a chemical reaction releases a lot of energy in the form of heat, light and sound, it is called **exothermic reactions.** When a chemical reaction uses energy from an external source, it is called **endothermic reactions** (fig 25)**.**

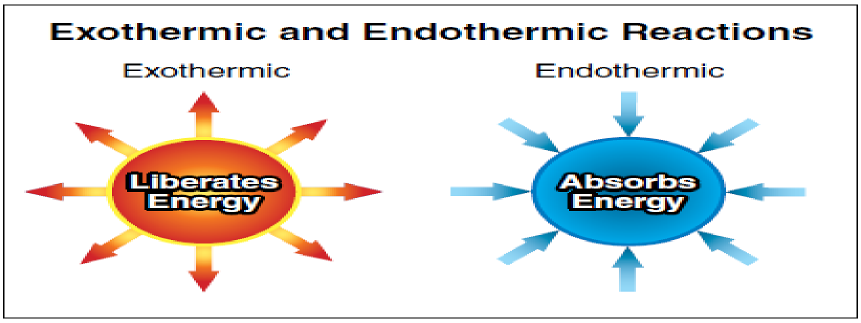


Fig.25 Exothermic and Endothermic reactions

During an exothermic reaction, large amount of energy is released; it can be very dangerous and leads to an explosion with heat and fire, causing severe damage to people and property. These chemicals may become hazardous when they react with one another than when it is not mixed. A reactive material can undergo chemical reaction only under certain conditions. The presence of air, water and even some normal conditions can trigger the conversion of chemicals to hazardous or toxic or explosive ones. Reactive materials can be shock-sensitive, friction-sensitive, pyrophoric or hypergolic.

Water reactive materials react with water; the severity of the reaction depends on the metal that reacts with water. A water reactive material is defined by the type of gas it gives like – toxic, corrosive or inflammable. Let us learn some of the reactions below:

Oxidation and combustion

In an oxidation-reduction reaction or redox reaction, electron is transferred from one atom, compound or molecule to another. Loss or gain of electrons takes place in all oxidation reduction reaction. The substance that loses electron is called a *reducing agent* (reducer). The substance that gains electron is called an *oxidizing agent* (oxidizer). Substances that either give up or gain electrons easily are called as strong reducing or oxidizing agents. Halogens are strong oxidizers, as they have seven electrons in the outer shell and needs one electron to fill it. As it requires only one electron to fill the outermost shell, halogens are very reactive compared to oxygen which needs two electrons to fill the outermost shell. After the oxidation-reduction reaction, the oxidizing agent is *reduced* and the reducing agent is *oxidized*.

Why oxidation-reduction reactions can become hazardous?

Oxidation is the chemical combination of an oxidizer (oxygen) with other substances. Oxidation-reduction potential is measure by how quick a material can give out its oxygen atom to react with other substance. The more readily a material gives out oxygen or any other oxidizer; it is more readily combustible and hazardous. The chemical reaction involving an oxidizer can be slow or fast depending on the chemical it reacts with.

Polymerization

Chemical reactions in which small compounds called **monomers** react with one another to form a large compound called polymers is known as **polymerization**. Or in other words, polymers are repeating units of monomers.

Laboratory Hazards

Hazards in the laboratory can be classified into chemical hazards, electrical hazards, fire hazards and biohazards

Chemical hazards

 ISBN: 978-93-94198-49-4

The chemicals stored in the laboratory can pose danger to the people working in the laboratory. They can be toxic, poisonous, corrosive, flammable and explosives.

It is the responsibility of everyone who is working or handling chemicals in the laboratory to conduct all the experiments in a safe manner to protect the people and environment. Some kind of safety measures have to be identified and adopted by all to prevent any unwanted incidents.

Understanding the chemical hazards

There are many harmful toxic hazardous chemicals or chemical products in the laboratory to which people are exposed when working in the laboratory. These materials can be very dangerous and create physically or health hazards. Physical hazards can be caused by fire or explosives that can lead to burns, rashes or even death. Chemical hazards can lead to toxicity or serious health hazards like cancer, heart disease or can damage the internal organs like liver, kidney, lungs. Though chemicals are available everywhere in many forms, the proper usage of chemicals, preventive measures, usage of personal protective equipments and correct way of disposal of chemicals and hazardous materials will prevent any kind of mishappening.

Chemicals can be classified as hazardous when it is capable of producing adverse effects on humans or the environment. It is very much important to know the chemicals and its harmful effects before working with it and follow all precautionary measures when handling the chemicals.

Physical hazards of chemicals

The chemicals that pose physical hazards can be unstable, oxidative, water reactive, compressed, flammable, combustible, explosive etc.

Flammable or combustible chemicals

The chemical that are capable of combustion or that has a tendency to explode is called flammable or combustible chemicals. They evaporate quickly and produce vapours which can easily get ignited. The flammable substances can be a solid, liquid or gas, but mostly it is a flammable liquid available in the chemical laboratory. The flammable chemicals can be classified based on the flash point, boiling point and ignition temperature.

Flash point (FP) is the lowest temperature at which the flammable liquid gives off sufficient vapor to ignite.

Boiling point (BP) is the temperature at which the vapor pressure of the liquid is equal to the atmospheric pressure under which the liquid vaporizes. Liquids with low boiling point are highly flammable.

Ignition temperature (IT) is the lowest temperature at which a chemical will ignite and burn. The lower the ignition temperature, the faster it will get inflamed.

These flammable or combustible chemicals can react with oxidizers and can cause a fire or explosion resulting in cloudy vapors. The flammable and combustible chemicals should have the label of hazard and all the details of the flash point, boiling point and ignition temperature mentioned. Some flammable chemicals can also be toxic and can cause serious hazards to the lungs if inhaled. The flammable liquids that have a flash point below the room temperature pose severe danger. Eg. Organic solvents like ether, alcohol, Gasoline, acetone, benzene, acetaldehyde, heptanes, hexane etc. are highly flammable. Certain other chemicals can turn toxic when they come in contact with other substances or when it is converted to another substance. Eg. Chloroform which is non-flammable can become flammable when exposed to heat or flame and can produce carbon monoxide, chlorine, phosgene which is highly toxic. Care should be taken while handling volatile and flammable substances; it should always be kept in a secluded place with good ventilation away from other chemicals and used inside a fume hood. Certain flames are invisible to the naked eye, which is difficult to trace and when it comes in contact with an ignition source, can catch fire. The best way to prevent a fire is to avoid any ignition source, reduce oxygen supply and avoid leakage.

Oxidizers

Oxidizers are compounds that lead to the oxidation of a compound; it is an electronegative atom (acceptor) which can accept electrons from another electropositive atom (donor) called as reducers. The electrons are either transferred from atom to another or shared between two atoms with a bond. In other words, oxidizers are chemicals that react with other substances, reducing agents or fuels. The energy released during the transfer of electrons can lead to increase in temperature and can cause an explosion or damage to the surroundings. For eg. When gasoline reacts with air, it can produce a 2000 ^{o}C increase in temperature and lead to a fire or explosion if not utilized properly. Most oxidizers like oxygen, fluorine, chlorine, chlorite, bromated, ozone, permanganate, nitric acid, chromic acid, hydrogen peroxide act in different ways based on their electronegativity.

Reactive chemicals

All chemicals are capable of reacting with other chemicals or the medium in which it is present except the inert ones. When chemicals react readily or vigorously with ordinary unreactive environment such as air, water, cellulose, protein or steel, it is called reactive. Certain chemicals are self-reactive without coming in contact with other substances. These reactive substances can spontaneously generate large quantities of heat, light, gases or toxic chemicals. The reaction rate will increase with increase in pressure, temperature or concentrations. Care should be taken when working with these kind of reactive chemicals such as explosives, acid sensitives, oxidizers, reducers, water sensitives and pyrophosphorics.

Water reactive chemicals

Chemicals that react violently with water, water vapor or moist air to produce heat and flammable gas that can result in ignition are called as water reactive chemicals. Apart from heat, the chemicals can also produce some toxic gases or any other hazardous substances. This is the reason that it is good practice to add acid to water, and not water to acid, which can be explosive. Adding the compound or acid to water can be safe, because the surrounding water will absorb the heat produced during the transfer.

Fire or combustion can take place when water is poured on the material that is highly water reactive, for eg. Sodium reacts with water and is highly explosive. Water reactive materials are alkali metals (lithium, sodium, potassium and their hydrides), calcium oxides, sulfur trioxides, silicon tetrachloride, aluminium chloride, phosphorus pentaoxide, tifluoromethane sulphonic anhydride and chlorosulfonic acid. As water is present everywhere it is very important not to dispose of these chemicals in the water directly, which can be very explosive and leads to high-risk danger to people and property. The possible damage can be spontaneous ignition with production of heat, release of toxic substances etc. It is not necessary to have plenty of water for spontaneous ignition; even a little amount of moisture in the air can penetrate the chemicals that are poorly preserved or not closed properly.

