Lecture Plan		
Duration hours	Topics to be covered	Support materials
Unit I		
1 hour	Overview	Tl- 293
	Colorimetry - spectra	
1 hour	Beer-Lambert law	Tl- 302
1 hour	Photoelectric colorimeter - principle, instrumentation	
1 hour	Spectroscopy - electromagnetic radiation - properties	Tl- 287,293
1 hour	UV -Vis Spectrophotometer, Spectrofluorimeter	Tl-307,312
1 hour	Atomic absorption spectroscopy	Tl- 354-357
1 hour	Maldi-TOF	
1 hour	Mass Spectroscopy	Tl- 368-369
1 hour	GC-MS, IR	
1 hour	FTIR	Tl- 340-342
	Total hours -10	
Unit II		
1 hour	Centrifugation - principles, types	
1 hour	Analytical and preparative - applications	Tl- 113-116,128-130
1 hour	Density gradient & Ultra centrifuge	TI-118-119
1 hour	Chromatography - types	Tl- 185
1 hour	Paper chromatography, TLC	TI-186-193
1 hour	Ion exchange and affinity chromatography	Tl- 203-209
1 hour	Gel filteration chromatography	Tl- 211-213
1 hour	HPLC&HPTLC	Tl- 227-237
1 hour	Unit I & II Possible questions discussion	
1 hour	Class test	
	Total hours -9	
Unit III		
1 hour	Electrophoresis - Principle, instrumentation, AGE - Application	Tl- 239-249
1 hour	SDS-PAGE	Tl- 254
1 hour	Native PAGE, Iso electric Focussing	Tl- 257-259
1 hour	Immuno, Pulsefield, Gel capillary electrophoresis	Tl- 263,250,243
1 hour	2D Gel electrophoresis	
1 hour	Gel documentation system	Tl- 260,271

1 hour	Class test	
	Total hours - 9	
Unit IV		
Ihour	Data collection - Classification - Presentation - tabulation	T2-31-39
1 hour	Measure of central tendency - mean	T2-121,123,228
1 hour	Median	
1 hour	Mode	
1 hour	Measure of disperssion - mean deviation	T2-229-260
	Standard deviation, standard error	T2-271-315
1 hour	Analysis of variance	
1 hour	Unit possible question disscussion	
	Total hours -9	
Unit V		
1 hour	Application of Biostatistics	T2- 676,679
1 hour	Probability & Distribution	
1 hour	Correlation & Regression	T2-359-321
	Theorems	T2-679-722
1 hour	Binomial, Normal, poison distribution	T2- 723-760
1 hour	Correlation coefficient	T2-359-321
1 hour	Simple and linear regression analysis	T2-359-321
1 hour	Test for significance - F, t test	T2-795-800
1 hour	DMRT & Chi-square test	T2-790-808
1 hour	Statistical & graphical softwares	T2-843-849
1 hour	Previous Year End semester question paper discussion	
1 hour	Previous Year End semester question paper discussion	
1 hour	Previous Year End semester question paper discussion	
	Total hours-10	

References

Tl) L.Veerakumari, Bioinstrumentation, 2009, impression MJP publishers, chennai.

T2) R.S.N. Pillai & Bagavathi, ,Statistics, 9th edition, S.Chand & company Ltd.

Semester -I

18BTP104 BIOINSTRUMENTATION AND BIOSTATISTICS 4H-4C

Total hours/week: **L: 3 T:1 P:0** **Marks:** Internal:**40** External:**60** Total: **100**

Course Objectives:

To impart technical information on Instrumentation related to Biotechnology and statistical analysis.

Course Outcomes:

This paper makes the student to understand the working principles of instruments and vital statistical techniques for analysis of biological data used in modern biology and biotechnology research.

UNIT - I

Colorimetry: Color and absorption spectra, Beer's and Lambert's law. Principle of photoelectric colorimeter, Spectroscopy – Properties of electromagnetic radiations, Instrumentation and applications of – UV Visible light spectroscopy, Spectrofluorimeter, atomic spectroscopy, NMR spectroscopy and MALDI –TOF, Mass spectroscopy GC – MS, IR and FTIR.

UNIT - II

Centrifugation: Principle, types of centrifuges, Principles and applications of analytical- and preparative centrifuge, density gradient and ultra centrifuge. **Chromatography:** Principles, Type – Paper, thin layer, ion-exchange, affinity, gel filtration, HPLC and HPTLC

UNIT - III

Electrophoresis: Principle, instrumentation and applications of agarose gel electrophoresis, sodium dodecyl sulphate – polyacrylamide gel (SDS-PAGE), native PAGE, isoelectric focusing, immuno, pulsefield, gel, capillary, 2D electrophoresis, gel documentation.

UNIT- IV

Biostatistics: Data collection, classification and presentation of tabulation. Measures of central tendency – mean, median and mode. Measures of dispersion – mean deviation, standard deviation, standard error and analysis of variance.

UNIT- V

Applications of biostatistics: Probability and probability distribution – theorems, binomial, poisson and normal distribution. Correlation and regression – simple correlation, correlation co-efficient, simple and linear regression analysis. Test of significance -F, t, DMRT and chi-square test. Randomized block design. Statistical and graphical software.

References

Glover,T., & Mitchell, H. (2008). *An Introduction to Biostatistics*. (2nd ed.) Boston: Mc Graw- Hill Co. Inc.
Friedfelder,D. (2001). *Physical Biochemistry* (5th ed.). New York: Oxford Publishers.

Sharma, B.K. (2004). *Instrumental Methods of Chemical Analysis* (24th ed.). Meerut: Goel Publishing House.

Chatwal, G.R., & Anand, S.K. (2003). *Instrumental Methods of Chemical Analysis* (5th ed.). Mumbai: Himalaya Publishing House.

Boyer, R. (2000). *Modern Experimental Biochemistry* (3rd ed.). New Delhi: Addision Wesley Longman.

Sawhney, S.K., & Singh,R. (2000). *Introductory practical Biochemistry*. New Delhi: Narosa Publishing House.

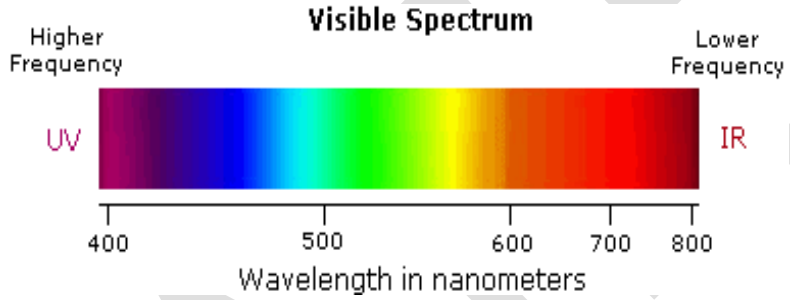
Wilson, K., & Walker, J. (2006). *Principles and Techniques of Biochemistry and Molecular Biology*. (7 th ed.) India : Cambridge University Press.

Sawhney, S.K., & Singh, R. (Eds.). (2005). *Introductory Practical Biochemistry*. Alpha Science International Ltd.

Colorimetry:

Colors and absorption Spectra

The electromagnetic spectrum is comprised of all known types of electromagnetic radiation, including X-rays and radio waves, and is organized by wavelength, which is determined by the energy of the radiation. Shorter wavelengths correspond to higher energy radiation. Visible light is the very small portion of the spectrum (380 nm to 800 nm) that the human eye is sensitive to. The colors of most substances arise from the interaction between visible light and the electrons within molecular bonds. Specifically, the light at specific wavelengths is absorbed, causing these electrons to occupy different energy levels.



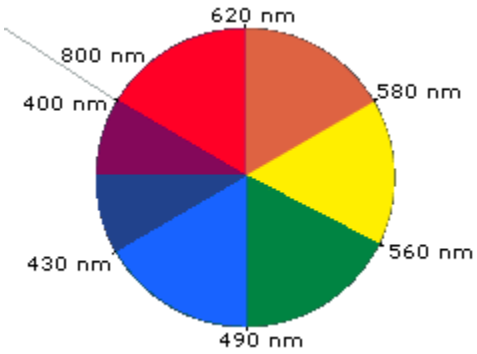
Wavelengths that the molecule doesn't absorb are transmitted, or reflected, to the observer's eye as the perceived color of the molecule, especially when it is dissolved in a transparent solvent such as water. For example, nickel sulfate, NiSO4, dissolved in water appears greenish-blue because it transmits those wavelengths associated with blue-green-yellow portion of the visible spectrum. All other visible wavelengths are absorbed.

- **Violet:** 400 - 420 nm
- **Indigo:** 420 - 440 nm
- **Blue:** 440 - 490 nm
- **Green:** 490 - 570 nm
- **Yellow:** 570 - 585 nm
- **Orange:** 585 - 620 nm
- **Red:** 620 - 780 nm

When white light passes through or is reflected by a colored substance, a characteristic portion of the mixed wavelengths is absorbed. The remaining light will then assume the complementary color to the wavelength(s) absorbed. The color wheel shown on the right demonstrates this relationship. Here, complementary colors are diametrically opposite each other. Thus, absorption of 420-430 nm lights

renders a substance yellow, and absorption of 500-520 nm lights makes it red. Green is unique in that it can be created by absoption close to 400 nm as well as absorption near 800 nm.

Early humans valued colored pigments, and used them for decorative purposes. Many of these were inorganic minerals, but several important organic dyes were also known. These included the crimson pigment, kermesic acid, the blue dye, indigo, and the yellow saffron pigment, crocetin. A rare dibromo-indigo derivative, punicin, was used to color the robes of the royal and wealthy. The deep orange hydrocarbon carotene is widely distributed in plants, but is not sufficiently stable to be used as permanent pigment, other than for food coloring. A common feature of all these colored compounds, displayed below, is a system of extensively conjugated pi-electrons.



The operation of colorimeters is based on the property of colored solutions of absorbing light passing through them. The absorption increases with increasing concentration *c* of the coloring substance. All colorimeter measurements are taken in monochromatic light in the region of the spectrum that is most strongly absorbed by the particular substance in the solution and most weakly absorbed by the solution's other components. Therefore, colorimeters are equipped with light filters; the use of various light filters with narrow spectral ranges for transmitted light makes possible separate determination of the concentration of the various components of the same solution.

Colorimeters are divided into visual and objective (photoelectric) types. In visual colorimeters the light passing through the solution being measured illuminates one part of the field of vision, and light passing through a solution with a known concentration of the same substance is incident on the other part. By changing the thickness *I* of the layer of one of the solutions being compared or the intensity *I* of the light beam, the viewer attempts to make the color tones of the two halves of the field of vision indistinguishable to the eye. The concentration of the solution under study may be determined from the known relationships for *I*, *l*, and *c* (the Bouguer-Lambert-Beer law).

Beer's Law

According to Beer's law when monochromatic light passes through the colored solution, the amount of light transmitted decreases exponentially with increase in concentration of the colored substance.

$$I_t = I_o^{e\text{-}KC}$$

Lambert's Law

According to Lambert's law the amount of light transmitted decreases exponentially with increase in thickness of the colored solution.

$$I_t = I_o^{e\text{-}kt}$$

Therefore, together Beer-Lambert's law is:

$$I_E/I_o = e^{-KCT}$$

where,

I_E = intensity of emerging light

I_o = intensity of incident light

e = base of neutral logarithm

K = a constant

C = concentration

T = thickness of the solution

Photoelectric Colorimeter

Photoelectric colorimeters give higher accuracy of measurement than the visual type. Selenium and vacuum photocells, photomultipliers, photoresistors, and photodiodes are used as light detectors in photoelectric colorimeters. The strength of the photocurrent of the detectors is determined by the intensity of the incident light and thus by the extent of absorption of the light beam in the solution (absorption increases with increasing concentration). In addition to photoelectric colorimeters with reading of the photocurrent strength, compensating colorimeters are also common; in this type the difference in the signals corresponding to the standard and test solutions is set to zero (compensated) by an electric or optical compensator (for example, a photometer wedge). The reading in this case is taken from the compensator scale. Compensation makes possible minimization of the effect on accuracy of measurement conditions, such as temperature and instability of properties of the colorimeter's parts. The readings of colorimeters do not immediately give concentration values for the substances studied. Conversion to concentration values requires the use of calibration curves obtained by measuring solutions of known concentrations.

Colorimeter measurements are relatively simple and rapid. The accuracy of such measurements is often not less than that of more complicated methods of chemical analysis. The lower limits for determinable concentration are 10^{-3} to 10^{-18} moles per liter, depending on the type of substance.

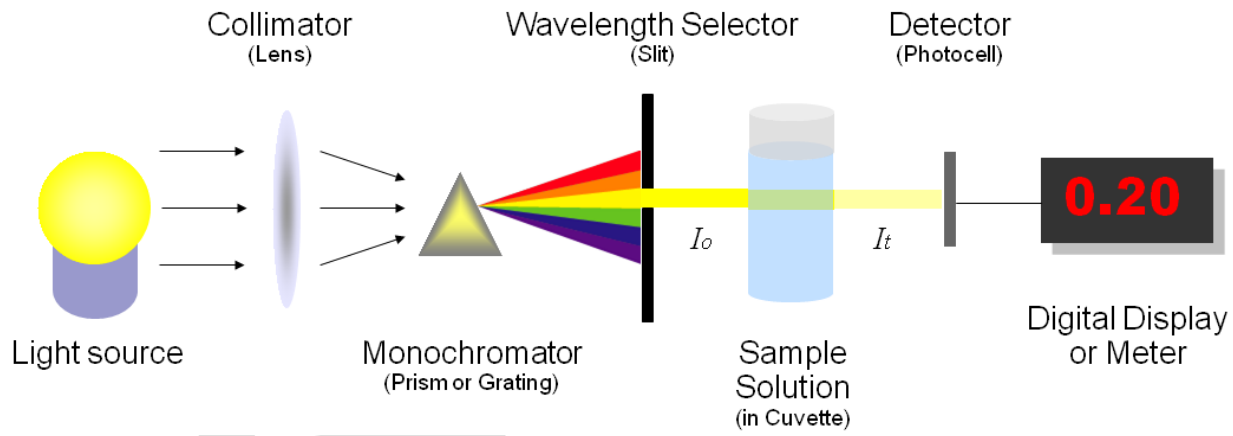
Principles:

Colorimetry is the techniques that is frequently used in biochemical investigations. This involves the quantitative estimation of colors. This measure the quantity of a substance in a mixture, you could use the technique of colorimetry, by allowing the substance to bind with color forming chromogens. The difference in color results in the difference in the absorption of light, which is made use of here in this technique called colorimetry.

- Light from a suitable source is passed through a light filter to select the most appropriate wavelength of light, some of which is then absorbed by the solution held in a special glass cuvette (a sort of 'test tube').
- The amount of light absorbed is called, and measured as, the absorbance which is a function of the coloured solute concentration.
- Most expensive instruments use a double beam system of two cuvets, one is a 'blank' of water and one the actual coloured solution under test, two photocells and sophisticated optics of lenses and mirrors which need not concern as at all.
- Cheaper colorimeters (i.e. in school and illustrated above) allow you to put in a cuvet of 'colourless' water, zero the instrument i.e. set it to read zero absorbance, replace with a cuvet of the coloured solution and simply read of the 'absorbance'.

- The 'zeroing' is necessary because even the apparently 'colourless blank' of glass cuvet and water can absorbed a tiny amount of light. This procedure eliminates this error.
- The filter is chosen to select the band of wavelengths which are most strongly absorbed by the coloured solution e.g. this is illustrated on the diagram above, and in the table below, by using a yellow filter to use in measuring the concentration of a blue coloured solution like copper (II) sulphate or its ammine/amine complex.
- Although the table illustrates the 'complementary' colour relationship between the solution and the filter, in practice it is better to try several filters on a typical concentration of the solution under test to see which filter gives the highest absorption value i.e. gives you maximum sensitivity and hence maximum accuracy in your measurements.

The difference in color results in the difference in the absorption of light, which is made use of here in this technique called colorimetry.



Instrumentation

The instrument use for colorimetry is colorimeter. This appartus will comprise of the following parts:

1. light source
2. filter (the device that selects the desired wavelength)
3. cuvette chamber (the transmitted light passes through compartment wherein the solution containing the colored solution are kept in cuvette, made of glass or disposable plastic)
4. detector (this is a photosensitive element that converts light into electrical signals)
5. Galvanometer (measures electrical signal quantitatively)

Steps for operating the photoelectric colorimeter:

- Choose the glass filter recommended (see table below) in the procedure and insert in the filter.
- Fill two of the cuvette with blank solution to about three-fourth and place it in the cuvette slot.
- Switch on the instrument and allow it to warm up for 4 – 5 minutes.
- Adjust to zero optical density.
- Take the test solution i another cuvette and read the optical density.
- Take the standard solution in varying concentration and note down the optical density as S1, S2, S3, S4, S5 and so on.
- A graph is plotted taking concentration of standard solution versus the optical density.
- From the graph the concentration of the test solution or the unknown solution can be calculated.

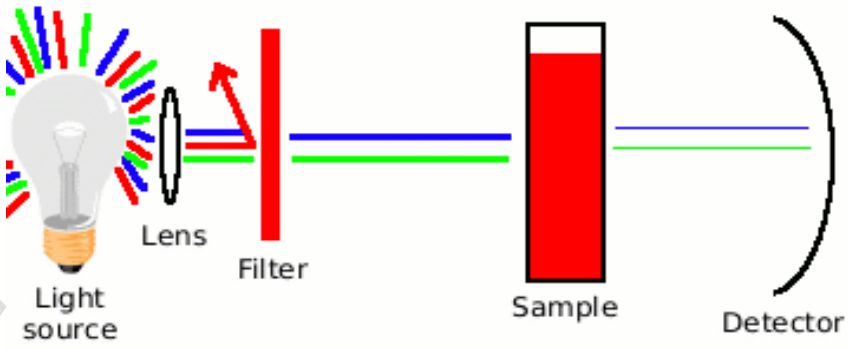


Table for choosing the wavelength of absorption:

Si. No.	Color of the Solution	Colour Absorbed	Wavelength of Absorption
1.	Yellow to Green	Violet	400 nm – 435 nm
2.	Yellow to Orange	Blue	435 nm – 490 nm
3.	Red	Blue to Green	490 nm – 500 nm
4.	Purple	Green	500 nm – 560 nm
5.	Violet	Yellow to Green	560 nm – 580 nm
6.	Blue to Green	Yellow to Orange	580 nm – 650 nm
7.	Bluish Green	Red	650 nm – 700 nm

SPECTROSCOPY

Properties of Electromagnetic Radiation

Electromagnetic radiation is characterized by a broad range of wavelengths and frequencies, each associated with a specific intensity (or amplitude) and quantity of energy. This interactive tutorial explores the relationship between frequency, wavelength, and energy, and enables the visitor to adjust the intensity of the radiation and to set the wave into motion.

The tutorial initializes with a visible light wave appearing in the window having a wavelength of 650 nanometers (red light) and amplitude of 61 candelas. Energies associated with waves in the tutorial appear beneath the window and are given in units of kJ/mole.

To adjust the wavelength (and simultaneously, the frequency) of the wave, translate either the Wavelength or Frequency sliders to the left or right. As the sliders are relocated, the new values for wavelength and frequency appear above the sliders, the wave color changes to match the value for visible light associated with the wavelength, and the energy associated with the wave appears beneath the tutorial window. The amplitude of the wave can be adjusted with the Amplitude slider, and the resulting intensity values will appear above the slider, measured in units of candelas. In order to stop propagation of the wave, click on the Propagation Stop button located in the lower right-hand side of the tutorial window. The wave can be restarted by again clicking on the button, which changes into a Start button when the wave is halted.

An electromagnetic wave moves or propagates in a direction that is at right angles to the vibrations of both the electric and magnetic oscillating field vectors, carrying energy from its radiation source to undetermined final destination. The two fields are mutually perpendicular. By convention, and to simplify illustrations, the vectors representing the electric and magnetic oscillating fields of electromagnetic waves are often omitted, although they are understood to still exist.

Whether transmitted to a radio from the broadcast station, heat radiating from the oven, furnace or fireplace, X-rays of teeth, or the visible and ultra-violet light emanating from the sun, the various forms of electromagnetic radiation all share fundamental wave-like properties. Every form of electromagnetic radiation, including visible light, oscillates in a periodic fashion with peaks and valleys, and displaying a characteristic amplitude, wavelength, and frequency that defines the direction, energy, and intensity of the radiation.

The standard unit for all electromagnetic radiation is the magnitude of the wavelength (in a vacuum), which is usually reported in terms of nanometers for the visible light portion of the spectrum. Each nanometer represents one-thousandth of a micrometer, and is measured by the distance between two successive peaks (see Figure 1). The corresponding frequency of the radiation wave, the number of

sinusoidal cycles (oscillations or complete wavelengths) that pass a given point per second, is proportional to the reciprocal of the wavelength. Frequency is usually measured in Hertz (Hz) or cycles per second (cps). Thus, longer wavelengths correspond to lower frequency radiation and shorter wavelengths correspond to higher frequency radiation.

The different wavelengths and frequencies of various forms of electromagnetic radiation are fundamentally similar in that they all travel at the same speed--about 186,000 miles per second (approximately 300,000 kilometers per second), commonly known as the speed of light (and identified with the variable **c**). Electromagnetic radiation (including visible light) travels 149 million kilometers (93 million miles) from the sun to Earth in about 8 minutes.

The wavelength of light, and all other forms of electromagnetic radiation, is related to the frequency by a relatively simple equation:

$$n = c/l$$

wherec is the speed of light (measured in meters per second), n is the frequency of the light in hertz (Hz), and l is the wavelength of the light measured in meters. From this relationship one can conclude that the wavelength of light is inversely proportional to frequency. An increase in frequency produces a proportional decrease in the wavelength of light with a corresponding increase in the energy of the photons that compose the light. Upon entering a new medium (such as glass or water from air), the speed and wavelength of light is reduced, although the frequency remains unaltered.

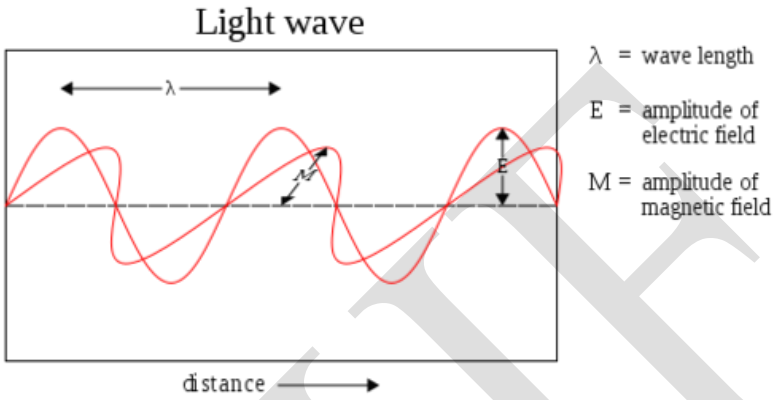
Electromagnetic radiation (EMR) is a form of energy that is produced by oscillating electric and magnetic disturbance, or by the movement of electrically charged particles traveling through a vacuum or matter. The electric and magnetic fields come at right angles to each other and combined wave moves perpendicular to both magnetic and electric oscillating fields thus the disturbance. Electron radiation is released as photons, which are bundles of light energy that travel at the speed of light as quantized harmonic waves. This energy is then grouped into categories based on its wavelength into the electromagnetic spectrum. These electric and magnetic waves travel perpendicular to each other and have certain characteristics, including amplitude, wavelength, and frequency.

General Properties of all electromagnetic radiation:

1. Electromagnetic radiation can travel through empty space. Most other types of waves must travel through some sort of substance. For example, sound waves need either a gas, solid, or liquid to pass through in order to be heard.

2. The speed of light is always a constant. (Speed of light : 2.99792458 x 10^8 m s^{-1})
3. Wavelengths are measured between the distances of either crests or troughs. It is usually characterized by the Greek symbol \(\lambda\).

Waves and their Characteristics



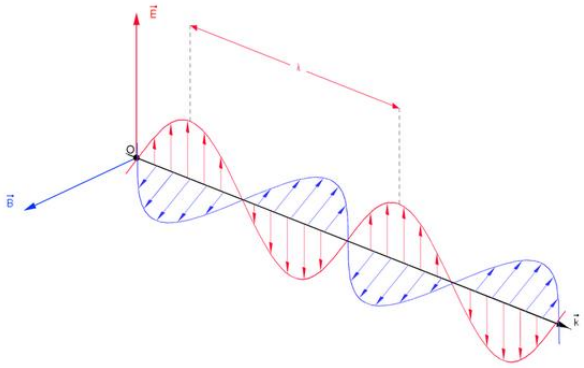


Fig. 1 & 2: Electromagnetic Waves

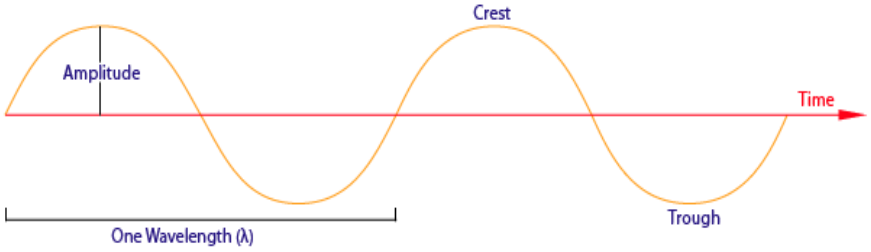
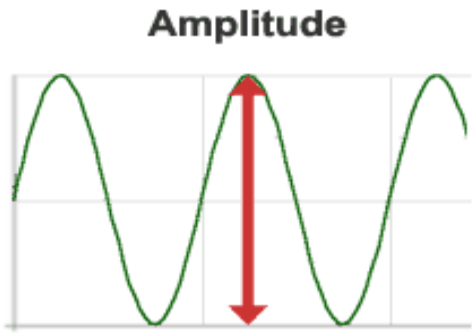


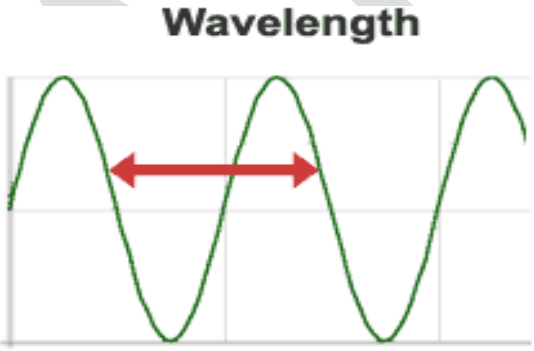
Fig. 3: An EM Wave

Amplitude



Amplitude is the distance from the maximum vertical displacement of the wave to the middle of the wave. This measures the magnitude of oscillation of a particular wave. In short, the amplitude is basically the height of the wave. Larger amplitude means higher energy and lower amplitude means lower energy. Amplitude is important because it tells you the intensity or brightness of a wave in comparison with other waves.

Wavelength



Wavelength (\(\lambda\)) is the distance of one full cycle of the oscillation. Longer wavelength waves such as radio waves carry low energy; this is why we can listen to the radio without any harmful consequences. Shorter wavelength waves such as x-rays carry higher energy that can be hazardous to our health. Consequently lead aprons are worn to protect our bodies from harmful radiation when we undergo x-rays. This wavelength frequently relationship is characterized by:

\[c = \lambda\nu \]

where

- c is the speed of light,
- \(\lambda \) is wavelength, and
- \(\nu \) is frequency.

Shorter wavelength means greater frequency, and greater frequency means higher energy. Wavelengths are important in that they tell one what type of wave one is dealing with.

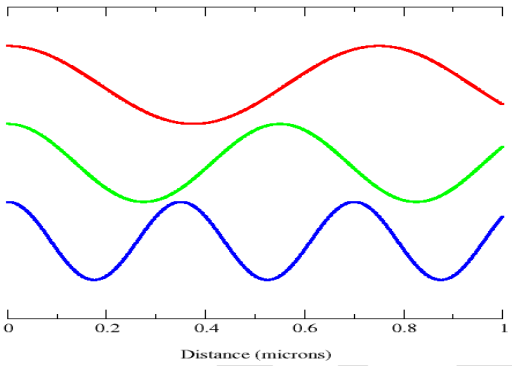
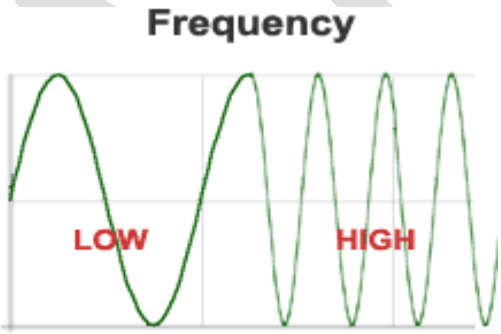


Fig. 4: Different Wavelengths and Frequencies

Frequency



Frequency is defined as the number of cycles per second, and is expressed as sec^{-1} or Hertz (Hz). Frequency is directly proportional to energy and can be express as:

\[E = h\nu \]

where

- E is energy,
- h is Planck's constant, (h= 6.62607 x 10^{-34} J), and
- \(\nu\) is frequency.

Period

Period (T) is the amount of time a wave takes to travel one wavelength. It is measured in seconds (s).

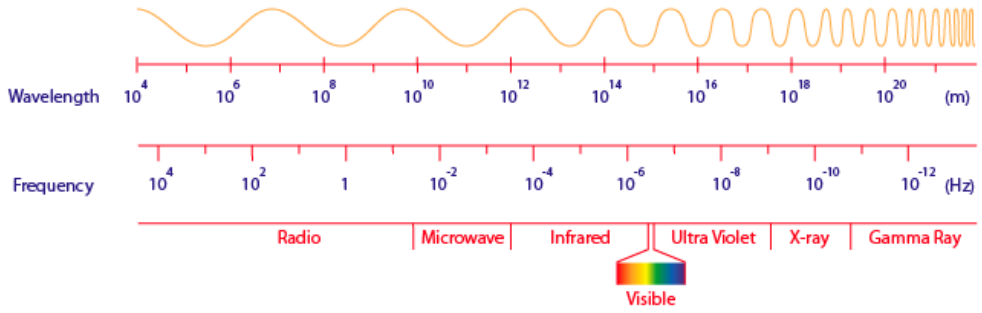
Velocity

The velocity of wave in general is expressed as:

\[velocity = \lambda\nu \]

For Electromagnetic wave, the velocity is 2.99x108m/s or 186,000 miles per second.

Electromagnetic spectrum



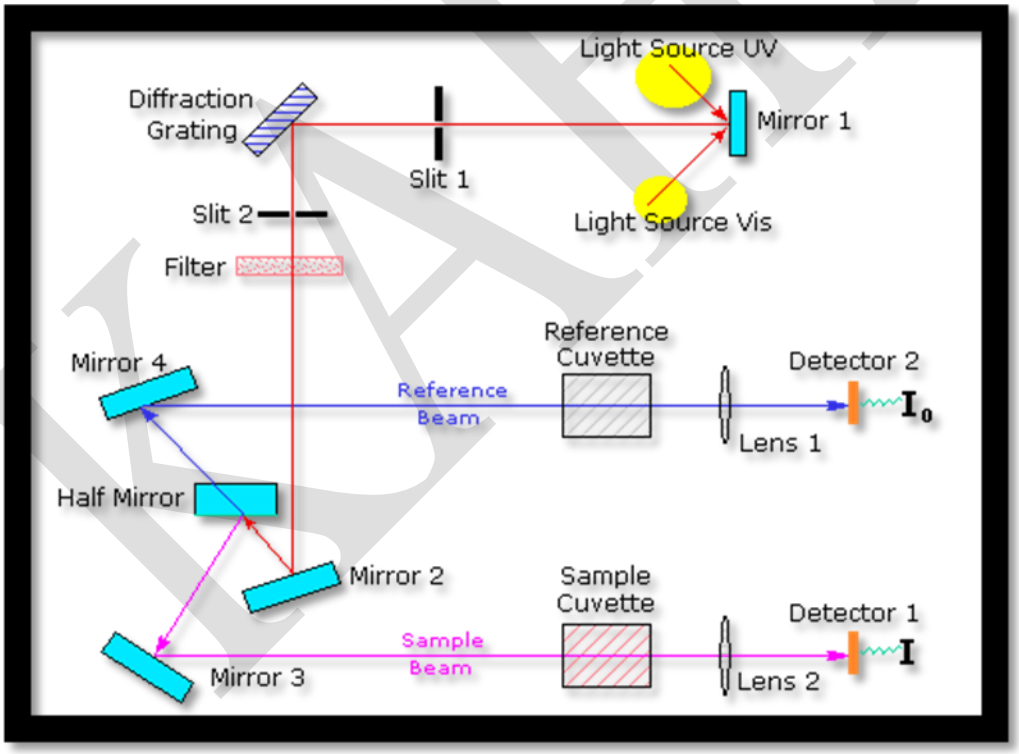
In general, as a wave's wavelength increases, the frequency decreases, and as wave's wavelength decreases, the frequency increases. When electromagnetic energy is released as the energy level increases, the wavelength decreases and frequency decreases. Thus, electromagnetic radiation is then grouped into categories based on its wavelength or frequency into the electromagnetic spectrum. The different types of electromagnetic radiation shown in the electromagnetic spectrum consists of radio waves, microwaves, infrared waves, visible light, ultraviolet radiation, X-rays, and gamma rays. The part of the electromagnetic spectrum that we are able to see is the visible light spectrum.

UV/Visible Spectroscopy

The absorption spectrum (plural, spectra), or more correctly the absolute absorption spectrum, of a compound may be shown as a plot of the light absorbed by that compound against wavelength. Such a plot for a colored compound will have one or more absorption maxima (λmax's) in the visible region of the spectrum (400 to 700 nm). Absorption spectra in the ultraviolet (200 to 400 nm) and visible regions

are due to energy transitions of both bonding and nonbonding outer electrons of the molecule. Usually delocalized electrons are involved such as the B bonding electrons of C=C and the lone pairs of nitrogen and oxygen. Since most of the electrons in a molecule are in the ground state at room temperature, spectra in this region give information about this state and the next higher one. As the wavelengths of light absorbed are determined by the actual transitions occurring, specific absorption peaks may be recorded and related to known molecular substructures.

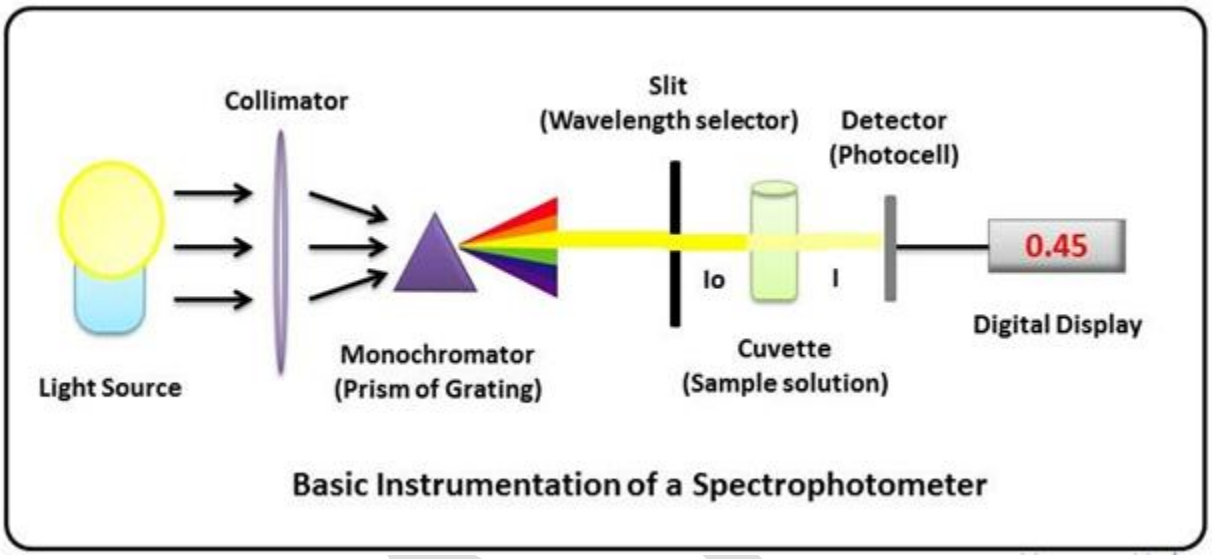
The term chromophore is given to that part of a molecule that gives rise independently to distinct parts of an absorption spectrum, for example the carbonyl group. Conjugated double bonds lower the energy required for electronic transitions and results in an increase in the wavelength at which a chromophore absorbs. This is referred to as a bathochromic shift, whereas a decrease in conjugation, caused for example by protonating a ring nitrogen atom, causes a hypochromic shift which leads to a decrease in wavelength. Hyperchromic and hypochromic effects refer to an increase and a decrease in absorbance respectively.



Instrumentation.

To obtain an absorption spectrum, the absorbance of a substance must be measured at a series of wavelengths. Absorption in the visible and ultraviolet regions can be measured by a UV/visible

spectrophotometer. UV/Vis spectrometers consist of three basic components, (i) a light source and a mechanism to select a specific wavelength of light in the UV/visible region of the spectrum, (ii) a chamber where a cuvette containing a test solution can be introduced into the light path, and (iii) a photocell that can determine the amount of light absorbed by the sample (or the intensity of light transmitted through the sample).



Basic Instrumentation of a Spectrophotometer

The light source is usually a tungsten lamp for the visible region of the spectrum, and either ahydrogen or deuterium lamp for ultraviolet wavelengths. Cuvettes are optically transparent cells that hold the material(s) under study and are used to introduce samples into the light path. A reference cuvette optically identical to, and containing the same solvent (and impurities) as the test cuvette is always required for setting the spectrophotometer to read zero absorbance at each wavelength used. For accurate work, the optical matching of the two cuvettes should always be checked. Glass and plastic absorb strongly below 310 nm and are not useful for measuring absorbance below that wavelength. Quartz or silica cells are used when measuring absorption of ultraviolet wavelengths by a solution since they are transparent to wavelengths greater than 180 nm.