Pyrophosphoric chemicals

 ISBN: 978-93-94198-49-4

Chemicals that ignite on exposure to air at 54° C or below are called as pyrophosphoric chemicals. These pyrophosphoric chemicals can be gas, solid or liquid, volatile or non-volatile. Oxygen reacts with the gas phase of the chemical to result in combustion of compounds like boranes, phosphorus or triethylaluminium, metal powders etc. This is a very rapid reaction and gives enough energy for the ignition. Many metallic powders are pytophosphoric and the reactivity depends on the particle size. Solids may be pyophosphoric when they are finely divided into small particles or when they are volatile. Oxygen can bind to a metal ion, or an unshared electron pair on an alkali; water in air combines with an alkali metal and ignites the hydrogen to cause combustion. An alkali metal binds with water and produces a metal oxide or hydroxide and hydrogen to produce heat for ignition. Many active metals will directly react with oxygen can oxidize completely without causing ignition.

Peroxide forming chemicals

Peroxides or peroxide forming chemicals are sometimes hazardous at certain temperature and pressure. Some chemicals are converted to peroxides when exposed to air during use or storage. Organic peroxides are very unpredictable and can be explosive. Mostly liquids are prone to form peroxides, but sometimes solids and liquefied gas can be converted to peroxides under certain temperature. Peroxides can be highly explosive when shaken vigorously, or during rapid heating or during friction. The molecule can rearrange its structure during initiation and can be converted to non-peroxides or can be polymerized. Aldehydes, amines, mercaptans, sulfides and N-alkyl amides are non-peroxides and are not harmful. Styrene, vinyl acetate and ethyl acetate will be polymerized and converted to plastic when they are exposed to air.

Shock-sensitive and friction-sensitive explosives

Chemicals that can spontaneously explosive when struck, vibrated, dropped, heated or agitated and can releases a lot of energy and are called as shock-sensitive explosives. Shock-sensitive explosives include peroxides, azides, alkyl azides, carbides, acetylides, nitrides, nitrates, fulminates, oxides and hypophosphite. Fiction-sensitive explosives are chemicals that explode during friction caused when opening, spilling and are different from shock-sensitive explosives.

Compressed gases

Compressed gases can be explosive when opened or released accidently and can cause physical and chemical hazards. Gases like arsine, phosphine, phosgene, nitric oxide, chlorine, sulfur tetrafluoride are kept in cylinders; when they are released, dropped or when there is a leakage, it can be lethally flammable and explosive. The gas inside the cylinder is highly compressed at a maximum pressure of 150 atmospheres and can act as lethal projectiles when disturbed.

Cryogenic liquids

A liquid with a boiling point less than or equal to -90 ^{o}C at the atmospheric pressure is called cryogenic liquid. Cryogenic fluids are super-cooled fluids such as liquid air, liquid nitrogen, liquid oxygen are odorless, colorless, tasteless liquids obtained at extremely cold temperature. Cryogenic hydrogen, methane, acetylene, oxygen are flammable gases; oxygen can increase the flammability. Cryogenic liquids are hazardous because of the super-cooled state and are more concentrated than its vapor state and may rapidly evaporate. Cryogenic liquids can cause fire, explosion, asphyxiation, destroy structural materials and tissue destruction. The gas in the liquefied solution may evaporate leaving a dense flammable gas. Cryogenic liquid spill on a person or clothes is highly dangerous and can cause injury on the skin. The clothes could trap more amounts of cold liquid and can be transferred to the skin. Care should be taken while opening the cryogenic fluids; always opening the cryogenic cans in a well-ventilated place may prevent danger.

Health hazards

Chemicals that cause serious hazards on the health of a person can be named as a health hazard. Toxicity is the ability of a chemical substance to cause injury or toxicity when it touches the skin, or inhaled or ingested by the person. Hazard is the probability of the chemical substance that will pose a risk and cause injury when used in an unsafe manner. Many chemicals are used in various places in a day-to-day manner. If these hazardous chemicals are handled properly with utmost care and caution, various damages and health hazards can be avoided.

The health hazard of chemicals depends on the route of entry or exposure, time of exposure, dosage, type and frequency of exposure.

Routes of entry

The toxicity of the chemical can be avoided by using proper preventive and precautionary measures. Care must be taken to prevent the exposure of chemicals to the body.

Inhalation

Breathing or inhaling a chemical is the most common route of entry of the chemical when the chemical is a gas or volatile liquid or solid. People working in asbestos and silica factory are at risk as there is a high possibility of inhaling the chemical substances while breathing. The leakage of the chemicals can be easily detected when they have an odor. Overexposure to these chemicals leads to headache, mucus production, irritation of the eye, nose and throat, dizziness, drowsiness, confusion and collapse. The leakage of chemicals can be avoided by proper closing of the chemicals, proper ventilation and use of fume hoods.

Skin and eye contact

The second route of entry that is very common is the skin and eyes. Chemicals can easily come in contact with the skin and eyes and produce irritation, burn or rash. Chemicals can be absorbed by the skin and enter the bloodstream and lead to toxic conditions, which may affect the internal organs. Chemicals can enter the skin through cuts or lacerations.

Minimum exposure of the skin to chemicals, use of lab coats, gloves, masks, face shield and goggles can prevent the contact of the chemicals to the skin and eyes. If exposed, the chemicals can cause dryness, redness, swelling, blisters, rashes, itching etc. If there is any contact of the chemicals with the skin, the area should be washed with water and immediate medical attention should be given. If the chemicals enter the eye, it can cause itching, redness and even blindness, which should be immediately attended to and necessary first aid given to the person.

Ingestion

Ingestion or swallowing is the next most common route of entry of chemicals into the body. This can happen when we work with chemicals without gloves and do not wash hands after working in the laboratory. We may ingest the chemicals accidentally by eating with hands without washing, contamination of the food or beverages in the laboratory. It is very important to avoid storing food, eating, drinking or smoking in the laboratory. We should take all care to avoid the ingestion of chemicals. In case a person happens to consume any chemical accidently, he should be immediately taken to the hospital and treated.

 ISBN: 978-93-94198-49-4

Injection

Injection is another means of entry of chemicals into the body through the skin. Chemicals accidentally enter our body through needles, broken glasses or sharp objects that have been contaminated with chemicals. Caution should be taken while handling sharp objects in the laboratory. If there is any such incident, the area on the skin should be washed with soap and water and taken to the hospital immediately.

Committees like General safety committee, Chemical safety committee, Institutional biosafety committee, Occupational health committee, Radiation safety committee should be formed to monitor the policies and safety measures of a laboratory.

 ISBN: 978-93-94198-49-4

CHAPTER-III

ANALYTICAL DATA

 ISBN: 978-93-94198-49-4

INTRODUCTION TO ANALYTICAL DATA

Analytical data is a collection of data which is used to support decision making or research. They are recorded qualitative and/or quantitative results of a chemical, physical, biological, microbiological, or radiological determination. In other words It is the process of examining data sets in order to find trends and draw conclusions about the information they contain.

EVALUATION OF ANALYTICAL DATA

Evaluation of analytical data is the aspect which can answer for the quality assurance. The question of accuracy must, in general, collect attention both before and after an analysis. Evaluation is the integral and major part of Analytical method where statistical tests are done to determine confidence limits in acquired data. A number of statistical techniques are used to analyze the data. Descriptive statistical techniques are used to handle the statistical data effectively.

The purpose of this unit is to provide adequate information to enable examining the factors affecting the reliability of results and understand the contributions of errors, their types,

 ISBN: 978-93-94198-49-4

minimization of errors for accuracy and precision of the measurement and the proper use of significant figures.

IMPORTANCE AND STEPS INVOLVED IN EVALUATION OF DATA:

Evaluation provides us a systematic approach to study a program, observe, intervention, or initiative to understand how well it achieves its goals. Evaluations help establish what works well and what could be improved in a program or initiative.

Steps Involved:

ANALYTIC AL PROCESS

DEFINE PROBLEM

SELECT METHOD

OBTAIN SAMPLE

PERFORM MEASUREMENTS

ANALYZE DATA

PRESENT RESULTS

Key concepts of this unit:

Precision and Accuracy Reliability

Errors and its types

Statistical techniques such as mean, median, mode & Standard deviation &

Normal Distribution curve

PRECISION AND ACCURACY

Precision and accuracy are two ways that scientists reflect about error. Accuracy states to how close a measurement is to the true or accepted value. Precision defines to how close measurements of the same item are to each other. Precision is independent of accuracy which means it is possible to be very precise but not very accurate, and it is also possible to be accurate without being precise. If it is both accurate and precise, we can say that it is of the best quality scientific observations.

Accuracy and precision are only two essential concepts used in scientific measurements. Two other vital aids are significant figures and scientific notation. Scientists use percent error as one method of describing how accurate and precise a value is. It's a simple but useful calculation.

EXAMPLE

Accurate and precise:

- If a weather thermometer reads 75oF outside and it really is 75oF, the thermometer is accurate.

- If the thermometer consistently registers the exact temperature for several days in a row, the thermometer is also precise.

Precise, but not accurate:

- A refrigerator thermometer is read ten times and registers degrees Celsius as: 39.1, 39.4, 39.1, 39.2, 39.1, 39.2, 39.1, 39.1, 39.4, and 39.1.
- However, the real temperature inside the refrigerator is 37 degrees C.
- The thermometer isn’t accurate (it’s almost two degrees off the true value), but as the numbers are all close to 39.2, it is precise.