Application

- UV/Vis spectroscopy is routinely used in analytical chemistry for the quantitative determination of different analyses, such as transition metal ions, highly conjugated organic compounds, and biological macromolecules. Spectroscopic analysis is commonly carried out in solutions but solids and gases may also be studied.

- Solutions of transition metal ions can be colored (i.e., absorb visible light) because d electrons within the metal atoms can be excited from one electronic state to another.
- The colour of metal ion solutions is strongly affected by the presence of other species, such as certain anions or ligands. For instance, the colour of a dilute solution of copper sulfate is a very light blue; adding ammonia intensifies the colour and changes the wavelength of maximum absorption (λmax).
- Organic compounds, especially those with a high degree of conjugation, also absorb light in the UV or visible regions of the electromagnetic spectrum. The solvents for these determinations are often water for water-soluble compounds, or ethanol for organic-soluble compounds. (Organic solvents may have significant UV absorption; not all solvents are suitable for use in UV spectroscopy. Ethanol absorbs very weakly at most wavelengths.)
- Solvent polarity and pH can affect the absorption spectrum of an organic compound. Tyrosine, for example, increases in absorption maxima and molar extinction coefficient when pH increases from 6 to 13 or when solvent polarity decreases.
- While charge transfer complexes also give rise to colours, the colours are often too intense to be used for quantitative measurement.
- The wavelengths of absorption peaks can be correlated with the types of bonds in a given molecule and are valuable in determining the functional groups within a molecule.
- The spectrum alone is not, however, a specific test for any given sample. The nature of the solvent, the pH of the solution, temperature, high electrolyte concentrations, and the presence of interfering substances can influence the absorption spectrum. Experimental variations such as the slit width (effective bandwidth) of the spectrophotometer will also alter the spectrum.
- To apply UV/Vis spectroscopy to analysis, these variables must be controlled or accounted for in order to identify the substances present.

Mass Spectroscopy

In order to measure the characteristics of individual molecules, a mass spectrometer converts them to ions so that they can be moved about and manipulated by external electric and magnetic fields. The three essential functions of a mass spectrometer, and the associated components, are:

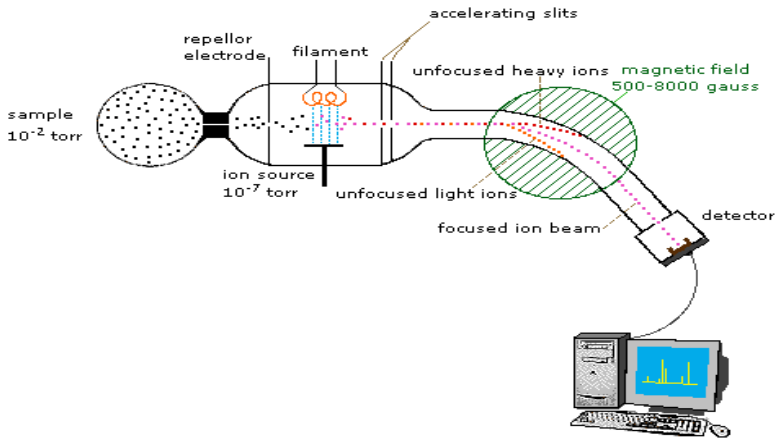
The Ion Source: A small sample is ionized, usually to cations by loss of an electron.

The Mass Analyzer: The ions are sorted and separated according to their mass and charge.

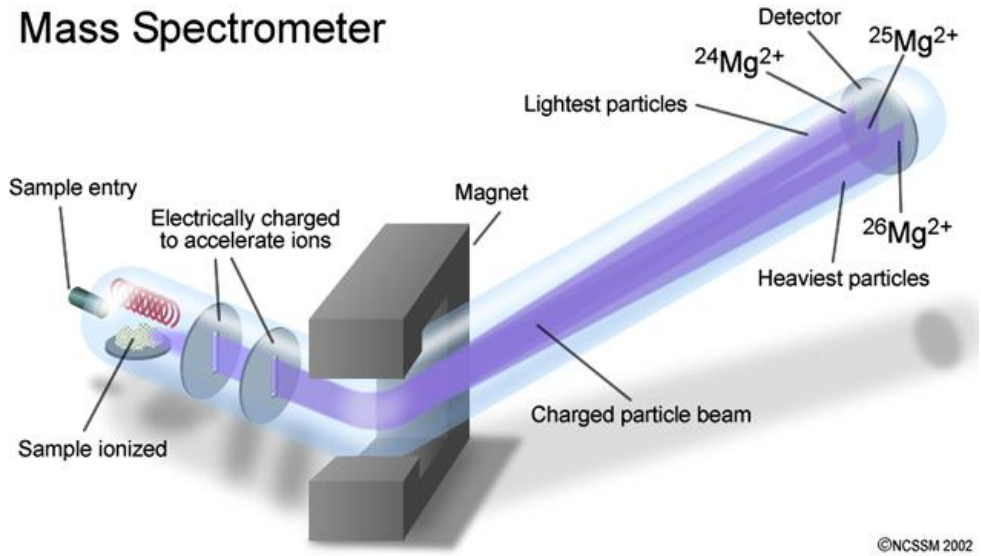
The Detector:The separated ions are then measured, and the results displayed on a chart.

Because ions are very reactive and short-lived, their formation and manipulation must be conducted in a vacuum. Atmospheric pressure is around 760 torr (mm of mercury). The pressure under which ions may be handled is roughly 10^{-5} to 10^{-8} torr (less than a billionth of an atmosphere). Each of the three tasks listed above may be accomplished in different ways. In one common procedure, ionization is effected by a high energy beam of electrons, and ion separation is achieved by accelerating and focusing the ions in a beam, which is then bent by an external magnetic field. The ions are then detected electronically and the resulting information is stored and analyzed in a computer. A mass spectrometer operating in this fashion is outlined in the following diagram. The heart of the spectrometer is the ion source. Here molecules of the sample (black dots) are bombarded by electrons (light blue lines) issuing from a heated filament. This is called an EI (electron-impact) source. Gases and volatile liquid samples are allowed to leak into the ion source from a reservoir (as shown). Non-volatile solids and liquids may be introduced directly.

Cations formed by the electron bombardment (red dots) are pushed away by a charged repeller plate (anions are attracted to it), and accelerated toward other electrodes, having slits through which the ions pass as a beam. Some of these ions fragment into smaller cations and neutral fragments. A perpendicular magnetic field deflects the ion beam in an arc whose radius is inversely proportional to the mass of each ion. Lighter ions are deflected more than heavier ions. By varying the strength of the magnetic field, ions of different mass can be focused progressively on a detector fixed at the end of a curved tube (also under a high vacuum).



When a high energy electron collides with a molecule it often ionizes it by knocking away one of the molecular electrons (either bonding or non-bonding). This leaves behind a molecular ion (colored red in the following diagram). Residual energy from the collision may cause the molecular ion to fragment into neutral pieces (colored green) and smaller fragment ions (colored pink and orange). The molecular ion is a radical cation, but the fragment ions may either be radical cations (pink) or carbocations (orange), depending on the nature of the neutral fragment. An animated display of this ionization process will appear if you click on the ion source of the mass spectrometer.



Nature of Mass Spectra

A mass spectrum will usually be presented as a vertical bar graph, in which each bar represents an ion having a specific mass-to-charge ratio (m/z) and the length of the bar indicates the relative abundance of the ion. The most intense ion is assigned an abundance of 100, and it is referred to as the base peak. Most of the ions formed in a mass spectrometer have a single charge, so the m/z value is equivalent to mass itself. Modern mass spectrometers easily distinguish (resolve) ions differing by only a single atomic mass unit (amu), and thus provide completely accurate values for the molecular mass of a compound. The highest-mass ion in a spectrum is normally considered to be the molecular ion, and lower-mass ions are fragments from the molecular ion, assuming the sample is a single pure compound.

Applications:

- Mass spectrometry has both qualitative and quantitative uses. These include identifying unknown compounds, determining the isotopic composition of elements in a molecule, and determining the structure of a compound by observing its fragmentation.
- Other uses include quantifying the amount of a compound in a sample or studying the fundamentals of gas phase ion chemistry (the chemistry of ions and neutrals in a vacuum).
- MS is now in very common use in analytical laboratories that study physical, chemical, or biological properties of a great variety of compounds.

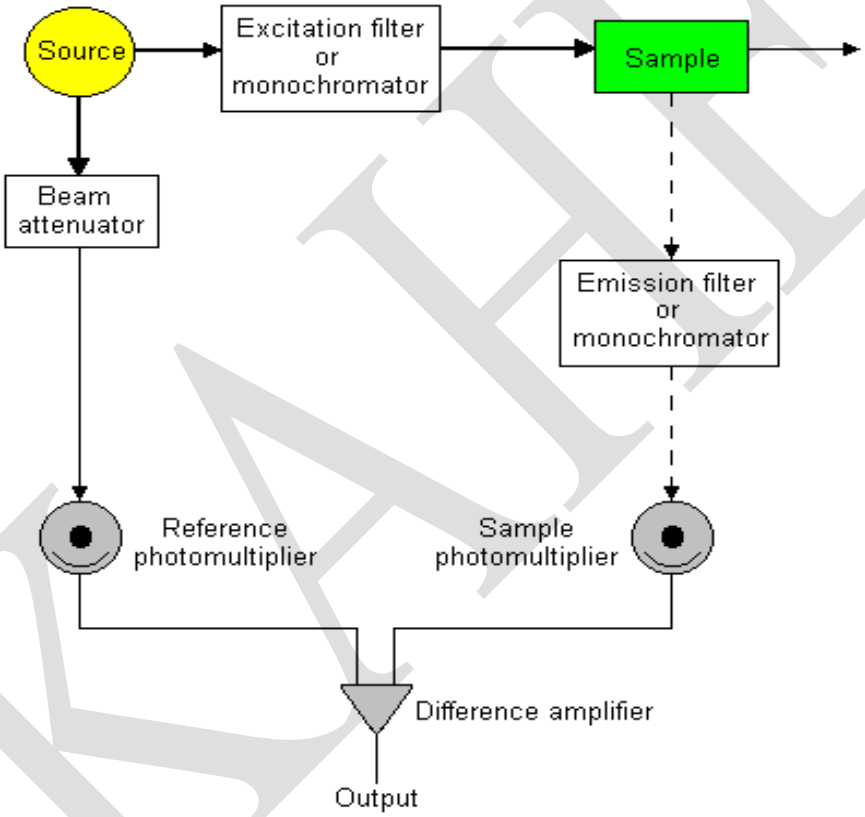
- As an analytical technique it possesses distinct advantages such as:
    - Increased sensitivity over most other analytical techniques because the analyzer, as a mass-charge filter, reduces background interference
    - Excellent specificity from characteristic fragmentation patterns to identify unknowns or confirm the presence of suspected compounds.
    - Information about molecular weight.
    - Information about the isotopic abundance of elements.
- Temporally resolved chemical data.

- A few of the disadvantages of the method is that often fails to distinguish between optical and geometrical isomers and the positions of substituent in o-, m- and p- positions in an aromatic ring.
- Also, its scope is limited in identifying hydrocarbons that produce similar fragmented ions.
- Mass spectrometry is also used to determine the isotopic composition of elements within a sample.

- Differences in mass among isotopes of an element are very small, and the less abundant isotopes of an element are typically very rare, so a very sensitive instrument is required.

Spectrofluorimeter

A spectrofluorimeter is an instrument which takes advantage of fluorescent properties of some compounds in order to provide information regarding their concentration and chemical environment in a sample. A certain excitation wavelength is selected, and the emission is observed either at a single wavelength, or a scan is performed to record the intensity versus wavelength, also called an emission spectra.



Instrumental components

Sources

Generally, the source must be more intense than that required for UV-Vis. absorption spectroscopy; magnitude of the emitted radiation is directly proportional to the power of the source.

Filter fluorometers often employ a low-pressure mercury vapour lamp. This source produces intense lines at certain wavelengths. One of these lines will usually be suitable for excitation of a fluorescent sample.

Spectrofluorometers, which need a continuous radiation source, are often equipped with a 75-450 W high-pressure xenon arc lamp.

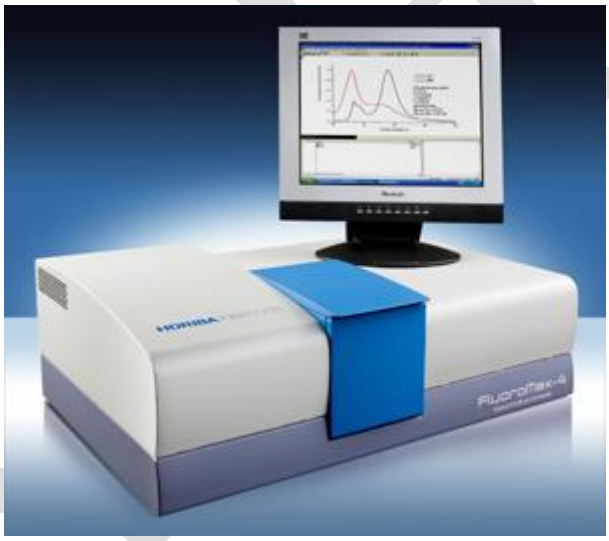
Lasers are sometimes used as excitation sources. A tunable dye laser, using a pulsed nitrogen laser as the primary source can produce monochromatic radiation between 360 and 650 nm. Since the radiation produced *is* monochromatic, there is no need for an excitation monochromator.

Filters and monochromators

Fluorometers use either interference or absorption filters. Spectrofluorometers are usually fitted with grating monochromators.

Detectors

Fluorescence signals are usually of low intensity. Photomultiplier tubes are in common use as detectors. Diode-array detectors are sometimes used.

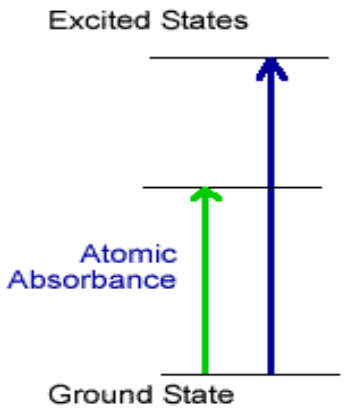


Applications:

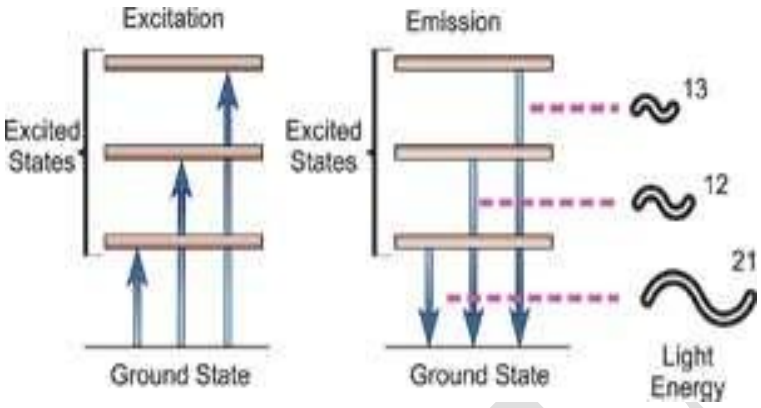
- Generally, spectrofluorometers use high intensity light sources to bombard a sample with as many photons as possible.
- This allows for the maximum number of molecules to be in an excited state at any one point in time.
- The light is either passed through a filter, selecting a fixed wavelength, or a monochromator, which allows a wavelength of interest to be selected for use as the exciting light.

Atomic Spectroscopy

Atomic spectroscopy exploits different energetic transitions experienced by atoms that are associated with either the absorption or emission of photons. When these transitions involve the excitation and relaxation of the valence (outer or bonding) shell electrons of metal atoms and ions, the corresponding photons have energies within the ultraviolet and visible regions of the spectrum. A good example of this is the dark absorption lines in the solar spectrum, which are caused by heavier elements present in the outer layers of the sun.



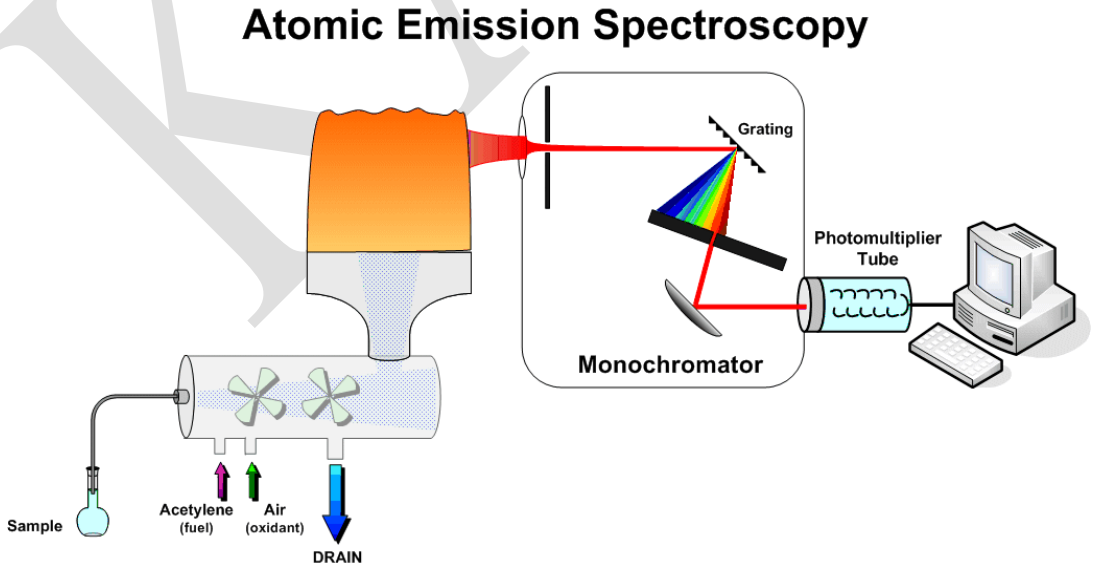
The figure shows a high energy photon with Ephoton = hν being absorbed, resulting in a 2s→3s electron excitation; similarly, a 3d→3p electron relaxation results in the emission of a lower energy photon. By convention, the change in electron energy ΔE = Ef − Ei, where f and i refer to the final and initial states, respectively; so ΔE = Ephoton, and the sign of Ephoton tells you whether the photon is being absorbed or emitted. Since Ef and Ei depend on the number electrons and protons within an atom (or monatomic ion), the wavelengths associated with atomic absorption and emission are considered characteristic for a particular element.



Absorption and Emission:

In atomic absorption (AA) spectroscopy, absorption of a photon results in excitation of an electron from a lower to higher energy atomic orbital (AO). An instrument measures the absorbance, A, which is defined as the logarithm of the ratio of incident to transmitted radiant power of the photon beam, A = log(P0 ÷ P), at a wavelength specific to the element of interest. Samples are typically analysed using a flame atomic absorption spectrophotometer.

In atomic emission (AE) spectroscopy, thermal or electrical energy from an arc, flame, spark, or plasma is used to excite and electron from a lower to higher energy AO; when the excited electron returns to its original AO (i.e. the ground state), it may do so by emitting a photon. The instrument measures the intensity, I, of these emitted photons as a function of wavelength.



Because AO energies are well-defined, atomic absorption and emission spectra consist of discrete, narrow lines. This allows the concentration of metallic elements in different samples to be determined selectively, with lower limits at or below 1 mg/L (1 ppm). Techniques such as graphite furnace atomic absorption spectrophotometry (GFAAS) allow concentration to be measured down to µg/L (ppb) levels. Actual limits-of-detection vary withelement, technique, and sample matrix.

Applications:

- Atomic spectroscopy has many useful applications. Since the emission spectrum is different for every element, it acts as an atomic fingerprint by which elements can be identified.
- Some elements were discovered by the analysis of their atomic spectrum.
- Helium, for example, was discovered while scientists were analyzing the absorption spectrum of the sun.

- Emission spectrum is especially useful to astronomers who use emission and absorption spectra to determine the makeup of far away stars and other celestial bodies.

CENTRIFUGATION TECHNIQUES:

Introduction:

- The centrifuge is an instrument, which is used to spin substances at high speed.
- It is often used to separate particles present in a liquid.
- The mixture is placed in a tube that pivots so that when the machine starts to move, the tube can swing out horizontally.
- The centrifugal force tries to push the mixture away from the centre of centrifuge.

BASIC PRINCIPLES OF SEDIMENTATION:

- This method is based on the principle of sedimentation.
- From everyday experience, the effect of sedimentation due to the influence of the Earth's gravitational field (g = 981 cms-2) versus the increased rate of sedimentation in centrifugal field (g>981 cms-2) is apparent.
- To give a simple but illustrative example, crude sand particles added to a bucket of water travel slowly to the bottom of the bucket by gravitation, but sediment much faster when the bucket is swung around in a circle.
- Similarly, biological structures exhibit a drastic increase in sedimentation when they undergo acceleration in a centrifugal field.
- The relative centrifugal field is usually expressed as a multiple of the acceleration due to gravity.
- Below is a short description of equations used in practical centrifugation classes.
- The more dense a biological structure is the faster it sediments in a centrifugal field.
- The more massive a biological particle is, the faster it moves in a centrifugal field.

- The denser the biological buffer system is, the slower the particle will move in a centrifugal field.
- The greater the frictional coefficient is, the slower a particle will move.
- The greater the centrifugal force is the faster the particle sediments.
- The sedimentation rate of a given particle will be zero when the density of the particle and the surrounding medium are equal.
- When particle sediments, it must displace some of the solution in which it is suspended, resulting in an up thrust on the particle equal to the weight of the liquid displaced.
- If a particle is assumed to be a sphere of known volume and density, then the net force(f) is experienced when the centrifugal force at an angular velocity of ω radians/sec is given by:

$$S = \text{Volume} \times \text{Density} \times \omega 2r$$

Or,

$$F = 4/3\ \Pi\ r \quad p3\ (\rho p - \rho m)\ \omega 2\ r$$

Where,

$4/3\ \Pi\ r\ p3$ = Volume of sphere of radius 'r'.

ρp = Density of the particle.

ρm = Density of the suspended medium.

ν = Distance of the particle from the center of rotation.

ω = Angular velocity of rotor.

DESKTOP CENTRIFUGES:

- Centrifuges are essential devices, each of the clinical laboratories.
- This may be a desktop or floor-type, refrigerated or cooling.
- Centrifuge is mainly used to create the fractionation due to fluid samples with high g forces of spin.

- Desktop centrifuge designed to meet the laboratory setting to limited space requirements.
- These are used to separate liquid solid;, also dealing with blood samples used.
- Desktop centrifuge is in a variety of different specifications, sizes and abilities.

Independent Electoral Commission Centra, Beckman and Adams is a leading manufacturer of desktop centrifuge.

Desktop centrifuge with innovative features

- Desktop centrifuge important feature is to break the power system strong guarantee automatic acceleration and deceleration.
- Another important feature is their silence and vibration free operation.
- The device is designed with a full timer, brushless motor and has a clear lid safety switch.
- It also has a maintenance-free brushless motor; with power interrupt the user's security design.
- The device and the cooling or not cooling at different speeds to choose from.
- Refrigeration unit is an independent unit, but the time dependence of their work, because it's centrifuge for power.
- The device has a multi-functional, four rotor, test tubes and bottles can accommodate all possible sizes.
- Desktop centrifuge have such characteristics:
- Removable rotor bowl for easy cleaning.
- Reagents, and consumables can be controlled.
- Speed control knob.
- Keep the programs running longer.
- Temperature, velocity and time display settings
- Find the real purchasing the best product distributor
- To analyze the working conditions before purchasing, quality and warranty desktop centrifuge specifications.

- A leading supplier of laboratory bench centrifuge at low cost high quality products, and to ensure service for years.

Desktop Centrifuges

LARGE CAPACITY REFRIGERATED CENTRIFUGES

- Controlled by microcomputer, touching panel and LED display.
- The compressor imported from Europe operating without Freon for environment protection, pre-cooling when power on.
- The brushless direct drive motor with high start torque and frequency inversion.
- Imbalance, over-speed, over-temperature and door interlocking protection function for the safety of people and instrument.
- The parameter can be changed at operating state.
- Automatic computing and setting RCF.adjustable rise-and-fail speed from grade 0 to 9
- Fitting to the adapters of 5ml.7ml .10ml.15ml.30ml.50ml, etc.
- Widely used in the filed of radicalization immunity, biochemistry, pharmacy, blood separation and purification
- Hold-on, transient centrifuge.

ANALYTICAL AND PREPARATIVE ULTRACENTRIFUGATION:

Ultracentrifuge:

- The ultracentrifuge is a centrifuge optimized for spinning a rotor at very high speeds, capable of generating acceleration as high as 2,000,000 g (approx. 19,600 km/s^2).

- There are two kinds of ultracentrifuges, the preparative and the analytical ultracentrifuge. Both classes of instruments find important uses in molecular biology, biochemistry and polymer science.

Analytical ultracentrifuge

- In an analytical ultracentrifuge, a sample being spun can be monitored in real time through an optical detection system, using ultraviolet light absorption and/or interference optical refractive index sensitive system.

- This allows the operator to observe the evolution of the sample concentration versus the axis of rotation profile as a result of the applied centrifugal field.

- With modern instrumentation, these observations are electronically digitized and stored for further mathematical analysis.

- Two kinds of experiments are commonly performed on these instruments: sedimentation velocity experiments and sedimentation equilibrium experiments.

- Sedimentation velocity experiments aim to interpret the entire time-course of sedimentation, and report on the shape and molar mass of the dissolved macromolecules, as well as their size-distribution.

- The size resolution of this method scales approximately with the square of the particle radii, and by adjusting the rotor speed of the experiment size-ranges from 100 Da to 10 GDa can be covered.

- Sedimentation velocity experiments can also be used to study reversible chemical equilibria between macromolecular species, by either monitoring the number and molar mass of macromolecular complexes, by gaining information about the complex composition from multi-signal analysis exploiting differences in each components spectroscopic signal, or by following the composition dependence of the sedimentation rates of the macromolecular system, as described in Gilbert-Jenkins theory.

- Sedimentation equilibrium experiments are concerned only with the final steady-state of the experiment, where sedimentation is balanced by diffusion opposing the concentration gradients, resulting in a time-independent concentration profile. Sedimentation equilibrium distributions in the centrifugal field are characterized by Boltzmann distributions.

- This experiment is insensitive to the shape of the macromolecule, and directly reports on the molar mass of the macromolecules and, for chemically reacting mixtures, on chemical equilibrium constants.

- The kinds of information that can be obtained from an analytical ultracentrifuge include the gross shape of macromolecules, the conformational changes in macromolecules, and size distributions of macromolecular samples.

- For macromolecules, such as proteins, which exist in chemical equilibrium with different non-covalent complexes, the number and subunit stoichiometry of the complexes and equilibrium constant constants can be studied.

Preparative ultracentrifuge

- Preparative ultracentrifuges are available with a wide variety of rotors suitable for a great range of experiments.

- Most rotors are designed to hold tubes that contain the samples.

- Swinging bucket rotors allow the tubes to hang on hinges so the tubes reorient to the horizontal as the rotor initially accelerates.

- Fixed angle rotors are made of a single block of metal and hold the tubes in cavities bored at a predetermined angle.

- Zonal rotors are designed to contain a large volume of sample in a single central cavity rather than in tubes.

- Some zonal rotors are capable of dynamic loading and unloading of samples while the rotor is spinning at high speed.

- Preparative rotors are used in biology for pelleting of fine particulate fractions, such as cellular organelles (mitochondria, microsomes, and ribosomes) and viruses.

- They can also be used for gradient separations, in which the tubes are filled from top to bottom with an increasing concentration of a dense substance in solution.

- Sucrose gradients are typically used for separation of cellular organelles.

- Gradients of caesium salts are used for separation of nucleic acids.
- After the sample has spun at high speed for sufficient time to produce the separation, the rotor is allowed to come to a smooth stop and the gradient is gently pumped out of each tube to isolate the separated components.

APPLICATIONS OF ULTRACENTRIFUGATION:

- The analytical ultracentrifuge has found the following wide applications in Biology.
- To determine relative molecular mass of macromolecules such as, proteins and DNA.
- To investigate the purity of DNA preparations, viruses and proteins.
- To detect conformational changes in macromolecules such as DNA and protein.
- Characterize assembly and disassembly mechanisms of bio molecular complexes
- Determine subunit stoichiometries
- Thermodynamic and hydrodynamic information
- Measure equilibrium constants and thermodynamic parameters for self- and hetero-associating systems.

Sedimentation velocity

- Sedimentation velocity is an analytical ultracentrifugation (AUC) method that measures the rate at which molecules move in response to centrifugal force generated in a centrifuge.
- This sedimentation rate provides information about both the molecular mass and the shape of molecules. In some cases this technique can also measure diffusion coefficients and molecular mass.
- In the biotechnology industry sedimentation velocity is used much more frequently than sedimentation equilibrium and thus when biotech scientists say "AUC" or "analytical ultracentrifugation" they typically really mean "sedimentation velocity".
- Sedimentation velocity is particularly valuable for:verifying whether a sample is entirely homogeneous in mass and conformation detecting aggregates in protein samples and quantifying the amount of aggregate comparing the conformations for samples from different lots, manufacturing processes, or expression systems (comparability studies), or comparing different

engineered variants of the same protein/peptide,establishing whether the native state of a protein or peptide is a monomer, dimer, trimer, etc.,determining the overall shape of non-glycosylated protein and peptide molecules in solution, measuring the distribution of sizes in samples which contain a very broad range of sizes, detecting changes in protein conformation, for example partial unfolding or transitions to "molten globule" states.studying the formation and stoichiometry of tight complexes between proteins (for example receptor-ligand or antigen-antibody complexes)

- In the sedimentation velocity method a sample is spun at very high speed (usually 40-60 K rpm) in an analytical ultracentrifuge. The high centrifugal force rapidly depletes all the protein from the region nearest the center of the rotor (the meniscus region at the air/solution interface), forming a boundary which moves toward the outside of the rotor with time (see example below), until finally all the protein forms a pellet at the outside of the cell.

- The concentration distribution across the cell at various times during the experiment is measured while the sample is spinning, using either absorbance or refractive index detection in our Beckman ProteomeLab XL-I.

- A major advantage of this method over sedimentation equilibrium is that experiments usually require only 3-5 hours, as opposed to the several days typical of sedimentation equilibrium. Thus sedimentation velocity can be used with samples that are too labile for sedimentation equilibrium.

- The major drawback relative to sedimentation equilibrium applies to interacting systems (proteins that reversibly self-associate or protein-protein complexes), where the non-equilibrium nature of the measurement can lead to significant changes in species distributions over the course of an experiment. Further, for interacting systems it is generally more difficult and less accurate to derive binding constants (Kd's) from sedimentation velocity data.

- An important strength of sedimentation velocity is its ability to study samples over a fairly wide range of pH and ionic strength conditions (and often directly in formulation buffers), and at temperatures from 4 to 40 °C. The amount of protein required depends on the application, but each sample is usually ~0.45 ml at typical protein concentrations of 0.1-1 mg/ml (45-450 micrograms total).

- Protein concentration can range as low as ~10 micrograms/ml or as high as ~40 mg/ml in some cases (but generally the concentration should be 2 mg/mL or below). Up to 3 samples can be run at one time.A sedimentation velocity case study: a monoclonal antibody.

- The graph below shows scans across the centrifuge cell, recording the absorbance at 280 nm versus position within the cell. These scans were taken starting at 13 minutes after initiating a

run at 45,000 rpm (the black data set in the graph), and then every ~12 minutes thereafter (blue, green, cyan, etc.).

- The sharp vertical spike at 6.02 cm indicates the position of the air-solution meniscus. In the first data set the sedimentation of the antibody has already depleted its concentration in the region near the meniscus and formed a sedimentation boundary.

- At later times in the run the depleted region expands and the boundary moves away from the center of the rotor, until by the time of the last data set the concentration of antibody has dropped to essentially zero throughout the upper half of the cell.

- The rate at which the sedimentation boundary moves is a measure of the sedimentation coefficient of the protein. The sedimentation coefficient depends on the molecular weight (larger proteins sediment faster) and also on molecular shape. Unfolded proteins or one with highly elongated shapes will experience more hydrodynamic friction, and thus will have smaller sedimentation coefficients than a folded, globular protein of the same molecular weight.

- The minimum width of the sedimentation boundary is related to the diffusion coefficient of the molecule; the presence of multiple species with similar sedimentation coefficients will cause the boundary to be broader than expected on the basis of diffusion alone. In this case the majority of the boundary is reasonably narrow, but the slow rise of the data on the right side of the boundary suggests the presence of some faster moving species.

- When viewed as in the graph above the data is difficult to interpret. What we often want to know is how much material is sedimenting at various sedimentation coefficients. By taking many scans close together in time (the graph above shows only a small number of the scans), subtracting them in pairs, and doing some mathematical manipulation these data can be transformed into the sedimentation coefficient distribution, g(s*), which is shown below.

- This distribution resembles a chromatogram, and in many ways is similar to a size-exclusion chromatogram except the peaks come in the opposite order. Like a chromatogram, the area under each peak gives the total amount of that species.

- For this antibody sample we see only one distinct peak, centered at a sedimentation coefficient of ~6.5 S, which corresponds to the native antibody 'monomer' (really a covalent heterotetramer of 2 light and 2 heavy chains). A sedimentation coefficient of 6.5 S is actually rather low for a 150 kDa species, which is consistent with high hydrodynamic friction from its highly asymmetric, non-globular 'Y' shape.

- Although we see only a single peak in the g(s*) distribution, a more detailed analysis quickly reveals that this sample is not homogeneous. The red curve is a fit of these data as a single species. This fit clearly fails to account for the data over the region from 8-12 S, indicating the presence of some dimer and possibly also some trimer.

- The fact that the distribution has not returned fully to zero at 12 S also suggests that there may also be small amounts of even faster sedimenting species present.

- From the width of the main peak we can also calculate the apparent diffusion coefficient of the monomer. In turn, from the ratio of sedimentation coefficient to diffusion coefficient we can calculate a mass of 151 kDa for this species, which matches the expected value well within 3-5% error expected for masses determined in this fashion.

Lastly, we can apply a new analysis algorithm to further enhance the resolution of the species distribution.

- In this figure the full distribution is shown in the main graph, while the inset magnifies the vertical scale by 10X in order to better show the minor components. With the enhanced resolution we now see a fully baseline-resolved dimer peak at ~9.4 S (4.7% of the total protein), and small peaks at ~13.7 S and 17.5 S (1.8% and 1.1%, probably trimer and tetramer). In addition there is 0.7% of a low mass contaminant at 2 S (possibly free light chain).

- Thus from this one velocity experiment we have been able to quantify the amount and mass of the main component, the content and sedimentation coefficients of 3 aggregates and a low mass contaminant, and to obtain information about the conformation of the main component.

Other applications

A sedimentation coefficient distribution for a sample of adenovirus (~150 MDa, used to deliver vectors for gene therapy). This illustrates the broad range of sizes and molecule types that can be studied using this technique. The inset shows a 20-fold expanded scale to allow the many minor peaks to be seen. This material turns out to be quite heterogeneous, having only 57% of the material as the main peak (virus monomer), plus many rapidly-sedimenting viral aggregates and some slowly-sedimenting species that may be fragments or empty capsids.

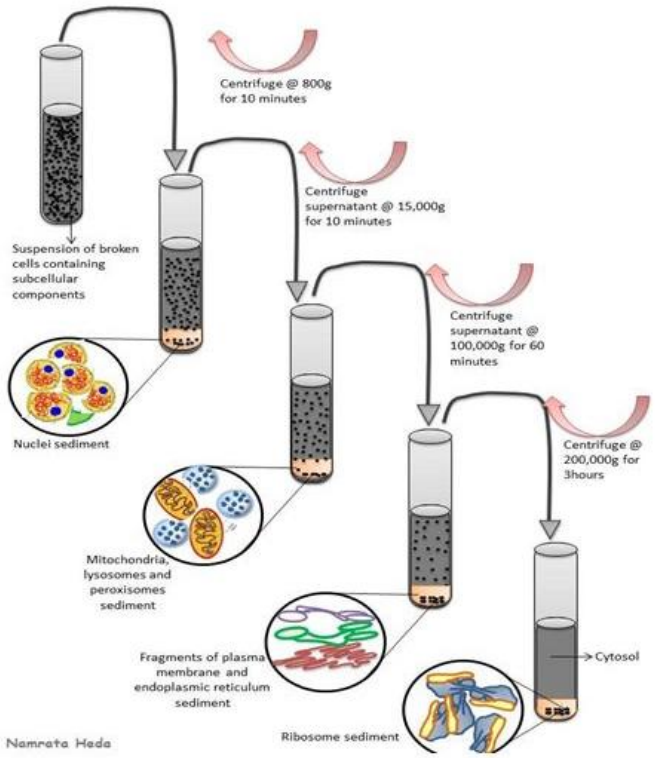
SEPARATION OF CELL ORGANELLES:

- To study the functions of any organelle in depth, it is first necessary to isolate it in a relatively pure form, free of concentration by other organelles.
- The process of separation of cell organelles is known as subcellular fractionation.
- As a first step toward isolating a specific organelle, organs (liver, brain, and kidney) are homogenized in a suitable homogenizing medium at 4º C.
- The resulting suspension, containing many contact organelles, is known as a homogenate.
- Fractionation of the contents of a homogenate is done by a classical biochemical technique called differential centrifugation.
- This methodis based upon the differences in the sedimentation rate of particles of different size and density.
- This method uses a series of four different centrifugation steps at successively greater speeds.
- Each step yields a pellet and a supernatant.
- The supernatant from each step is subjected to centrifugation in the next step.
- This procedure provides four pellets, namely, nuclear, mitochondrial, lysosomal and microsomal fractions.
- At the end of each step, the pellet is washed several times by resuspending in the homogenization medium followed by recentrifugation under the same conditions.
- This procedure minimizes contamination of other subcellular organelles and gives a fairly pure preparation of pellet fraction.
- The purity of organelles obtained by differential centrifugation is measured by estimating some marker activity.
- A marker is one that is almost exclusively present in one particular organelle.
- A marker may be an enzyme molecule or a biochemical compound.
- Various fractions, their functions and markers

S. No	Organelle	Function	Marker
1.	Plasma membrane	Regulates entry and exit of compounds	5' Nucleotidase

2.	Nucleus	Site of DNA-directed RNA synthesis	DNA
3.	Mitochondrion	Citric acid cycle, ammonia release	DHases
	for urea formation.		
4.	Lysosome	Site of many hydrolases.	Acid phosphatase
5.	ER	Oxidation of many xenobiotics	Glucose6-pase.
6.	Cytosol	Enzymes of glycolysis.	LDH

- The microsomal fraction contains mostly a mixture of smooth endoplasmic reticulum and free ribosomes.
- The contents of the final supernatant correspond approximately to those of Cytosol.