RELIABILITY

 ISBN: 978-93-94198-49-4

Research reliability means to whether research methods can reproduce the same results repeatedly. If research methods can produce steady results, then the methods are likely consistent and not inclined by external factors.

ASSESSING THE RELIABILITY IN RESEARCH

To assess whether the research methods are producing reliable results, either we have to carry out the same task many times or in many ways.

For example,

Performing the same experiment by different groups of people. Performing different experiments by the same group of people.

Both methods keep control by having one element exactly the same and changing other elements to confirm other factors don't impact the research results.

If a researcher repeats an experiment with a different set of people or a different batch of the same chemicals and gets very alike results then those results are said to be reliable. Reliability is measured by a percentage – if the obtained results are exactly the same resulting every time, then they are 100% reliable.

Quality and reliability are vital terms in chemistry. Reliable analytical information implies quality. Maintaining the quality in time shows reliability. Estimation of quality and reliability is important for the validation of a new method of analysis.

Learning good practice in chemistry will improve the quality of the laboratories of analysis, by implementing highly reliable analytical methods.

To ensure quality and reliability, sampling is one the most important tool in the analytical process. The sampling plan must support the goals of the analysis. For example, a material scientist interested in describing a metal's surface chemistry is more likely to choose a freshly exposed surface, created by cleaving the sample under a vacuum, than a surface previously exposed to the atmosphere. In a qualitative analysis, a sample is not required to be identical to the original substance provided there is sufficient analyte present to ensure its detection. In fact, if the goal of an analysis is to identify a trace-level component, it may be desirable to discriminate against major components when collecting samples.

ERRORS

If we measure the same object two different times, the two measurements may not be exactly the same. That difference between two measurements is called a variation or error.

In quantitative analysis, the final result depends on the various data observed throughout the experiment, and sometimes it might be wrong if there are any errors in the data considered for the experiment. These errors occur due to the apparatus or instruments used in experiments or due to the wrong observation carried out. If we measure the same object two different times, the two measurements may not be exactly the same. That difference between two measurements is called a variation or error.

Important Types of Error

Depending on the origin of errors they can be categorized into

1. Determinate errors or systematic errors
2. Random errors or indeterminate errors
3. Gross Errors

DETERMINATE ERRORS

Determinate errors are ones that can affect the results in a series of determinations with a particular sample or with a particular observer or with a particular method. Causes of determinate errors are Instrument errors, Method errors, and Personal errors.

Instrument errors can be corrected by calibration or proper instrumentation.

Method errors can be eliminated with proper method development.

Personal errors can be minimized with proper training and experience.

Example

Errors in reading the burette Errors in pipetting solutions Use of impure reagents RANDOM ERRORS

Any repeated calculation provides slightly differed values, even when carried out with maximum care and with similar conditions. It does not have any fixed value or sign which is known as random error or indeterminate error. These errors are due to limitations in making the observation; lack of care in making measurements etc., these errors may be eliminated by using high precision instruments with utmost care. Random or indeterminate errors cannot be completely eradicated.

For Eg. Setting the same weight on electronic scales many times and obtaining the results which vary in a random fashion from one reading to the next. The difference between this reading s and the actual weight correspond to the random error of the scale measurements.

GROSS ERRORS

Easily recognized since they involve a major breakdown in the analytical process (samples being spilt, wrong dilution, or wrong using of instrument), it is detectable by carrying out sufficient replicate measurements.

For example, writing down a value of 100 when the reading was actually 1.00 is a gross error.

EXPRESSION OF ERROR

The error can be expressed as a numerical difference between the measured value and the true value.

Error = True value ~ Measured value.

Absolute error (E): It is the difference between the measured value and the true value

$E = X_i – X_t$ where X_i is the measured value and X_t is the true value.

Relative error: It is the ratio of the absolute error to the true value.

$E_r = (X_i – X_t)/ X_t$.

The relative error is usually expressed in terms of a percent or parts per thousand.

$E_r = (X_i – X_t) \times 100 / X_t$. or $E_r = (X_i – X_t) \times 1000 / X_t$.

ERROR HANDLING:

Usually, they occur with statistical distribution and are treated by statistical methods. An experiment designed to reveal the presence of errors can be performed. Statistical tests can be applied to the data to provide an estimate of the likely value of that error.

Uses of Statistics in data evaluation in Chemistry:

One of the uses of statistics in analytical chemistry is therefore to provide an estimate of the likely value of that error;

in other words, to establish the uncertainty associated with the measurement

Definition of some statistical terms:

In statistics, we lean towards representing a set of data by a representative value which would approximately define the entire collection. This representative value is called the measure of central tendency, and the name suggests that it is a value around which the data is centered. These central tendencies are mean, median, and mode.

Mean is the most generally used measure of central tendency. It is the average of the given collection of data. It is applicable for both continuous and discrete data.

It is equal to the sum of all the values in the collection of data divided by the total number of values.

Suppose we have n number values in a set of data namely as x1, x2, x3… xn, then the mean of data is given by:

$\bar{x}$=(x1+x2+x3+……..+x) / n

Data may be given in raw form or tabular form. Let's find the mean in both cases.

Raw Data

Let x1, x2, x3 , . . . , xn be n observations.

We can find the arithmetic mean using the mean formula.

Mean, $\bar{x}$ = (x1 + x2 + ... + xn) / n

Example: 1.

If the heights of 5 people are 142 cm, 150 cm, 149 cm, 156 cm, and 153 cm. Find the mean height.

Solution:

Mean height, $\bar{x}$ = (142 + 150 + 149 + 156 + 153) / 5

= 750 / 5

= 150

Mean, $\bar{x}$ = 150 cm

Thus, the mean height is 150 cm.

Example: 2.

Find the mean of the first 10 odd integers.

Solution:

First 10 odd integers: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 Mean = Sum of the first 10 odd integers/Number of such integers

= (1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19) /10

= 100 /10

= 10

Frequency Distribution (Tabular) Form

When the data is present in tabular form, we use the following formula:

Mean, $\bar{x} = (x1f1 + x2f2 + ... + xnfn) / (f1 + f2 + ... + fn)$

Example 3:

Find the mean of the following distribution:

X	4	6	9	10	15
F	5	10	10	7	8

Solution:

Calculation table for arithmetic mean:

xi	fi	xifi
4	5	20

6	10	60
9	10	90
10	7	70
15	8	120
	$\sum fi = 40$	$\sum xi\ fi = 360$

Mean, $\bar{x} = (\sum xi\ fi) / (\sum fi)$

= 360 / 40

= 9

Thus the solution is 9

Example 4:

Here is an example where the data is in the form of class intervals. The following table indicates the data on the number of customers visiting a car service centre in a month. Find the average number of customers visiting the service centre in a day.

Number of customers Number of days visiting service centre

0-10 2

10-20 6

20-30 9

30-40 7

40-50 4

50-60 2

Solution:

In this case, we find the mid-point or class mark of a class for each class.

Note: Class mark = (lower limit + upper limit)/2

Let x1, x2, x3 , . . . , xn be the class marks of the respective classes.

Hence, we get the following table:

Class mark (xi) frequency (fi)

xifi

5	2	10
15	6	90
25	9	225
35	7	245
45	4	180
55	2	110

Total

$\sum fi = 30$ $\sum fixi = 860$

Mean, $\bar{x} = (\sum xifi)/(\sum fi)$

$= 860/30$

$= 28.67$

$\bar{x} = 28.67$

Example 5

Consider the following frequency distribution. Calculate the mean weight of students.

Weight (in kg) 31-
35 36 –
40 41 –
45 46 –
50 51 –
55 56 –
60 61 –
65 66 –
70 71 –
75

Number of Students 9 6 15 3 1 2 2 1
1

Solution:

The given distribution has discontinuous class intervals, so we need to make them continuous.

Class intervals	Number of students (fi)	Class mark (xi)	di = xi – a	fidi
30.5 – 35.5	9	33	-10	-90
35.5 – 40.5	6	38	-5	-25
40.5 – 45.5	15	43 = a	0	0

45.5 – 50.5	3	48	5	15
50.5 – 55.5	1	53	10	10
55.5 – 60.5	2	58	15	30
60.5 – 65.5	2	63	20	40
65.5 – 70.5	1	68	25	25
70.5 – 75.5	1	73	30	30
Total	∑fi = 40			∑fidi = 35

Here, ∑fi = 40 and ∑fidi = 35

By Assumed mean method, Mean = a + (∑fidi/∑fi)

= 43 + (35/40)

= 43 + 0.875

= 43.875

Therefore, the mean weight of the students is 43.875 kg.

MEDIAN

The value of the middlemost observation, obtained after arranging the data in ascending or descending order, is called the median of the data.

For example, consider the data: 4, 4, 6, 3, 2. let's arrange this data in ascending order: 2, 3, 4, 4, 6. There are 5 observations. Thus, median = middle value i.e. 4.

Case 1: Ungrouped Data

- Step 1: Arrange the data in ascending or descending order
- Step 2: Let the total number of observations be n.