Fractionation of Cells

Although biochemical analysis requires disruption of the anatomy of the cell, gentle fractionation techniques have been devised to separate the various cell components while preserving their individual functions. Just as a tissue can be separated into its living constituent cell types, so the cell can be separated into its functioning organelles and macromolecules. In this section we consider the methods that allow organelles and proteins to be purified and analyzed biochemically.

Organelles and Macromolecules Can Be Separated by Ultracentrifugation:

• Cells can be broken up in various ways: It can be subjected to osmotic shock or ultrasonic vibration, forced through a small orifice, or ground up in a blender. These procedures break many of the membranes of the cell (including the plasma membrane and membranes of the endoplasmic reticulum) into fragments that immediately reseal to form small closed vesicles. If

carefully applied, however, the disruption procedures leave organelles such as nuclei, mitochondria, the Golgi apparatus, lysosomes, and peroxisomes largely intact.

- The suspension of cells is thereby reduced to thick slurry (called a homogenate or extract) that contains a variety of membrane-enclosed organelles, each with a distinctive size, charge, and density. Provided that the homogenization medium has been carefully chosen (by trial and error for each organelle), the various components—including the vesicles derived from the endoplasmic reticulum, called microsomes—retain most of their original biochemical properties.

- The different components of the homogenate must then be separated. Such cell fractionations became possible only after the commercial development in the early 1940s of an instrument known as the preparative ultracentrifuge, in which extracts of broken cells are rotated at high speeds. This treatment separates cell components by size and density: in general, the largest units experience the largest centrifugal force and move the most rapidly.

- At relatively low speed, large components such as nuclei sediment to form a pellet at the bottom of the centrifuge tube; at slightly higher speed, a pellet of mitochondria is deposited; and at even higher speeds and with longer periods of centrifugation, first the small closed vesicles and then the ribosomes can be collected.

- All of these fractions are impure, but re-suspending the pellet and repeating the centrifugation procedure several times can remove many of the contaminants.

The preparative ultracentrifuge:

- The sample is contained in tubes that are inserted into a ring of cylindrical holes in a metal rotor. Rapid rotation of the rotor generates enormous centrifugal forces, which cause particles in the sample to sediment. The vacuum reduces friction, preventing heating of the rotor and allowing the refrigeration system to maintain the sample at 4°C.

- Centrifugation is the first step in most fractionations, but it separates only components that differ greatly in size. A finer degree of separation can be achieved by layering the homogenate in a thin band on top of a dilute salt solution that fills a centrifuge tube. When centrifuged, the various components in the mixture move as a series of distinct bands through the salt solution, each at a different rate, in a process called velocity sedimentation.

- For the procedure to work effectively, the bands must be protected from convective mixing, which would normally occur whenever a denser solution (for example, one containing

organelles) finds itself on top of a lighter one (the salt solution). This is achieved by filling the centrifuge tube with a shallow gradient of sucrose prepared by a special mixing device. The resulting density gradient—with the dense end at the bottom of the tube—keeps each region of the salt solution denser than any solution above it, and it thereby prevents convective mixing from distorting the separation.

- When sediment through such dilute sucrose gradients, different cell components separate into distinct bands that can be collected individually. The relative rate at which each component sediments depends primarily on its size and shape—being normally described in terms of its sedimentation coefficient, or s value. Present-day ultracentrifuges rotate at speeds of up to 80,000 rpm and produce forces as high as 500,000 times gravity. With these enormous forces, even small macromolecules, such as tRNA molecules and simple enzymes can be driven to sediment at an appreciable rate and so can be separated from one another by size.

- Measurements of sedimentation coefficients are routinely used to help in determining the size and subunit composition of the organized assemblies of macromolecules found in cells.The ultracentrifuge is also used to separate cellular components on the basis of their buoyant density, independently of their size and shape. In this case the sample is usually sedimented through a steep density gradient that contains a very high concentration of sucrose or cesium chloride.

- Each cellular component begins to move down the gradient but it eventually reaches a position where the density of the solution is equal to its own density. At this point the component floats and can move no farther. A series of distinct bands is thereby produced in the centrifuge tube, with the bands closest to the bottom of the tube containing the components of highest buoyant density. This method, called equilibrium sedimentation, is so sensitive that it is capable of separating macromolecules that have incorporated heavy isotopes, such as 13C or 15N, from the same macromolecules that contain the lighter, common isotopes (12C or 14N).

- In fact, the cesium-chloride method was developed in 1957 to separate the labeled from the unlabeled DNA produced after exposure of a growing population of bacteria to nucleotide precursors containing 15N; this classic experiment provided direct evidence for the semiconservative replication of DNA.

- Studies of organelles and other large subcellular components isolated in the ultracentrifuge have contributed enormously to our understanding of the functions of different cellular components. Experiments on mitochondria and chloroplasts purified by centrifugation, for example, demonstrated the central function of these organelles in converting energy into forms that the cell can use.

- Similarly, resealed vesicles formed from fragments of rough and smooth endoplasmic reticulum (microsomes) have been separated from each other and analyzed as functional models of these compartments of the intact cell.

- An extension of this approach makes it possible to study many other biological processes free from all of the complex side reactions that occur in a living cell, by using purified cell-free systems. In this case, cell homogenates are fractionated with the aim of purifying each of the individual macromolecules that are needed to catalyze a biological process of interest.

- For example, the mechanisms of protein synthesis were deciphered in experiments that began with a cell homogenate that could translate RNA molecules to produce proteins. Fractionation of this homogenate, step by step, produced in turn the ribosomes, tRNAs, and various enzymes that together constitute the protein-synthetic machinery.

- Once individual pure components were available, each could be added or withheld separately to define its exact role in the overall process. A major goal today is the reconstitution of every biological process in a purified cell-free system, so as to be able to define all of its components and their mechanism of action.

CHROMATOGRAPHY

Definition:

- Chromatography (from Greek χρῶμα chroma "color" and γράφειν graphein "to write") is the collective term for a set of laboratory techniques for the separation of mixtures.

- The mixture is dissolved in a fluid called the "mobile phase", which carries it through a structure holding another material called the "stationary phase".

- The various constituents of the mixture travel at different speeds, causing them to separate.

- The separation is based on differential partitioning between the mobile and stationary phases.

- Subtle differences in a compound's partition coefficient result in differential retention on the stationary phase and thus changing the separation.

- Chromatography may be preparative or analytical.

- The purpose of preparative chromatography is to separate the components of a mixture for more advanced use (and is thus a form of purification).

- Analytical chromatography is done normally with smaller amounts of material and is for measuring the relative proportions of analytes in a mixture. The two are not mutually exclusive.

THE PRINCIPLE OF CHROMATOGRAPHY:

- In chromatography a liquid is pumped through a bed of particles.
- The liquid is called the mobile phase and the particles the stationary phase.
- A mixture of the molecules that shall be separated is introduced into the mobile phase.
- In the animation below the blue molecules shall be separated from the red molecules and a mixture containing these two types of molecules is introduced into the mobile phase in front of the stationary phase.
- The mixture of the red and blue molecules is then transported by the mobile phase through the stationary phase.
- The molecules in the mixture that adsorbs the most to the stationary phase, in this particular case the red molecules, is moving slowest through the particle bed. The red molecules become separated from the blue.

Partition co-efficient:In the physical sciences, a partition-coefficientis the ratio of concentrations of a compound in a mixture of two immisciblephases at equilibrium. These coefficients are a measure of the difference in solubility of the compound in these two phases. The two phases are often restricted to mean two immiscible solvents. In this context, a partition coefficient is the ratio of concentrations of a compound in the two phases of a mixture of two immiscible liquids at equilibrium.Normally one of the solvents chosen is aqueous while the second is hydrophobic such as 1-octanol. Hence both the partition and distribution coefficient are measures of how hydrophilic ("water-loving") or hydrophobic ("water-fearing") a chemical substance is. Partition coefficients are useful in estimating the distribution of drugs within the body. Hydrophobic drugs with high octanol/water partition coefficients are preferentially distributed to hydrophobic compartments such as the lipid bilayers of cells while hydrophilic drugs (low octanol/water partition coefficients) preferentially are found in aqueous compartments such as blood serum.

TYPES OF CHROMATOGRAPHY:

Adsorption Chromatography

- Adsorption chromatography is probably one of the oldest types of chromatography around.

- It utilizes a mobile liquid or gaseous phase that is adsorbed onto the surface of a stationary solid phase.

- The equilibration between the mobile and stationary phase accounts for the separation of different solutes.

Partition Chromatography

- This form of chromatography is based on a thin film formed on the surface of a solid support by a liquid stationary phase.

- Solute equilibriates between the mobile phase and the stationary liquid.

Ion Exchange Chromatography

- In this type of chromatography, the use of a resin (the stationary solid phase) is used to covalently attach anions or cations onto it.

- Solute ions of the opposite charge in the mobile liquid phase are attracted to the resin by electrostatic forces.

Molecular Exclusion Chromatography

- Also known as gel permeation or gel filtration, this type of chromatography lacks an attractive interaction between the stationary phase and solute.

- The liquid or gaseous phase passes through a porous gel, which separates the molecules according to its size.

- The pores are normally small and exclude the larger solute molecules, but allow smaller molecules to enter the gel, causing them to flow through a larger volume.
- This causes the smaller molecules to pass through the column at a faster rate than the larger molecules.

Affinity Chromatography

- This is the most selective type of chromatography employed.
- It utilizes the specific interaction between one kind of solute molecule and a second molecule that is immobilized on a stationary phase.
- For example, the immobilized molecule may be an antibody to some specific protein.
- When this molecule passes solutes containing a mixture of proteins, only the specific protein is reacted to this antibody, binding it to the stationary phase.
- This protein is later extracted by changing the ionic strength or pH.

PAPER CHROMATOGRAPHY:

- Paper chromatography is one of the most common types of chromatography in which filter paper serves as a support for immobile liquid phase.
- Removing liquid flows between the fibers of the cellulose but these are not the stationary phase.
- The true stationary phase is the very thin film of liquid usually water adhering o the surface of the fibers. (Water is adsorbed on the fibers/ cellulose by strong hydrogen bonds with –OH of the cellulose).
- The substrate to be separated is distributed between the two liquids, stationary liquid that is held on the fibers of the paper and moving liquid in developing solvent.
- It uses a strip of paper and capillary action is used to pull the solvents up through the paper to separate the solutes.

- A small concentrated spot of solution that contains the sample is applied to a strip of chromatography paper about 2 cm away from the base of the plate, usually using a capillary tube for maximum precision.
- This sample is absorbed onto the paper and may form interactions with it.
- Any substance that reacts or bonds with the paper cannot be measured using this technique.
- The paper is then dipped in to a suitable solvent, such as ethanol or water, taking care that the spot is above the surface of the solvent, and placed in a sealed container.
- The solvent moves up the paper by capillary action, which occurs as a result of the attraction of the solvent molecules to the paper, also this can be explained as differential absorption of the solute components into the solvent.
- As the solvent rises through the paper it meets and dissolves the sample mixture, which will then travel up the paper with the solvent solute sample.
- Different compounds in the sample mixture travel at different rates due to differences in solubility in the solvent, and due to differences in their attraction to the fibers in the paper.
- This method has been largely replaced by thin layer chromatography

SEPARATION OF AMINO ACIDS BY ASCENDING CHROMATOGRAPHY:

- Cut the chromatography sheet carefully to a convenient size (40 x 24cm). Draw a line with pencil across the sheet about 5cm away from one end. Mark a number of points at intervals of 3cm.
- Apply a small volume (say, 25mL) of each amino acid as a separate small spot using a microsyringe. A stream of hot air from a hair-dryer facilitates fast drying of spot. The spot should be as small as possible for better resolution.
- Similarly spot different known aliquots of sample extract.
- After spotting, place the sheet in a stainless steel trough in the chromatography chamber; firmly hold it by placing a long steel rod over the sheet. The spot-end of the sheet should be in the trough (descending chromatography). Otherwise, the sheet may be rolled as a cylinder, tied together with fine thread and placed upright with the spots as the bottom in a large Petridish for upward movement of solvent (ascending chromatography).

- Add the organic (phase) solvent to the trough/petri dish and close the chamber airtight. Develop the chromatogram, preferably overnight or longer, until the solvent moves almost to the other end.

- Note the solvent front and dry the chromatogram free of solvent in a fume chamber.

- Spray the chromatogram with the ninhydrin reagent using an atomizer. Dry the paper for about 5 min at room temp followed by at 100°C in an oven for 2-3 min.

Amino acids appear as purple spots; hydroxyproline and proline give yellow colored spots.Mark all the spots and calculate their Rfvalues by the formula.

Rf = Distance (cm) moved by the solute from the origin

Distance (cm) moved by the solvent from the origin

- The amino acids present in the sample are then identified by comparing the Rf values with that of the authentic amino acids, co-chromatographed.

- For quantitative estimation, cut each spot into several small bits and transfer to the bottom of the test tube. Add 3mL of elution mixture. Shake the tubes vigorously for 15 min. Decant the liquid and elute the pieces with another 2mL of elution mixture. Repeat the elution with small aliquots until the bits are colorless. Combine and clear the eluate by centrifuging at 10,000rpm for 10 min. Read the intensity of purple color at 570nm in a colorimeter. Use the spot of leucine (50mg) run as standard for comparison.

THIN LAYER CHROMATOGRAPHY:

- The surface of the silica gel is polar and, because of the -OH groups, can form hydrogen bonds with suitable compounds around it as well as Van der Waals dispersion forces and dipole-dipole attractions.

- The other commonly used stationary phase is alumina - aluminium oxide. The aluminium atoms on the surface of this also have -OH groups attached.

- Spot the material at the origin (bottom) of the TLC plate.

- Place the plate into a glass jar with a small amount of a solvent in the glass jar.

- This solvent acts as the moving phase.

- Remove the plate from the bottle when the solvent is close to the top of the plate.

Visualization of the spots

- Non-polar compounds are less strongly attracted to the plate and spend more time in the moving phase.
- This compound will move faster and will appear closer to the top of the plate.
- Polar compounds will be more strongly attracted to the plate and will spend less time in the moving phase and appear lower on the plate.
- It is used to detect pesticide or insecticide residues in food.
- Thin-layer chromatography is also used in forensics to analyze the dye composition of fibers.

PRINCIPLE AND SEPARATION OF PHOSPHOLIPIDS:

Principle:

- Similar to other chromatographic methods TLC is also based on the principle of separation.
- The separation depends on the relative affinity of compounds towards stationary and mobile phase.
- The compounds under the influence of mobile phase (driven by capillary action) travel over the surface of stationary phase.
- During this movement the compounds with higher affinity to stationary phase travel slowly while the others travel faster.
- Thus separation of components in the mixture is achieved.
- Once separation occurs individual components are visualized as spots at respective level of travel on the plate.
- Their nature or character are identified by means of suitable detection techniques

Separation of phospholipids:

- A particularly frequent approach is to obtain information on the various phospholipids components of the lipid extract under investigation.

- This can be achieved easily and efficiently by subjecting a sample to thin-layer chromatography (TLC).

- There is a huge number of TLC techniques described for that purpose, some of them using one-dimensional migration others using two successive migrations in two orthogonal directions (two-dimensional).

- We describe below some techniques that have proven easy and reliable in our hands but others could be found in specialized books and scientific journals.

- The proposed simple one-dimensional TLC procedure is routinely used in our laboratory and may be considered as efficient as common two-dimensional techniques.

- It has the peculiarity to allow a good separation of phosphatidylinositol and phosphatidylserine owing to the impregnation of TLC plates with boric acid.

- Other one-dimensional techniques are used for the separation of less frequent phospholipid forms (polyphosphoinositides, mono-and dimethyl phosphatidylethanolamine) or for quantitative estimation by densitometry.

- For some specialized purposes, a two-dimensional procedure is also described.

ADSORPTION CHROMATOGRAPHY:

- Chromatography in which separation is based mainly on differences between the adsorption affinities of the sample components for the surface of an active solid.

Chromatography Adsorbents Description

- Chromatography adsorbents are regularly used in pharmaceutical & chemical manufacturing units, where prime concern of chromatography is to make final pure compounds or make an impurity profile studies for pharmaceutical & Herbal products.

- Column chromatography techniques help in extractions, synthesis & purifications of natural products & Active Pharmaceutical Ingredients.

- Sorbead India with technical tie up with Swambe chemicals manufactures various types of Silica & alumina for chromatographic separations, with an aim to offer various particle sizes & various pore diameters products, which is required for complex chromatographic separations.

- The Silica Gel & Alumina Brockmann I-II grades, which are manufactured, are as Normal phase, Flash grade & gravity grade with different pore diameters.

- Alumina is being offered as basic, acidic and neutral grades.

- Sorbead India measures the pore diameter using the BET method, which guarantees the best and most accurate results during manufacturing process.

- Sorbead manufactures the following chromatography Adsorbents for industrial and laboratory use

1. Aluminum Oxide- Aluminum oxide is a whitish colored powder which is used for chromatography and is highly porous and water adsorbing. With a bulk density of 800-920 gms per liter and a surface area of 180-240 square meters per gram, these adsorbents are suited for various applications like column chromatographic separations, Food Colours, Dyes & Spectroscopic solvents, Herbal extractions of Natural products, isolation and antibiotics purification.

2. Silica Gels – Silica gels have a different mesh size, which are between 35-800. The Silica gel particle size distribution offers an excellent flow rates and a high level of stability. The typical surface area of these gels is 400-800 square meters per gm. Sorbead India adopts a highest quality standard for such a sensitive & accurate product, which ensures a batch-to-batch reproducible result for any chromatographic separations.

Separation:

- Classical column chromatography and TLC were originally used for determination of carotenoids.

- However, these methods are time-consuming, need large sample amounts, their separation efficiency is not particularly high, and they suffer from a poor reproducibility of results and low recoveries of the analytes.

- Among the high-performance separation methods, gas chromatography (GC) is unsuitable, primarily because of low volatility and thermo ability of Carotenoids (however, the volatility of carotenoids can be increased by reducing the double bonds).

- Capillary zone electrophoresis (CZE) is inapplicable because of the absence of charge on the carotenoid molecules.

Adsorption Chromatography

- Therefore, the most common method used in the analysis of carotenoids is HPLC employing various detection techniques.
- Bothnormal- and reversed-phase systems are used, either in isocratic or gradient elution modes.
- Reversed-phase systems several disadvantages, namely, lower column stability.
- In reversed-phase systems, non-aqueous mobile phases are recommended, in view of the pronounced hydrophobicity of carotenoids that makes their separations in mobile phases containing water difficult or impossible. Various mixtures of Solvents, mostly of methanol, acetonitrile and tetrahydrofuran, have been successfully applied to the purpose.
- Antioxidants, such as BHT, are added to the mobile phase, and the temperature of the HPLC column should be maintained low and constant (around 20 0C), to prevent decomposition of carotenoid samples during the HPLC analysis and improve the reproducibility of quantitative analysis.
- An example is the analysis of carotenoids in orange juice (a- and b-carotenes, lutein, zeaxanthin, b-cryptoxanthin) on a C-18 reversed phase, using a ternary mobile phase consisting of a mixture of acetonitrile - methanol – dichloromethane (60:35:5), with additions of the BHT antioxidant (0.1 %), triethylamine (0.1 %), and ammonium acetate (0.05 mol.L-1 solution in methanol).
- Triethylamine and ammonium acetate minimize the effects of acidity generated by the free silanol groups present on the silica support.
- One of the great problems of carotenoid analyses lies in the unavailability of standard compounds caused by natural instability of carotenoids.
- Some carotenoids, e.g., b-carotene, can be obtained commercially but their purity is insufficient for their use as chromatographic standards and thus they must be purified under spectrophotometric control.
- However, the purity of standard solutions must also be verified chromatographically (HPLC).

- The importance of certified materials in analyses for carotenoids in vegetables has been emphasized in an inter laboratory study.

- Carotenoids can be purified using classical column chromatography on alumina, silica gel, magnesium oxide or carbonate, calcium hydroxide or carbonate, Cellite and further adsorbents, with various solvent systems.

- Preparative TLC can be used but preparative HPLC is most common Degradation and isomerization of carotenoids must be prevented during the preparation of standards, by preparing and storing stock solutions in darkness, at -20oC, under a protective atmosphere (nitrogen, argon) and using solvents containing antioxidants.

- An interesting procedure, based on recrystallization and fractional dissolution, has been used for the obtaining of pure all-trans b-carotene.

- This paper also describes the spectral characteristics for the pure and a partially degraded product and points out a high reactivity with atmospheric and dissolved oxygen. The procedure is based on recrystallization and fractional dissolution.

- So far the best separations of various carotenoids have been attained on a C-30 chemically bonded phase.

ION EXCHANGE CHROMATOGRAPHY:

- Imagine if we had a tube whose surfaces were coated with an immobilized cation.

- These would have electrostatic attraction for anions.

- If a solution containing a mixture of positively and negatively charged groups flows through this tube, the anions would preferentially bind, and the cations in the solution would flow through this is the basis of ion exchange chromatography.

- The example above is termed "anion exchange" because the inert surface is interacting with anions

- If the immobile surface was coated with anions, then the chromatography would be termed "cation exchange" chromatography (and cations would selectively bind and be removed from the solution flowing through strength of binding can be affected by pH, and salt concentration of the buffer.

- The ionic species "stuck" to the column can be removed (i.e. "eluted") and collected by changing one of these conditions.
- For example, we could lower the pH of the buffer and protonate anions.
- This would eliminate their electrostatic attraction to the immobilized cation surface. Or, we could increase the salt concentration of the buffer, the anions in the salt would "compete off" bound anions on the cation surface.

PRINCIPLE:

- To optimize binding of all charged molecules, the mobile phase is generally a low to medium conductivity (i.e., low to medium salt concentration) solution.
- The adsorption of the molecules to the solid support is driven by the ionic interaction between the oppositely charged ionic groups in the sample molecule and in the functional ligand on the support.
- The number and location of the charges on the molecule and on the functional group determine the strength of the interaction.
- By increasing the salt concentration (generally by using a linear salt gradient) the molecules with the weakest ionic interactions start to elute from the column first.
- Molecules that have a stronger ionic interaction require a higher salt concentration and elute later in the gradient.
- The binding capacities of ion exchange resins are generally quite high.
- This is of major importance in process scale chromatography, but is not critical for analytical scale separations.

Buffer pH

- As a rule, the pH of the mobile phase buffer must be between the pI (isoelectric point) or pKa (acid dissociation constant) of the charged molecule and the pKa of the charged group on the solid support.
- For example, in cation exchange chromatography, using a functional group on the solid support with a pKa of 1.2, a sample molecule with a pI of 8.2 may be run in a mobile phase buffer of pH 6.0.

- In anion exchange chromatography a molecule with a pI of 6.8 may be run in a mobile phase buffer at pH 8.0 when the pKa of the solid support is 10.3.

Salt Gradients

- As in most other modes of chromatography (SEC being the exception) a protein sample is injected onto the column under conditions where it will be strongly retained.
- A gradient of linearly increasing salt concentration is then applied to elute the sample components from the column.
- An alternative to using a linear gradient is to use a step gradient.
- This requires less complicated equipment and can be very effective to elute different fractions if the appropriate concentrations of salt are known, usually from linear gradient experiments.

Varying pH

- Many chromatographers also use changes in pH to affect a separation.
- In cation exchange chromatography, raising the pH of the mobile phase buffer will cause the molecule to become less protonated and hence less positively charged.
- The result is that the protein no longer can form an ionic interaction with the negatively charged solid support, which ultimately results in the molecule to elute from the column.
- In anion exchange chromatography, lowering the pH of the mobile phase buffer will cause the molecule to become more protonated and hence more positively (and less negatively) charged.
- The result is that the protein no longer can form an ionic interaction with the positively charged solid support, which causes the molecule to elute from the column.

DIFFERENT TYPES OF RESINS:

- In ion-exchange chromatography, four basic types of resins are commonly used.

Strongly acidic cation exchange resins:

- Sulphonated polystyrene resins belong to this class.
- They are useful in the pH range from 1 to 14.
- They are used mainly in the fractionation of cations, inorganic separations, vitamins, peptides and amino acids.

Strongly basic anion exchange resins:

- Quaternary ammonium polystyrene resins belong to this class and are effective between pH 0 and 12.
- They are useful in the fraction of anions, halogens, alkaloids, B-complex vitamins, fatty acids, etc.

Weakly acidic cation exchange resins:

- Carboxylic polymethacrylate is an example of this group.
- The resins are effective between pH 5 and 14.
- They are useful in the fractionation of cations, transition elements, amino acids, antibiotics and organic bases.

Weakly basic anion exchange resins:

- Phenol, formaldehyde and polyamine polystyrene resins belong to this class.
- They are effective in the pH range 0 to 9.
- They are useful in the fractionation of anionic complexes of metals, anions of different valencies, vitamins and amino acid.

SEPARATION OF AMINO ACIDS:

- The strength of the acidity or basicity of these groups and their number per unit volume of resin determine the type and strength of binding of an exchanger.

- Fully ionized acidic groups such as sulfonic acids result in an exchanger with a negative charge, which binds cations very strongly.

- Weakly acidic or basic groups yield resins whose charge (and binding capacity) depends on the pH of the eluting solvent.

- The choice of the appropriate resin depends on the strength of binding desired.

- The bare charges on such solid phases must be counterbalanced by oppositely charged ions in solution ("counterions").

- Washing a cation exchange resin, such as Dowex-50, which has strongly acidic phenyl-SO3- groups, with a NaCl solution results in the formation of the so-called sodium form of the resin. Operation of a cation exchange column for the separation of the amino acids Asp, Ser and Lys.

The separation of amino acids on a cation exchange column.

- When the mixture whose separation is desired is added to the column, the positively charged solute molecules displace the Na+ ions and bind to the resin.

- A gradient of an appropriate salt is then applied to the column, and the solute molecules are competitively (and sequentially) displaced (eluted) from the column by the rising concentration of cations in the gradient, in an order that is inversely related to their affinities for the column.

- The separation of a mixture of amino acids on such a column

- Operation of a cation exchange column, separating a mixture of Asp, Ser, and Lys.

- (a) The cation exchange resin in the beginning, Na+ form.

- (b) A mixture of Asp, Ser, and Lys is added to the column containing the resin.

- (c) A gradient of the eluting salt (e.g., NaCl) is added to the column. Asp, the least positively charged amino acid, is eluted first.

- (d) As the salt concentration increases, Ser is eluted.

- (e) As the salt concentration is increased further, Lys, the most positively charged of the three amino acids, is eluted last

GEL FILTERATION:

- Gel filtration does not rely on any chemical interaction with the protein; rather it is based on a physical property of the protein - that being the effective molecular radius (which relates to mass for most typical globular proteins).
- Gel filtration resin can be thought of as beads, which contain pores of a defined size range.
- Large proteins, which cannot enter these pores, pass around the outside of the beads.
- Smaller proteins, which can enter the pores of the beads, have a longer, tortuous path before they exit the bead.
- Thus, a sample of proteins passing through a gel filtration column will separate based on molecular size: the big ones will elute first and the smallest ones will elute last (and "middle" sized proteins will elute in the middle).
- If your protein is unusually "small" or "large"in comparison to contaminating proteins then gel filtration may work quite well.

Principle:

- Gel media A gel is a heterogeneous phase system in which a continuous liquid phase, usually aqueous, is contained within the pores of a continuous solid phase,

the gel matrix.

- In gels made for gel filtration, the pores have a carefully

controlled range of sizes, and the matrix is chosen for its chemical and

physical stability, and inertness (lack of adsorptive properties).

- Gels may be formed from polymers by cross-linking to form a

three-dimensional network; for example Sephadex that is formed by cross-linking dextran. Some polymers, like agarose, form gels spontaneou-

sly under the appropriate conditions.

- Composite gels may be prepared by, for example, grafting a second polymer onto a pre-formed matrix.

- Superdex is such a gel. Dextran chains arecovalently bonded to a highly cross-linked agarose gel matrix.

- Compositegels are of interest since they can combine valuable properties from morethan one gel-forming system

ESTIMATION OF MOLECULAR SIZE AND MOLECULAR WEIGHT OF A BIOMACROMOLECULE:

Molecular size:

- Most molecules are far too small to be seen with the naked eye, but there are exceptions. DNA, a macromolecule, can reach macroscopic sizes, as can molecules of many polymers.

- The smallest molecule is the diatomic hydrogen (H2), with a bond length of 0.74 Å.

- Molecules commonly used as building blocks for organic synthesis have a dimension of a few Å to several dozen Å.

- Single molecules cannot usually be observed by light (as noted above), but small molecules and even the outlines of individual atoms may be traced in some circumstances by use of an atomic force microscope.

- Some of the largest molecules are macromolecules or supermolecules

Molecular Weight:

- The log of a molecule's molecular weight is proportional to the distance that molecule has migrated.

- Therefore, the first step is to generate a standard curve using molecules of know size (the molecular weight markers).

- When using semilog paper (see the next page), the molecular weights (in bp for DNA and kiloDaltons, kDa, for proteins) is plotted on the Y-axis and the distance the molecule migrated is plotted on the X-axis.
- When generating a standard curve, you will obtain a straight line (use a best-fit line).
- Once your standard curve is ready, measure the distance traveled by your molecule of interest.
- Find that distance on the X-axis, and go up until you intersect with your standard curve.
- Move over to the Y-axis and that will indicate the molecular weight of the molecule you are studying.
- Use the graph paper below and the DNA gel shown to the right to determine the molecular weight of the unknown band indicated with an arrow.

AFFINITY CHROMATOGRAPHY:

- Affinity chromatography separates proteins on the basis of a reversible interaction between a protein (or group of proteins) and a specific ligand coupled to a chromatography matrix.
- The technique offers high selectivity, hence high resolution, and usually high capacity for the protein(s) of interest.
- Purification can be in the order of several thousand fold and recoveries of active material are generally very high.
- Affinity chromatography is unique in purification technology since it is the only technique that enables the purification of a biomolecule on the basis of its biological function or individual chemical structure.
- Purification that would otherwise be time-consuming, difficult or even impossible using other techniques can often be easily achieved with affinity chromatography.
- The technique can be used to separate active biomolecules from denatured or functionally different forms, to isolate pure substances present at low concentration in large volumes of crude sample and also to remove specific contaminants.

PRINCIPLE AND SEPARATION OF AN ENZYME:

Principle

- The immobile phase is typically a gel matrix, often of agarose; a linear sugar molecule derived from algae.
- Usually the starting point is an undefined heterogeneous group of molecules in solution, such as a cell lysate, growth medium or blood serum.
- The molecule of interest will have a well-known and defined property which can be exploited during the affinity purification process.
- The process itself can be thought of as an entrapment, with the target molecule becoming trapped on a solid or stationary phase or medium.
- The other molecules in solution will not become trapped, as they do not possess this property.
- The solid medium can then be removed from the mixture, washed and the target molecule released from the entrapment in a process known as elution.
- Possibly the most common use of affinity chromatography is for the purification of recombinant proteins.

Separation

- The goal of affinity chromatography is to separate all the molecules of a particular specificity from the whole gamut of molecules in a mixture such as a blood serum.
- For example, the antibodies in a serum sample specific for a particular antigenic determinant can be isolated by the use of affinity chromatography.

Step 1:

- An immunoadsorbent is prepared. This consists of a solid matrix to which the antigen (shown in blue) has been coupled (usually covalently).
- Agarose, sephadex, derivatives of cellulose, or other polymers can be used as the matrix.

Step 2:

- The serum is passed over the immunoadsorbent.
- As long as the capacity of the column is not exceeded, those antibodies in the mixture specific for the antigen (shown in red) will bind (noncovalently) and be retained. Antibodies of other specificities (green) and other serum proteins (yellow) will pass through unimpeded.

Step 3: Elution

- A reagent is passed into the column to release the antibodies from the immunoadsorbent.
- Buffers containing a high concentration of salts and/or low pH are often used to disrupt the noncovalent interactions between antibodies and antigen.
- A denaturing agent, such as 8 M urea, will also break the interaction by altering the configuration of the antigen-binding site of the antibody molecule.
- Another, gentler, approach is to elute with a soluble form of the antigen.
- These compete with the immunoadsorbent for the antigen-binding sites of the antibodies and release the antibodies to the fluid phase.

Step 4:Dialysis

- The eluate is then dialyzed against, for example, buffered saline in order to remove the reagent used for elution.

UNIT III Electrophoresis: Principle, instrumentation and applications of agarose gel electrophoresis, sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE), native PAGE, isoelectric focusing, immunoelectrophoresis, 2D gel electrophoresis. Pulse field gel electrophoresis, capillary electrophoresis, gel documentation – Applications. Blotting techniques.

ELECTROPHORESIS:

- Electrophoresis is the motion of dispersed particles relative to a fluid under the influence of a spatially uniform electric field
- A technique for separating the components of a mixture of charged molecules (proteins, DNAs, or RNAs) in an electric field within a gel or other support.
- The movement of electrically charged molecules in an electric field often resulting in their separation.

PRINCIPLES:

- Migration of charged particles in an electric field.
- The rate of travel of the particle depends upon the following major factors.

a) The charge of the particle

b) Applied electrical field

c) Temperature and

d) Nature of the suspended medium.

- Many biologically important molecules such as amino acids, peptides, proteins, nucleotides and nucleic acids possess ionizable groups.
- Under the influence of an electric field, these charged particles will migrate either to the cathode or to the anode, depending on the nature of their net charge.
- There are two main types of electrophoretic methods.
- When the separation is carried out in absence of a supporting or stabilizing medium, the method is called free solution method.

- When the separation is carried out in presence of a stabilizing medium, such as paper gel, the technique is known as zone electrophoresis.

TYPES OF ELECTROPHORESIS:

Agarose Gel Electrophoresis

- Agarose gels are commonly used to sort DNA and RNA molecules based on size.
- The agarose gel concentration can be varied, based on the size of the molecules that need to be isolated.

SDS-PAGE Electrophoresis

- Sodium dodecyl sulfate - polyacrylamide gel electrophoresis is used to separate proteins based on size.
- The proteins are unfolded, or denatured, using SDS detergent, and run on a polyacrylamide gel.

PAPER ELECTROPHORESIS:

- Paper electrophoresis technique is a simple and less expensive one.
- It requires micro quantities of plasma for separation.
- The serum under investigation is mixed with bromophenol blue, a blue coloured stain, and spotted at the centre of a strip of a special filter paper, saturated with barbitone of pH 8.6.

Graphical representation of the concentration of serum protein fractions

- When an electric current of proper amperage and voltage is passes through the paper, charged protein fractions bearing different charges migrate at different rates.
- If the pH of the serum is adjusted by the addition of a proper buffer to a value alkaline to the isoelectric points of all the fractions of plasma protein, they all will carry negative charges, but of different magnitudes.

- The different fraction\s of plasma will migrate toward the anode at characteristically different rates.
- After a run of about 5 to 6 hours, the paper is dried and stained with a solution containing bromophenol blue.
- In human serum, five different bands can be identified on paper electrophoresis.
- They are designated in the order of decreasing mobility as albumin, alpha1 – globulin, alpha2 – globulin, beta globulin, fibrinogen and gamma – globulin.
- Albumin being the fastest moving fraction of the proteins of plasma forms the last band of the paper.
- Gamma globulin, which is the slowest moving protein, forms a band at the other end.
- The rest of the fractions take their positions in between theses two bands.

AGAROSE GEL ELECTROPHORESIS:

Agarose gels provide a simple method for analyzing preparations of DNA. Although the base compositions of individual DNA molecules vary, the basic chemical structure of DNA is the same for all DNA molecules. DNA molecules share the same charge/mass ratio, which imparts similar electrophoretic properties to DNAs of widely varying lengths.

Procedure

- An electrophoresis chamber and power supply
- Gel casting trays, which are available in a variety of sizes and composed of UV-transparent plastic. The open ends of the trays are closed with tape while the gel is being cast, then removed prior to electrophoresis.
- Sample combs, around which molten agarose is poured to form sample wells in the gel.
- Electrophoresis buffer, usually Tris-acetate-EDTA (TAE) or Tris-borate-EDTA (TBE).
- Loading buffer, which contains something dense (e.g. glycerol) to allow the sample to "fall" into the sample wells, and one or two tracking dyes, which migrate in the gel and allow visual monitoring or how far the electrophoresis has proceeded.

- Ethidium bromide, a fluorescent dye used for staining nucleic acids. NOTE: Ethidium bromide is a known mutagen and should be handled as a hazardous chemical - wear gloves while handling.

- Transilluminator (an ultraviolet lightbox), which is used to visualize ethidium bromide-stained DNA in gels. NOTE: always wear protective eyewear when observing DNA on a transilluminator to prevent damage to the eyes from UV light.

- To pour a gel, agarose powder is mixed with electrophoresis buffer to the desired concentration, then heated in a microwave oven until completely melted. Most commonly, ethidium bromide is added to the gel (final concentration 0.5 ug/ml) at this point to facilitate visualization of DNA after electrophoresis. After cooling the solution to about 60C, it is poured into a casting tray containing a sample comb and allowed to solidify at room temperature or, if you are in a big hurry, in a refrigerator.