To find the median, we need to consider if n is even or odd. If n is odd, then use the formula:

Median = (n + 1)/2th observation

Example 6: Let's consider the data: 56, 67, 54, 34, 78, 43, and 23. What is the median?

Solution:

Arranging in ascending order, we get: 23, 34, 43, 54, 56, 67, and

78. Here, n (number of observations) = 7 So, (7 + 1) / 2 = 4

∴ Median = 4th observation Median = 54

 ISBN: 978-93-94198-49-4

If n is even, then use the formula:

Median = [(n/2)th obs.+ ((n/2) + 1)th obs.] / 2.

Example 7: Let's consider the data: 50, 67, 24, 34, 78 and 43. What is the median?

Solution:

Arranging in ascending order, we get: 24, 34, 43, 50, 67 and 78. Here, n (no. of observations) = 6
6/2 = 3

Using the median formula, Median = (3rd obs. + 4th obs.)/2
= (43 + 50) / 2

Median = 46.5

Case 2: Grouped Data

When the data is continuous and in the form of a frequency distribution, the median is found as shown below:

 ISBN: 978-93-94198-49-4

Step 1: Find the median class.

Let n = total number of observations i.e. ∑ fi

Note: Median Class is the class where (n/2) lies.

Step 2: Use the following formula to find the median.

Median = l + ((n/2 – c)/f) x h where,

- l = lower limit of median class
- c = cumulative frequency of the class preceding the median class
- f = frequency of the median class
- h = class size

Let's consider the following example to understand this better.

Example: 8. Find the median marks for the following distribution:

Classes	0-10	10-20	20-30	30-40	40-50
Frequency	2	12	22	8	6

Solution:

We need to calculate the cumulative frequencies to find the median.

Calculation table:

Classes	Number of students	Cumulative frequency
0-10	2	2
10-20	12	2 + 12 = 14
20-30	22	14 + 22 = 36
30-40	8	36 + 8 = 44
40-50	6	44 + 6 = 50

N = 50

N/2 = 50/2 = 25

 ISBN: 978-93-94198-49-4

Median Class = (20 - 30)

l = 20, f = 22, c = 14, h = 10

Using Median formula:

Median = l + ((n/2 – c)/f) x h

= 20 + (25 - 14)/22 × 10

= 20 + (11/22) × 10

= 20 + 5 = 25 ∴ Median = 25

Example 9:

If the median of a distribution given below is 28.5, then find the value of x and y.

CI 0 –

10 10 –

20 20 –

30 30 –

40 40 –

50 50 –

 ISBN: 978-93-94198-49-4

60 Total

Frequency 5 x 20 15 y 5 60

Solution:

From the given, N/2 = 60 / 2 = 30

Median of the given data = 28.5

Median class is 20 – 30 with a cumulative frequency = 25 + x.

Lower limit of median class = l = 20

Frequency of the median class = f = 20

Cumulative frequency of the class preceding the median class = cf = 5 + x

Class size = h = 10

Median = l + ((n/2 – c)/f) x h

$28.5 = 20 + [(30 - 5 - x)/ 20] \times 10$

$28.5 - 20 = (25 - x)/2$

$8.5 \times 2 = 25 - x$

$17 = 25 - x$

$x = 25 - 17$

Implies x = 8

Also,

$60 = 5 + 20 + 15 + 5 + x + y$

$60 = 5 + 20 + 15 + 5 + 8 + y$

$y = 60 - 53$

 ISBN: 978-93-94198-49-4

Implies y = 7

Solution is x = 8 and y = 7.

MODE

The value which appears most often in the given data i.e. the observation with the highest frequency is called a mode of data.

Case 1: Ungrouped Data

For ungrouped data, we just need to identify the observation which occurs maximum times.

Mode = Observation with maximum frequency

For example in the data: 6, 8, 9, 3, 4, 6, 7, 6, 3, the value 6 appears the most number of times. Thus, mode = 6. An easy way to remember mode is: Most Often Data Entered.

Note: A data may have no mode, 1 mode, or more than 1 mode. Depending upon the number of modes the data has, it can be called unimodal, bimodal, trimodal, or multimodal.

The example discussed above has only 1 mode, so it is unimodal.

Case 2: Grouped Data

 ISBN: 978-93-94198-49-4

When the data is continuous, the mode can be found using the following steps:

- Step 1: Find modal class i.e. the class with maximum frequency.
- Step 2: Find mode using the following formula:

Mode = l + [(fm – f1)/(2 fm – f1 – f2)] x h where,

- l = lower limit of modal class,
- fm = frequency of modal class,
- f1 = frequency of class preceding modal class,
- f2 = frequency of class succeeding modal class,
- h = class width

Consider the following example to understand the formula.

Example: 10. Find the mode of the given data:

Marks Obtained

0-20 20-

40

40-60 60-

80 80-

100

Number of students

5

10

12

6

3

Solution:

The highest frequency = 12, so the modal class is 40-60. l = lower limit of modal class = 40

fm = frequency of modal class = 12

f1 = frequency of class preceding modal class = 10 f2 = frequency of class succeeding modal class = 6 h = class width = 20

Using the mode formula,

Mode = l + [(fm – f1) / (2 fm – f1 – f2)] x h

$= 40 + [(12-10)/(2\times12-10-6)]\times20$

$= 40 + (2/8) \times 20$

$= 45$

$\therefore$ Mode = 45

Relation between Mean, Median, and Mode

The three measures of central values i.e. mean, median, and mode are closely connected by the following relations called an empirical relationship

2Mean + Mode = 3Median

For instance, if we are asked to calculate the mean, median, and mode of continuous grouped data, then we can calculate the mean and median using the formulas as discussed in the previous sections and then find the mode using the empirical relation.

Example11:

If we have data whose mode = 65 and median = 61.6 then find the mean.

We can find the mean using the above mean, median, and mode relation.

2Mean + Mode = 3 Median

$\therefore$2Mean = 3 × 61.6 - 65

$\therefore$2Mean = 119.8

 ISBN: 978-93-94198-49-4

⇒ Mean = 119.8/2

⇒ Mean = 59.9

Example: 12.

If the mean of the given frequency distribution is 35, then find the missing frequency y. Also, calculate the median and mode for the distribution.

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	2	4	7	y	1

Solution:

Class	Frequency (fi)	Class mark (xi)	fixi
10 – 20	2	15	30
20 – 30	4	25	100
30 – 40	7	35	245
40 – 50	y	45	45y
50 – 60	1	55	55
Total	∑fi = 14 + y		∑fixi = 430 + 45y

 ISBN: 978-93-94198-49-4

As we know, Mean = ∑fixi / ∑fi

Given that the mean of the distribution is 35. So, (430 + 45y) / (14 + y) = 35

430 + 45y = (14 + y) 35

430 + 45y = 490 + 35y

45y – 35y = 490 – 430

10y = 60

y = 6

Thus, the missing frequency is 6.

Now, we can calculate the mode as follows:

Class	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	2	4 = f0	7 = f1	6 = f2	1

Here, l = 30

h = 10

Mode = l + [(f1 – f0) / (2f1 – f0 – f2)] × h Mode = 30 + [(7 – 4) / (14 – 4 – 6)] × 10

= 30 + (30/4)

= 30 + 7.5

= 37.5

Now, using the formula Mode = 3Median – 2 Mean, we can get the value of median.

37.5 = 3 Median – 2 (35) 3 Median = 37.5 + 70 Median = 107.7 / 3 = 35.9

STANDARD DEVIATION

Standard deviation is the positive square root of the variance. Standard deviation is one of the basic methods of statistical analysis. Standard deviation is commonly abbreviated as SD and denoted by 'σ' and it tells about the value and how much it has deviated from the mean value.

Steps to Calculate Standard Deviation

- Find the mean, which is the arithmetic mean of the observations.
- Find the squared differences from the mean. (The data value - mean)2

- Find the average of the squared differences. (Variance = The sum of squared differences ÷ the number of observations)
- Find the square root of variance. (Standard deviation = √Variance)

The formula for Calculating Standard Deviation

The population standard deviation formula is given as:

- σ=√

N

∑ (Xi−μ)2 which ∑ varies from 1 to N

Here,

- σ = Population standard deviation
- μ = Assumed mean

Similarly, the sample standard deviation formula is:

- s =√

 ISBN: 978-93-94198-49-4

$n-1$

$\sum (Xi-\overline{X})2$ where $\sum$ varies from 1 to n

Here,

s = Sample standard deviation

$\overline{X}$ = Arithmetic mean of the observations

Example: 13.

Let's calculate the standard deviation for the data given below:

xi	6	10	12	14	24
fi	2	3	4	5	4

Calculate mean ($\overline{X}$) : (6+8 +10+12+ 14) / 5 = 10

xi	fi	fixi	xi- $\overline{X}$		$(xi-\overline{X})2$	$fi(xi-\overline{X})2$
6	2	12	-4	16	32	
8	3	24	-2	4	12	
10	4	40	0	0	0	
12	5	60	2	4	20	
14	4	56	4	16	64	
	18	192			128	

N = 18, $\sum$ fixi = 192, $\sum fi(xi-\bar{x})2 = 128$

Calculate variance: $\sigma 2 = 1$

N

$\sum (Xi-\mu)2$ where $\sum$ varies from 1 to N

$= 1/18 \times 128 = 7.1$

Calculate SD: $\sigma = \sqrt{Variance} = \sqrt{7.1} = 2.66$

Example 14:

During a survey, 6 students were asked how many hours per day they study on an average. Their answers were as follows: 2, 6, 5, 3, 2, 3. Evaluate the standard deviation.

Solution:

Find the mean of the data: (2+6+5+3+2+3)/6

= 3.5

Step 2: Construct the table:

x1	x1 − x̄	(x1 − x̄)2
2	-1.5	2.25
6	2.5	6.25
5	1.5	2.25
3	-0.5	0.25
2	-1.5	2.25
3	-0.5	0.25

= 13.5

Step 3: Now, use the Standard Deviation formula

Sample Standard Deviation = s =√ 1

n−1

∑ (Xi−$\overline{X}$)2 where

∑ varies from 1 to n. Implies

=√(13.5/[6-1])

=√[2.7] s =1.643

NORMAL DISTRIBUTION CURVE

A normal distribution which is also known as the Gaussian distribution is a

The probability distribution that is symmetric about the mean, shows that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

Characteristics of a normal distribution

The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side.

- The mean, median, and mode are all equal.
- The curve is known to be symmetric at the center, which is around the mean.
- Exactly 1/2 of all the values are known to be to the left of the center whereas exactly half of all the values are to the right of the center.
- The area under the normal distribution curve represents probability and the total area under the curve sums to one.

DISTRIBUTION OF RANDOM ERRORS

Random errors are distributed normally which means it follows a normal distribution

 ISBN: 978-93-94198-49-4

P(x) = 1

$\sigma\sqrt{2\pi}$

μ = Mean

$-1/2(x-\mu)$

$e \quad \sigma$

σ = Standard deviation

x = Normal random variable

Example: 15

Find the probability density function for the normal distribution where mean = 4 and standard deviation = 2 and x = 3.

Solution:

Given:

Mean, μ = 4

Standard deviation, σ = 2

Random variable, x = 3.

We know that the normal distribution formula is:

 ISBN: 978-93-94198-49-4

P(x) = 1

$\sigma\sqrt{2\pi}$

$-1/2(x-\mu)$

$e \quad \sigma$

Now, substitute the values in the formula, we get

f(x)=0.19947 × e−0.125 f(x)=0.19947 × 0.882496 f(x)=0.17603

Therefore, the probability density function for the normal distribution is 0.17603.

For a perfectly normal distribution the mean, median and mode will be the same value, visually represented by the peak of the curve.

The normal distribution has

Mean = Median = Mode

Symmetric about the center

50% of values less than the mean and 50% greater than the mean.

REFERENCES:

1. Random Errors in Chemical Analysis , Source slideplayer.com/Fundamentals of Analytical Chemistry, F.J. Holler, S.R.Crouch.

2. Analysis of random errors in physical measurements: A PRIMER, Dr. Pramod Joag,Department of Physics, University of Pune.

3. Analytical Chemistry, David Harvey, Source URL: http://www.asdlib.org/onlineArticles/ecourseware/Analytical %2 0Chemistry%202.0/Text_Files.html

5. Chemical Analysis, A series of Monographs on Analytical Chemistry and its Applications, J.D. Winefordner, Volume 153, A Wiley – Inter Science Publication.

6. Using Data Analysis to Evaluate and Compare Chemical Syntheses, Dustin Kaiser, Jianbo Yang, ACS publications.

7. Evaluation of Water quality in the Tigris River within Baghdad, Iraq using Multivariate Statistical Techniques Salwan Ali Abed, Salam Hussein Ewaid and Nadhir Al-Ansari, Journal of Physics.

ISBN: 978-93-94198-49-4

CHAPTER–IV

STATISTICAL TREATMENT OF FINITE SAMPLES

 ISBN: 978-93-94198-49-4

STATISTICAL TREATMENT OF FINITE SAMPLES:

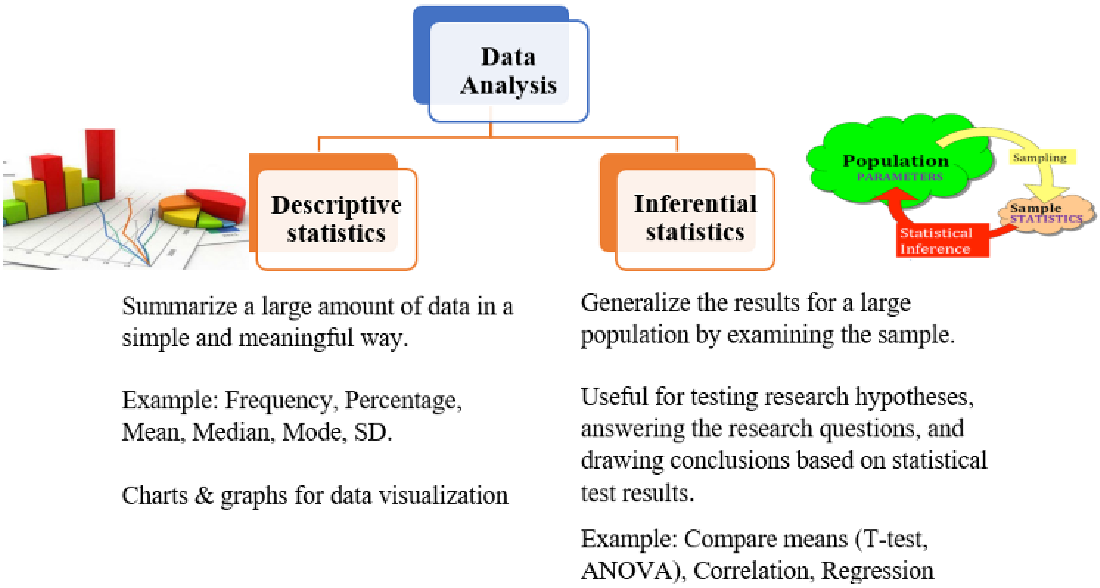
Statistical treatment of data is when you apply some form of statistical method to a data set to transform it from a group of meaningless numbers into meaningful output. For example, consider a medical study that is investigating the effect of a drug on the human population. As the drug can affect different people in different ways based on parameters such as gender, age and rac e, the researchers would want to group the data into different subgroups based on these parameters to determine how each one affects the effectiveness of the drug. Categorizing the data in this way is an example of performing basic statistical treatment.

In another example, a researcher wants to investigate the biochemical content (total proteins, cellulose, lignin, and starch) in plant leaves related to their optical properties through statistical relationships. The collection of data based on sample selection from dry or fresh leaves, their biochemical contents along with physical measurements (blade thickness, water content, specific leaf area) will help in inferring the results for whole population.

A sample is a subset (a finite number of measures) selected from the population, which is the entire set of measurements (which can be infinitely vast).

Greek letters (such as μ, s) are used by statisticians for population parameters, while English letters (such as x and s) are used for finite samples.

Statistical treatment can be either descriptive statistics, which describes the relationship between variables in a population, or inferential statistics, which tests a hypothesis by making inferences from the collected data.



Several types of descriptive measures can be computed from a set of data. In this chapter, however, we limit discussion to measures of central tendency and measures of dispersion. Measures of central tendency convey information regarding the average value of a set of values. The three most commonly used measures of central tendency are the mean, the median, and the mode.

Mean: The *mean* is the most widely used measure of the central value. The mean of a finite number of measurements, x_1, x_2, x_3, x_4, ..., x_n, is commonly represented as *x*. Population mean is represented by μ	$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 \ldots\ldots + x_n}{n} = \frac{\sum_{i=1}^{i=n} x_1}{n}$
Median: It represents the middle score of the distribution and relatively unaffected by extreme scores. The value of the **middlemost observation**, obtained after arranging the data in ascending or descending order, is called the **median** of the data.	If **n is odd** Median = $(n + 1)/2^{th}$ observa If **n is even** Median = $[(n/2)^{th}$ obs.+ $((n/2) + 1)^{th}$ obs.
Mode It is the most frequently occurring value in	Mode = 3 Median

 ISBN: 978-93-94198-49-4

a set of data and particularly useful in the study of popular sizes. Mode is the average to be used to find the ideal size in a series.	– 2 Mean
Average Deviation (or Mean Deviation): It is the average of the differences between the individual results and the mean. It is regarded as a measure of variability. In the case of a small number of observations the average deviation is found to be not quite significant statistically.	$\bar{d} = \frac{\sum_{i=1}^{i=n}[x_1 - \bar{x}]}{n}$
Standard Deviation: It is the distance from the mean to the point of inflexion of the normal distribution curve. In comparison to the average deviation the 'standard deviation' is usually considered to be much more useful and meaningful statistically. For a finite number of values it is normally symbolised as 'S'. The standard deviation is related to normal distribution in the following way. The mean	$S = \sqrt{\frac{\sum_{i=1}^{i=n}[i - \bar{x}]^2}{n-1}}$