- After the gel has solidified, the comb is removed, using care not to rip the bottom of the wells. The gel, still in its plastic tray, is inserted horizontally into the electrophoresis chamber and just covered with buffer. Samples containing DNA mixed with loading buffer are then pipeted into the sample wells, the lid and power leads are placed on the apparatus, and a current is applied. You can confirm that current is flowing by observing bubbles coming off the electrodes.DNA will migrate towards the positive electrode, which is usually colored red.

- When adequate migration has occured, DNA fragments are visualized by staining with ethidium bromide. This fluorescent dye intercalates between bases of DNA and RNA. It is often incorporated into the gel so that staining occurs during electrophoresis, but the gel can also be stained after electrophoresis by soaking in a dilute solution of ethidium bromide. To visualize DNA or RNA, the gel is placed on a ultraviolet transilluminator. Be aware that DNA will diffuse within the gel over time, and examination or photography should take place shortly after cessation of electrophoresis.

Migration of DNA Fragments in Agarose

Fragments of linear DNA migrate through agarose gels with a mobility that is inversely proportional to the log10 of their molecular weight. In other words, if you plot the distance from the well that DNA fragments have migrated against the log10 of either their molecular weights or number of base pairs, a roughly straight line will appear.

Agarose Concentration: By using gels with different concentrations of agarose, one can resolve different sizes of DNA fragments. Higher concentrations of agarose facilite separation of small DNAs, while low agarose concentrations allow resolution of larger DNAs.

The image to the right shows migration of a set of DNA fragments in three concentrations of agarose, all of which were in the same gel tray and electrophoresed at the same voltage and for identical times. Notice how the larger fragments are much better resolved in the 0.7% gel, while the small fragments separated best in 1.5% agarose. The 1000 bp fragment is indicated in each lane.

Voltage: As the voltage applied to a gel is increased, larger fragments migrate proportionally faster that small fragment. For that reason, the best resolution of fragments larger than about 2 kb is attained by applying no more than 5 volts per cm to the gel (the cm value is the distance between the two electrodes, not the length of the gel).

Electrophoresis Buffer: Several different buffers have been recommended for electrophoresis of DNA. The most commonly used for duplex DNA are TAE (Tris-acetate-EDTA) and TBE (Tris-borate-EDTA). DNA fragments will migrate at somewhat different rates in these two buffers due to differences in ionic strength. Buffers not only establish a pH, but provide ions to support conductivity. If you mistakenly use water instead of buffer, there will be essentially no migration of DNA in the gel! Conversely, if you use concentrated buffer (e.g. a 10X stock solution), enough heat may be generated in the gel to melt it.

Effects of Ethidium Bromide: Ethidium bromide is a fluorescent dye that intercalates between bases of nucleic acids and allows very convenient detection of DNA fragments in gels, as shown by all the images on this page. As described above, it can be incorporated into agarose gels, or added to samples of DNA before loading to enable visualization of the fragments within the gel. As might be expected, binding of ethidium bromide to DNA alters its mass and rigidity, and therefore its mobility.

Applications

- Estimation of the size of DNA molecules following restriction enzyme digestion, e.g. in restriction mapping of cloned DNA.
- Analysis of PCR products, e.g. in molecular genetic diagnosis or genetic fingerprinting
- Separation of DNA fragments for extraction and purification.
- Separation of restricted genomic DNA prior to Southern transfer, or of RNA prior to Northern transfer.

Agarose gels are easily cast and handled compared to other matrices and nucleic acids are not chemically altered during electrophoresis. Samples are also easily recovered. After the experiment is finished, the resulting gel can be stored in a plastic bag in a refrigerator.

Electrophoresis is performed in buffer solutions to reduce pH changes due to the electric field, which is important because the charge of DNA and RNA depends on pH, but running for too long can exhaust the buffering capacity of the solution. Further, different preparations of genetic material may not migrate consistently with each other, for morphological or other reasons.

POLYACRYLAMIDE GEL ELECTROPHORESIS:

- To separate proteins on the basis of their size and charge

Theory

- PAGE (Polyacrylamide Gel Electrophoresis), is an analytical method used to separate components of a protein mixture based on their size.
- The technique is based upon the principle that a charged molecule will migrate in an electric field towards an electrode with opposite sign.
- The general electrophoresis techniques cannot be used to determine the molecular weight of biological molecules because the mobility of a substance in the gel depends on both charge and size.
- To overcome this, the biological samples needs to be treated so that they acquire uniform charge, then the electrophoretic mobility depends primarily on size.
- For this different protein molecules with different shapes and sizes, needs to be denatured(done with the aid of SDS) so that the proteins lost their secondary, tertiary or quaternary structure.
- The proteins being covered by SDS are negatively charged and when loaded onto a gel and placed in an electric field, it will migrate towards the anode (positively charged electrode) are separated by a molecular sieving effect based on size.
- After the visualization by a staining (protein-specific) technique, the size of a protein can be calculated by comparing its migration distance with that of a known molecular weight ladder (marker).

Principle behind separation:

- Separation of charged molecules in an electric field is based on the relative mobility of charged species which is related to frictional resistance

Charge of the species:

PAGE is working upon the principle in which, the charged molecule will migrate towards the appositively charged electrode through highly cross linked matrix.

- Separation occurs due to different rates of migration occurs by the magnitude of charge and frictional resistance related to the size.

Relative Mobility:

where,

Z = charge on the molecule

E = Voltage applied and ,

f = frictional resistance

Rf is measured by:

Direction of movement is determined from Z : -

if $Z < 0$, then $\rightarrow +$

if $Z > 0$, then $\rightarrow -$

if $Z = 0$, then no movement

- The gel is used is divided into an upper "stacking" gel of low percentage (with large pore size) and low pH (6.8) , where the protein bands get squeezed down as a thin layer migrating toward the anode and a resolving gel (pH 8.8) with smaller pores.

- Cl - is the only mobile anion present in both gels. When electrophoresis begins, glycine present in the electrophoresis buffer, enters the stacking gel, where the equilibrium favors zwitter ionic form with zero net charge.

- The glycine front moves through the stacking gel slowly, lagging behind the strongly charged, Cl- ions.

- Since these two current carrying species separate, a region of low conductivity, with high voltage drop, is formed between them.

- This zone sweeps the proteins through the large pores of the stacking gel, and depositing it at the top of the resolving gel as a narrow band.

Stacking gel interactions:

- Stacking occurs by the differential migration of ionic species, which carry the electric current through the gel.

- When an electrical current is applied to the gel,the negatively charged molecules start migrating to the positively charged electrode .

- Cl- ions, having the highest charge/mass ratio move faster, being depleted and concentrated at anode end.

- SDS coated proteins has a higher charge/mass ratio than glycine so it moves fast, but slower than Cl-.

- When protein encounters resolving gel it slows the migration because of increased frictional resistance, allowing the protein to stack in the gel.

Resolving Gel Interactions:

- When glycine reaches resolving gel it becomes positively charged and migrates much faster than protein due to higher charge/mass ratio.

- Now proteins are the main carrier of current and separate according to their molecular mass by the sieving effect of pores in gel.

Materials Required For PAGE

- Acrylamide solutions (for resolving & stacking gels).

- Isopropanol /distilled water.

- Gel loading buffer.
- Running buffer.
- Staining, destaining solutions.
- Protein samples.
- Molecular weight markers.

The equipment and supplies necessary for conducting SDS-PAGE includes:

- An electrophoresis chamber and power supply.
- Glass plates(a short and a top plate).
- Casting frame .
- Casting stand.
- Combs.

Objectives:

Separation of proteins based on their molecular weight.

1. Gloves should be worn, while performing SDS-PAGE.
2. To ensure proper alignment,all the requirements should be clean.
3. Special attention should be paid while using acrylamide (since it is a neurotoxin).

Stock solutions

- 30% Polyacrylamide solution(29g acrylamide+1g bisacrylamide in 50 mL of water, dissolve completely using a magnetic stirrer, make the volume upto 100mL).Keep the solution away from sunlight.
- 5 M Tris, pH 8.8 (Add 121.14g of Tris in 100 ml water. Adjust the pH to 8.8 and make up the final volume to 200ml).
- 5 M Tris, pH 6.8 (Add 60.57g of Tris in 40 ml water. Adjust the pH to 6.8 and themake up the final volume to 100 ml).
- 10% SDS (10 g SDS in 100mL distilled water).

- 10% ammonium persulfate (0.1 g in 1 ml water). It should be freshly prepared.
- 10x SDS running buffer (pH ~8.3) - Take 60.6 g Tris base,288g Glycine and 20g SDS in separate beakers and dissolve them using distilled water. When properly dissolved, mix three of them and make upto 2L.(working standard is 1X buffer)

Gel loading buffer:

To make 10 ml of 4X stock:

- 2.0 ml 1M Tris-HCl pH 6.8.
- 0.8 g SDS.
- 4.0 ml 100% glycerol.
- 0.4 ml 14.7 M
- β-mercaptoethanol.
- 0.5 M EDTA.
- 8 mg bromophenol Blue.

Staining solution:

- Weigh 0.25g of Coomassie Brilliant Blue R250 in a beaker. Add 90 ml methanol: water (1:1 v/v) and 10ml of Glacial acetic acid ,mix properly using a magnetic stirrer. (When properly mixed, filter the solution through a Whatman No. 1 filter to remove any particulate matter and store in appropriate bottles)

Destaining solution:

- Mix 90 ml methanol: water (1:1 v/v) and 10ml of Glacial acetic acid using a magnetic stirrer and store in appropriate bottles.

Resolving gel (10%)	Stacking gel (5%)
dH20 4.0 ml	
30% acrylamide mix 3.3 ml	
1.5M Tris pH8.8 2.5 ml	
10% SDS 0.1 ml	

10% ammonium persulfate 0.1 ml

TEMED 0.004ml

dH20 5.65 ml

30% acrylamide mix 1.65 ml

1.0M Tris pH 6.8 2.5 ml

10% SDS 0.1 ml

10% ammonium persulfate 0.1 ml

TEMED 0.004ml

UNIT 4

Introduction

In the modern world of computers and information technology, the importance of statistics is very well recogonised by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

Origin and Growth of Statistics:

The word ‘ Statistics’ and ‘ Statistical’ are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of scientific method is of comparatively recent growth. Research particularly into the mathematical theory of statistics is rapidly proceeding and fresh discoveries are being made all over the world.

Meaning of Statistics:

Statistics is concerned with scientific methods for collecting, organising, summarising, presenting and analysing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis. Statistics is concerned with the systematic collection of numerical data and its interpretation.

The word ‘ statistic’ is used to refer to
1. Numerical facts, such as the number of people living inparticular area.
2. The study of ways of collecting, analysing and interpreting the facts.

Definition by Croxton and Cowden:

Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis. It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one.
According to this definition there are four stages:
1. Collection of Data: It is the first step and this is the foundation upon

which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of correct method.

2. Presentation of data: The mass data collected should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

3. Analysis of data: The data presented should be carefully analysed for making inference from the presented data such as measures of central tendencies, dispersion, correlation, regression etc.,

4. Interpretation of data: The final step is drawing conclusion from the data collected. A valid conclusion must be drawn on the basis of analysis. A high degree of skill and experience is necessary for the interpretation.

Functions of Statistics:

There are many functions of statistics. Let us consider the following five important functions.

1. Condensation:

Generally speaking by the word ‘ to condense’ , we mean to reduce or to lessen. Condensation is mainly applied at embracing the understanding of a huge mass of data by providing only few observations. If in a particular class in Chennai School, only marks in an examination are given, no purpose will be served. Instead if we are given the average mark in that particular examination, definitely it serves the better purpose. Similarly the range of marks is also another measure of the data. Thus, Statistical measures help to reduce the complexity of the data and consequently to understand any huge mass of data.

2. Comparison:

Classification and tabulation are the two methods that are used to condense the data. They help us to compare data collected from different sources. Grand totals, measures of central tendency measures of dispersion, graphs and diagrams, coefficient of correlation etc provide ample scope for comparison.

If we have one group of data, we can compare within itself. If the rice production (in Tonnes) in Tanjore district is known, then we can compare one

region with another region within the district. Or if the rice production (in Tonnes) of two different districts within Tamilnadu is known, then also a comparative study can be made. As statistics is an aggregate of facts and figures, comparison is always possible and in fact comparison helps us to understand the data in a better way.

3. Forecasting:

By the word forecasting, we mean to predict or to estimate before hand. Given the data of the last ten years connected to rainfall of a particular district in Tamilnadu, it is possible to predict or forecast the rainfall for the near future. In business also forecasting plays a dominant role in connection with production, sales, profits etc. The analysis of time series and regression analysis plays an important role in forecasting.

4. Estimation:

One of the main objectives of statistics is drawn inference about a population from the analysis for the sample drawn from that population. The four major branches of statistical inference are

1. Estimation theory
2. Tests of Hypothesis
3. Non Parametric tests
4. Sequential analysis

In estimation theory, we estimate the unknown value of the population parameter based on the sample observations. Suppose we are given a sample of heights of hundred students in a school, based upon the heights of these 100 students, it is possible to estimate the average height of all students in that school.

5.Tests of Hypothesis:

A statistical hypothesis is some statement about the probability distribution, characterising a population on the basis of the information available from the sample observations. In the formulation and testing of hypothesis, statistical methods are extremely useful. Whether crop yield has increased because of the use of new fertilizer or whether the new medicine is effective in eliminating a particular disease are some examples of statements of hypothesis and these are tested by proper statistical tools.

6. Scope of Statistics:

Statistics is not a mere device for collecting numerical data, but as a means of developing sound techniques for their handling, analysing and drawing valid inferences from them. Statistics is applied in every sphere of human activity – social as well as physical – like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics cannot be applied.

Limitations of statistics:

Statistics with all its wide application in every sphere of human activity has its own limitations. Some of them are given below.

1.Statistics is not suitable to the study of qualitative phenomenon: Since statistics is basically a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements. As a matter of fact,qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed numerically and any statistical analysis cannot be directly applied on these qualitative phenomenons. Nevertheless, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.

2. Statistics does not study individuals:

Statistics does not give any specific importance to the individual items, in fact it deals with an aggregate of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.

3. Statistical laws are not exact:

It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.

4. Statistics table may be misused:

Statistics must be used only by experts; otherwise, statistical methods are the most dangerous tools on the hands of the inexpert. The use of statistical

tools by the inexperienced and untraced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. As King says aptly ‘ statistics are like clay of which one can make a God or Devil as one pleases’ .

5. Statistics is only, one of the methods of studying a problem:

Statistical method do not provide complete solution of the problems because problems are to be studied taking the background of the countries culture, philosophy or religion\ into consideration. Thus the statistical study should be supplemented by other evidences.

COLLECTION OF DATA, CLASSIFICATION AND TABULATION

Nature of data:

It may be noted that different types of data can be collected for different purposes. The data can be collected in connection with time or geographical location or in connection with time and location. The following are the three types of data:

1. Time series data.
2. Spatial data
3. Spacio-temporal data.

1 Time series data:

It is a collection of a set of numerical values, collected over a period of time. The data might have been collected either at regular intervals of time or irregular intervals of time.

Example 1:

The following is the data for the three types of expenditures in rupees for a family for the four years 2001,2002,2003,2004.

Year	Food	Education	Others	Total
2001	3000	2000	3000	8000
2002	3500	3000	4000	10500
2003	4000	3500	5000	12500
2004	5000	5000	6000	16000

2 Spatial Data:

If the data collected is connected with that of a place, then it is termed as spatial data. For example, the data may be

1. Number of runs scored by a batsman in different test matches in a test series at different places
2. District wise rainfall in Tamilnadu
3. Prices of silver in four metropolitan cities

Example 2:

The population of the southern states of India in 1991.

State	Population
Tamilnadu	5,56,38,318
Andhra Pradesh	6,63,04,854
Karnataka	4,48,17,398
Kerala	2,90,11,237
Pondicherry	7,89,416

3.2.3 Spacio Temporal Data:

If the data collected is connected to the time as well as place then it is known as spacio temporal data.

Example 3:

State	Population	
	1981	1991
Tamil Nadu	4,82,97,456	5,56,38,318
Andhra Pradesh	5,34,03,619	6,63,04,854
Karnataka	3,70,43,451	4,48,17,398
Kerala	2,54,03,217	2,90,11,237
Pondicherry	6,04,136	7,89,416

3.3 Categories of data:

Any statistical data can be classified under two categories depending upon the sources utilized.

These categories are,

1. Primary data
2. Secondary data

3.3.1 Primary data:

Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.

Example 4:

If a researcher is interested to know the impact of noon-meal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

The primary data can be collected by the following five methods.

1. Direct personal interviews.
2. Indirect Oral interviews.
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.

3.3.2 Secondary Data:

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency. According to W.A.Neiswanger, ' A primary source is a publication in which the data are published by the same authority which gathered and analysed them. A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible' .

Sources of Secondary data:

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. The sources of secondary data can broadly be classified under two heads:

1. Published sources, and
2. Unpublished sources.

3.4 Classification:

The collected data, also known as raw data or ungrouped data are always in an un organised form and need to be organised and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible and assimilable form. The process of grouping into different classes or sub classes according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data. Thus classification is the first step in tabulation.

For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore, Mumbai etc.,

Objects of Classification:

The following are main objectives of classifying the data:

1. It condenses the mass of data in an easily assimilable form.
2. It eliminates unnecessary details.
3. It facilitates comparison and highlights the significant aspect of data.
4. It enables one to get a mental picture of the information and helps in drawing inferences.
5. It helps in the statistical treatment of the information collected.

Types of classification:

Statistical data are classified in respect of their characteristics. Broadly there are four basic types of classification namely

a) Chronological classification
b) Geographical classification
c) Qualitative classification
d) Quantitative classification

a) Chronological classification:

In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc., The data is generally classified in ascending order of time. For example, the data related with population, sales of a firm, imports and exports of a country are always subjected to chronological classification.

Example 5:

The estimates of birth rates in India during 1970 – 76 are

Year	1970	1971	1972	1973	1974	1975	1976
Birth Rate	36.8	36.9	36.6	34.6	34.5	35.2	34.2

b) Geographical classification:

In this type of classification the data are classified according to geographical region or place. For instance, the production of paddy in different states in India, production of wheat in different countries etc.,

Example 6:

Country	America	China	Denmark	France	India
Yield of wheat in (kg/acre)	1925	893	225	439	862

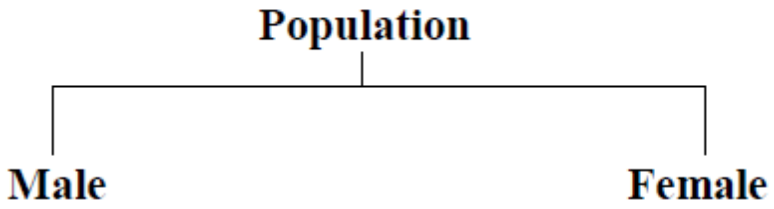
c) Qualitative classification:

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc., Such attributes cannot be measured along with a scale.

For example, if the population to be classified in respect to one attribute, say sex, then we can classify them into two namely that of males and females. Similarly, they can also be classified into 'employed' or 'unemployed' on the basis of another attribute 'employment'.

Thus when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is called simple or dichotomous classification.

A simple classification may be shown as under



d) Quantitative classification:

Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc., For example the students of a college may be classified according to weight as given below.

Weight (in lbs)	No of Students
90-100	50
100-110	200
110-120	260
120-130	360
130-140	90
140-150	40
Total	1000

In this type of classification there are two elements, namely (i) the variable (i.e) the weight in the above example, and (ii) the frequency in the number of students in each class. There are 50 students having weights ranging from 90 to 100 lb, 200 students having weight ranging between 100 to 110 lb and so on.

3.5 Tabulation:

Tabulation is the process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information. A table is a systematic arrangement of classified data in columns and rows. Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious.Classification and 'Tabulation', as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation data are classified and then displayed under different columns and rows of a table.

Advantages of Tabulation:

Statistical data arranged in a tabular form serve following objectives:

1. It simplifies complex data and the data presented are easily understood.
2. It facilitates comparison of related facts.
3. It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
4. It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided. Moreover, the needed information can be easily located.
5. Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

Preparing a Table:

The making of a compact table itself an art. This should contain all the information needed within the smallest possible space. What the purpose of tabulation is and how the tabulated information is to be used are the main points to be kept in mind while preparing for a statistical table. An ideal table should consist of the following main parts:

1. Table number
2. Title of the table
3. Captions or column headings
4. Stubs or row designation
5. Body of the table
6. Footnotes
7. Sources of data

A model structure of a table is given below:

Table Number **Title of the Table**

Sub Heading	Caption Headings / Caption Sub-Headings	Total
Stub Sub- Headings	Body	
Total		

Requirements of a Good Table:

A good statistical table is not merely a careless grouping of columns and rows but should be such that it summarizes the total information in an easily accessible form in minimum possible space. Thus while preparing a table, one must have a clear idea of the information to be presented, the facts to be compared and he points to be stressed.

Though, there is no hard and fast rule for forming a table yet a few general point should be kept in mind:

1. A table should be formed in keeping with the objects of statistical enquiry.
2. A table should be carefully prepared so that it is easily understandable.
3. A table should be formed so as to suit the size of the paper. But such an adjustment should not be at the cost of legibility.
4. If the figures in the table are large, they should be suitably rounded or approximated. The method of approximation and units of measurements too should be specified.
5. Rows and columns in a table should be numbered and certain figures to be stressed may be put in ' box' or ' circle' or in bold letters.
6. The arrangements of rows and columns should be in a logical and systematic order. This arrangement may be alphabetical, chronological or according to size.
7. The rows and columns are separated by single, double or thick lines to represent various classes and sub-classes used. The corresponding proportions or percentages should be given in adjoining rows and columns to enable comparison. A vertical expansion of the table is generally more convenient than the horizontal one.
8. The averages or totals of different rows should be given at the right of the table and that of columns at the bottom of the table. Totals for every sub-class too should be mentioned.

9. In case it is not possible to accommodate all the information in a single table, it is better to have two or more related tables.

Type of Tables:

Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:

1. Simple or one-way table
2. Two way table
3. Manifold table

Simple or one-way Table:

A simple or one-way table is the simplest table which contains data of one characteristic only. A simple table is easy to construct and simple to follow. For example, the blank table given below may be used to show the number of adults in different occupations in a locality.

The number of adults in different occupations in a locality

Occupations	No. Of Adults
Total	

Two-way Table:

A table, which contains data on two characteristics, is called a two-way table. In such case, therefore, either stub or caption is divided into two co-ordinate parts. In the given table, as an example the caption may be further divided in respect of ' sex' . This subdivision is shown in two-way table, which now contains two characteristics namely, occupation and sex.

The umber of adults in a locality in respect of occupation and sex

Occupation	**No. of Adults**		**Total**
	Male	**Female**	
Total			

Manifold Table:

Thus, more and more complex tables can be formed by including other characteristics. For example, we may further classify the caption sub-headings in the above table in respect of "marital status", " religion" and "socio-economic status" etc. A table ,which has more than two characteristics of data is considered as a manifold table. For instance , table shown below shows three characteristics namely, occupation, sex and marital status.

Occupation	No. of Adults						Total
	Male			Female			
	M	**U**	**Total**	**M**	**U**	**Total**	
Total							

Foot note: M Stands for Married and U stands for unmarried.

Manifold tables, though complex are good in practice as these enable full information to be incorporated and facilitate analysis of all related facts. Still, as a normal practice, not more than four characteristics should be represented in one table to avoid confusion. Other related tables may be formed to show the remaining characteristics

a) Discrete (or) Ungrouped frequency distribution:

Example 1:

In a survey of 40 families in a village, the number of children per family was recorded and the following data obtained.

1	0	3	2	1	5	6	2
2	1	0	3	4	2	1	6
3	2	1	5	3	3	2	4
2	2	3	0	2	1	4	5
3	3	4	4	1	2	4	5

Represent the data in the form of a discrete frequency distribution.

Solution:

Frequency distribution of the number of children

Number of Children	Tally Marks	Frequency
0	|||	3
1	𝍸 ||	7
2	𝍸 𝍸	10
3	𝍸 |||	8
4	𝍸 |	6
5	||||	4
6	||	2
	Total	40

b) Continuous frequency distribution:

In this form of distribution refers to groups of values. This becomes necessary in the case of some variables which can take any fractional value and in which case an exact measurement is not possible. Hence a discrete variable can be presented in the form of a continuous frequency distribution.

Wage distribution of 100 employees

Weekly wages (Rs)	Number of employees
50-100	4
100-150	12
150-200	22
200-250	33
250-300	16
300-350	8
350-400	5
Total	100

Cumulative frequency table:

Example 3:

Age group (in years)	Number of women	Less than Cumulative frequency	More than cumulative frequency
15-20	3	3	64
20-25	7	10	61
25-30	15	25	54
30-35	21	46	39
35-40	12	58	18
40-45	6	64	6

(a) Less than cumulative frequency distribution table

End values upper limit	less than Cumulative frequency
Less than 20	3
Less than 25	10
Less than 30	25
Less than 35	46
Less than 40	58
Less than 45	64

(b) More than cumulative frequency distribution table

End values lower limit	Cumulative frequency more than
15 and above	64
20 and above	61
25 and above	54
30 and above	39
35 and above	18
40 and above	6

DIAGRAMATIC AND GRAPHICAL REPRESENTATION

5.1 Introduction:

In the previous chapter, we have discussed the techniques of classification and tabulation that help in summarising the collected data and presenting them in a systematic manner. However, these forms of presentation do not always prove to be interesting to the common man. One of the most convincing and appealing ways in which statistical results may be presented is through diagrams and graphs. Just one diagram is enough to represent a given data more effectively than thousand words.

Moreover even a layman who has nothing to do with numbers can also understands diagrams. Evidence of this can be found in newspapers, magazines, journals, advertisement, etc. An attempt is made in this chapter to illustrate some of the major types of diagrams and graphs frequently used in presenting statistical data.

5.2 Diagrams:

A diagram is a visual form for presentation of statistical data, highlighting their basic facts and relationship. If we draw diagrams on the basis of the data collected they will easily be understood and appreciated by all. It is readily intelligible and save a considerable amount of time and energy.

5.3 Significance of Diagrams and Graphs:

Diagrams and graphs are extremely useful because of the following reasons.

1. They are attractive and impressive.
2. They make data simple and intelligible.
3. They make comparison possible
4. They save time and labour.
5. They have universal utility.
6. They give more information.
7. They have a great memorizing effect.

5.4 General rules for constructing diagrams:

The construction of diagrams is an art, which can be acquired through practice. However, observance of some general guidelines can help in making them more attractive and effective. The diagrammatic presentation of statistical facts will be advantageous provided the following rules are observed in drawing diagrams.

1. A diagram should be neatly drawn and attractive.
2. The measurements of geometrical figures used in diagram should be accurate and proportional.
3. The size of the diagrams should match the size of the paper.
4. Every diagram must have a suitable but short heading.
5. The scale should be mentioned in the diagram.
6. Diagrams should be neatly as well as accurately drawn with the help of drawing instruments.
7. Index must be given for identification so that the reader can easily make out the meaning of the diagram.
8. Footnote must be given at the bottom of the diagram.
9. Economy in cost and energy should be exercised in drawing diagram.

5.5 Types of diagrams:

In practice, a very large variety of diagrams are in use and new ones are constantly being added. For the sake of convenience and simplicity, they may be divided under the following heads:

1. One-dimensional diagrams
2. Two-dimensional diagrams
3. Three-dimensional diagrams
4. Pictograms and Cartograms

5.5.1 One-dimensional diagrams:

In such diagrams, only one-dimensional measurement, i.e height is used and the width is not considered. These diagrams are in the form of bar or line charts and can be classified as

1. Line Diagram
2. Simple Diagram
3. Multiple Bar Diagram
4. Sub-divided Bar Diagram
5. Percentage Bar Diagram

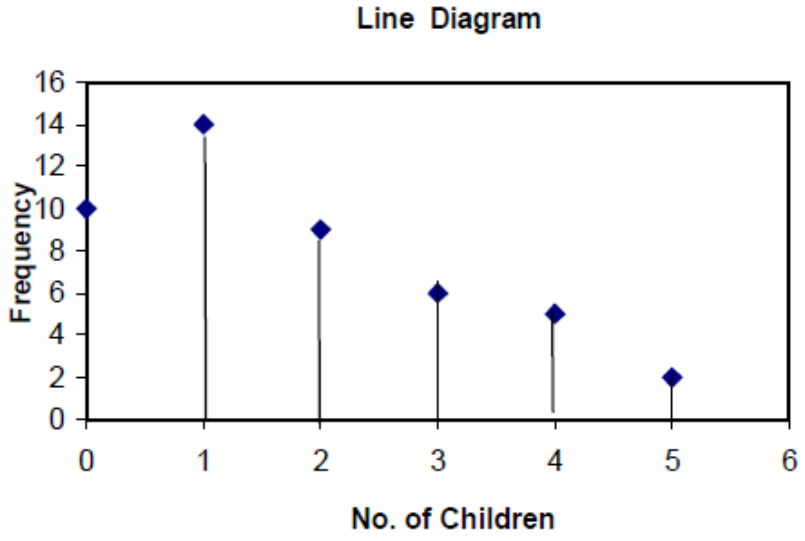
Line Diagram:

Line diagram is used in case where there are many items to be shown and there is not much of difference in their values. Such diagram is prepared by drawing a vertical line for each item according to the scale. The distance between lines is kept uniform. Line diagram makes comparison easy, but it is less attractive.

Example 1:

Show the following data by a line chart:

No. of children	0	1	2	3	4	5
Frequency	10	14	9	6	4	2

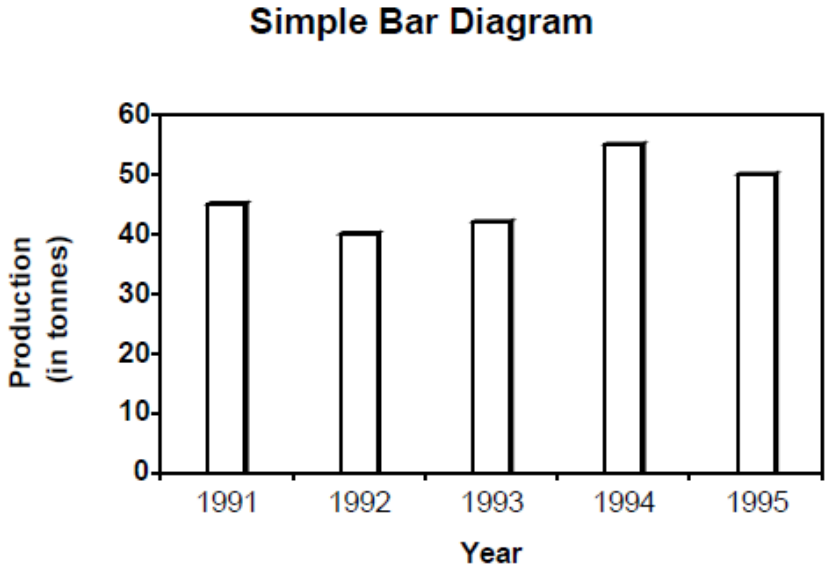


Simple Bar Diagram:

Example 2:
Represent the following data by a bar diagram.

Year	Production (in tones)
1991	45
1992	40
1993	42
1994	55
1995	50

Solution:



Multiple Bar Diagram:

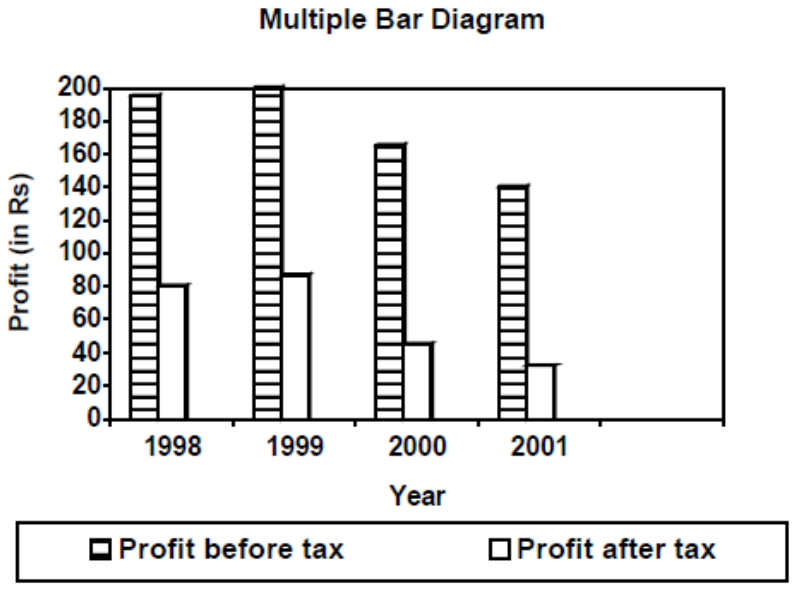
Multiple bar diagram is used for comparing two or more sets of statistical data. Bars are constructed side by side to represent the set of values for comparison. In order to distinguish bars, they may be either differently coloured or there should be different types of crossings or dotting, etc. An index is also prepared to identify the meaning of different colours or dottings.

Example 3:

Draw a multiple bar diagram for the following data.

Year	Profit before tax (in lakhs of rupees)	Profit after tax (in lakhs of rupees)
1998	195	80
1999	200	87
2000	165	45
2001	140	32

Solution:



Sub-divided Bar Diagram:

In a sub-divided bar diagram, the bar is sub-divided into various parts in proportion to the values given in the data and the whole bar represent the total. Such diagrams are also called Component Bar diagrams. The sub divisions are distinguished by different colours or crossings or dottings.

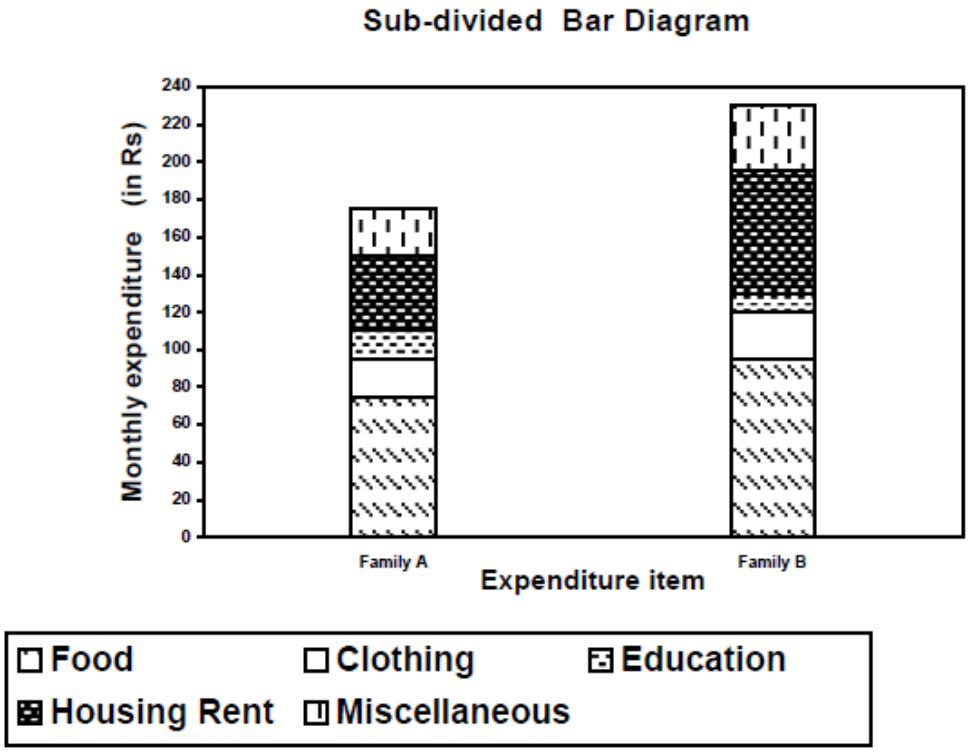
The main defect of such a diagram is that all the parts do not have a common base to enable one to compare accurately the various components of the data.

Example 4:

Represent the following data by a sub-divided bar diagram.

Expenditure items	Monthly expenditure (in Rs.)	
	Family A	Family B
Food	75	95
Clothing	20	25
Education	15	10
Housing Rent	40	65
Miscellaneous	25	35

Solution:



5.5.2 Two-dimensional Diagrams:

In one-dimensional diagrams, only length 9 is taken into account. But in two-dimensional diagrams the area represent the data and so the length and breadth have both to be taken into account. Such diagrams are also called area diagrams or surface diagrams. The important types of area diagrams are:

1. Rectangles 2. Squares 3. Pie-diagrams

Rectangles:

Rectangles are used to represent the relative magnitude of two or more values. The area of the rectangles are kept in proportion to the values. Rectangles are placed side by side for comparison. When two sets of figures are to be represented by rectangles, either of the two methods may be adopted.

We may represent the figures as they are given or may convert them to percentages and then subdivide the length into various components. Thus the percentage sub-divided rectangular diagram is more popular than sub-divided rectangular since it enables comparison to be made on a percentage basis.

Example 6:
Represent the following data by sub-divided percentage rectangular diagram.

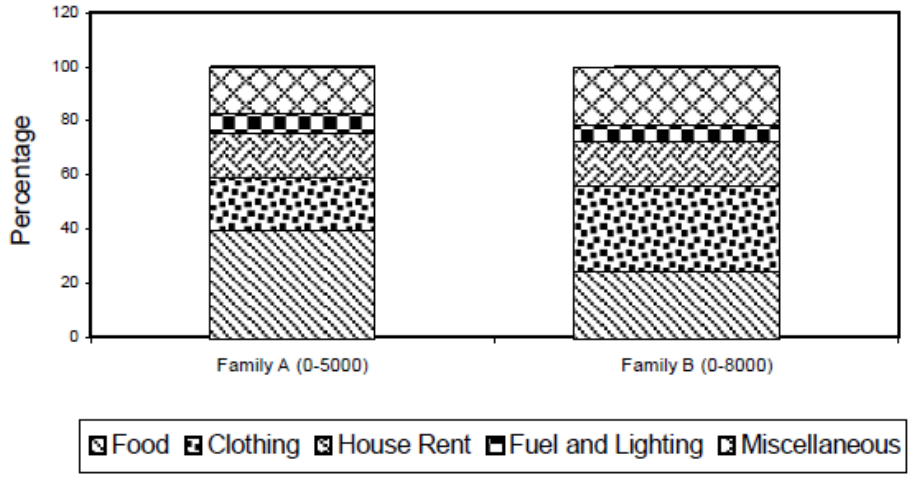
Items of Expenditure	Family A (Income Rs.5000)	Family B (income Rs.8000)
Food	2000	2500
Clothing	1000	2000
House Rent	800	1000
Fuel and lighting	400	500
Miscellaneous	800	2000
Total	5000	8000

Solution:

The items of expenditure will be converted into percentage as shown below:

Items of Expenditure	Family A		Family B	
	Rs.	Y	Rs.	Y
Food	2000	40	2500	31
Clothing	1000	20	2000	25
House Rent	800	16	1000	13
Fuel and Lighting	400	8	500	6
Miscellaneous	800	16	2000	25
Total	5000	100	8000	100

SUBDIVIDED PERCENTAGE RECTANGULAR DIAGRAM



Pie Diagram or Circular Diagram:

Another way of preparing a two-dimensional diagram is in the form of circles. In such diagrams, both the total and the component parts or sectors can be shown. The area of a circle is proportional to the square of its radius.

While making comparisons, pie diagrams should be used on a percentage basis and not on an absolute basis. In constructing a pie diagram the first step is to prepare the data so that various components values can be transposed into corresponding degrees on the circle.

The second step is to draw a circle of appropriate size with a compass. The size of the radius depends upon the available space and other factors of presentation. The third step is to measure points on the circle and representing the size of each sector with the help of a protractor.

Example 8:

Draw a Pie diagram for the following data of production of sugar in quintals of various countries.

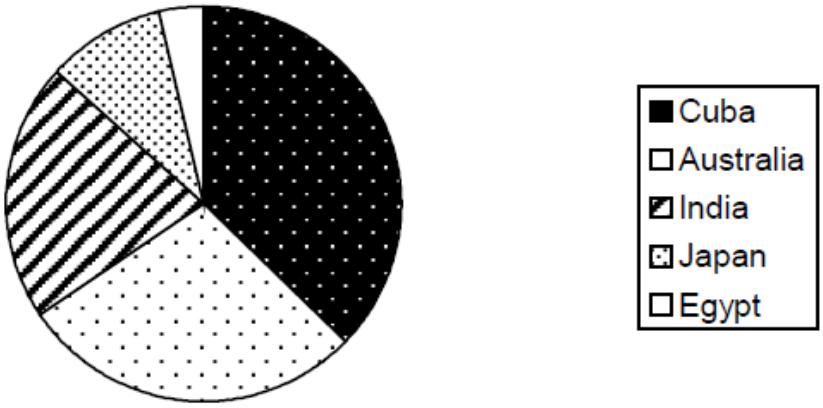
Country	Production of Sugar (in quintals)
Cuba	62
Australia	47
India	35
Japan	16
Egypt	6

Solution:

The values are expressed in terms of degree as follows.

Country	Production of Sugar	
	In Quintals	In Degrees
Cuba	62	134
Australia	47	102
India	35	76
Japan	16	35
Egypt	6	13
Total	166	360

Pie Diagram



5.5.3 Three-dimensional diagrams:

Three-dimensional diagrams, also known as volume diagram, consist of cubes, cylinders, spheres, etc. In such diagrams three things, namely length, width and height have to be taken into account. Of all the figures, making of cubes is easy. Side of a cube is drawn in proportion to the cube root of the magnitude of data.

Cubes of figures can be ascertained with the help of logarithms. The logarithm of the figures can be divided by 3 and the antilog of that value will be the cube-root.

Example 9:

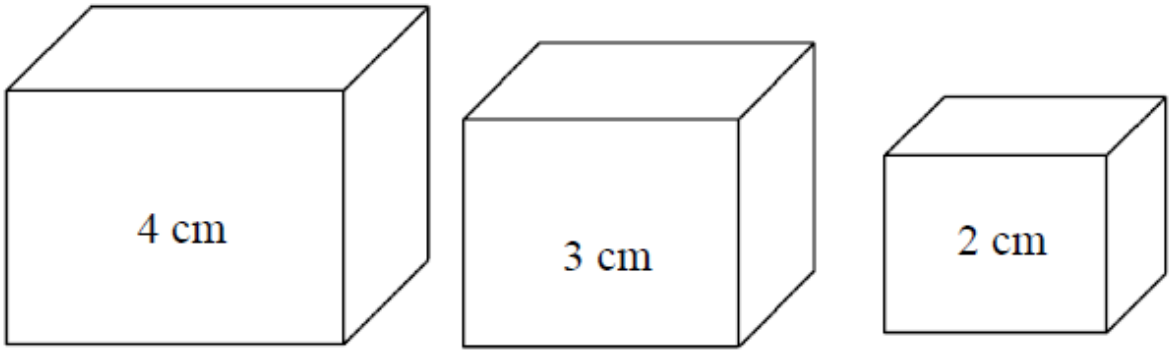
Represent the following data by volume diagram.

Category	Number of Students
Under graduate	64000
Post graduate	27000
Professionals	8000

Solution:

The sides of cubes can be determined as follows

Category	Number of students	Cube root	Side of cube
Undergraduate	64000	40	4 cm
Postgraduate	27000	30	3 cm
Professional	8000	20	2 cm



Undergraduate **Postgraduate** **professional**

5.5.4 Pictograms and Cartograms:

Pictograms are not abstract presentation such as lines or bars but really depict the kind of data we are dealing with. Pictures are attractive and easy to comprehend and as such this method is particularly useful in presenting statistics to the layman. When Pictograms are used, data are represented through a pictorial symbol that is carefully selected.

Cartograms or statistical maps are used to give quantitative information as a geographical basis. They are used to represent spatial distributions. The quantities on the map can be shown in many ways such as through shades or colours or dots or placing pictogram in each geographical unit.

5.6 Graphs:

A graph is a visual form of presentation of statistical data. A graph is more attractive than a table of figure. Even a common man can understand the message of data from the graph. Comparisons can be made between two or more phenomena very easily with the help of a graph.

However here we shall discuss only some important types of graphs which are more popular and they are

1.Histogram 2. Frequency Polygon
3.Frequency Curve 4. Ogive 5. Lorenz Curve

5.6.1 Histogram:

A histogram is a bar chart or graph showing the frequency of occurrence of each value of the variable being analysed. In histogram, data are plotted as a series of rectangles. Class intervals are shown on the 'X-axis' and the frequencies on the 'Y-axis'.

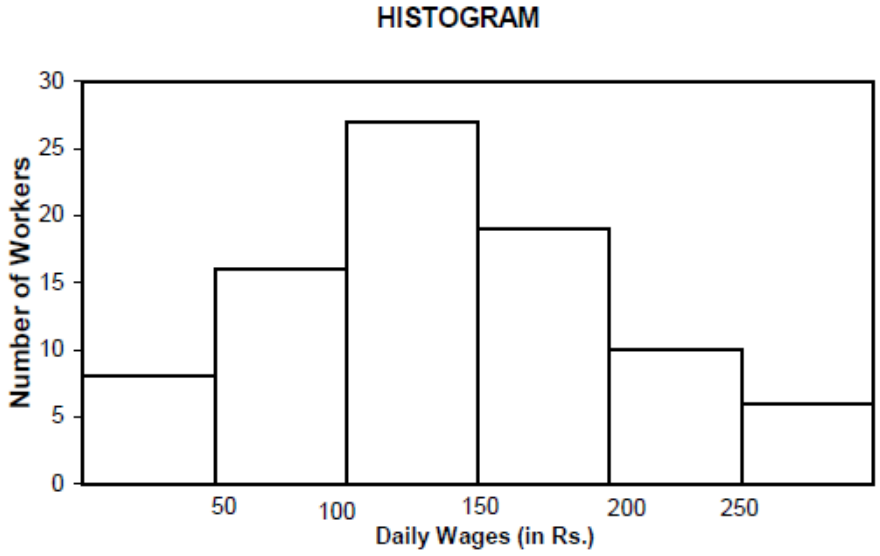
The height of each rectangle represents the frequency of the class interval. Each rectangle is formed with the other so as to give a continuous picture. Such a graph is also called staircase or block diagram.

Example 10:

Draw a histogram for the following data.

Daily Wages	Number of Workers
0-50	8
50-100	16
100-150	27
150-200	19
200-250	10
250-300	6

Solution:



Example 11:

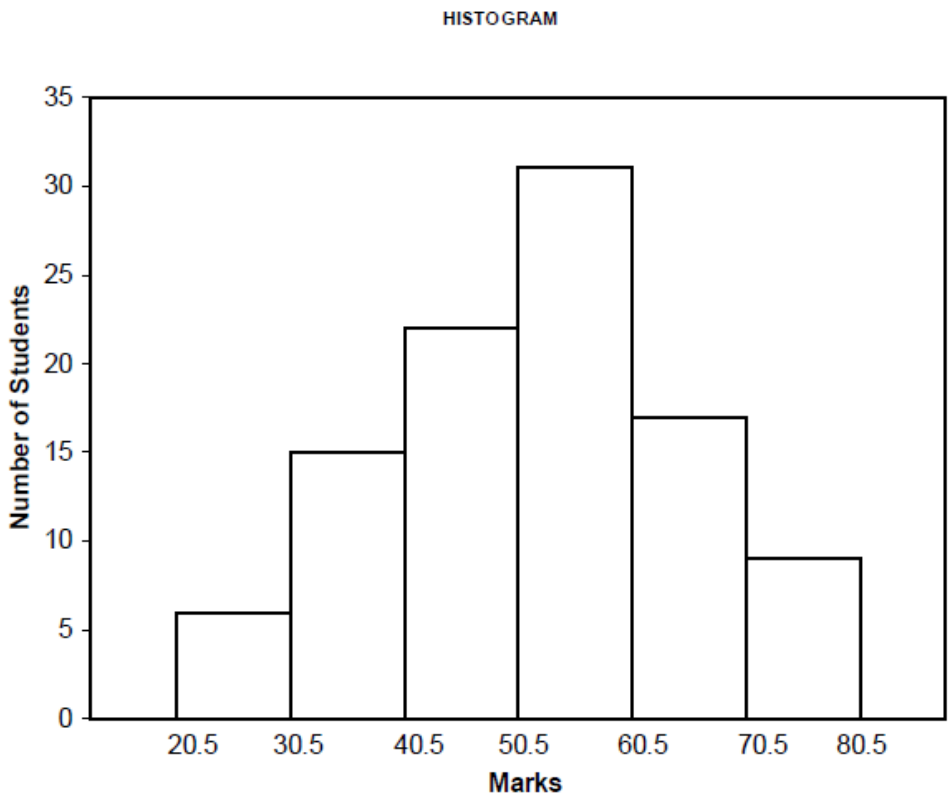
For the following data, draw a histogram.

Marks	Number of Students
21-30	6
31-40	15
41-50	22
51-60	31
61-70	17
71-80	9

Solution:

For drawing a histogram, the frequency distribution should be continuous. If it is not continuous, then first make it continuous as follows.

Marks	Number of Students
20.5-30.5	6
30.5-40.5	15
40.5-50.5	22
50.5-60.5	31
60.5-70.5	17
70.5-80.5	9



Example 12:

Draw a histogram for the following data.

Profits (in lakhs)	Number of Companies
0-10	4
10-20	12
20-30	24
30-50	32
50-80	18
80-90	9
90-100	3

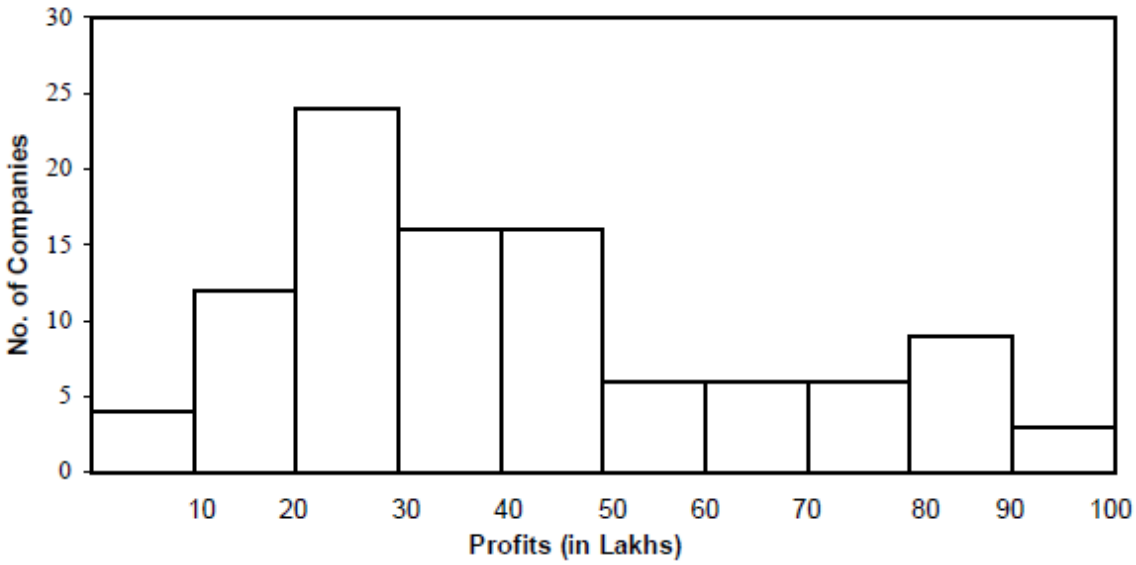
Solution:

When the class intervals are unequal, a correction for unequal class intervals must be made. The frequencies are adjusted as follows: The frequency of the class 30-50 shall be divided by two since the class interval is in double. Similarly the class interval 50-80 can be divided by 3. Then draw the histogram.

Now we rewrite the frequency table as follows.

Profits (in lakhs)	Number of Companies
0-10	4
10-20	12
20-30	24
30-40	16
40-50	16
50-60	6
60-70	6
70-80	6
80-90	9
90-100	3

HISTOGRAM

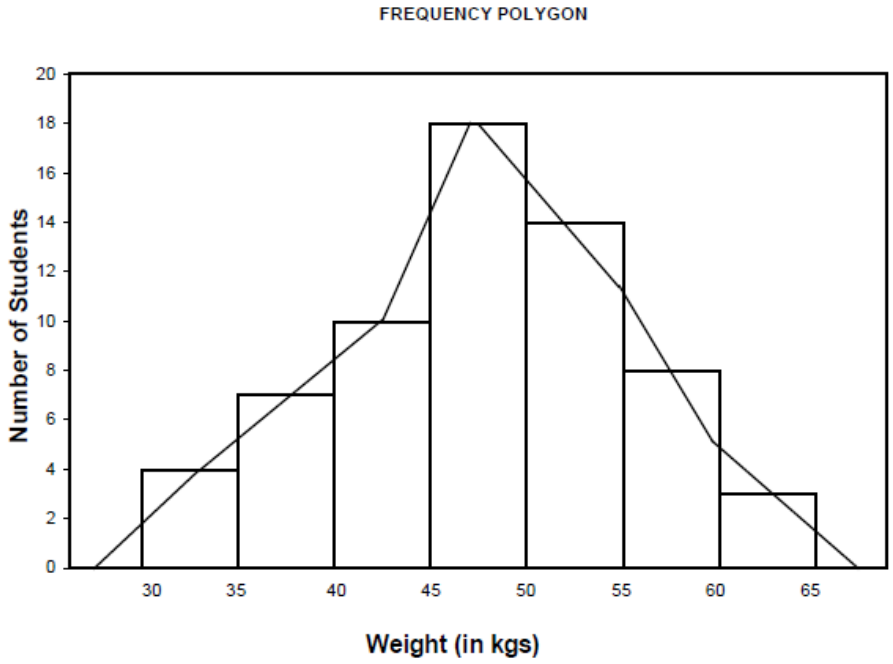


5.6.2 Frequency Polygon:

If we mark the midpoints of the top horizontal sides of the rectangles in a histogram and join them by a straight line, the figure so formed is called a Frequency Polygon. This is done under the assumption that the frequencies in a class interval are evenly distributed throughout the class. The area of the polygon is equal to the area of the histogram, because the area left outside is just equal to the area included in it.

Example 13:
Draw a frequency polygon for the following data.

Weight (in kg)	Number of Students
30-35	4
35-40	7
40-45	10
45-50	18
50-55	14
55-60	8
60-65	3



5.6.3 Frequency Curve:

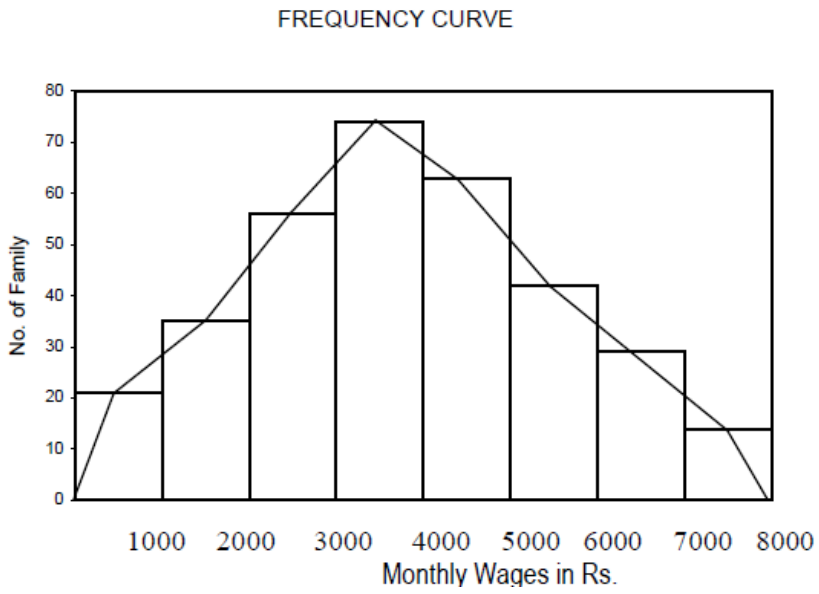
If the middle point of the upper boundaries of the rectangles of a histogram is corrected by a smooth freehand curve, then that diagram is called frequency curve. The curve should begin and end at the base line.

Example 14:

Draw a frequency curve for the following data.

Monthly Wages (in Rs.)	No. of family
0-1000	21
1000-2000	35
2000-3000	56
3000-4000	74
4000-5000	63
5000-6000	40
6000-7000	29
7000-8000	14

Solution:



Measures of Central Tendency:

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

Characteristics for a good or an ideal average :

The following properties should possess for an ideal average.

1. It should be rigidly defined.
2. It should be easy to understand and compute.
3. It should be based on all items in the data.
4. Its definition shall be in the form of a mathematical formula.
5. It should be capable of further algebraic treatment.
6. It should have sampling stability.
7. It should be capable of being used in further statistical computations or processing.

Arithmetic mean or mean :

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values x_1, x_2 .. x_n then the mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

Example 1 :

Calculate the mean for 2, 4, 6, 8, 10

Solution:

$$\bar{x} = \frac{2+4+6+8+10}{5}$$

$$= \frac{30}{5} = 6$$

Short-Cut method :

Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

Example 2 :

A student' s marks in 5 subjects are 75, 68, 80, 92, 56. Find his average mark.

Solution:

	X	d=x-A
	75	7
A	68	0
	80	12
	92	24
	56	-12
Total		31

$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 68 + \frac{31}{5}$$

$$= 68 + 6.2$$

$$= 74.2$$

Grouped Data :

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{\sum fx}{N}$$

where x = the mid-point of individual class

f = the frequency of individual class

N = the sum of the frequencies or total frequencies.

Short-cut method :

$$\bar{x} = A + \frac{\sum fd}{N} \times c$$

where $d = \frac{x - A}{c}$

A = any value in x

N = total frequency

c = width of the class interval

Example 3:

Given the following frequency distribution, calculate the arithmetic mean

Marks	: 64	63	62	61	60	59
Number of Students	: 8	18	12	9	7	6

Solution:

X	F	fx	d=x-A	fd
64	8	512	2	16
63	18	1134	1	18
62	12	744	0	0
61	9	549	–1	–9
60	7	420	–2	–14
59	6	354	–3	–18
	60	3713		- 7

Direct method

$$\bar{x} = \frac{\sum fx}{N} = \frac{3713}{60} = 61.88$$

Short-cut method

$$\bar{x} = A + \frac{\sum fd}{N} = 62 - \frac{7}{60} = 61.88$$

Example 4 *:*

Following is the distribution of persons according to different income groups. Calculate arithmetic mean.

Income Rs(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Solution:

Income C.I	Number of Persons (f)	Mid X	$d = \frac{x - A}{c}$	Fd
0-10	6	5	-3	-18
10-20	8	15	-2	-16
20-30	10	25	-1	-10
30-40	12	A 35	0	0
40-50	7	45	1	7
50-60	4	55	2	8
60-70	3	65	3	9
	50			-20

$$\text{Mean} = \bar{x} = A + \frac{\sum fd}{N}$$

$$= 35 - \frac{20}{50} \times 10$$

$$= 35 - 4$$

$$= 31$$

Merits and demerits of Arithmetic mean :

Merits:

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

Demerits:

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

Median :

The median is that value of the variate which divides the group into two equal parts, one part comprising all values greater, and the other, all values less than median.

Ungrouped or Raw data :

Arrange the given values in the increasing or decreasing order. If the number of values are odd, median is the middle value .If the number of values are even, median is the mean of middle two values.

By formula

$$\text{Median} = \text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

Example 11:

When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

Solution:

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the 5^{th} item i.e., 25 is the median

Using formula

$$\text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{9+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{10}{2}\right)^{\text{th}} \text{item}$$

$$= 5^{\text{th}} \text{ item}$$

$$= 25$$

Example 12 :

When even number of values are given. Find median for the following data

5, 8, 12, 30, 18, 10, 2, 22

Solution:

Arranging the data in the increasing order 2, 5, 8, 10, 12, 18, 22, 30

Here median is the mean of the middle two items (ie) mean of (10,12) ie

$$= \left(\frac{10+12}{2}\right) = 11$$

$$\therefore \text{median} = 11.$$

Using the formula

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{8+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{9}{2}\right)^{th} \text{item} = 4.5^{th} \text{item}$$

$$= 4^{th} \text{item} + \left(\frac{1}{2}\right)(5^{th} \text{item} - 4^{th} \text{item})$$

$$= 10 + \left(\frac{1}{2}\right)[12\text{-}10]$$

$$= 10 + \left(\frac{1}{2}\right) \times 2$$

$$= 10 + 1$$

$$= 11$$

Example 13:

The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and Accountancy.

Serial No	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	53	55	52	32	30	60	47	46	35	28
Marks (Accountancy)	57	45	24	31	25	84	43	80	32	72

Indicate in which subject is the level of knowledge higher ?

Solution:

For such question, median is the most suitable measure of central tendency. The mark in the two subjects are first arranged in increasing order as follows:

Serial No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	28	30	32	35	46	47	52	53	55	60
Marks in Accountancy	24	25	31	32	43	45	57	72	80	84

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{item} = \left(\frac{10+1}{2}\right)^{th} \text{item} = 5.5^{th} \text{item}$$

$$= \frac{Value\ of\ 5^{th}\ item + value\ of\ 6^{th}\ item}{2}$$

$$\text{Md (Statistics)} = \frac{46+47}{2} = 46.5$$

$$\text{Md (Accountancy)} = \frac{43+45}{2} = 44$$

There fore the level of knowledge in Statistics is higher than that in Accountancy.

Grouped Data:

In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution , cumulative frequencies have to be calculated to know the total number of items.

Cumulative frequency : (cf)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Discrete Series:

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N+1}{2}\right)$

Step3: See in the cumulative frequencies the value just greater than $\left(\frac{N+1}{2}\right)$

Step4: Then the corresponding value of x is median.

Example 14:

The following data pertaining to the number of members in a family. Find median size of the family.

Number of members **x**	1	2	3	4	5	6	7	8	9	10	11	12
Frequency F	1	3	5	6	10	13	9	5	3	2	2	1

Solution:

X	f	cf
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	60	

Median = size of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item

$= \text{size of}\left(\frac{60+1}{2}\right)^{\text{th}} \text{item}$

$= 30.5^{\text{th}}$ item

The cumulative frequencies just greater than 30.5 is 38.and the value of x corresponding to 38 is 6.Hence the median size is 6 members per family.

Continuous Series:

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N}{2}\right)$

Step3: See in the cumulative frequency the value first greater than $\left(\frac{N}{2}\right)$, Then the corresponding class interval is called the Median class. Then apply the formula

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where l = Lower limit of the median class

m = cumulative frequency preceding the median

c = width of the median class

f =frequency in the median class.

N=Total frequency.

Example 15:

The following table gives the frequency distribution of 325 workers of a factory, according to their average monthly income in a certain year.

Income group (in Rs)	Number of workers
Below 100	1
100-150	20
150-200	42
200-250	55
250-300	62
300-350	45
350-400	30
400-450	25
450-500	15
500-550	18
550-600	10
600 and above	2
	325

Calculate median income

Solution:

Income group (Class-interval)	Number of workers (Frequency)	Cumulative frequency c.f
Below 100	1	1
100-150	20	21
150-200	42	63
200-250	55	118
250-300	62	180
300-350	45	225
350-400	30	255
400-450	25	280
450-500	15	295
500-550	18	313
550-600	10	323
600 and above	2	325
	325	

$$\frac{N}{2}=\frac{325}{2}=162.5$$

Here $l = 250$, N = 325, f = 62, c = 50, m = 118

$$\text{Md} = 250+\left(\frac{162.5-118}{62}\right)\times 50$$

$$= 250+35.89$$

$$= 285.89$$

Example 16:

Following are the daily wages of workers in a textile. Find the median.

Wages (in Rs.)	Number of workers
less than 100	5
less than 200	12
less than 300	20
less than 400	32
less than 500	40
less than 600	45
less than 700	52
less than 800	60
less than 900	68
less than 1000	75

Solution :

We are given upper limit and less than cumulative frequencies. First find the class-intervals and the frequencies. Since the values are increasing by 100, hence the width of the class interval equal to 100.

Class interval	f	c.f
0-100	5	5
100-200	7	12
200-300	8	20
300- 400	12	32
400-500	8	40
500-600	5	45
600-700	7	52
700-800	8	60
800-900	8	68
900-1000	7	75
	75	

$$\left(\frac{N}{2}\right) = \left(\frac{75}{2}\right) = 37.5$$

$$\text{Md} = l + \left(\frac{\frac{N}{2} - m}{f}\right) \times \text{c}$$

$$= 400 + \left(\frac{37.5 - 32}{8}\right) \times 100 = 400 + 68.75 = 468.75$$

Merits of Median :

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open-end intervals.
3. Median can be located even if the data are incomplete.
4. Median can be located even for qualitative factors such as ability, honesty etc.

Demerits of Median :

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in mean deviation.
4. It is not taken into account all the observations.

Mode :

The mode refers to that value in a distribution, which occur most frequently. It is an actual value, which has the highest concentration of items in and around it.

Computation of the mode:

Ungrouped or Raw Data:

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

Example 29:

2 , 7, 10, 15, 10, 17, 8, 10, 2

$\therefore$ Mode = $M_0 = 10$

In some cases the mode may be absent while in some cases there may be more than one mode.

Example 30:

1. 12, 10, 15, 24, 30 (no mode)
2. 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

$\therefore$ the modes are 7 and 10

Grouped Data:

For Discrete distribution, see the highest frequency and corresponding value of X is mode.

Continuous distribution :

See the highest frequency then the corresponding value of class interval is called the modal class. Then apply the formula.

$$\text{Mode} = M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

l = Lower limit of the model class

$$\triangle_1 = f_1 - f_0$$
$$\triangle_2 = f_1 - f_2$$

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

The above formula can also be written as

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Example 31:

Calculate mode for the following :

C- I	f
0-50	5
50-100	14
100-150	40
150-200	91
200-250	150
250-300	87
300-350	60
350-400	38
400 and above	15

Solution:

The highest frequency is 150 and corresponding class interval is 200 – 250, which is the modal class.

Here l=200,f_1=150,f_0=91, f_2=87, C=50

$$\text{Mode} = M_0 = 1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

$$= 200 + \frac{150\text{-}91}{2\times150-91-87}\times 50$$

$$= 200 + \frac{2950}{122}$$

$$= 200 + 24.18 = 224.18$$

Determination of Modal class :

For a frequency distribution modal class corresponds to the maximum frequency. But in any one (or more) of the following cases

i. If the maximum frequency is repeated
ii. If the maximum frequency occurs in the beginning or at the end of the distribution
iii. If there are irregularities in the distribution, the modal class is determined by the method of grouping.

Steps for Calculation :

We prepare a grouping table with 6 columns

1. In column I, we write down the given frequencies.
2. Column II is obtained by combining the frequencies two by two.
3. Leave the 1^{st} frequency and combine the remaining frequencies two by two and write in column III
4. Column IV is obtained by combining the frequencies three by three.
5. Leave the 1st frequency and combine the remaining frequencies three by three and write in column V
6. Leave the 1st and 2^{nd} frequencies and combine the remaining frequencies three by three and write in column VI

Mark the highest frequency in each column. Then form an analysis table to find the modal class. After finding the modal class use the formula to calculate the modal value.

Example 32:

Calculate mode for the following frequency distribution.

Class interval	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Grouping Table

C I	f	2	3	4	5	6
0- 5	9					
5-10	12	21		36		
10-15	15		27		**43**	
15-20	16	31				**48**
20-25	**17**		**33**	**48**		
25-30	15	**32**			42	38
30-35	10		25			
35-40	13	23				

Analysis Table

Columns	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
1					1			
2					1	1		
3				1	1			
4				1	1	1		
5		1	1	1				
6			1	1	1			
Total		1	2	4	5	2		

The maximum occurred corresponding to 20-25, and hence it is the modal class.

$$\text{Mode} = \text{Mo} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

Here $l = 20$; $\Delta_1 = f_1 - f_0 = 17 - 16 = 1$

$\Delta_2 = f_1 - f_2 = 17 - 15 = 2$

$$\therefore M_0 = 20 + \frac{1}{1+2} \times 5$$

$$= 20 + 1.67 = 21.67$$

MEASURES OF DISPERSION

Characteristics of a good measure of dispersion:

An ideal measure of dispersion is expected to possess the following properties

1.It should be rigidly defined

2. It should be based on all the items.

3. It should not be unduly affected by extreme items.

4. It should lend itself for algebraic manipulation.

5. It should be simple to understand and easy to calculate

Absolute and Relative Measures :

There are two kinds of measures of dispersion, namely

1.Absolute measure of dispersion

2.Relative measure of dispersion.

The various absolute and relative measures of dispersion are listed below.

Absolute measure	**Relative measure**
1. Range	1.Co-efficient of Range
2.Quartile deviation	2.Co-efficient of Quartile deviation
3.Mean deviation	3. Co-efficient of Mean deviation
4.Standard deviation	4.Co-efficient of variation

7.3 Range and coefficient of Range:

7.3.1 Range:

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

In symbols, Range $= L - S$.

Where L = Largest value.

S = Smallest value.

In individual observations and discrete series, L and S are easily identified. In continuous series, the following two methods are followed.

Method 1:

L = Upper boundary of the highest class

S = Lower boundary of the lowest class.

Method 2:

L = Mid value of the highest class.

S = Mid value of the lowest class.

7.3.2 Co-efficient of Range :

$$\text{Co-efficient of Range} = \frac{L - S}{L + S}$$

Example1:

Find the value of range and its co-efficient for the following data.

7, 9, 6, 8, 11, 10, 4

Solution:

L=11, S = 4.

Range $= L - S = 11 - 4 = 7$

$$\text{Co-efficient of Range} = \frac{L - S}{L + S}$$

$$= \frac{11 - 4}{11 + 4}$$

$$= \frac{7}{15} = 0.4667$$

Example 2:

Calculate range and its co efficient from the following distribution.

Size:	60-63	63-66	66-69	69-72	72-75
Number:	5	18	42	27	8

Solution:

L = Upper boundary of the highest class.

= 75

S = Lower boundary of the lowest class.
= 60

Range = L – S = 75 – 60 = 15

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$

$$= \frac{75-60}{75+60}$$

$$= \frac{15}{135} = 0.1111$$

7.3.3 Merits and Demerits of Range :

Merits:

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, et c., range is most widely used.

Demerits:

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure.

7.6 Standard Deviation and Coefficient of variation:

7.6.1 Standard Deviation :

Karl Pearson introduced the concept of standard deviation in 1893. It is the most important measure of dispersion and is widely used in many statistical formulae. Standard deviation is also called Root-Mean Square Deviation. The reason is that it is the square–root of the mean of the squared deviation from the arithmetic mean. It provides accurate result. Square of standard deviation is called Variance.

Definition:

It is defined as the positive square-root of the arithmetic mean of the Square of the deviations of the given observation from their arithmetic mean.

The standard deviation is denoted by the Greek letter σ (sigma)

7.6.2 Calculation of Standard deviation-Individual Series :

There are two methods of calculating Standard deviation in an individual series.

a) Deviations taken from Actual mean
b) Deviation taken from Assumed mean

a) Deviation taken from Actual mean:

This method is adopted when the mean is a whole number.

Steps:

1. Find out the actual mean of the series ($\bar{x}$)
2. Find out the deviation of each value from the mean ($x = X - \overline{X}$)
3. Square the deviations and take the total of squared deviations $\sum x^2$

4. Divide the total (Σx^2) by the number of observation $\left(\frac{\Sigma x^2}{n}\right)$

The square root of $\left(\frac{\Sigma x^2}{n}\right)$ is standard deviation.

Thus $\sigma = \sqrt{\left(\frac{\Sigma x^2}{n}\right)}$ or $\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

b) Deviations taken from assumed mean:

This method is adopted when the arithmetic mean is fractional value.

Taking deviations from fractional value would be a very difficult and tedious task. To save time and labour, We apply short –cut method; deviations are taken from an assumed mean. The formula is:

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

Where d-stands for the deviation from assumed mean = (X-A)

Steps:

1. Assume any one of the item in the series as an average (A)
2. Find out the deviations from the assumed mean; i.e., X-A denoted by d and also the total of the deviations Σd
3. Square the deviations; i.e., d^2 and add up the squares of deviations, i.e, Σd^2
4. Then substitute the values in the following formula:

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

Note: We can also use the simplified formula for standard deviation.

$$\sigma = \frac{1}{n}\sqrt{n\sum d^2 - (\sum d)^2}$$

For the frequency distribution

$$\sigma = \frac{c}{N}\sqrt{N\sum fd^2 - (\sum fd)^2}$$

Example 9:

Calculate the standard deviation from the following data.

14, 22, 9, 15, 20, 17, 12, 11

Solution:

Deviations from actual mean.

Values (X)	$X - \overline{X}$	$(X - \overline{X})^2$
14	-1	1
22	7	49
9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
120		140

$$\overline{X} = \frac{120}{8} = 15$$

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{140}{8}}$$

$$= \sqrt{17.5} = 4.18$$

Example 10:

The table below gives the marks obtained by 10 students in statistics. Calculate standard deviation.

Student Nos :	1	2	3	4	5	6	7	8	9	10
Marks :	43	48	65	57	31	60	37	48	78	59

Solution: (Deviations from assumed mean)

Nos.	Marks (x)	d=X-A (A=57)	d^2
1	43	-14	196
2	48	-9	81
3	65	8	64
4	57	0	0
5	31	-26	676
6	60	3	9
7	37	-20	400
8	48	-9	81
9	78	21	441
10	59	2	4
n = 10		Σd=-44	Σd^2 =1952

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

$$= \sqrt{\frac{1952}{10} - \left(\frac{-44}{10}\right)^2}$$

$$= \sqrt{195.2 - 19.36}$$

$$= \sqrt{175.84} = 13.26$$

7.6.3 Calculation of standard deviation:

Discrete Series:

There are three methods for calculating standard deviation in discrete series:

(a) Actual mean methods

(b) Assumed mean method

(c) Step-deviation method.

(a) Actual mean method:

Steps:

1. Calculate the mean of the series.
2. Find deviations for various items from the means i.e.,
   $x - \bar{x} = d$.
3. Square the deviations ($= d^2$) and multiply by the respective frequencies(f) we get fd^2
4. Total to product ($\sum fd^2$) Then apply the formula:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}}$$

(b) Assumed mean method:

Here deviation are taken not from an actual mean but from an assumed mean. Also this method is used, if the given variable values are not in equal intervals.

Steps:

1. Assume any one of the items in the series as an assumed mean and denoted by A.
2. Find out the deviations from assumed mean, i.e, X-A and denote it by d.
3. Multiply these deviations by the respective frequencies and get the $\sum fd$
4. Square the deviations (d^2).
5. Multiply the squared deviations ($d^{2)}$ by the respective frequencies (f) and get $\sum fd^2$.
6. Substitute the values in the following formula:

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

Where $d = X - A$, $N = \sum f$.

Example 11:

Calculate Standard deviation from the following data.

X :	20	22	25	31	35	40	42	45
f :	5	12	15	20	25	14	10	6

Solution:

Deviations from assumed mean

x	f	d = x –A (A = 31)	d^2	fd	fd^2
20	5	-11	121	-55	605
22	12	-9	81	-108	972
25	15	-6	36	-90	540
31	20	0	0	0	0
35	25	4	16	100	400
40	14	9	81	126	1134
42	10	11	121	110	1210
45	6	14	196	84	1176
	N=107			$\Sigma fd=167$	$\Sigma fd^2 =6037$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{6037}{107} - \left(\frac{167}{107}\right)^2}$$

$$= \sqrt{56.42 - 2.44}$$

$$= \sqrt{53.98} = 7.35$$

7.6.4 Calculation of Standard Deviation –Continuous series:

In the continuous series the method of calculating standard deviation is almost the same as in a discrete series. But in a continuous series, mid-values of the class intervals are to be found out. The step- deviation method is widely used.