 ISBN: 978-93-94198-49-4

plus or minus one standard deviation includes 68% of the population. The mean ±2SD includes 95% of the population and mean ±3SD includes 99% of the population.	
Coefficient of Variation (v) The coefficient of variation (v) is simply the standard deviation(s) expressed as a percentage of the mean (x')	$v = \frac{s}{\bar{x}} \times 100$
Variance (s^2) The spread of the data set — how far apart the numbers are in relation to the mean The variance is the square of the standard deviation(s) *i.e.*, s^2	$s^2 = \frac{\sum (X - \bar{X})^2}{N}$

Example: Calculate central tendency measures and deviations for following data set:

30,55,65,60,25,45,35,70,60,50

 ISBN: 978-93-94198-49-4

- Mean = 30 + 55 + 65 + 60 + 25 + 45+ 35+ 70 + 60 + 50/10 = 49.5

For Median, first arrange data into descending or ascending

25,30,35,45,50,55,60,60,65,70

- Median = (50+55)/2= 52.5

- Mode = 60

- Average deviation = Sum of differences/ number of sample

  = [-19.5 +5.5 + 15.5 + 10.5+ (-24.5) + (-4.5) + (-14.5) + 20.5+ +10.5 + 0.5] / 10

  = 12.6

- Standard deviation (sample) = $\sqrt{}[(30-49.5)^2+(55-49.5)^2+(65-49.5)^2+(60-49.5)^2+(25-49.5)^2+(45-49.5)^2+(35-49.5)^2+(70-49.5)^2+(60-49.5)^2+(50-49.5)^{2]}$ / (10-1)

Standard deviation = 15.3568

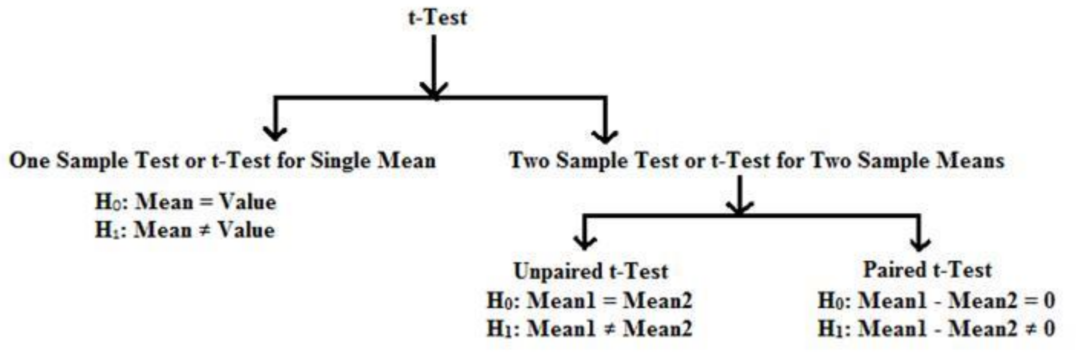
- Variance = $(\text{Standard deviation})^2 = (15.3568)^2 = 235.83$

- Coefficient of variation = (15.3568/ 49.5) *100 = 31.02

STUDENT T-TEST:

William Sealy Gosset (a brewmaster) had invented a new method for determining how large a sample of persons should be used in the taste-testing of beer. In 1908, he published his findings in the journal Biometrika under the pseudonym 'student.' This is how t- test got name as 'student's t. T-test is a method of testing hypothesis about the mean of small sample drawn from a normally distributed population when the standard deviation for the sample is unknown. It can also be used to test the difference between the mean of two sets of data's ($\overline{\ }_1$ and $\bar{x}_2$). The purpose of the test is to compare the mean of samples with some standard value and to express some level of confidence in the significance of comparison. It helps us understand if the difference between two sample means is real or simply due to chance.

Degree of freedom: It may be defined as the number of individual observations that could be allowed to vary under the condition that $\bar{x}$ and s, once determined, be held constant.



Important Formula with calculations:

I. **One-sample t-test:** The one-sample t-test is a statistical method for determining whether a sample's mean deviates significantly from the population mean. Sample should be continuous variable and normally distributed.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Where:

- $\bar{x}$ is sample mean
- μ is population mean
- s sample standard deviation
- n is number of observations

Problem statement: The cholesterol levels (mg/100 of serum) in a group of 7 individuals after taking a drug are 231, 258, 245,199, 195, 208 and 205. Find whether the drug is effective in reducing the cholesterol level in man. The normal cholesterol level in human is 190mg/100 of serum.

Hypothesis:

H_0 : μ = 190mg/100 of serum

H_1 : **μ** ≠ 190mg/100 of serum

Decision rule: the degree of freedom is (n-1) = 6. The t table value at 0.05 alpha level is ±2.45. The null hypothesis will be rejected if T observed value less than -2.45 and greater than +2.45.

T = (220.14 – 190) / (24.58/√7) = 3.2446

Result: The observed value, t = 3.245, is greater than the critical table value, t = +2.45. The null hypothesis is rejected. There is a significant difference in the mean cholesterol levels of sample mean from normal cholesterol levels. As the mean sample value is higher, inferring that drug is not effective in reducing the cholesterol levels.

 ISBN: 978-93-94198-49-4

II. **Independent t-test or Unpaired t -test:** is an inferential statistic that tests if there is a statistically significant difference between the means of two unrelated (independent) groups. This test is performed using a categorical variable with two classes (Grouping variable) and a continuous normally distributed variable (Test variable).

Example: The unpaired T-test would be used to determine if there is a significant difference between the control and treated enzyme activities.

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Where:

- **x1** is the mean of sample 1
- **s1** is the standard deviation of sample 1
- **n1** is the sample size of sample 1
- **x2** is the mean of sample 2
- **s2** is the standard deviation of sample 2
- **n2** is the sample size in sample 2

Problem statement: Transverse diameter measurements on the hearts of adult males and females gave the following results:

Group	Sample size	$\bar{x}$ (cm)	S (cm)
Male	15	13.21	1.05

 ISBN: 978-93-94198-49-4

Female	12	11.00	1.01

Hypothesis:

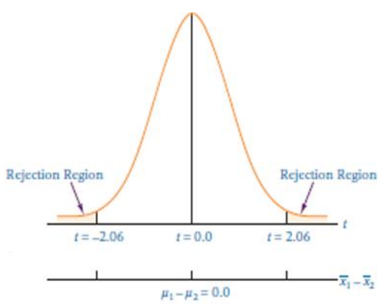
$H_0: \mu_1 - \mu_2 = 0$
$H_a: \mu_1 - \mu_2 \neq 0$

Decision rule: Because the hypotheses are = and ≠ this test is two tailed. The degrees of freedom are 25 (15 + 12 - 2 = 25) and alpha is .05. The t table requires an alpha value for one tail only, and, because it is a two-tailed test, alpha is split from .05 to .025 to obtain the table t value: $t_{.025,25} = \pm 2.060$.

The null hypothesis will be rejected if the observed t value is less than -2.060 or greater than +2.060.

	One-sided α											
	.25	.10	.05	.025	.01	.005	.0025	.001	.0005	.00025	.0001	.00005
	Two-sided α											
	.50	.20	.10	.05	.02	.01	.005	.002	.001	.0005	.0002	.0001
df												
1	1.00	3.08	6.31	12.71	31.82	63.66	127.32	318.31	636.62	1273.24	3183.10	6366.20
2	.82	1.89	2.92	4.30	6.96	9.22	14.09	22.33	31.60	44.70	70.70	99.99
3	.76	1.64	2.35	3.18	4.54	5.84	7.45	10.21	12.92	16.33	22.20	28.00
4	.74	1.53	2.13	2.78	3.75	4.60	5.60	7.17	8.61	10.31	13.03	15.54
5	.73	1.48	2.02	2.57	3.37	4.03	4.77	5.89	6.87	7.98	9.68	11.18
6	.72	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96	6.79	8.02	9.08
23	.68	1.32	1.71	2.07	2.50	2.81	3.10	3.49	3.77	4.05	4.42	4.69
24	.68	1.32	1.71	2.06	2.49	2.80	3.09	3.47	3.75	4.02	4.38	4.65
25	.68	1.32	1.71	2.06	2.49	2.79	3.08	3.45	3.73	4.00	4.35	4.62
26	.68	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71	3.97	4.32	4.59

Table available at https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/



$t = (13.21 - 11)/\sqrt{[(1.05)^2/15] + (1.01)^2/12]} = 5.5261$

 ISBN: 978-93-94198-49-4

Result: The observed value, t = 5.5261, is greater than the critical table value, t = +2.06, the observed value of t is in the rejection region. The null hypothesis is rejected. There is a significant difference in the mean scores of the two tests.

III. **Paired T-test or Dependent samples t-test :** Used to test if there is a statistically significant difference between the means of two sets of paired observations. The same subjects are measured twice, or observed in two different ways, in this experiment. This test requires the use of paired variables (pre and post observations of the same participants), which must be continuous and normally distributed.

Example: The paired T-test would be used to determine if there is a significant difference between the pre- and post-treatment blood pressures.

$$t = \frac{(\sum D)/N}{\sqrt{\frac{\sum D^2 - (\frac{(\sum D)^2}{N})}{(N-1)(N)}}}$$

Where:

- ΣD is the sum of the differences
- D is difference per paired value
- N is number of samples

Degree of freedom (df) = n-1

 ISBN: 978-93-94198-49-4

Problem statement: The table below shows the hours of relief provided by two analgesic drugs in 5 patients suffering from arthritis. Is there any evidence that one drug provides longer relief than the other?

Patients	Drug A- Relief in hours	Drug B- Relief in hours	D	D^2
1	3	5	2	4
2	3.6	5.7	2.1	4.41
3	7.4	10.2	2.8	7.84
4	7	6.5	-0.5	0.25
5	2	4	2	4
	Mean = 4.6	Mean = 6.28	$\Sigma D = 8.4$ $\Sigma D/N = 1.68$	$\Sigma D^2 = 20.5$

Hypothesis: The relief provided by Drug A and B is equal.

Decision rule: The degree of freedom = n-1 = 5-1 = 4. T table value at df=4 at 0.05 alpha = 2.78

T = 1.68 / √[20.5 – (8.4)2/5)] / 4*5 = 1.68/0.565 = **2.97**

Result: The observed value, t = 2.97, is greater than the critical table value, t = 2.78, the observed value of t is in the rejection region. The null hypothesis is rejected. There is a significant difference in the mean scores relief in hours patients are getting. Based on the mean value it can be concluded that **Drug B** is more effective compared to drug A in providing longer hours relief.

F-TEST

A statistical technique used to compare the means between three or more groups is known as ANOVA or F test. The F test is based on the null hypothesis that the two population variances under consideration are equal.

$$F = \frac{V_1}{V_2} = \frac{s_1^2}{s_2^2}$$

Decision rule:

To test null hypothesis for the two sets of data by F test, the calculated value of F is compared with the corresponding tabular value of F. On comparison

 ISBN: 978-93-94198-49-4

(i) If Fcalc < Ftab,

that is if the experimental F is smaller than the corresponding tabular value of F, then no statistically significant difference is indicated between s1 and s2 (i.e. between two sets of data), and the null hypothesis is valid, and

(ii) If Fcalc > Ftab, then s1 is significantly greater than s2, and the null hypothesis is not valid.

For example, suppose in two series of observations, one of 6 observations having standard deviation 0.04 and another of 8 observations of standard deviation 0.06. We have to test whether there is significant difference between the two standard deviations.

In above case consider greater standard deviation is s_1= 0.06 and s_2= 0.04. Hence, degree of freedom (df1) =8-1= 7 and df2= 6-1= 5.

$F_{calc} = V1/V2 = s_1^2/s_2^2 = (0.06)^2/ (0.04)^2 = 2.25$

$F_{tab} =$ (at $\alpha = 0.05$) = 4.88

 ISBN: 978-93-94198-49-4

The F - Distribution with $\alpha = 0.05$

$v_2 \backslash v_1$	2	3	4	5	6	7	8	9
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

The above findings proved that Fcalc < Ftab, it is inferred that null hypothesis is valid and there is no statistically significant difference between the standard deviations of the two sets of data or statistically no significant difference is observed between the precisions of the two sets of data.

One way ANOVA:

For current study we will illustrate one way ANOVA. The One-way ANOVA is also called a single factor analysis of variance because there is only one independent variable or factor. The One-way ANOVA is extension of independent samples *t* test where, means are compared among three or more independent groups. It is not advisable to run multiple t-tests as there is always the potential for making a Type I error. The typical margin of error is 5%. The probability of "making a mistake" doubles to 10% when two t-tests are performed on the same data.

For one way ANOVA, the dependent variable should measure on continuous scale and one categorical independent variable

 ISBN: 978-93-94198-49-4

has at least three categories. For example, reduction in weight using three diets (Diet A, B & C).

Post-hoc test in ANOVA: One-way ANOVA is an omnibus test statistic; therefore it does not tell you which groups were statistically different from one other; instead, it shows you that there were at least two groups with different means. Whether your study has three, four, five, or more groups, it is crucial to identify the differences between them. A post hoc test will allow you to do this.

$$F = \frac{\text{Estimate of population variance based on between samples variance}}{\text{Estimate of population variance based on within samples variance}} = \frac{MS \text{ between}}{MS \text{ within}}$$

Problem statement: The Chloride content (mg/lit) of a river at a particular site collected from three different seasons are given below. Find the mean and standard deviation. Analyze the significance of variance. Infer the variance in oxygen content for different seasons.

Hypothesis:

H_0 : μ1 = μ2 = μ3 (“ The mean Chloride content of all the 3 seasons are equal.”)

 ISBN: 978-93-94198-49-4

Winter	Monsoon	Post monsoon
70	68	65
74	62	87
61	78	92
53	85	58
82	72	73
$\bar{X}_1$ = 68	$\bar{X}2$ = 73	$\bar{X}3$ = 75
Grand mean = 68+73+75/3 = 72		

Step 2: Calculate SSC (sum of square column/between)

$$SS \text{ between} = n_1\left(\bar{X}_1 - \bar{\bar{X}}\right)^2 + n_2\left(\bar{X}_2 - \bar{\bar{X}}\right)^2 + n_3\left(\bar{X}_3 - \bar{\bar{X}}\right)^2$$

SSC = 5 $(68-72)^2$ + $5(73-72)^2$ + $5(75-72)^2$ = 130

Step 3: Calculate SSW (Sum of squares for variance within samples (or *SS* within).

$$SS \text{ within} = \Sigma\left(X_{1i} - \bar{X}_1\right)^2 + \Sigma\left(X_{2i} - \bar{X}_2\right)^2 + \ldots + \Sigma\left(X_{ki} - \bar{X}_k\right)^2$$
$$i = 1, 2, 3, \ldots$$

 ISBN: 978-93-94198-49-4

SSW = { $(70-68)^2+(74-68)^2+(61-68)^2+(53-68)^2+(82-68)^2$} + {$(68-73)^2+(62-73)^2+(78-73)^2+(85-73)^2+(72-73)^2$} + {$(65-75)^2+(87-75)^2+(92-75)^2+(58-75)^2+(73-75)^2$ }

SSW= 1652

Step 4: Sum of square Total (SST)

SS for total variance = *SS* between + *SS* within = 1782

Step 5: Degree of Freedom and Mean sum of square

df numerator = k-1= 3-1 = 2, df denominator = n-k = 15-3 = 12, df total = n-1=14

MS between = SS between/ df numerator = 130/2 =65

MS within = SS within/ df denominator = 1652/12 = 137.67

F-test = MSbetween/Mswithin

F = 65/137.67 = 0.472

F-table value at 5% = 3.8853

	DF1	α = 0.05				
DF2	1	2	3	4	5	6
1	161.45	199.5	215.71	224.58	230.16	233.99
2	18.513	19	19.164	19.247	19.296	19.33
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.866
8	5.3177	4.459	4.0662	3.8379	3.6875	3.5806
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738
10	4.9646	4.1028	3.7083	3.478	3.3258	3.2172
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905

Interpretation: The F calculate value is 0.47 which is less than the table value of 3.89 at 5% level with d.f. being $v1 = 2$ and $v2 = 12$ and hence could have arisen due to chance. This analysis supports the null hypothesis of no difference in sample means. We may, therefore, conclude that the difference in Chloride content in river for three seasons is insignificant and is just a matter of chance.

CRITERIA FOR REJECTION OF AN OBSERVATION

One or more values in a set of replicate measurements of a physical or chemical quantity may deviate significantly from the majority of the others. In such a circumstance, there is always good reason to exclude the outlying numbers from further analysis (for example, when calculating the mean value or the standard deviation). Only questionable values that can be "legitimately" labelled as outliers are eligible for this treatment.

Outliers are observations from a different model or distribution from the main "body" of data. Rejecting suspicious observations must always be done based on hard data and not on gut feelings. The use of reliable statistical tests for "outlier detection" is one method for accomplishing this.

Q-test is a statistical tool used to identify an outlier within a data set. Dixon’s Q test, is a way to find outliers in very small, normally distributed, data sets.

The rejection quotient is defined as the ratio of the divergence of the doubtful value from its nearest neighbor vwhen the values are arranged in a sequence. If the value of Q is greater than the Q value given in the table at the desired confidence level for a given number of observations the suspect value is rejected.

$$Q_{exp} = \frac{x_2 - x_1}{x_n - x_1}$$

x1 is the smallest (suspect) value, x2 is the second smallest value, and xn is the largest value.

Example: Is 152 an outlier in this set of data? Test at the 95% confidence Level (i.e. at an alpha level of 5%).

162, 166, 152, 174, 170, 165

The Q test is applied as follows:

Step 1: Arrange observations in ascending or descending order.

152, 162, 165, 166, 170, 174

Step 2: Insert the value in formula

Q = (162 – 152) / 174 – 152 = 10/22 = 0.455.

 ISBN: 978-93-94198-49-4

Step 3: Find the Q critical value in the Q table. For sample 6 at 5% alpha level Q critical =0.625

Critical Rejection Values for Identifying an Outlier: Q-test

	Q_{crit}		
N	90% CL	95% CL	99% CL
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

Step 4: Compare the Q statistic from Step 2 with the Q critical value in Step 3. If the Q statistic is greater than the Q critical value, the point is an outlier. Qstatistic =0.455.

Qcritical value = 0.625.