The formula is,

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$$

$$d = \frac{m - A}{C}$$, C- Class interval.

Steps:

1. Find out the mid-value of each class.
2. Assume the center value as an assumed mean and denote it by A
3. Find out $d = \frac{m - A}{C}$
4. Multiply the deviations d by the respective frequencies and get Σfd
5. Square the deviations and get d^2
6. Multiply the squared deviations (d^2) by the respective frequencies and get Σfd^2
7. Substituting the values in the following formula to get the standard deviation

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

Example 13:

The daily temperature recorded in a city in Russia in a year is given below.

Temperature C^0	No. of days
-40 to –30	10
-30 to –20	18
-20 to –10	30
-10 to 0	42
0 to 10	65
10 to 20	180
20 to 30	20
	365

Calculate Standard Deviation.

Solution:

Temperature	Mid value (m)	No. of days f	$d = \frac{m - (-5^n)}{10^n}$	fd	fd^2
-40 to -30	-35	10	-3	-30	90
-30 to -20	-25	18	-2	-36	72
-20 to -10	-15	30	-1	-30	30
-10 to - 0	-5	42	0	0	0
0 to 10	5	65	1	65	65
10 to 20	15	180	2	360	720
20 to 30	25	20	3	60	180
		N=365		Σfd = 389	Σfd^2 =1157

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$$

$$= \sqrt{\frac{1157}{365} - \left(\frac{389}{365}\right)^2} \times 10$$

$$= \sqrt{3.1699 - 1.1358} \times 10$$

$$= \sqrt{2.0341} \times 10$$

$$= 1.4262 \times 10$$

$$= 14.26^\circ c$$

7.6.6 Merits and Demerits of Standard Deviation:

Merits:

1. It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.
2. As it is based on arithmetic mean, it has all the merits of arithmetic mean.
3. It is the most important and widely used measure of dispersion.
4. It is possible for further algebraic treatment.
5. It is less affected by the fluctuations of sampling and hence stable.
6. It is the basis for measuring the coefficient of correlation and sampling.

Demerits:

1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

7.6.7 Coefficient of Variation :

The Standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of students cannot be compared with the standard deviation of weights of students, as both are expressed in different units, i.e heights in centimeter and weights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation.

The coefficient of variation is obtained by dividing the standard deviation by the mean and multiply it by 100. symbolically,

$$\text{Coefficient of variation (C.V)} = \frac{\sigma}{\bar{X}} \times 100$$

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the C.V. is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogeneous.

Example 15:

In two factories A and B located in the same industrial area, the average weekly wages (in rupees) and the standard deviations are as follows:

Factory	Average	Standard Deviation	No. of workers
A	34.5	5	476
B	28.5	4.5	524

1. Which factory A or B pays out a larger amount as weekly wages?
2. Which factory A or B has greater variability in individual wages?

Solution:

Given $N_1 = 476$, $\overline{X}_1 = 34.5$, $\sigma_1 = 5$

$N_2 = 524$, $\overline{X}_2 = 28.5$, $\sigma_2 = 4.5$

1. Total wages paid by factory A

$= 34.5 \times 476$

$=$ Rs.16.422

Total wages paid by factory B

$= 28.5 \times 524$

$=$ Rs.14,934.

Therefore factory A pays out larger amount as weekly wages.

2. C.V. of distribution of weekly wages of factory A and B are

$$C.V.(A) = \frac{\sigma_1}{\bar{X}_1} \times 100$$

$$= \frac{5}{34.5} \times 100$$

$$= 14.49$$

$$C.V\ (B) = \frac{\sigma_2}{\bar{X}_2} \times 100$$

$$= \frac{4.5}{28.5} \times 100$$

$$= 15.79$$

Factory B has greater variability in individual wages, since C.V. of factory B is greater than C.V of factory A

Example 16:

Prices of a particular commodity in five years in two cities are given below:

Price in city A	Price in city B
20	10
22	20
19	18
23	12
16	15

Which city has more stable prices?

Solution:
Actual mean method

City A			City B		
Prices (X)	Deviations from $\bar{X}$=20 dx	dx^2	Prices (Y)	Deviations from $\bar{Y}$ =15 dy	dy^2
20	0	0	10	-5	25
22	2	4	20	5	25
19	-1	1	18	3	9
23	3	9	12	-3	9
16	-4	16	15	0	0
Σx=100	Σdx=0	Σdx^2=30	Σy=75	Σdy=0	Σdy^2 =68

City A: $\bar{X} = \frac{\Sigma x}{n} = \frac{100}{5} = 20$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma dx^2}{n}}$$

$$= \sqrt{\frac{30}{5}} = \sqrt{6} = 2.45$$

$$C.V(x) = \frac{\sigma_x}{\bar{x}} \times 100$$

$$= \frac{2.45}{20} \times 100$$

$$= 12.25\ \%$$

City B: $\bar{Y} = \frac{\Sigma y}{n} = \frac{75}{5} = 15$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\frac{\Sigma dy^2}{n}}$$

$$= \sqrt{\frac{68}{5}} = \sqrt{13.6} = 3.69$$

$$C.V.(y) = \frac{\sigma_y}{\bar{y}} \text{ x } 100$$

$$= \frac{3.69}{15} \times 100$$

$$= 24.6\ \%$$

City A had more stable prices than City B, because the coefficient of variation is less in City A.

UNIT 5

CORRELATION

Introduction:

The term correlation is used by a common man without knowing that he is making use of the term correlation. For example when parents advice their children to work hard so that they may get good marks, they are correlating good marks with hard work.

The study related to the characteristics of only variable such as height, weight, ages, marks, wages, etc., is known as univariate analysis. The statistical Analysis related to the study of the relationship between two variables is known as Bi-Variate Analysis. Some times the variables may be inter-related. In health sciences we study the relationship between blood pressure and age, consumption level of some nutrient and weight gain, total income and medical expenditure, etc., The nature and strength of relationship may be examined by correlation and Regression analysis.

Thus Correlation refers to the relationship of two variables or more. (e-g) relation between height of father and son, yield and rainfall, wage and price index, share and debentures etc.

Correlation is statistical Analysis which measures and analyses the degree or extent to which the two variables fluctuate with reference to each other. The word relationship is important. It indicates that there is some connection between the variables. It measures the closeness of the relationship. Correlation does not indicate cause and effect relationship. Price and supply, income and expenditure are correlated.

Definitions:

1. Correlation Analysis attempts to determine the degree of relationship between variables- Ya-Kun-Chou.
2. Correlation is an analysis of the covariation between two or more variables.- A.M.Tuttle.

Correlation expresses the inter-dependence of two sets of variables upon each other. One variable may be called as (subject)

independent and the other relative variable (dependent). Relative variable is measured in terms of subject.

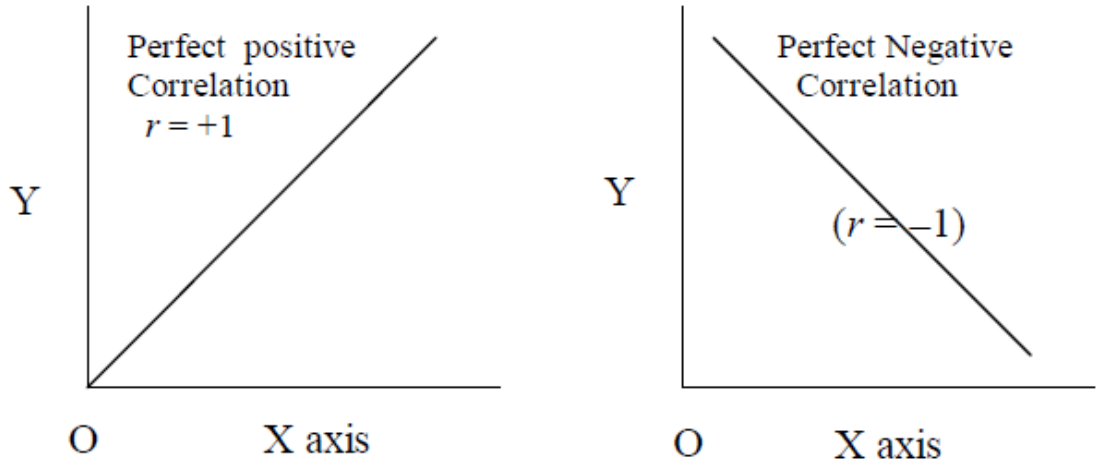
Uses of correlation:

1. It is used in physical and social sciences.
2. It is useful for economists to study the relationship between variables like price, quantity etc. Businessmen estimates costs, sales, price etc. using correlation.
3. It is helpful in measuring the degree of relationship between the variables like income and expenditure, price and supply, supply and demand etc.
4. Sampling error can be calculated.
5. It is the basis for the concept of regression.

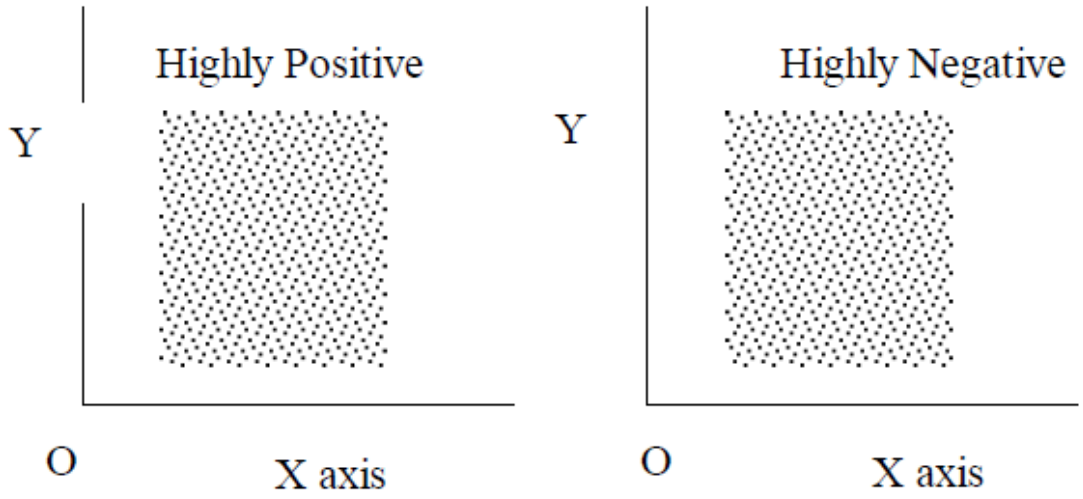
Scatter Diagram:

It is the simplest method of studying the relationship between two variables diagrammatically. One variable is represented along the horizontal axis and the second variable along the vertical axis. For each pair of observations of two variables, we put a dot in the plane. There are as many dots in the plane as the number of paired observations of two variables. The direction of dots shows the scatter or concentration of various points. This will show the type of correlation.

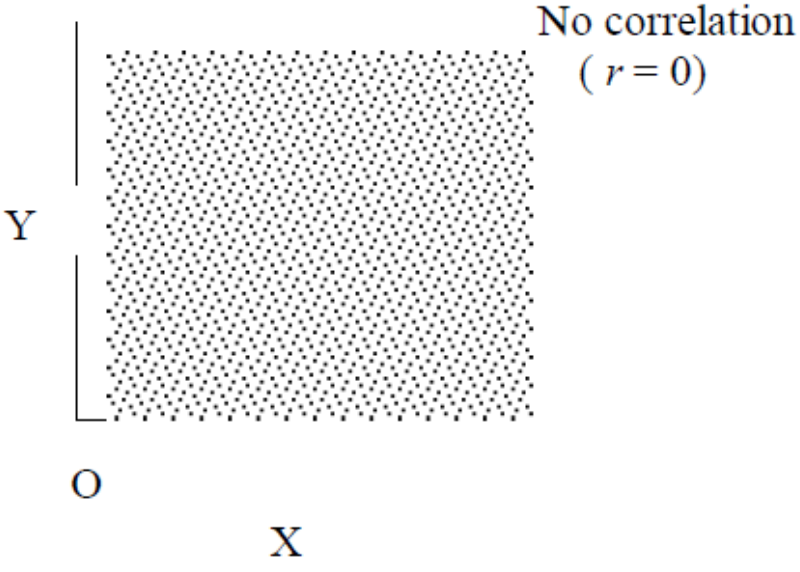
1. If all the plotted points form a straight line from lower left hand corner to the upper right hand corner then there is
   Perfect positive correlation. We denote this as $r = +1$



1. If all the plotted dots lie on a straight line falling from upper left hand corner to lower right hand corner, there is a perfect negative correlation between the two variables. In this case the coefficient of correlation takes the value $r = -1$.
2. If the plotted points in the plane form a band and they show a rising trend from the lower left hand corner to the upper right hand corner the two variables are highly positively correlated.



1. If the points fall in a narrow band from the upper left hand corner to the lower right hand corner, there will be a high degree of negative correlation.
2. If the plotted points in the plane are spread all over the diagram there is no correlation between the two variables.



Merits:

1. It is a simplest and attractive method of finding the nature of correlation between the two variables.
2. It is a non-mathematical method of studying correlation. It is easy to understand.
3. It is not affected by extreme items.
4. It is the first step in finding out the relation between the two variables.
5. We can have a rough idea at a glance whether it is a positive correlation or negative correlation.

Demerits:

By this method we cannot get the exact degree or correlation between the two variables.

Types of Correlation:

Correlation is classified into various types. The most important ones are

i) Positive and negative.
ii) Linear and non-linear.
iii) Partial and total.
iv) Simple and Multiple.

Positive and Negative Correlation:

It depends upon the direction of change of the variables. If the two variables tend to move together in the same direction (ie) an increase in the value of one variable is accompanied by an increase in the value of the other, (or) a decrease in the value of one variable is accompanied by a decrease in the value of other, then the correlation is called positive or direct correlation. Price and supply, height and weight, yield and rainfall, are some examples of positive correlation.

If the two variables tend to move together in opposite directions so that increase (or) decrease in the value of one variable is accompanied by a decrease or increase in the value of the other variable, then the correlation is called negative (or) inverse correlation. Price and demand, yield of crop and price, are examples of negative correlation.

Linear and Non-linear correlation:

If the ratio of change between the two variables is a constant then there will be linear correlation between them.

Consider the following.

X	2	4	6	8	10	12
Y	3	6	9	12	15	18

Here the ratio of change between the two variables is the same. If we plot these points on a graph we get a straight line.

If the amount of change in one variable does not bear a constant ratio of the amount of change in the other. Then the relation is called Curvi-linear (or) non-linear correlation. The graph will be a curve.

Simple and Multiple correlation:

When we study only two variables, the relationship is simple correlation. For example, quantity of money and price level, demand and price. But in a multiple correlation we study more than two variables simultaneously. The relationship of price, demand and supply of a commodity are an example for multiple correlation.

Partial and total correlation:

The study of two variables excluding some other variable is called **Partial correlation**. For example, we study price and demand eliminating supply side. In total correlation all facts are taken into account.

Computation of correlation:

When there exists some relationship between two variables, we have to measure the degree of relationship. This measure is called the measure of correlation (or) correlation coefficient and it is denoted by ' r' .

Co-variation:

The covariation between the variables x and y is defined as

Cov(x,y) = $\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$ where $\bar{x}, \bar{y}$ are respectively means of x and y and ' n' is the number of pairs of observations.

Karl pearson' s coefficient of correlation:

Karl pearson, a great biometrician and statistician, suggested a mathematical method for measuring the magnitude of linear relationship between the two variables. It is most widely used method in practice and it is known as pearsonian coefficient of correlation. It is denoted by 'r'. The formula for calculating 'r' is

(i) $r = \dfrac{Cov(x, y)}{\sigma_x . \sigma_y}$ where σ_x, σ_y are S.D of x and y respectively.

(ii) $r = \dfrac{\Sigma xy}{n\ \sigma_x\ \sigma_y}$

(iii) $r = \dfrac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$, $X = x - \bar{x}$, $Y = y - \bar{y}$

when the deviations are taken from the actual mean we can apply any one of these methods. Simple formula is the third one.

The third formula is easy to calculate, and it is not necessary to calculate the standard deviations of x and y series respectively.

Steps:

1. Find the mean of the two series x and y.
2. Take deviations of the two series from x and y.
   $X = x - \bar{x}$, $Y = y - \bar{y}$
3. Square the deviations and get the total, of the respective squares of deviations of x and y and denote by ΣX^2 , ΣY^2 respectively.
4. Multiply the deviations of x and y and get the total and Divide by n. This is covariance.
5. Substitute the values in the formula.

$$r = \frac{cov(x, y)}{\sigma x . \sigma y} = \frac{\Sigma (x - \bar{x})\ (y\ -\ \bar{y})/n}{\sqrt{\dfrac{\Sigma (x - \bar{x})^2}{n}} \sqrt{\dfrac{\Sigma (y - \bar{y})^2}{n}}}$$

The above formula is simplified as follows

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}, \quad X = x - \bar{x}\ , Y = y - \bar{y}$$

Example 1:

Find Karl Pearson's coefficient of correlation from the following data between height of father (x) and son (y).

X	64	65	66	67	68	69	70
Y	66	67	65	68	70	68	72

Comment on the result.

Solution:

x	Y	$X = x - \bar{x}$ $X = x - 67$	X^2	$Y = y - \bar{y}$ $Y = y - 68$	Y^2	XY
64	66	-3	9	-2	4	6
65	67	-2	4	-1	1	2
66	65	-1	1	-3	9	3
67	68	0	0	0	0	0
68	70	1	1	2	4	2
69	68	2	4	0	0	0
70	72	3	9	4	16	12
469	476	0	28	0	34	25

$$\bar{x} = \frac{469}{7} = 67 \ ; \ \bar{y} = \frac{476}{7} = 68$$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}} = \frac{25}{\sqrt{28 \times 34}} = \frac{25}{\sqrt{952}} = \frac{25}{30.85} = 0.81$$

Since r = + 0.81, the variables are highly positively correlated. (ie) Tall fathers have tall sons.

Working rule (i)

We can also find r with the following formula

We have $r = \frac{Cov(x, y)}{\sigma_x . \sigma_y}$

$$\text{Cov}(x,y) = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n} = \frac{\Sigma(xy + \bar{x}\bar{y} - \bar{y}x - \bar{x}\bar{y})}{n}$$

$$= \frac{\Sigma xy}{n} - \frac{\bar{y}\Sigma x}{n} - \frac{\bar{x}\Sigma y}{n} + \frac{\Sigma \bar{x}\bar{y}}{n}$$

$$\text{Cov}(x,y) = \frac{\Sigma xy}{n} - \cancel{\bar{y}\bar{x}} - \bar{x}\bar{y} + \cancel{\bar{x}\bar{y}} = \frac{\Sigma xy}{n} - \bar{x}\bar{y}$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 \quad , \quad \sigma_y^2 = \frac{\Sigma y^2}{n} - \bar{y}^2$$

$$\text{Now } r = \frac{Cov(x,y)}{\sigma_x . \sigma_y}$$

$$r = \frac{\frac{\Sigma xy}{n} - \bar{x}\bar{y}}{\sqrt{\left(\frac{\Sigma x^2}{n} - \bar{x}^2\right)} \sqrt{\left(\frac{\Sigma y^2}{n} - \bar{y}^2\right)}}$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

Note: In the above method we need not find mean or standard deviation of variables separately.

Example 2:

Calculate coefficient of correlation from the following data.

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

x	y	x^2	y^2	xy
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135
45	108	285	1356	597

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$r = \frac{9 \times 597 - 45 \times 108}{\sqrt{\left(9 \times 285 - (45)^2\right).(9 \times 1356 - (108)^2)}}$$

$$r = \frac{5373 - 4860}{\sqrt{(2565 - 2025).(12204 - 11664)}}$$

$$= \frac{513}{\sqrt{540 \times 540}} = \frac{513}{540} = 0.95$$

Working rule (ii) (shortcut method)

We have $r = \frac{Cov(x, y)}{\sigma_x . \sigma_y}$

where Cov(x,y) = $\frac{\Sigma(x - \bar{x})(y - \bar{y})}{n}$

Take the deviation from x as x – A and the deviation from y as y – B

$$\text{Cov(x,y)} = \frac{\Sigma\ [(x - A) - (\bar{x} - A)]\ [(y - B) - (\bar{y} - B)]}{n}$$

$$= \frac{1}{n}\ \Sigma\ [(x - A)\ (y - B) - (x - A)\ (\bar{y} - B) - (\bar{x} - A)(y - B) + (\bar{x} - A)(\bar{y} - B)]$$

$$= \frac{1}{n}\ \Sigma\ [(x - A)\ (y - B) - (\bar{y} - B)\ \frac{\Sigma(x - A)}{n} - (\bar{x} - A)\frac{\Sigma(y - B)}{n} + \frac{\Sigma(\bar{x} - A)(\bar{y} - B)}{n}$$

$$= \frac{\Sigma(x - A)(y - B)}{n} - (\bar{y} - B)\ (\bar{x} - \frac{nA}{n}) - (\bar{x} - A)\ (\bar{y} - \frac{nB}{n}) + (\bar{x} - A)\ (\bar{y} - B)$$

$$= \frac{\Sigma(x - A)(y - B)}{n} - (\bar{y} - B)\ (\bar{x} - A)$$

$$- \cancel{(\bar{x} - A)\ (\bar{y} - B)} + \cancel{(\bar{x} - A)\ (\bar{y} - B)}$$

$$= \frac{\Sigma(x - A)(y - B)}{n} - (\bar{x} - A)\ (\bar{y} - B)$$

Let $x - A = u$; $y - B = v$; $\bar{x} - A = \bar{u}$; $\bar{y} - B = \bar{v}$

$$\therefore \text{Cov}\ (x,y) = \frac{\Sigma uv}{n} - \bar{u}\bar{v}$$

$$\sigma\sigma_x^2 = \frac{\Sigma u^2}{n} - \bar{u}^2 = \sigma u^2$$

$$\sigma\sigma_y^2 = \frac{\Sigma v^2}{n} - \bar{v}^2 = \sigma v^2$$

$$\therefore r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right].\left[(n\Sigma v^2) - (\Sigma v)^2\right]}}$$

Example 3:

Calculate Pearson' s Coefficient of correlation.

X	45	55	56	58	60	65	68	70	75	80	85
Y	56	50	48	60	62	64	65	70	74	82	90

X	Y	u = x-A	v = y-B	u^2	v^2	uv
45	56	-20	-14	400	196	280
55	50	-10	-20	100	400	200
56	48	-9	-22	81	484	198
58	60	-7	-10	49	100	70
60	62	-5	-8	25	64	40
65	64	0	-6	0	36	0
68	65	3	-5	9	25	-15
70	70	5	0	25	0	0
75	74	10	4	100	16	40
80	82	15	12	225	144	180
85	90	20	20	400	400	400
		2	-49	1414	1865	1393

$$r = \frac{n\Sigma uv - (\Sigma u)\ (\Sigma v)}{\sqrt{[n\Sigma u^2 - (\Sigma u^2)]\ [n\Sigma v^2 - (\Sigma v)^2]}}$$

$$r = \frac{11 \times 1393 - 2 \times (-49)}{\sqrt{(1414 \times 11 - (2)^2) \times (1865 \times 11 - (-49)^2)}}$$

$$= \frac{15421}{\sqrt{15550 \times 18114}} = \frac{15421}{16783.11} = +0.92$$

Limitations:

1. Correlation coefficient assumes linear relationship regardless of the assumption is correct or not.
2. Extreme items of variables are being unduly operated on correlation coefficient.
3. Existence of correlation does not necessarily indicate cause-effect relation.

Interpretation:

The following rules helps in interpreting the value of ' r' .

1. When r = 1, there is perfect +ve relationship between the variables.
2. When r = -1, there is perfect –ve relationship between the variables.
3. When r = 0, there is no relationship between the variables.
4. If the correlation is +1 or –1, it signifies that there is a high degree of correlation. (+ve or –ve) between the two variables.

If r is near to zero (ie) 0.1,-0.1, (or) 0.2 there is less correlation.

Rank Correlation:

It is studied when no assumption about the parameters of the population is made. This method is based on ranks. It is useful to study the qualitative measure of attributes like honesty, colour, beauty, intelligence, character, morality etc.The individuals in the group can be arranged in order and there on, obtaining for each individual a number showing his/her rank in the group. This method was developed by Edward Spearman in 1904. It is defined as $r = 1 - \frac{6\Sigma D^2}{n^3 - n}$ r = rank correlation coefficient.

Note: Some authors use the symbol ρ for rank correlation.

ΣD^2 = sum of squares of differences between the pairs of ranks.

n = number of pairs of observations.

The value of r lies between –1 and +1. If r = +1, there is complete agreement in order of ranks and the direction of ranks is also same. If r = -1, then there is complete disagreement in order of ranks and they are in opposite directions.

Computation for tied observations: There may be two or more items having equal values. In such case the same rank is to be given. The ranking is said to be tied. In such circumstances an average rank is to be given to each individual item. For example if the value so is repeated twice at the 5^{th} rank, the common rank to be assigned to each item is $\frac{5+6}{2} = 5.5$ which is the average of 5 and 6 given as 5.5, appeared twice.

If the ranks are tied, it is required to apply a correction factor which is $\frac{1}{12}$ (m^3-m). A slightly different formula is used when there is more than one item having the same value.

The formula is

$$r = 1 - \frac{6[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + ...]}{n^3 - n}$$

Where m is the number of items whose ranks are common and should be repeated as many times as there are tied observations.

Example 6:

In a marketing survey the price of tea and coffee in a town based on quality was found as shown below. Could you find any relation between and tea and coffee price.

Price of tea	88	90	95	70	60	75	50
Price of coffee	120	134	150	115	110	140	100

Price of tea	Rank	Price of coffee	Rank	D	D^2
88	3	120	4	1	1
90	2	134	3	1	1
95	1	150	1	0	0
70	5	115	5	0	0
60	6	110	6	0	0
75	4	140	2	2	4
50	7	100	7	0	0
					$\Sigma D^2 = 6$

$$r = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 6}{7^3 - 7}$$

$$= 1 - \frac{36}{336} = 1 - 0.1071$$

$$= 0.8929$$

The relation between price of tea and coffee is positive at 0.89. Based on quality the association between price of tea and price of coffee is highly positive.

Example 7:

In an evaluation of answer script the following marks are awarded by the examiners.

1^{st}	88	95	70	960	50	80	75	85
2^{nd}	84	90	88	55	48	85	82	72

Do you agree the evaluation by the two examiners is fair?

x	R1	y	R2	D	D^2
88	2	84	4	2	4
95	1	90	1	0	0
70	6	88	2	4	16
60	7	55	7	0	0
50	8	48	8	0	0
80	4	85	3	1	1
85	3	75	6	3	9
					30

$$r = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 30}{8^3 - 8}$$

$$= 1 - \frac{180}{504} = 1 - 0.357 = 0.643$$

r = 0.643 shows fair in awarding marks in the sense that uniformity has arisen in evaluating the answer scripts between the two examiners.

Example 8:

Rank Correlation for tied observations. Following are the marks obtained by 10 students in a class in two tests.

Students	A	B	C	D	E	F	G	H	I	J
Test 1	70	68	67	55	60	60	75	63	60	72
Test 2	65	65	80	60	68	58	75	63	60	70

Calculate the rank correlation coefficient between the marks of two tests.

Student	Test 1	R1	Test 2	R2	D	D^2
A	70	3	65	5.5	-2.5	6.25
B	68	4	65	5.5	-1.5	2.25
C	67	5	80	1.0	4.0	16.00
D	55	10	60	8.5	1.5	2.25
E	60	8	68	4.0	4.0	16.00
F	60	8	58	10.0	-2.0	4.00
G	75	1	75	2.0	-1.0	1.00
H	63	6	62	7.0	-1.0	1.00
I	60	8	60	8.5	0.5	0.25
J	72	2	70	3.0	-1.0	1.00
						50.00

60 is repeated 3 times in test 1.
60,65 is repeated twice in test 2.
m = 3; m = 2; m = 2

$$r = 1 - \frac{6[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)}{n^3 - n}$$

$$= 1 - \frac{6[50 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)]}{10^3 - 10}$$

$$= 1 - \frac{6[50 + 2 + 0.5 + 0.5]}{990}$$

$$= 1 - \frac{6 \times 53}{990} = \frac{672}{990} = 0.68$$

REGRESSION

9.1 Introduction:

After knowing the relationship between two variables we may be interested in estimating (predicting) the value of one variable given the value of another. The variable predicted on the basis of other variables is called the "dependent" or the ' explained' variable and the other the ' independent' or the ' predicting' variable. The prediction is based on average relationship derived statistically by regression analysis. The equation, linear or otherwise, is called the regression equation or the explaining equation.

For example, if we know that advertising and sales are correlated we may find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales.

The relationship between two variables can be considered between, say, rainfall and agricultural production, price of an input and the overall cost of product, consumer expenditure and disposable income. Thus, regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

9.1.1 Definition:

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

9.2 Types Of Regression:

The regression analysis can be classified into:

a) Simple and Multiple
b) Linear and Non –Linear
c) Total and Partial

a) Simple and Multiple:

In case of simple relationship only two variables are considered, for example, the influence of advertising expenditure on sales turnover. In the case of multiple relationship, more than

two variables are involved. On this while one variable is a dependent variable the remaining variables are independent ones.

For example, the turnover (y) may depend on advertising expenditure (x) and the income of the people (z). Then the functional relationship can be expressed as $y = f(x,z)$.

b) Linear and Non-linear:

The linear relationships are based on straight-line trend, the equation of which has no-power higher than one. But, remember a linear relationship can be both simple and multiple. Normally a linear relationship is taken into account because besides its simplicity, it has a better predective value, a linear trend can be easily projected into the future. In the case of non-linear relationship curved trend lines are derived. The equations of these are parabolic.

c) Total and Partial:

In the case of total relationships all the important variables are considered. Normally, they take the form of a multiple relationships because most economic and business phenomena are affected by multiplicity of cases. In the case of partial relationship one or more variables are considered, but not all, thus excluding the influence of those not found relevant for a given purpose.

9.3 Linear Regression Equation:

If two variables have linear relationship then as the independent variable (X) changes, the dependent variable (Y) also changes. If the different values of X and Y are plotted, then the two straight lines of best fit can be made to pass through the plotted points. These two lines are known as regression lines. Again, these regression lines are based on two equations known as regression equations. These equations show best estimate of one variable for the known value of the other. The equations are linear.

Linear regression equation of Y on X is

$$Y = a + bX \ldots\ldots(1)$$

And X on Y is

$$X = a + bY \ldots\ldots(2)$$

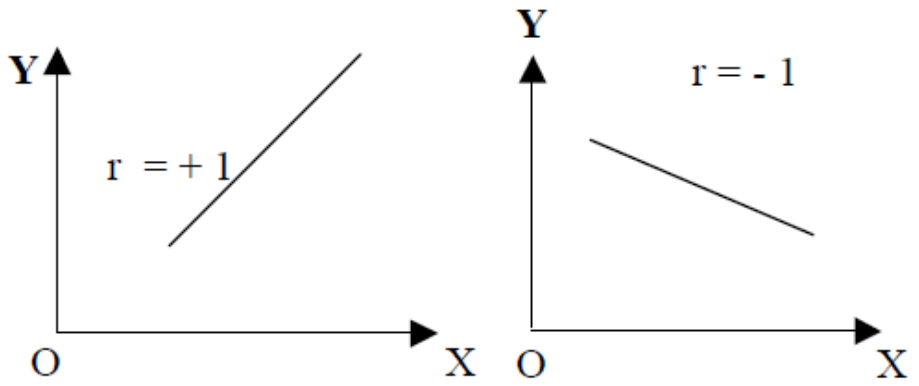
a, b are constants.

From (1) We can estimate Y for known value of X.
(2) We can estimate X for known value of Y.

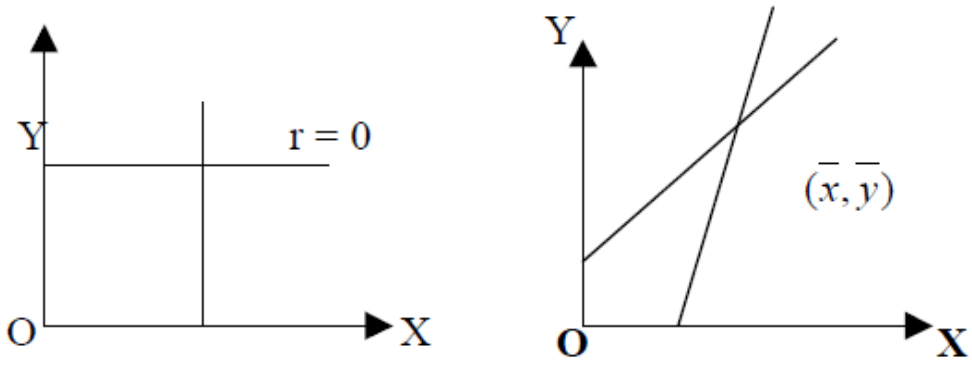
9.3.1 Regression Lines:

For regression analysis of two variables there are two regression lines, namely Y on X and X on Y. The two regression lines show the average relationship between the two variables.

For perfect correlation, positive or negative i.e., $r = \pm 1$, the two lines coincide i.e., we will find only one straight line. If $r = 0$, i.e., both the variables are independent then the two lines will cut each other at right angle. In this case the two lines will be parallel to X and Y-axes.



Lastly the two lines intersect at the point of means of X and Y. From this point of intersection, if a straight line is drawn on X-axis, it will touch at the mean value of x. Similarly, a perpendicular drawn from the point of intersection of two regression lines on Y-axis will touch the mean value of Y.



9.3.2 Principle of ' Least Squares' :

Regression shows an average relationship between two variables, which is expressed by a line of regression drawn by the method of "least squares". This line of regression can be derived graphically or algebraically. Before we discuss the various methods let us understand the meaning of least squares.

A line fitted by the method of least squares is known as the line of best fit. The line adapts to the following rules:

(i) The algebraic sum of deviation in the individual observations with reference to the regression line may be equal to zero. i.e.,

$\sum(X - Xc) = 0$ or $\sum (Y - Yc) = 0$

Where Xc and Yc are the values obtained by regression analysis.

(ii) The sum of the squares of these deviations is less than the sum of squares of deviations from any other line. i.e.,

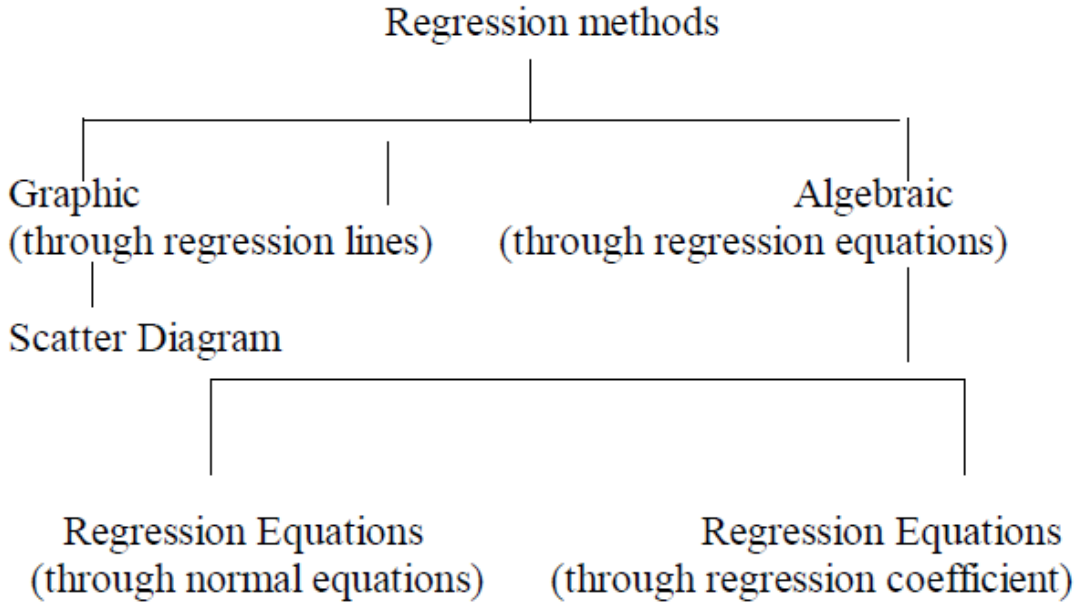
$\sum(Y - Yc)^2 < \sum (Y - Ai)^2$

Where Ai = corresponding values of any other straight line.

(iii) The lines of regression (best fit) intersect at the mean values of the variables X and Y, i.e., intersecting point is $\bar{x}, \bar{y}$.

9.4 Methods of Regression Analysis:

The various methods can be represented in the form of chart given below:



9.4.1 Graphic Method:

Scatter Diagram:

Under this method the points are plotted on a graph paper representing various parts of values of the concerned variables. These points give a picture of a scatter diagram with several points spread over. A regression line may be drawn in between these points either by free hand or by a scale rule in such a way that the squares of the vertical or the horizontal distances (as the case may be) between the points and the line of regression so drawn is the least. In other words, it should be drawn faithfully as the line of best fit leaving equal number of points on both sides in such a manner that the sum of the squares of the distances is the best.

9.4.2 Algebraic Methods:

(i) Regression Equation.

The two regression equations

for X on Y; $X = a + bY$

And for Y on X; $Y = a + bX$

Where X, Y are variables, and a,b are constants whose values are to be determined

For the equation, $X = a + bY$

The normal equations are

$$\sum X = na + b \sum Y \text{ and}$$

$$\sum XY = a\sum Y + b\sum Y^2$$

For the equation, $Y = a + bX$, the normal equations are

$$\sum Y = na + b\sum X \text{ and}$$

$$\sum XY = a\sum X + b\sum X^2$$

From these normal equations the values of a and b can be determined.