Solution: 0.455 is not greater than 0.625, so this point is not an outlier at an alpha level of 5%.

REGRESSION ANALYSIS

Regression analysis is a way of predicting an outcome variable (Y) from one predictor variable (simple regression) or several predictor variables (multiple regression). The best-fitting line is called a *regression line*.

 ISBN: 978-93-94198-49-4

Regression analysis is used to estimate the relationship that exists, on the average, between the dependent variable and the explanatory variable. It also determines the effect of each of the explanatory variables on the dependent variable, controlling the effects of all other explanatory variables. Further, Predict the value of dependent variable for a given value of the explanatory variable.

Regression analysis aims to create a model describing a set of experimental x and y data and to predict unknown x values using created model. This generates some estimation errors. For a given set of data there may be a large number of regression models, but in order to obtain reliable results, the model showing the smallest deviation from experimental data should be chosen.

Important Formulas:

 ISBN: 978-93-94198-49-4

$$\hat{y} = b_0 + b_1 x$$

Where

b_0 = the sample intercept
b_1 = the sample slope

$$b_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$b_0 = \bar{y} - b_1\bar{x} = \frac{\Sigma y}{n} - b_1\frac{(\Sigma x)}{n}$$

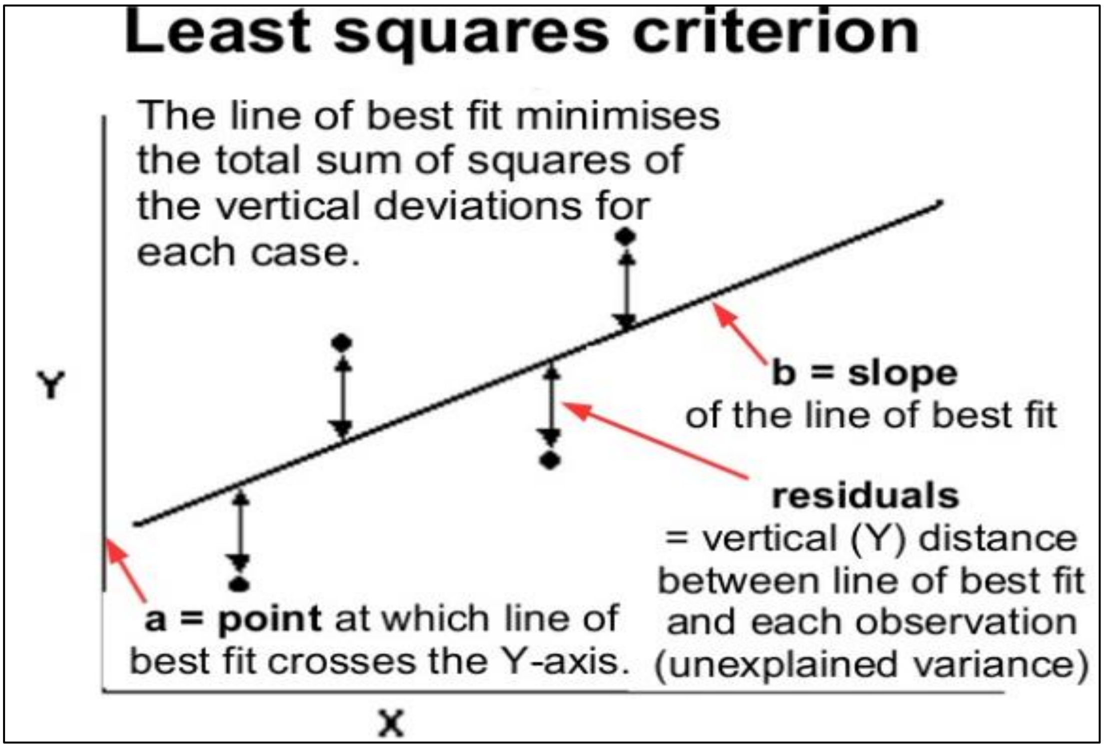
$$SS_{xy} = \Sigma(x - \bar{x})(y - \bar{y}) = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$$SS_{xx} = \Sigma(x - \bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

To determine the equation of the regression line for a sample of data, the researcher must determine the values for *b*0 and *b*1. This process is sometimes referred to as least squares analysis.

Least squares analysis is *a process whereby a regression model is developed by producing the minimum sum of the squared error values.* Based on this premise and calculus, a particular set of equations has been developed to produce components of the regression model.

 ISBN: 978-93-94198-49-4



Examine the regression line fit through the points in above Figure. Observe that the line does not actually pass through any of the points. The vertical distance from each point to the line is the error of the prediction. In theory, an infinite number of lines could be constructed to pass through these points in some manner. The least squares regression line is the regression line that results in the smallest sum of errors squared.

Problem statement: Formulate regression equation for the following data:

 ISBN: 978-93-94198-49-4

S.No.	X	Y	X^2	XY
1	29	69	841	2001
2	35	95	1225	3325
3	42	118	1764	4956
4	50	125	2500	6250
5	64	158	4096	10112
	ΣX=220	ΣY=565	ΣX^2=10426	ΣXY=26644

Using these values, the researcher solved for the sample slope (b_1) and the sample y-intercept (b_0).

$SS_{XY} = \Sigma XY – (\Sigma X) (\Sigma Y)/N = 26644 - (220) (565)/5 = 1784$

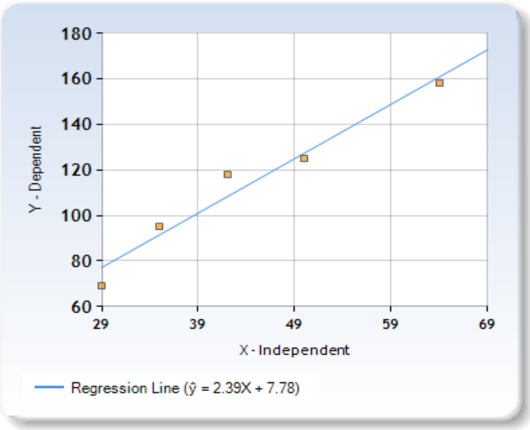
$SS_{XX} = \Sigma X^2 – (\Sigma X)^2/N = 10426 – (220)^2/5 = 746$

$b_1 = SS_{XY} / SS_{XX} = 2.39142$

$b_0 = \Sigma Y/N - b_1 (\Sigma X/N) = 113 - (2.39*44) = 7.77748$

The least squares equation of the regression line is

$\hat{y}$ **= 2.39142*X* + 7.77748**



SIGNIFICANT FIGURES

The number of significant figures in a result is simply the number of figures that are known with some degree of reliability. The number of significant digits, or "significant figures," in a scientifically expressed coefficient of expression ranges from zero to nine. How confidently or precisely an engineer or scientist expresses a number is shown by the number of significant figures included in the expression.

Analytical concentration reports need a sufficient power-resolution, which is achieved by collecting a sufficient number of relevant data points. The number of significant digits needed can be calculated using several different approaches or criteria. In most cases, three key figures are sufficient.

Accuracy refers to how close the measured value is to the true of accepted value.

Precision refers to how close the measurements are to each other.

To determine the number of significant figures in a number use the following rules:

Rule 1: Non-zero digits are always significant

Example	No. of significant digits
2.4	2
426	3
389.32	5
2419	4
1236.47	6

Here we can see any digit that is a non-zero digit is considered significant. This is true both in the presence or absence of a decimal.

Rule 2: Any zeros between two significant digits are significant

Example	No. of significant digits
4.02	3
1025	4
140079	6
3209.3	5
429.23004	8

Rule 3: Zeros before a decimal point that precede non-zero digits are not significant

Example	No. of significant digits
001.02	3

0031.256	5
0425.4	4

Rule 4: Zeros to the right of a decimal after a non-zero number are significant

Example	No. of significant digits
1.40	3
0.007	1
8.2000	5
0.0040	2

Significant Figures in Calculations

- For addition and subtraction, the answer should have the same number of decimal places as the term with the fewest decimal places.
- For multiplication and division, the answer should have the same number of significant figures as the term with the fewest number of significant figures.

Addition and Subtraction. 21.753 + 3.48 + 2.25694 = 27.48994

- With significant figures it is **27.49** since **3.48** has 2 decimal places, which is the least number of decimal places.

Multiplication and Division. 57.837/0.36 = 160.658333

- With significant figures, the final value should be reported as 1.6×10^2 since 0.36 has only 2 significant figures. Notice that 160 would be ambiguous, so scientific notation is necessary in this situation.

Rules for Rounding off numbers

If the digit to be dropped is greater than 5, the last retained digit is increased by one. Round 5.8475 to 4 significant figures: 5.8475 becomes 5.848.

If the digit to be dropped is less than 5, the last remaining digit is left as it is. Round 3.7464 to 4 significant figures: 3.7464 is 3.746.

References:

Analysis of biochemical data. https://dspace.cvut.cz/bitstream/handle/10467/85627/F3-D-2019-Anyz-Jiri-dizertace_anyz.pdf?sequence=-1&isAllowed=y.

https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/.

BIOSTATISTICS A Foundation for Analysis in the Health Sciences.

Black, K. (2019). Business statistics: for contemporary decision making. John Wiley & Sons.