Example 1:

Find the two regression equations from the following data:

X:	6	2	10	4	8
Y:	9	11	5	8	7

Solution:

X	Y	X^2	Y^2	XY
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
30	40	220	340	214

Regression equation of Y on X is $Y = a + bX$ and the normal equations are

$$\sum Y = na + b\sum X$$
$$\sum XY = a\sum X + b\sum X^2$$

Substituting the values, we get

$$40 = 5a + 30b \ldots\ldots (1)$$
$$214 = 30a + 220b \ldots\ldots (2)$$

Multiplying (1) by 6

$$240 = 30a + 180b \ldots\ldots (3)$$

(2) – (3) $\quad -26 = 40b$

$$\text{or } b = -\frac{26}{40} = -0.65$$

Now, substituting the value of 'b' in equation (1)

$$40 = 5a - 19.5$$
$$5a = 59.5$$
$$a = \frac{59.5}{5} = 11.9$$

Hence, required regression line Y on X is $Y = 11.9 - 0.65\,X$.

Again, regression equation of X on Y is

$X = a + bY$ and

The normal equations are

$$\sum X = na + b\sum Y \text{ and}$$
$$\sum XY = a\sum Y + b\sum Y^2$$

Now, substituting the corresponding values from the above table, we get

$$30 = 5a + 40b \ldots(3)$$
$$214 = 40a + 340b \ldots(4)$$

Multiplying (3) by 8, we get

$$240 = 40a + 320\, b \ldots(5)$$

(4) – (5) gives

$$-26 = 20b$$
$$b = -\frac{26}{20} = -1.3$$

Substituting $b = -1.3$ in equation (3) gives

$$30 = 5a - 52$$
$$5a = 82$$
$$a = \frac{82}{5} = 16.4$$

Hence, Required regression line of X on Y is

$X = 16.4 - 1.3Y$

(ii) Regression Co-efficents:

The regression equation of Y on X is $y_e = \bar{y} + r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$

Here, the regression Co.efficient of Y on X is

$$b_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x}$$
$$y_e = \bar{y} + b_1(x - \bar{x})$$

The regression equation of X on Y is

$$X_e = \bar{x} + r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

Here, the regression Co-efficient of X on Y

$$b_2 = b_{xy} = r\frac{\sigma_x}{\sigma_y}$$
$$X_e = \bar{X} + b_2(y - \bar{y})$$

If the deviation are taken from respective means of x and y

$$b_1 = b_{yx} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (X-\overline{X})^2} = \frac{\sum xy}{\sum x^2} \quad \text{and}$$

$$b_2 = b_{xy} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sum (Y-\overline{Y})^2} = \frac{\sum xy}{\sum y^2}$$

where $x = X - \overline{X}, y = Y - \overline{Y}$

If the deviations are taken from any arbitrary values of x and y (short – cut method)

$$b_1 = b_{yx} = \frac{n\sum uv - \sum u \sum v}{n\sum u^2 - \left(\sum u\right)^2}$$

$$b_2 = b_{xy} = \frac{n\sum uv - \sum u \sum v}{n\sum v^2 - \left(\sum v\right)^2}$$

where u = x – A : v = Y-B

A = any value in X

B = any value in Y

9.5 Properties of Regression Co-efficient:

1. Both regression coefficients must have the same sign, ie either they will be positive or negative.
2. correlation coefficient is the geometric mean of the regression coefficients ie, $r = \pm\sqrt{b_1 b_2}$
3. The correlation coefficient will have the same sign as that of the regression coefficients.
4. If one regression coefficient is greater than unity, then other regression coefficient must be less than unity.
5. Regression coefficients are independent of origin but not of scale.
6. Arithmetic mean of b_1 and b_2 is equal to or greater than the coefficient of correlation. Symbolically $\frac{b_1 + b_2}{2} \geq r$

7. If r=0, the variables are uncorrelated , the lines of regression become perpendicular to each other.
8. If r= ± 1, the two lines of regression either coincide or parallel to each other
9. Angle between the two regression lines is $\theta = \tan^{-1}\left[\frac{m_1 - m_2}{1 + m_1 m_2}\right]$

   where m_1 and,m_2 are the slopes of the regression lines X on Y and Y on X respectively.
10. The angle between the regression lines indicates the degree of dependence between the variables.

Example 2:

If 2 regression coefficients are $b_1=\frac{4}{5}$ and $b_2=\frac{9}{20}$ .What would be the value of r?

Solution:

The correlation coefficient , $r = \pm\sqrt{b_1 b_2}$

$$= \sqrt{\frac{4}{5} \text{x} \frac{9}{20}}$$

$$= \sqrt{\frac{36}{100}} = \frac{6}{10} = 0.6$$

Example 3:

Given $b_1 = \frac{15}{8}$ and $b_2 = \frac{3}{5}$, Find r

Solution:

$r = \pm\sqrt{b_1 b_2}$

$$= \sqrt{\frac{15}{8} \text{x} \frac{3}{5}}$$

$$= \sqrt{\frac{9}{8}} = 1.06$$

It is not possible since r, cannot be greater than one. So the given values are wrong

Example 4:

Compute the two regression equations from the following data.

X	1	2	3	4	5
Y	2	3	5	4	6

If x =2.5, what will be the value of y?

Solution:

X	Y	$x = X - \overline{X}$	$y = Y - \overline{Y}$	x^2	y^2	xy
1	2	-2	-2	4	4	4
2	3	-1	-1	1	1	-1
3	5	0	1	0	1	0
4	4	1	0	1	0	0
5	6	2	2	4	4	4
15	20	20		10	10	9

$$\overline{X} = \frac{\Sigma X}{n} = \frac{15}{5} = 3$$

$$\overline{Y} = \frac{\Sigma Y}{n} = \frac{20}{5} = 4$$

Regression Co efficient of Y on X

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{9}{10} = 0.9$$

Hence regression equation of Y on X is

$$Y = \overline{Y} + b_{yx}(X - \overline{X})$$

$$= 4 + 0.9\ (X - 3)$$

$$= 4 + 0.9X - 2.7$$

$$= 1.3 + 0.9X$$

when X = 2.5

$Y = 1.3 + 0.9 \times 2.5$

$= 3.55$

Regression co efficient of X on Y

$$b_{xy} = \frac{\sum xy}{\sum y^2} = \frac{9}{10} = 0.9$$

So, regression equation of X on Y is

$$X = \overline{X} + b_{xy}(Y - \overline{Y})$$

$$= 3 + 0.9\,(Y - 4)$$

$$= 3 + 0.9Y - 3.6$$

$$= 0.9Y - 0.6$$

Example 6:

In a correlation study, the following values are obtained

	X	Y
Mean	65	67
S.D	2.5	3.5

Co-efficient of correlation = 0.8

Find the two regression equations that are associated with the above values.

Solution:

Given,

$\overline{X} = 65$, $\overline{Y} = 67$, $\sigma_x = 2.5$, $\sigma_y = 3.5$, $r = 0.8$

The regression co-efficient of Y on X is

$$b_{yx} = b_1 = r\frac{\sigma_y}{\sigma_x}$$

$$= 0.8 \times \frac{3.5}{2.5} = 1.12$$

The regression coefficient of X on Y is

$$b_{xy} = b_2 = r\,\frac{\sigma_x}{\sigma_y}$$

$$= 0.8 \times \frac{2.5}{3.5} = 0.57$$

Hence, the regression equation of Y on X is

$$Y_e = \overline{Y} + b_1(X - \overline{X})$$

$$= 67 + 1.12 \text{ (X-65)}$$

$$= 67 + 1.12 \text{ X} - 72.8$$

$$= 1.12\text{X} - 5.8$$

The regression equation of X on Y is

$$X_e = \overline{X} + b_2(Y - \overline{Y})$$

$$= 65 + 0.57 \text{ (Y-67)}$$

$$= 65 + 0.57\text{Y} - 38.19$$

$$= 26.81 + 0.57\text{Y}$$

9.7 Uses of Regression Analysis:

1. Regression analysis helps in establishing a functional relationship between two or more variables.
2. Since most of the problems of economic analysis are based on cause and effect relationships, the regression analysis is a highly valuable tool in economic and business research.
3. Regression analysis predicts the values of dependent variables from the values of independent variables.
4. We can calculate coefficient of correlation (r) and coefficient of determination (r^2) with the help of regression coefficients.
5. In statistical analysis of demand curves, supply curves, production function, cost function, consumption function etc., regression analysis is widely used.

9.8 Difference between Correlation and Regression:

S.No	Correlation	Regression
1.	Correlation is the relationship between two or more variables, which vary in sympathy with the other in the same or the opposite direction.	Regression means going back and it is a mathematical measure showing the average relationship between two variables
2.	Both the variables X and Y are random variables	Here X is a random variable and Y is a fixed variable. Sometimes both the variables may be random variables.
3.	It finds out the degree of relationship between two variables and not the cause and effect of the variables.	It indicates the causes and effect relationship between the variables and establishes functional relationship.
4.	It is used for testing and verifying the relation between two variables and gives limited information.	Besides verification it is used for the prediction of one value, in relationship to the other given value.
5.	The coefficient of correlation is a relative measure. The range of relationship lies between –1 and +1	Regression coefficient is an absolute figure. If we know the value of the independent variable, we can find the value of the dependent variable.
6.	There may be spurious correlation between two variables.	In regression there is no such spurious regression.

7.	It has limited application, because it is confined only to linear relationship between the variables.	It has wider application, as it studies linear and non-linear relationship between the variables.
8.	It is not very useful for further mathematical treatment.	It is widely used for further mathematical treatment.
9.	If the coefficient of correlation is positive, then the two variables are positively correlated and vice-versa.	The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable.

CHAPTER – V

TEST OF SIGNIFICANCE (Basic Concepts)

Parameter and Statistic:

The statistical constants of the population such as mean, (μ), variance (σ^2), correlation coefficient (ρ) and proportion (P) are called ' Parameters' .

Statistical constants computed from the samples corresponding to the parameters namely mean ($\bar{x}$), variance (S^2), sample correlation coefficient (r) and proportion (p) etc, are called statistic.

Parameters are functions of the population values while statistic are functions of the sample observations. In general, population parameters are unknown and sample statistics are used as their estimates.

Sl.No	Statistic	Standard Error
1.	Sample mean $\bar{x}$	$\frac{\sigma}{\sqrt{n}}$
2.	Observed sample proportion p	$\sqrt{\frac{PQ}{n}}$
3.	Difference between of two samples means ($\bar{x}_1 - \bar{x}_2$)	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
4.	Difference of two sample proportions $p_1 - p_2$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

Uses of standard error

i) Standard error plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

ii) The magnitude of the S.E gives an index of the precision of the estimate of the parameter.

iii) The reciprocal of the S.E is taken as the measure of reliability or precision of the sample.

iv) S.E enables us to determine the probable limits within which the population parameter may be expected to lie.

Null Hypothesis and Alternative Hypothesis

Hypothesis testing begins with an assumption called a Hypothesis, that we make about a population parameter. A hypothesis is a supposition made as a basis for reasoning. The conventional approach to hypothesis testing is not to construct a single hypothesis about the population parameter but rather to set up two different hypothesis. So that of one hypothesis is accepted, the other is rejected and vice versa.

Null Hypothesis:

A hypothesis of no difference is called null hypothesis and is usually denoted by H_0 " Null hypothesis is the hypothesis" which is tested for possible rejection under the assumption that it is true " by Prof. R.A. Fisher. It is very useful tool in test of significance. For example: If we want to find out whether the special classes (for Hr. Sec. Students) after school hours has benefited the students or not. We shall set up a null hypothesis that "H_0: special classes after school hours has not benefited the students".

Alternative Hypothesis:

Any hypothesis, which is complementary to the null hypothesis, is called an alternative hypothesis, usually denoted by H_1, For example, if we want to test the null hypothesis that the population has a specified mean μ_0 (say),

i.e., : Step 1: null hypothesis H_0: $\mu = \mu_0$

then 2. Alternative hypothesis may be

i) $H_1 : \mu \neq \mu_0$ (ie $\mu > \mu_0$ or $\mu < \mu_0$)

ii) $H_1 : \mu > \mu_0$

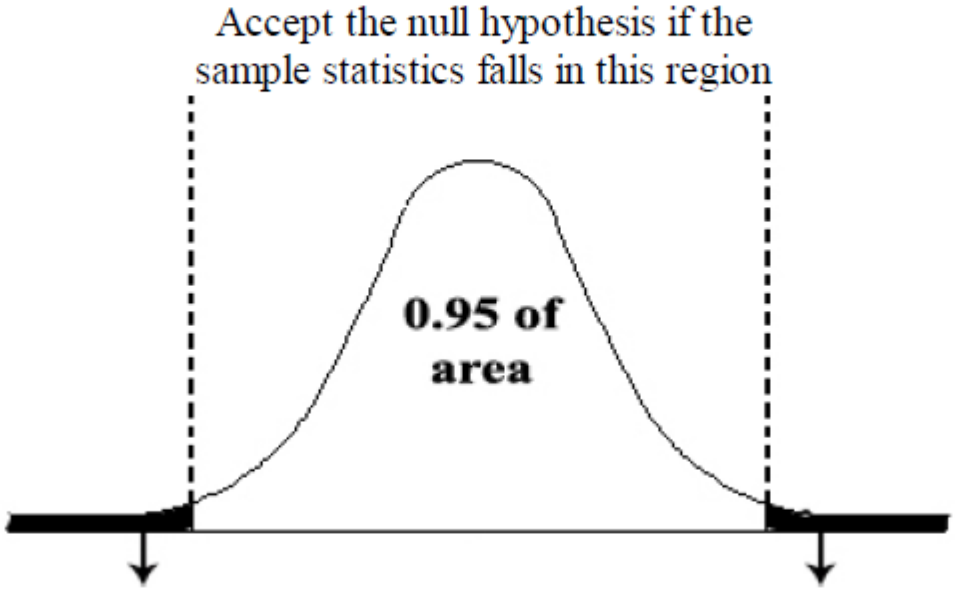
iii) $H_1 : \mu < \mu_0$

the alternative hypothesis in (i) is known as a two – tailed alternative and the alternative in (ii) is known as right-tailed (iii) is known as left –tailed alternative respectively. The settings of alternative hypothesis is very important since it enables us to decide whether we have to use a single – tailed (right or left) or two tailed test.

Level of significance:

In testing a given hypothesis, the maximum probability with which we would be willing to take risk is called level of significance of the test. This probability often denoted by "α" is generally specified before samples are drawn.

The following diagram illustrates the region in which we could accept or reject the null hypothesis when it is being tested at 5 % level of significance and a two-tailed test is employed.



Reject the null hypothesis if the sample
Statistics falls in these two region

Critical Value:

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value. It depends upon i) the level of significance used and ii) the alternative hypothesis, whether it is two-tailed or single-tailed

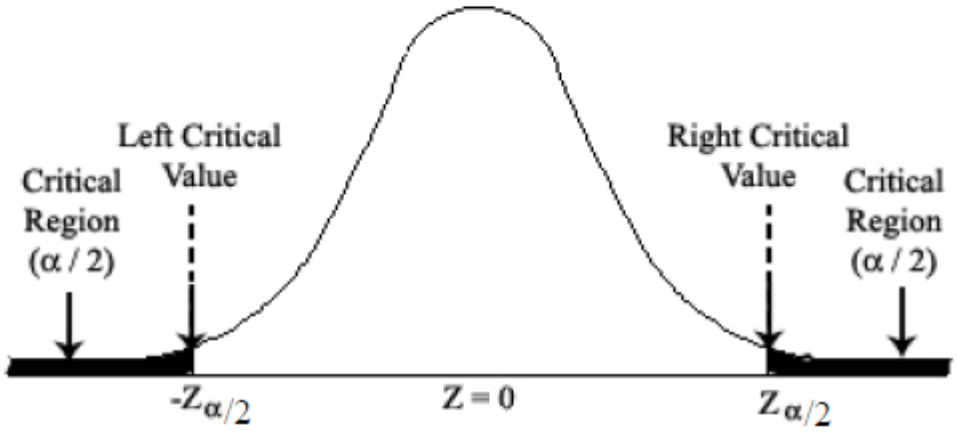
For large samples the standard normal variate corresponding to the statistic t, $Z = \left| \frac{t - E(t)}{S.E.(t)} \right| \sim N(0,1)$

asymptotically as $n \rightarrow \infty$

The value of z under the null hypothesis is known as test statistic. The critical value of the test statistic at the level of significance α for a two - tailed test is given by $Z_{\alpha/2}$ and for a one tailed test by Z_α. where Z_α is determined by equation $P(|Z| > Z_\alpha) = \alpha$

$Z\alpha$ is the value so that the total area of the critical region on both tails is α . $\therefore$ $P(Z > Z_\alpha) = \frac{\alpha}{2}$. Area of each tail is $\frac{\alpha}{2}$.

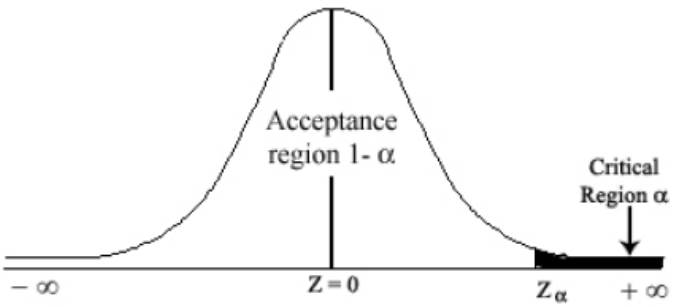
$Z\alpha$ is the value such that area to the right of $Z\alpha$ and to the left of $-Z\alpha$ is $\frac{\alpha}{2}$ as shown in the following diagram.



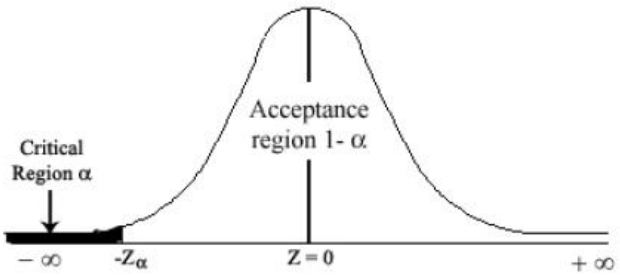
In any test, the critical region is represented by a portion of the area under the probability curve of the sampling distribution of the test statistic.

One tailed test: A test of any statistical hypothesis where the alternative hypothesis is one tailed (right tailed or left tailed) is called a one tailed test.

Right tailed test:



Left tailed test:



Two tailed test:

A test of statistical hypothesis where the alternative hypothesis is two tailed such as, $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1: \mu \neq \mu_0$ ($\mu > \mu_0$ and $\mu < \mu_0$) is known as two tailed test and in such a case the critical region is given by the portion of the area lying in both the tails of the probability curve of test of statistic.

Critical values ($Z\alpha$) of Z

Level of significance α	0.05 or 5%		0.01 or 1%	
	Left	Right	Left	Right
Critical values of Z_α for one tailed Tests	–1.645	1.645	–2.33	2.33
Critical values of $Z_{\alpha/2}$ for two tailed tests	–1.96	1.96	–2.58	2.58

Type I and Type II Errors:

When a statistical hypothesis is tested there are four possibilities.

1. The hypothesis is true but our test rejects it (Type I error)
2. The hypothesis is false but our test accepts it (Type II error)
3. The hypothesis is true and our test accepts it (correct decision)
4. The hypothesis is false and our test rejects it (correct decision)

In a statistical hypothesis testing experiment, a Type I error is committed by rejecting the null hypothesis when it is true. On the other hand, a Type II error is committed by not rejecting (accepting) the null hypothesis when it is false.

If we write ,

α = P (Type I error) = P (rejecting H_0 | H_0 is true)

β = P (Type II error) = P (Not rejecting H_0 | H_0 is false)

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad. Thus we find ourselves in the situation which is described in the following table.

	Accept H_0	Reject H_0
H_0 is true	Correct decision	Type I Error
H_0 is false	Type II error	Correct decision

Test Procedure :

Steps for testing hypothesis is given below. (for both large sample and small sample tests)

1. Null hypothesis : set up null hypothesis H_0.
2. Alternative Hypothesis: Set up alternative hypothesis H_1, which is complementry to H_0 which will indicate whether one tailed (right or left tailed) or two tailed test is to be applied.
3. Level of significance : Choose an appropriate level of significance (α), α is fixed in advance.
4. Test statistic (or test of criterian):

   Calculate the value of the test statistic, $Z = \dfrac{t - E(t)}{S.E.(t)}$ under the null hypothesis, where t is the sample statistic
5. Inference: We compare the computed value of Z (in absolute value) with the significant value (critical value) $Z\alpha/2$ (or $Z\alpha$). If $|Z| > Z\alpha$, we reject the null hypothesis H_0 at α % level of significance and if $|Z| \leq Z\alpha$, we accept H_0 at α % level of significance.

TEST OF SIGNIFICANCE
(Large Sample)

Large samples (n > 30):

The tests of significance used for problems of large samples are different from those used in case of small samples as the assumptions used in both cases are different. The following assumptions are made for problems dealing with large samples:

(i) Almost all the sampling distributions follow normal asymptotically.

(ii) The sample values are approximately close to the population values.

The following tests are discussed in large sample tests.

(i) Test of significance for proportion

(ii) Test of significance for difference between two proportions

(iii) Test of significance for mean

(iv) Test of significance for difference between two means.

1. Test of significance for mean:

Let x_i (i = 1,2....n) be a random sample of size n from a population with variance σ^2, then the sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots x_n)$$

$$E(\bar{x}) = \mu$$

$$V(\bar{x}) = V\left[\frac{1}{n}(x_1 + x_2 + \ldots x_n)\right]$$

$$= \frac{1}{n^2}[(V(x_1) + V(x_2) + \ldots V(x_n)]$$

$$= \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}$$

$$\therefore \text{S.E}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Test Procedure:

Null and Alternative Hypotheses:

$H_0: \mu = \mu_0$.

$H_1: \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$)

Level of significance:

Let $\alpha = 0.05$ or 0.01

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left| \frac{\bar{x} - E(\bar{x})}{S.E(\bar{x})} \right| = \left| \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right|$$

Expected value:

$$Z_e = \left| \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right| \sim N(0,1)$$

$= 1.96$ for $\alpha = 0.05$ (1.645)

or

$= 2.58$ for $\alpha = 0.01$ (2.33)

Inference :

If $Z_0 \leq Z_e$, we accept our null hypothesis and conclude that the sample is drawn from a population with mean $\mu = \mu_0$

If $Z_0 > Z_e$ we reject our H_0 and conclude that the sample is not drawn from a population with mean $\mu = \mu_0$

Example 6:

The mean lifetime of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean lifetime of all the bulbs produced by the company, test the hypothesis μ=1600 hours against the alternative hypothesis $\mu \neq 1600$ hours using a 5% level of significance.

Solution:

We are given

$\bar{x}$ = 1570 hrs μ = 1600hrs s =120 hrs n=100

Null hypothesis:

H_0: μ= 1600.ie There is no significant difference between the sample mean and population mean.

Alternative Hypothesis:

H_1: $\mu \neq 1600$ (two tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistics

Under H_0, the test statistic is

$$Z_0 = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right|$$

$$= \left|\frac{1570 - 1600}{\frac{120}{\sqrt{100}}}\right|$$

$$= \frac{30 \times 10}{120}$$

$$= 2.5$$

Expected value:

$$Z_0 = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$= 1.96 \text{ for } \alpha = 0.05$$

Inference :

Since $Z_0 > Z_e$ we reject our null hypothesis at 5% level of significance and say that there is significant difference between the sample mean and the population mean.

Example 7:

A car company decided to introduce a new car whose mean petrol consumption is claimed to be lower than that of the existing car. A sample of 50 new cars were taken and tested for petrol consumption. It was found that mean petrol consumption for the 50 cars was 30 km per litre with a standard deviation of 3.5 km per litre. Test at 5% level of significance whether the company' s claim that the new car petrol consumption is 28 km per litre on the average is acceptable.

Solution:

We are given $\bar{x} = 30$; $\mu = 28$; $n = 50$; $s = 3.5$

Null hypothesis:

H_0: $\mu = 28$. i.e The company' s claim that the petrol consumption of new car is 28km per litre on the average is acceptable.

Alternative Hypothesis:

H_1: $\mu < 28$ (Left tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistic:

Under H_0 the test statistics is

$$Z_0 = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right|$$

$$= \left| \frac{30 - 28}{\frac{3.5}{\sqrt{50}}} \right|$$

$$= \frac{2 \times \sqrt{50}}{3.5}$$

$$= 4.04$$

Expected value:

$$Z_e = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| \sim N(0,1) \text{ at } \alpha = 0.05$$

$$= 1.645$$

Inference :

Since the calculated $Z_0 > Z_e$ we reject the null hypothesis at 5% level of significance and conclude that the company' s claim is not acceptable.

5.5 Test of significance for difference between two means:

Test procedure

Set up the null and alternative hypothesis

H_0: $\mu_1 = \mu_2$; **H_1:** $\mu_1 \neq \mu_2$ ($\mu_1 > \mu_2$ or $\mu_1 < \mu_2$)

Level of significance:

Let α%

Calculation of statistic:

Under H_0 the test statistic is

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (ie) If the samples have been drawn from the population with common S.D σ then under $H_0 : \mu_1 = \mu_2$

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

Expected value:

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{S.E(\bar{x}_1 - \bar{x}_2)} \right| \sim N(0,1)$$

Inference:

(i) If $Z_0 \leq Z_e$ we accept the H_0 (ii) If $Z_0 > Z_e$ we reject the H_0

Example 8:

A test of the breaking strengths of two different types of cables was conducted using samples of $n_1 = n_2 = 100$ pieces of each type of cable.

Cable I	Cable II
$\bar{x}_1 = 1925$	$\bar{x}_2 = 1905$
$\sigma_1 = 40$	$\sigma_2 = 30$

Do the data provide sufficient evidence to indicate a difference between the mean breaking strengths of the two cables? Use 0.01 level of significance.

Solution:

We are given

$\bar{x}_1 = 1925$ $\bar{x}_2 = 1905$ $\sigma_1 = 40$ $\sigma_2 = 30$

Null hypothesis

$H_0: \mu_1 = \mu_2$ .ie There is no significant difference between the mean breaking strengths of the two cables.

$H_1 : \mu_1 \neq \mu_2$ (Two tailed alternative)

Level of significance:

Let $\alpha = 0.01$ or 1%

Calculation of statistic:

Under H_0 the test statistic is

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

$$= \left| \frac{1925 - 1905}{\sqrt{\frac{40^2}{100} + \frac{30^2}{100}}} \right| = \frac{20}{5} = 4$$

Expected value:

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| \sim N(0,1) = 2.58$$

Inference:

Since $Z_0 > Z_e$, we reject the H_0. Hence the formulated null hypothesis is wrong ie there is a significant difference in the breaking strengths of two cables.

Example 9:

The means of two large samples of 1000 and 2000 items are 67.5 cms and 68.0cms respectively. Can the samples be regarded as drawn from the population with standard deviation 2.5 cms. Test at 5% level of significance.

Solution:

We are given

$n_1 = 1000$; $n_2 = 2000$ $\quad \bar{x}_1 = 67.5$ cms ; $\bar{x}_2 = 68.0$ cms $\quad \sigma = 2.5$ cms

Null hypothesis

H_0: $\mu_1 = \mu_2$ (i.e.,) the sample have been drawn from the same population.

Alternative Hypothesis:

H_1: $\mu_1 \neq \mu_2$ (Two tailed alternative)

Level of significance:

$\alpha = 5\%$

Calculation of statistic:

Under H_o the test statistic is

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

$$= \left| \frac{67.5 - 68.0}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} \right|$$

$$= \frac{0.5 \times 20}{2.5\sqrt{3/5}}$$

$$= 5.1$$

Expected value:

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0,1) = 1.96$$

Inference :

Since $Z_0 > Z_e$ we reject the H_0 at 5% level of significance and conclude that the samples have not come from the same population.

TESTS OF SIGNIFICANCE
(Small Samples)

6.1 t - statistic definition:

If $x_1, x_2, \ldots x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2, then Student' s t-statistic is defined as $t = \dfrac{\bar{x} - \mu}{\dfrac{S}{\sqrt{n}}}$

where $\bar{x} = \dfrac{\sum x}{n}$ is the sample mean

and $S^2 = \dfrac{1}{n-1}\sum(x - \bar{x})^2$

is an unbiased estimate of the population variance σ^2 It follows student' s t-distribution with $\nu = n - 1$ d.f

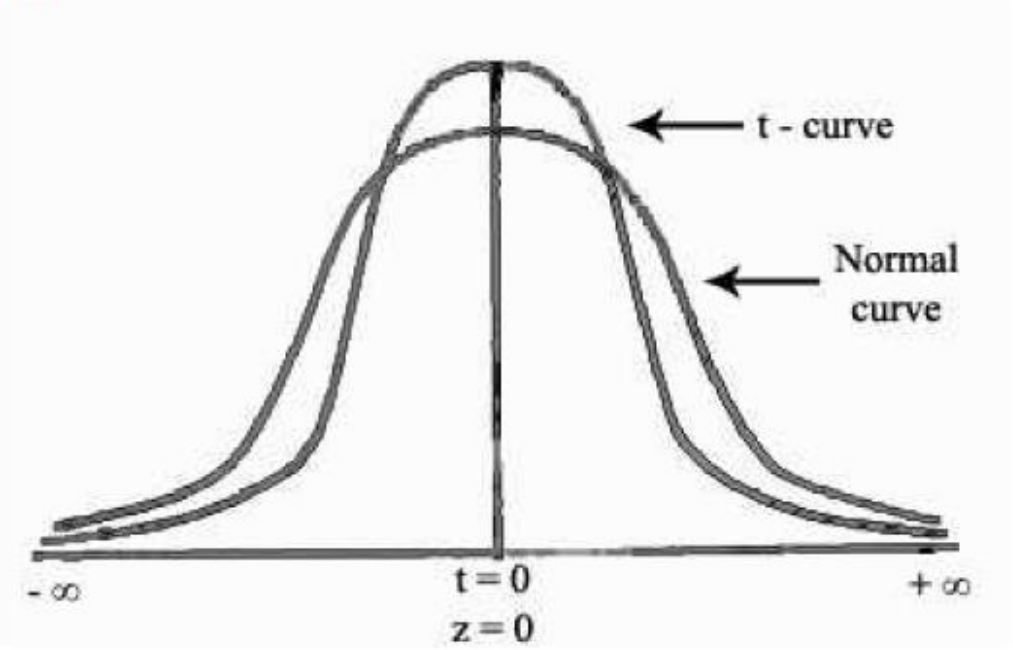
6.1.1 Assumptions for students t-test:

1. The parent population from which the sample drawn is normal.
2. The sample observations are random and independent.
3. The population standard deviation σ is not known.

6.1.2 Properties of t- distribution:

1. t-distribution ranges from $-\infty$ to ∞ just as does a normal distribution.
2. Like the normal distribution, t-distribution also symmetrical and has a mean zero.
3. t-distribution has a greater dispersion than the standard normal distribution.
4. As the sample size approaches 30, the t-distribution, approaches the Normal distribution.

Comparison between Normal curve and corresponding t - curve:



6.2 Test of significance for Mean:

We set up the corresponding null and alternative hypotheses as follows:

H_0: $\mu = \mu_0$; There is no significant difference between the sample mean and population Mean.

H_1: $\mu \neq \mu_0$ ($\mu < \mu_0$ (or) $\mu > \mu_0$)

Level of significance:

5% or 1%

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \left| \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \right| \quad \text{or} \quad \left| \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \right|$$

where $\bar{x} = \frac{\sum x}{n}$ is the sample mean

and $S^2 = \frac{1}{n-1}\sum(x - \bar{x})^2$ (or) $s^2 = \frac{1}{n}\sum(x - \bar{x})^2$

Expected value :

$$t_e = \left| \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \right| \sim \text{student's t-distribution with (n-1) d.f}$$

Inference :

If $t_0 \leq t_e$ it falls in the acceptance region and the null hypothesis is accepted and if $t_o > t_e$ the null hypothesis H_0 may be rejected at the given level of significance.

Example 1:

Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kg) as follows:

50 49 52 44 45 48 46 45 49 45

Test if the average packing can be taken to be 50 kg.

Solution:

Null hypothesis:

$H_0 : \mu = 50$ kgs in the average packing is 50 kgs.

Alternative Hypothesis:

$H_1 : \mu \neq 50$kgs (Two -tailed)

Level of Significance:

Let $\alpha = 0.05$

Calculation of sample mean and S.D

X	d = x –48	d^2
50	2	4
49	1	1
52	4	16
44	–4	16
45	–3	9
48	0	0
46	–2	4
45	–3	9
49	+1	1
45	–3	9
Total	–7	69

$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 48 + \frac{-7}{10}$$

$$= 48 - 0.7 = 47.3$$

$$S^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]$$

$$= \frac{1}{9}\left[69 - \frac{(7^2)}{10}\right]$$

$$= \frac{64.1}{9} = 7.12$$

Calculation of Statistic:

Under H_0 the test statistic is :

$$t_0 = \left|\frac{\bar{x} - \mu}{\sqrt{S^2/n}}\right|$$

$$= \left|\frac{47.3 - 50.0}{\sqrt{7.12/10}}\right|$$

$$= \frac{2.7}{\sqrt{0.712}} = 3.2$$

Expected value:

$t_e = \left| \frac{\bar{x} - \mu}{\sqrt{S^2 / n}} \right|$ follows t distribution with (10–1) d.f

$= 2.262$

Inference:

Since $t_0 > t_e$, H_0 is rejected at 5% level of significance and we conclude that the average packing cannot be taken to be 50 kgs.

Example 2:

A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?

Solution:

We are given

$n = 26;$ $\bar{x} = 147$dozens; $s = 16$

Null hypothesis:

H_0: $\mu = 140$ dozens i.e. Advertisement is not effective.

Alternative Hypothesis:

H_1: $\mu > 140$kgs (Right -tailed)

Calculation of statistic:

Under the null hypothesis H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \mu}{s / \sqrt{n-1}} \right|$$

$$= \left| \frac{147 - 140}{16 / \sqrt{25}} \right| = \frac{7 \times 5}{16} = 2.19$$

Expected value:

$t_e = \left| \frac{\bar{x} - \mu}{\sqrt{S^2 / n}} \right|$ follows t distribution with (10–1) d.f

$= 2.262$

Inference:

Since $t_0 > t_e$, H_0 is rejected at 5% level of significance and we conclude that the average packing cannot be taken to be 50 kgs.

Example 2:

A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?

Solution:

We are given

$n = 26;$ $\bar{x} = 147$dozens; $s = 16$

Null hypothesis:

H_0: $\mu = 140$ dozens i.e. Advertisement is not effective.

Alternative Hypothesis:

H_1: $\mu > 140$kgs (Right -tailed)

Calculation of statistic:

Under the null hypothesis H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \mu}{s / \sqrt{n-1}} \right|$$

$$= \left| \frac{147 - 140}{16 / \sqrt{25}} \right| = \frac{7 \times 5}{16} = 2.19$$

Expected value:

$t_e = \left| \frac{\bar{x} - \mu}{s / \sqrt{n-1}} \right|$ follows t-distribution with (26-1) = 25d.f

$= 1.708$

Inference:

Since $t_0 > t_e$, H_0 is rejected at 5% level of significance. Hence we conclude that advertisement is certainly effective in increasing the sales.

6.3 Test of significance for difference between two means:

6.3.1 Independent samples:

Suppose we want to test if two independent samples have been drawn from two normal populations having the same means, the population variances being equal. Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2, \ldots y_{n_2}$ be two independent random samples from the given normal populations.

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. the samples have been drawn from the normal populations with same means.

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ $(\mu_1 < \mu_2$ or $\mu_1 > \mu_2)$

Test statistic:

Under the H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

where $\bar{x} = \frac{\sum x}{n_1}$; $\bar{y} = \frac{\sum y}{n_2}$

and $S^2 = \frac{1}{n_1 + n_2 - 2} [\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2] = \frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}$

Expected value:

$$t_e = \left| \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

follows t-distribution with $n_1 + n_2 - 2$ d.f

Inference:

If the $t_0 < t_e$ we accept the null hypothesis. If $t_0 > t_e$ we reject the null hypothesis.

Example 3:

A group of 5 patients treated with medicine 'A' weigh 42, 39, 48, 60 and 41 kgs: Second group of 7 patients from the same hospital treated with medicine 'B' weigh 38, 42 , 56, 64, 68, 69 and 62 kgs. Do you agree with the claim that medicine 'B' increases the weight significantly?

Solution:

Let the weights (in kgs) of the patients treated with medicines A and B be denoted by variables X and Y respectively.

Null hypothesis:

$H_0 : \mu_1 = \mu_2$

i.e. There is no significant difference between the medicines A and B as regards their effect on increase in weight.

Alternative Hypothesis:

$H_1 : \mu_1 < \mu_2$ (left-tail) i.e. medicine B increases the weight significantly.

Level of significance : Let $\alpha = 0.05$

Computation of sample means and S.Ds

Medicine A

X	$x - \bar{x}$ $(\bar{x} = 46)$	$(x - \bar{x})^2$
42	– 4	16
39	–7	49
48	2	4
60	14	196
41	– 5	25
230	**0**	**290**

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{230}{5} = 46$$

Medicine B

Y	$y - \bar{y}$ $(\bar{y} = 57)$	$(y - \bar{y})^2$
38	–19	361
42	–15	225
56	–1	1
64	7	49
68	11	121
69	12	144
62	5	25
399	**0**	**926**

$$\bar{y} = \frac{\Sigma y}{n_2} = \frac{399}{7} = 57$$

$$S^2 = \frac{1}{n_1 + n_2 - 2} [\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2]$$

$$= \frac{1}{10} [290 + 926] = 121.6$$

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

$$= \left| \frac{46 - 57}{\sqrt{121.6\left(\frac{1}{5} + \frac{1}{7}\right)}} \right|$$

$$= \frac{11}{\sqrt{121.6 \times \frac{12}{35}}}$$

$$= \frac{11}{6.57} = 1.7$$

Expected value:

$t_e = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$ follows t-distribution with (5+7–2) =10 d.f

= 1.812

Inference:

Since $t_0 < t_e$ it is not significant. Hence H_0 is accepted and we conclude that the medicines A and B do not differ significantly as regards their effect on increase in weight.

Example 4:

Two types of batteries are tested for their length of life and the following data are obtained:

	No of samples	Mean life (in hrs)	Variance
Type A	9	600	121
Type B	8	640	144

Is there a significant difference in the two means?

Solution:

We are given

$n_1=9$; $\bar{x}_1=600$hrs; $s_1^2=121$; $n_2=8$; $\bar{x}_2=640$hrs; $s_2^2=144$

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. Two types of batteries A and B are identical i.e. there is no significant difference between two types of batteries.

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ (Two- tailed)

Level of Significance:

Let $\alpha = 5\%$

Calculation of statistics:

Under H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

$$\text{where } S^2 = \frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}$$

$$= \frac{9 \times 121 + 8 \times 144}{9 + 8 - 2}$$

$$= \frac{2241}{15} = 149.4$$

$$\therefore \quad t_0 = \left| \frac{600 - 640}{\sqrt{149.4\left(\frac{1}{9} + \frac{1}{8}\right)}} \right|$$

$$= \frac{40}{\sqrt{149.4\left(\frac{17}{72}\right)}} = \frac{40}{5.9391} = 6.735$$

Expected value:

$$t_e = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$ follows t-distribution with 9+8–2 =15 d.f

$= 2.131$

Inference:

Since $t_0 > t_e$ it is highly significant. Hence H_0 is rejected and we conclude that the two types of batteries differ significantly as regards their length of life.

6.3.2 Related samples –Paired t-test:

In the t-test for difference of means, the two samples were independent of each other. Let us now take a particular situations where

(i) The sample sizes are equal; i.e., $n_1 = n_2 = n$(say), and

(ii) The sample observations $(x_1, x_2, \ldots\ldots x_n)$ and $(y_1, y_2, \ldots\ldots y_n)$ are not completely independent but they are dependent in pairs.

That is we are making two observations one before treatment and another after the treatment on the same individual. For example a business concern wants to find if a particular media of promoting sales of a product, say door to door canvassing or advertisement in papers or through T.V. is really effective. Similarly a pharmaceutical company wants to test the efficiency of a particular drug, say for inducing sleep after the drug is given. For testing of such claims gives rise to situations in (i) and (ii) above, we apply paired t-test.

Paired – t –test:

Let $d_i = X_i - Y_i$ $(i = 1, 2, \ldots n)$ denote the difference in the observations for the i^{th} unit.

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ ie the increments are just by chance

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ $(\mu_1 > \mu_2$ (or) $\mu_1 < \mu_2)$

Calculation of test statistic:

$$t_0 = \left| \frac{\bar{d}}{S/\sqrt{n}} \right|$$

where $\bar{d} = \frac{\sum d}{n}$ and $S^2 = \frac{1}{n-1}\sum (d - \bar{d})^2 = \frac{1}{n-1}[\sum d^2 - \frac{(\sum d)^2}{n}]$

Expected value:

$$t_e = \left| \frac{\bar{d}}{S/\sqrt{n}} \right| \text{ follows t-distribution with } n-1 \text{ d.f}$$

Inference:

By comparing t_0 and t_e at the desired level of significance, usually 5% or 1%, we reject or accept the null hypothesis.

Example 5:

To test the desirability of a certain modification in typists desks, 9 typists were given two tests of as nearly as possible the same nature, one on the desk in use and the other on the new type. The following difference in the number of words typed per minute were recorded:

Typists	A	B	C	D	E	F	G	H	I
Increase in number of words	2	4	0	3	–1	4	–3	2	5

Do the data indicate the modification in desk promotes speed in typing?

Solution:

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. the modification in desk does not promote speed in typing.

Alternative Hypothesis:

$H_1 : \mu_1 < \mu_2$ (Left tailed test)

Level of significance: Let $\alpha = 0.05$

Typist	d	d^2
A	2	4
B	4	16
C	0	0
D	3	9
E	–1	1
F	4	16
G	–3	9
H	2	4
I	5	25
	$\Sigma d = 16$	$\Sigma d^2 = 84$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{16}{9} = 1.778$$

$$S = \sqrt{\frac{1}{n-1}[\Sigma d^2 - \frac{(\Sigma d)^2}{n}]}$$

$$= \sqrt{\frac{1}{8}[84 - \frac{(16)^2}{9}]} = \sqrt{6.9} = 2.635$$

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \left|\frac{\bar{d}.\sqrt{n}}{S}\right| = \frac{1.778 \times 3}{2.635} = 2.024$$

Expected value:

$t_e = \left|\frac{\bar{d}.\sqrt{n}}{S}\right|$ follows t- distribution with $9 - 1 = 8$ d.f

$= 1.860$

Inference:

When $t_0 < t_e$ the null hypothesis is accepted. The data does not indicate that the modification in desk promotes speed in typing.

Example 6:

An IQ test was administered to 5 persons before and after they were trained. The results are given below:

Candidates	I	II	III	IV	V
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

Test whether there is any change in IQ after the training programme (test at 1% level of significance)

Solution:

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. there is no significant change in IQ after the training programme.

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ (two tailed test)

Level of significance :

$\alpha = 0.01$

x	110	120	123	132	125	Total
y	120	118	125	136	121	-
d = x–y	–10	2	–2	–4	4	–10
d^2	100	4	4	16	16	140

$$\bar{d} = \frac{\sum d}{n} = \frac{-10}{5} = -2$$

$$S^2 = \frac{1}{n-1}[\sum d^2 - \frac{(\sum d)^2}{n}]$$

$$= \frac{1}{4}[140 - \frac{100}{5}] = 30$$

Calculation of Statistic:

Under H_0 the test statistic is

$$t_0 = \left|\frac{\bar{d}}{S/\sqrt{n}}\right|$$

$$= \left|\frac{-2}{\sqrt{30/5}}\right|$$

$$= \frac{2}{2.45}$$

$$= 0.816$$

Expected value:

$$t_e = \left|\frac{\bar{d}}{\sqrt{S^2/n}}\right|$$ follows t-distribution with $5 - 1 = 4$ d.f

$$= 4.604$$

Inference:

Since $t_0 < t_e$ at 1% level of significance we accept the null hypothesis. We therefore, conclude that there is no change in IQ after the training programme.

6.4 Chi square statistic:

Various tests of significance described previously have mostly applicable to only quantitative data and usually to the data which are approximately normally distributed. It may also happens that we may have data which are not normally distributed. Therefore there arises a need for other methods which are more appropriate for studying the differences between the expected and observed frequencies. The other method is called Non-parametric or distribution free test. A non- parametric test may be defined as a statistical test in which no hypothesis is made about specific values of parameters. Such non-parametric test has assumed great importance in statistical analysis because it is easy to compute.

6.4.1 Definition:

The Chi- square (χ^2) test (Chi-pronounced as ki) is one of the simplest and most widely used non-parametric tests in statistical work. The χ^2 test was first used by Karl Pearson in the year 1900. The quantity χ^2 describes the magnitude of the discrepancy between theory and observation. It is defined as

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(Oi - Ei)^2}{Ei}\right]$$

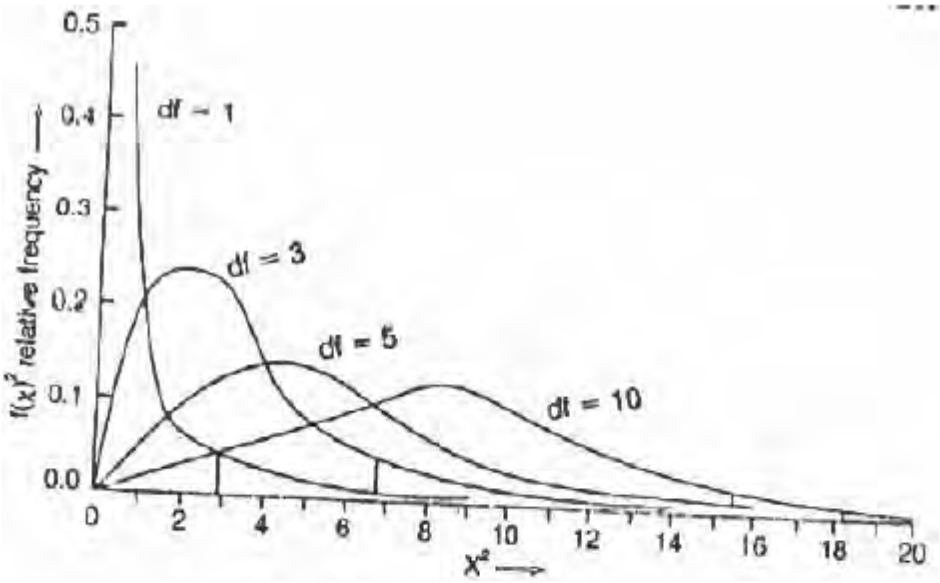
Where O refers to the observed frequencies and E refers to the expected frequencies.

Note:

If χ^2 is zero, it means that the observed and expected frequencies coincide with each other. The greater the discrepancy between the observed and expected frequencies the greater is the value of χ^2.

Chi square - Distribution:

The square of a standard normal variate is a Chi-square variate with 1 degree of freedom i.e., If X is normally distributed with mean μ and standard deviation σ, then $\left(\frac{x-\mu}{\sigma}\right)^2$ is a Chi-square variate (χ^2) with 1 d.f. The distribution of Chi-square depends on the degrees of freedom. There is a different distribution for each number of degrees of freedom.



6.4.2 properties of Chi-square distribution:

1. The Mean of χ^2 distribution is equal to the number of degrees of freedom (n)
2. The variance of χ^2 distribution is equal to 2n
3. The median of χ^2 distribution divides, the area of the curve into two equal parts, each part being 0.5.
4. The mode of χ^2 distribution is equal to (n–2)
5. Since Chi-square values always positive, the Chi square curve is always positively skewed.
6. Since Chi-square values increase with the increase in the degrees of freedom, there is a new Chi-square distribution with every increase in the number of degrees of freedom.
7. The lowest value of Chi-square is zero and the highest value is infinity ie $\chi^2 \geq 0$.
8. When Two Chi- squares ${\chi_1}^2$ and ${\chi_2}^2$ are independent χ^2 distribution with n_1 and n_2 degrees of freedom and their sum ${\chi_1}^2 + {\chi_2}^2$ will follow χ^2 distribution with $(n_1 + n_2)$ degrees of freedom.
9. When n (d.f) > 30, the distribution of $\sqrt{2\chi^2}$ approximately follows normal distribution. The mean of the distribution $\sqrt{2\chi^2}$ is $\sqrt{2n-1}$ and the standard deviation is equal to 1.

6.4.3 Conditions for applying χ^2 test:

The following conditions should be satisfied before applying χ^2 test.

1. N, the total frequency should be reasonably large, say greater than 50.
2. No theoretical cell-frequency should be less than 5. If it is less than 5, the frequencies should be pooled together in order to make it 5 or more than 5.
3. Each of the observations which makes up the sample for this test must be independent of each other.
4. χ^2 test is wholly dependent on degrees of freedom.

6.5 Testing the Goodness of fit (Binomial and Poisson Distribution):

Karl Pearson in 1900, developed a test for testing the significance of the discrepancy between experimental values and the theoretical values obtained under some theory or hypothesis. This test is known as χ2-test of goodness of fit and is used to test if the deviation between observation (experiment) and theory may be attributed to chance or if it is really due to the inadequacy of the theory to fit the observed data.

Under the null hypothesis that there is no significant difference between the observed and the theoretical values. Karl Pearson proved that the statistic

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(Oi - Ei)^2}{Ei}\right]$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \ldots\ldots\frac{(O_n - E_n)^2}{E_n}$$

Follows χ^2-distribution with $\nu = n - k - 1$ d.f. where 0_1, 0_2, ...0_n are the observed frequencies, E_1 , E_2. E_n, corresponding to the expected frequencies and k is the number of parameters to be estimated from the given data. A test is done by comparing the computed value with the table value of χ^2 for the desired degrees of freedom.

Example 7:

Four coins are tossed simultaneously and the number of heads occurring at each throw was noted. This was repeated 240 times with the following results.

No. of heads	0	1	2	3	4
No. of throws	13	64	85	58	20

Fit a Binomial distribution assuming under the hypothesis that the coins are unbiased.

Solution:

Null Hypothesis:

H_0: The given data fits the Binomial distribution. i.e the coins are unbiased.

$p = q = 1/2$ $\quad$ $n = 4$ $\quad$ $N = 240$

Computation of expected frequencies:

No. of heads	$P(X= x) = 4\,C_x\,p^x\,q^{n-x}$	Expected Frequency N. P(X = x)
0	$4Co\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \left(\frac{1}{16}\right)$	$\left(\frac{1}{16}\right)$ x 240 = 15
1	$4C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \left(\frac{4}{16}\right)$	$\left(\frac{4}{16}\right)$ x 240 = 60
2	$4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \left(\frac{6}{16}\right)$	$\left(\frac{6}{16}\right)$ x 240 = 90
3	$4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \left(\frac{4}{16}\right)$	$\left(\frac{4}{16}\right)$ x 240 = 60
4	$4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \left(\frac{1}{16}\right)$	$\left(\frac{1}{16}\right)$ x 240 = 15
		240

Computation of chi square values

Observed Frequency O	Expected Frequency E	$(O-E)^2$	$\left(\frac{(O-E)^2}{E}\right)$
13	15	4	0.27
64	60	16	0.27
85	90	25	0.28
58	60	4	0.07
20	15	25	1.67
			2.56

$$\chi_0^2 = \Sigma\left(\frac{(O-E)^2}{E}\right) = 2.56$$

Expected Value:

$\chi_e^2 = \Sigma\left(\frac{(O-E)^2}{E}\right)$ follows χ^2-distribution with (n–k–1) d.f.

(Here k = 0, since no parameter is estimated from the data)

= 9.488 for $\nu = 5-1 = 4$ d.f.

Inference:

Since $\chi_0^2 < \chi_e^2$ we accept our null hypothesis at 5% level of significance and say that the given data fits Binomial distribution.

Example 8:

The following table shows the distribution of goals in a foot ball match.

No. of goals	0	1	2	3	4	5	6	7
No. of matches	95	158	108	63	40	9	5	2

Fit a Poisson distribution and test the goodness of fit.

Solution:

Null Hypothesis :

The given data fits the Poisson distribution.

Level of significance :

Let $\alpha = 0.05$

Computation of expected frequencies:

$$m = \frac{812}{480} = 1.7$$

$$P(0) = e^{-1.7}\frac{(1.7)^0}{0!} = 0.183.$$

$$f(o) = N.P(o) = 480 \times 0.183 = 87.84$$

The other expected frequencies will be obtained by using the recurrence formula

$$f(x+1) = \frac{m}{x+1} \times f(x)$$

Putting x = 0, 1, 2, ... we obtain the following frequencies.

$$f(1) = 1.7 \times 87.84$$
$$= 149.328$$

$$f(2) = \frac{1.7}{2} \times 149.328$$
$$= 126.93$$

$$f(3) = \frac{1.7}{3} \times 126.93$$
$$= 71.927$$

$$f(4) = \frac{1.7}{4} \times 71.927$$
$$= 30.569$$

$$f(5) = \frac{1.7}{5} \times 30.569$$
$$= 10.393$$

$$f(6) = \frac{1.7}{6} \times 10.393$$
$$= 2.94$$

$$f(7) = \frac{1.7}{7} \times 2.94$$
$$= 0.719$$

No. of goals	0	1	2	3	4	5	6	7	Total
Expected frequency	88	149	127	72	30	10	3	1	480

Computation of statistic:

Observed Frequency O	Expected Frequency E	$(O-E)^2$	$\left(\frac{(O-E)^2}{E}\right)$
95	88	49	0.56
158	150	64	0.43
108	126	324	2.57
63	72	81	1.13
40	30	100	3.33
9 5 2 } 16	10 3 1 } 14	4	0.29
			8.31

$$\chi_o^2 = \Sigma \left(\frac{(O-E)^2}{E}\right) = 8.31$$

Expected Value:

$$\chi_e^2 = \Sigma \left(\frac{(O-E)^2}{E}\right)$$ χ^2 -distribution with $(n-k-1)$ d.f

$= 9.488$ for $\nu = 6-1-1 = 4$ d.f.

Inference:

Since $\chi_0^2 < \chi_e^2$, we accept our null hypothesis at 5% level of significance and say that the given data fits the Poisson distribution.

6.6 Test of independence

Let us suppose that the given population consisting of N items is divided into r mutually disjoint (exclusive) and exhaustive classes $A_1, A_2, \ldots A_r$ with respect to the attribute A so that randomly selected item belongs to one and only one of the attributes $A_1, A_2, \ldots A_r$ Similarly let us suppose that the same population is divided into c mutually disjoint and exhaustive classes $B_1, B_2, \ldots B_c$ w.r.t another attribute B so that an item selected at random possess one and only one of the attributes $B_1, B_2, \ldots B_c$. The frequency distribution of the items belonging to

the classes $A_1, A_2, \ldots A_r$ and $B_1, B_2, \ldots B_c$ can be represented in the following r × c manifold contingency table.

r × c manifold contingency table

B / A	B_1	B_2	…	B_j	…	B_c	Total
A_1	(A_1B_1)	(A_1B_2)	…	(A_1B_j)	…	(A_1B_c)	(A_1)
A_2	(A_2B_1)	(A_2B_2)	…	(A_2B_j)	…	(A_2B_c)	(A_2)
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_i	(A_iB_1)	(A_iB_2)	…	(A_iB_j)	…	(A_iB_c)	(A_i)
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_r	(A_rB_1)	(A_rB_2)	…	(A_rB_j)	…	(A_rB_c)	(A_r)
Total	(B_1)	(B_2)	…	(B_j)	…	(B_c)	$\Sigma Ai = \Sigma Bj = N$

(A_i) is the number of persons possessing the attribute A_i, (i=1,2,..r), (Bj) is the number of persons possing the attribute B_j,(j=1,2,3,...c) and $(A_i\ B_j)$ is the number of persons possessing both the attributes A_i and B_j (i=1,2,..r. ,j=1,2,..c).

Also $\Sigma A_i = \Sigma B_j = N$

Under the null hypothesis that the two attributes A and B are independent, the expected frequency for (A_iB_j) is given by

$$= \frac{(Ai)(Bj)}{N}$$

Calculation of statistic:

Thus the under null hypothesis of the independence of attributes,the expected frequencies for each of the cell frequencies of the above table can be obtained on using the formula

$$\chi_0^2 = \Sigma\left(\frac{(O_i - E_i)^2}{E_i}\right)$$

Expected value:

$\chi_e^2 = \Sigma \left(\frac{(O_i - E_i)^2}{E_i} \right)$ follows χ^2-distribution with (r−1) (c−1) d.f

Inference:

Now comparing χ_0^2 with χ_e^2 at certain level of significance ,we reject or accept the null hypothesis accordingly at that level of significance.

6.6.1 2×2 contingency table :

Under the null hypothesis of independence of attributes, the value of χ^2 for the 2×2 contingency table

			Total
	a	b	a+b
	c	d	c+d
Total	a+c	b+d	N

is given by

$$\chi_o^2 = \frac{N(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

Example 9:

1000 students at college level were graded according to their I.Q. and the economic conditions of their homes. Use χ^2 test to find out whether there is any association between economic condition at home and I.Q.

Economic Conditions	IQ		Total
	High	Low	
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

Solution:

Null Hypothesis:

There is no association between economic condition at home and I.Q. i.e. they are independent.

$$E_{11} = \frac{(A)(B)}{N} = \frac{600 \times 700}{1000} = 420$$

The table of expected frequencies shall be as follows.

			Total
	420	180	600
	280	120	400
Total	700	300	1000

Observed Frequency O	Expected Frequency E	$(O - E)^2$	$\left(\frac{(O - E)^2}{E}\right)$
460	420	1600	3.81
240	280	1600	5.714
140	180	1600	8.889
160	120	1600	13.333
			31.746

$$\chi_o^2 = \Sigma \left(\frac{(O - E)^2}{E}\right) = 31.746$$

Expected Value:

$\chi_e^2 = \Sigma\left(\frac{(O - E)^2}{E}\right)$ follow χ^2 distribution with (2–1) (2–1) = 1 d.f

= 3.84

Inference :

$\chi_o^2 > \chi_e^2$, hence the hypothesis is rejected at 5 % level of significance. ∴ there is association between economic condition at home and I.Q.

Example 10:

Out of a sample of 120 persons in a village, 76 persons were administered a new drug for preventing influenza and out of them, 24 persons were attacked by influenza. Out of those who were not administered the new drug ,12 persons were not affected by influenza.. Prepare

(i) 2x2 table showing actual frequencies.

(ii) Use chi-square test for finding out whether the new drug is effective or not.

Solution:

The above data can be arranged in the following 2 x 2 contingency table.

Table of observed frequencies

New drug	Effect of Influenza		Total
	Attacked	Not attacked	
Administered	24	76 – 24 = 52	76
Not administered	44 –12 = 32	12	120 – 76 = 44
Total	120 – 64 = 56 24 + 32 = 56	52 + 12 = 64	120

Null hypothesis:

'Attack of influenza' and the administration of the new drug are independent.

Computation of statistic:

$$\chi_o^2 = \frac{N(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

$$= \frac{120(24 \times 12 - 52 \times 32)^2}{56 \times 64 \times 76 \times 44}$$

$$= \frac{120(-1376)^2}{56 \times 64 \times 76 \times 44} = \frac{120(1376)^2}{56 \times 64 \times 76 \times 44}$$

= Anti log [log 120 + 2log1376 –(log56 +log64+log76+log44)]

= Antilog (1.2777) = 18.95

Expected value:

$\chi_e^2 = \Sigma \left(\frac{(O - E)^2}{E} \right)$ follows χ^2 distribution with (2–1) × (2–1) d.f

= 3.84

Inference:

Since $\chi_o^2 > \chi_e^2$, H_0 is rejected at 5 % level of significance. Hence we conclude that the new drug is definitely effective in controlling (preventing) the disease (influenza).

Example 11:

Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence levels. The results are as follows

Researchers	No. of Students				Total
	Below average	Average	Above average	Genius	
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

Would you say that the sampling techniques adopted by the two researchers are independent?

Solution:

Null Hypothesis:

The sampling techniques adopted by the two researchers are independent.

$$E(86) = \frac{126 \times 200}{300} = 84$$

$$E(60) = \frac{93 \times 200}{300} = 62$$

$$E(44) = \frac{69 \times 200}{300} = 46$$

The table of expected frequencies is given below.

	Below average	Average	Above average	Genius	Total
X	84	62	46	200–192 = 8	200
Y	126 – 84 = 42	93 – 62 = 31	69 – 46 = 23	12 – 8 = 4	100
Total	126	93	69	12	300

Computation of chi-square statistic:

Observed Frequency O	Expected Frequency E	(O – E)	$(O - E)^2$	$\left(\frac{(O - E)^2}{E}\right)$
86	84	2	4	0.048
60	62	–2	4	0.064
44	46	– 2	4	0.087
10	8	2	4	0.500
40	42	–2	4	0.095
33	31	2	4	0.129
$\begin{bmatrix}25\\2\end{bmatrix}$ 27	$\begin{bmatrix}23\\4\end{bmatrix}$ 27	0	0	0
300	**300**	**0**		**0.923**

$$\chi_o^2 = \Sigma \left(\frac{(O - E)^2}{E} \right) = 0.923$$

Expected value:

$\chi_e^2 = \Sigma \left(\frac{(O - E)^2}{E} \right)$ follows χ^2 distribution with (4–1)(2–1)

$= 3 - 1 = 2$ df

$= 5.991$

Inference:

Since $\chi_o^2 < \chi_e^2$, we accept the null hypothesis at 5 % level of significance. Hence we conclude that the sampling techniques by the two investigators, do not differ significantly.

Sl.No.	Questions	opt1	opt2
1	β-sheets are stabilized by	hydrophobic bonds	ionic bonds
2	The 21st amino acid is	hydroxyl proline	selenocysteine
3	All of the below mentioned amino acids can participate in hydrogen bonding except one	Serine	Cysteine
4	Which would be best to separate a protein that binds strongly to its substrate?	Gel filtration	Affinity chromatography
5	In nuclear research …… is used to determine uranium in salts.	colorimeter	spectrophotometer
6	On oxidation of thiamine (vitamin B1) it forms thiochrome which emits----------	purple	blue
7	Both qualitative and quantitative analysis of sample can be done using …….	NMR	fluorimeter
8	Based on Lambert's law the amount of light absorbed is directly proportional to ----------	length of the medium	concentration of the su
9	When the number of light absorbing molecules increases in the medium, the intensity of light	decreased	increased
10	Beer Lambert's law is the principle followed in ……..	MS	spectrophotometer
11	A monochromator consists of ----------	grating	prism
12	A quartz cuvette will have a optical path of ……	1cm	0.5cm
14	Half silvered mirror is used in---------- instrument	double beam spectrosco	single beam spectrosc
15	Light energy is converted to electrical energy by …….	monochromator	photomultiplier tube
16	In spectrophotometer, after passing through cuvette the transmitted light will fall on ……	photomultiplier tube	monochromator
17	In electromagnetic spectrum ……. will have the higher wavelength.	X-rays	visible
18	If a sample absorbs all wavelengths in the visible region of the spectrum, it will appear--------	blue	white
19	The color we see in a sample of solution is due to ……	adsorption	absorption
20	If a sample does not absorbs any wavelengths in the visible region of the spectrum, it will ap	colourless	white
21	A colorimeter will contain ----------	thermosensor	filter
22	In a colorimeter monochromatic light is produced by	filter	photo cell
23	In a photomultiplier tube electrons produced is amplified by …….	amplifier	cathode
24	Estimation of cadmium is done by ……..	fluorimeter	colorimeter
25	The wavelength of visible spectrum of light ranges between	600-720nm	576-580nm
26	Prisms which are made up of quartz is for----------	gamma rays	X-rays
27	The condensing lens renders light rays into ……. beam before it falls on monochromat	perpendicular	parallel
28	Photo multiplier tube consist of ……. dynodes	4	6

29	The light source of fluorimeter is …….. lamp	sodium	tungsten
30	In fluorimeter light from sample pass through ……. before PMT	primary filter	condensing lens
31	Grating is superior to prism because of …… of the spectrum	monochromatic light	linear resolution
32	Monochromatic light consists of---------- wave length	single	linear
33	The negative logarithm of transmittance with inverse relationship is ……	concentration	adsorption
34	An instrument which separates electromagnetic radiation into wavelengths and selectively m	spectrophotometer	fluorimeter
35	Grating which comprises numerous equi-distant parallel lines ruled on a plane surface is …..	refraction	diffraction
36	Light that cannot be separated into components is …….	monocromatic light	linear resolution
37	A spectrophotometer with two dispersing elements is called ……..	colorimeter	single beam spectroph
38	The spectrum can be reunited to give the original white light by focusing the components ba	reversed prism	grating
39	The wide range of wave length that the light source capable to produce is ……	polychromatic	intensity
40	On heating sodium metal emits …… colour.	blue	yellow
41	In an electromagnetic spectrum---------- have less wave length	microwave	infra red
42	When a beam of light is incident on certain substance they emit visible light which is called	luminescence	fluorescence
44	The light coming out of tungsten lamp will contain ……	polychromatic	monochromatic
45	When the length of the medium is increased, then the optical density of the solution ……	decreases	increases
46	In the equation a=E x c x l, E stands for	exotic coefficient	extinction coefficient
47	**Unit 2**		
48	Centrifuge produce a strong ___	gravitational force	centripetal force
49	The rate of sedimentation depends upon the	relative centrifugal force	weight of particle
50	In centrifuge, centrifugal force is directed	radialy outwards	radially inwards
51	Rotor speed is expressed in terms of	km/hr	rpm
52	The value of earth's gravitational force 'g' is	98.1cm sec ^{2}	981cm sec ^{3}
53	The relative centrifugal force is commonly reffered as	number of rotation	rpm
54	Relative centrifugal force depends upon the	rpm and angle of rotatio	rpm and radius of rotat
55	Relative centrifugal force is mathematically expressed as	RCF=(1.118x10 ^{3})(rev. /	RCF=(1.118x10)(rev
56	The angular velocity of the centrifuge is mathematically expressed as	(ω=2л rev. min ^{1})/60	ω=2л rev. min ^{2}/60
57	If the density of the particle and medium is equal, then the sedimentation rate becomes	maximum	zero
58	When a centrifugal force field increases, then the sedimentation rate	decreases	increases
59	When the viscosity of the medium increases, then the sedimentation rate	increases	decreases
60	Sedimentation coefficient is also called as	Einstein Unit	Centrifuge Unit

61	The sedimentation constant temperature of the water is	37˚C	42˚C
62	The basic unt of sedimentation coefficient is	1x10 ^{13}min	1x10 ^{12}sec
63	If the particle size is larger, then the sedimentation is	slower	faster
64	The range of Svedberg unit of viruses are	50S to 1,000S	45S to 1,200S
65	The range of Svedberg unit of lysosomes is	40,000S	60,000S
66	The range of Svedberg unit of mitochondira are	20x10^{2}S to 70x10^{3}S	20x10^{3}S to 70x10^{3}S
67	The purity of a solute collected between two times t1 and t2 during chromatographic separation is	Amount of solute eluted - amount of impurity	Amount of solute eluted - amount of
67	Hand centrifuge consist of _____ tube holders	three	two
68	The maximum speed of clinical centrifuge is about	3,000 rev. min	6,000 rev. min ^{1}
69	In a clinical centrifuge, rotors are mounted on a rigid shaft	horizontally	vertically
70	In a clinical centrifuge, the centrifugal tubes must be placed ____ to each other	adjacent	alternative
71	The maximum centrifugal field of large capacity refrigerated centrifuge is	6,000g -	950g
72	The maximum speed of large capacity refrigerated centrifuge is	3,000rev. min ^{1}	6,000 rev. min ^{1}
73	The maximum speed of high speed refrigerated centrifuge are	20,000 rev. min ^{1}	25,000rev. min ^{1}
74	The relative centrifugal field of high speed refrigerated centrifuge is about	60,000g	30,000g
75	The long and tubular rotor is present in	Small bench centrifuge	clinical centrifuge
76	Ultra centrifuges are broadly classified into	Three types	Two types
77	The maximum speed of preparative Ultra Centrifuge is	60,000 rev. min ^{1}	70,000 rev. min ^{1}
78	The Preparative Ultra Centrifuge produce a relative centrifugal field of about	50,000g	25,000g
79	For the safety reasons, the rotor chambers of both high speed and ultra centrifuges are enclos	a metal cover	Heavy Armour platting
80	An air driven, table top Preparative Ultra Centrifuge is called	Air outlet	Air driver
81	The diameter of the rotor for the Airfuge is	8.1cm	6.9cm
82	The rotor speed of the airfuge is	1,00,000rev. min	1,00,000 rev. min ^{1}
83	The relative centrifugal field of airfuge is about	1,00,000g	70,000g
84	The rotor chamber is refrigerated and sealed in	large capacity refrigerate	High speed refrigerate
85	In an Ultra Centrifuge cell, the optical system for recording the distribution of the sample is p	Analytical Ultra Centrifu	High Speed Refrigerate
86	In Analytical Ultra Centrifuge, tip of the rotor contains	Thermistor	Thermometer
87	In Analytical Ultra Centrifuge, the rotor chamber contains an upper and lower____ lens respe	Condensing and Diverging	Condensing and Collin
88	The rotor of an analytical ultra centrifuge contains	Three cells	Two Cells
89	In Ultra centrifuge, there are ___ types of optical systems available	Three	Two

90	Double sector cell is present in the	Ulta violet light absorpti	Schlieren optical syste
91	Plotting of refractive index gradient against the distance along the analytical cell for sedimen	Ultra violet light absorpt	Infra red light absorpti
92	In an Ultra violet absorption system of the analytical centrifuge, the intensity of light transmi	Photo electric plate	Photographic plate
93	Rotors are made of alloys of aluminium and	Platinum	Uranium
94	The vertical tube rotor is a	movable zero angle roto	vertically movable rotc
95	In a fixed angle rotor, the angle of the tubes in the holes are between	14˚ to 40˚	10˚ to 14˚
96	The type of rotor having the bucket is	Vertical tube rotor	Fixed angle rotor
97	The swinging bucket rotors are perpendicular to its	Direction of rotation	Angle of rotation
98	The swinging bucket rotors are parallel to its	gravitational force	Centrifugal force
99	Zonal rotors are classified into	Three types	Two types
100	Zonal rotors are specially designed to minimize the	Heat producing	Wastage of sample
101	The recesses to hold a single conical separation chamber is present in the	Zonal Rotors	Elutriator Rotors
102	Based on the purpose, the centrifugation is classified into	Three types	Two types
103	Optical method is used in	Preparative Ultra Centrif	Analytical Ultra Centri
104	The relative molecular mass can be determined by	Sedimentation Homogen	Sedimentation equilibr
105	The process of separation of cell organelles is called as	Sub cellular fractionatio	Cellular disintegration
106	After homogenising, the cell suspension containing many intact organelles is known as	Supernautant	Homogenate
107	The purity of organelles obtained by differential centrifugation is determined by	Marker	Pellet
108	In differential centrifugation, the marker for Plasma membrane is	Lactate Dehydrogenase	5' Nucleotidase
109	In differential centrifugation, the marker for Nucleus is	DNA	Nucleotidase
110	In differential centrifugation, the marker for Mitochondrion is	Glutamate dehydrogenas	Glucose 6 Phosphate
111	In differential centrifugation, the marker for Lysosome is	Glucose 6 Phosphate	Nucleotidase
112	In differential centrifugation, the marker for Endoplasmic reticulum is	Acid Phosphotase	Glucose 6 Phosphate
113	In differential centrifugation, the marker for Cytosol is	DNA	Lactate Dehydrogenas
114	Isopycnic centrifugation depends upon the	shape of the particle	buoyant density of the
115	The separation of DNA-RNA Hybrids can be done by	Rate zonal centrifugatior	Isopycnic centrifugatio
116	The gradient material used for fracionation of Lipoproteins is	Caesium bromide	Caesium sulphate
117	The gradient material used for fracionation of viruses and whole cell is	Ficoll	Caesium bromide
118	The gradient material used forfracionation of bonding membrane fragments and protein is	Caesium bromide	Glycerol
119	Bovine serum albumin is a gradient material used in	Fractionation of Lipopro	Separation of whole ce
120	The gradient material used for the purification of Proteoglycans is	Glyserol	Caesium sulphate

121	The gradient material used for bonding of microsome is	Sodium iodide	Bovine serum albumin
122	Citric acid cycle and releasing of ammonia for urea formation takes place in	Lysosome	Nucleus
123	Purification of proteins can be done by ……. Chromatography.	ion-exchange	affinity
124	Series of symmetric peaks in chromatography is	chromatogram	spectrum
125	The Rf value is always	more than 2	more than 1
126	Silica gel is the stationary phase in	PAGE	GLC
127	The stationary phase used in TLC is	cellulose	silica gel
128	Stationary phase used in paper chromatography is	filter paper	silica gel
129	Forces involved in paper chromatography is	capillary forces	van der Waals forces
130	Stationary phase used in TLC is for separation of plant pigments is	cellulose	silica gel
131	Solvent system of amino acids in paper chromatography is	petroleum ether	acetic acid
132	Molecules with higher solubility will migrate to	lesser than Rf	greater than Rf
133	Usually low moleculalr weight compounds are separated using …….. chromatography	partition	adsoprtion
134	Impurities present in paper are removed by washin with	0.1 N HCl	1 N HCl
135	In paper chromatography, amino acids are viewed in purple or blue by spraying	ninhydrin	bromine water
136	The stationary phase used in column chromatography is known as	gradient	filter paper
137	Gel matrix cellulose has ……. units.	β-1,4 linked fucose	β-1,4 linked galactose
138	Gel matrix dextran has ……. units.	α-1,6 linked galact ose	α-1,6 linked gluc ose
139	Gel matrix agarose has ……. units.	D- galact ose	D-glucose
140	The stationary phase silica is made up of	sulphuric acid	acetic acid
141	The peaks obtained during column chromatography is	EEG	electrophoretogram
142	Ion exchange chromatography is as process based on	neutrally charged	oppositely charged par
143	When a gel matrix exchanges positive ions, it is called as	cation exchanger	anion exchanger
144	Example of strong cationic exchanger is ……..	epoxyamine	cellulose
145	Removal of sample from solid matrix using solvent is ……...	bed volume	effluent
146	In chromatography, volume of mobile phase is	bed volume	void volume
147	Time taken for each material to emerge from coumn is ……	bed volume	void volume
148	Column development using single solvent as mobile phase is ……	bed volume	void volume
149	Column chromatography involves ……. phenomenon.	3	2
150	Adsorption chromatography was developed by …….	D.T. Day	Sorensen
151	Adsorption chromatography is used mainly for separation of ……..	clinical samples	soil samples

152	Scientit who used adsorption chromatography for separating plant pigment is …….	M.S.Tswett	Sorensen
153	Powdered charcoal can be prepared using ……..	coal	nuts
154	Fuller's earth is mixture of ……. from clay deposits	vitamins	minerals
155	Hydroxyapatite is ……	MgCl2	KCL
156	Adsorption chromatography is used for separation of ……	chlorides	proteins
157	Commerical name of strong cationic exchanger is ……	AG 3	Sephadex AG 50
158	Commerical name of weak cationic exchanger is ……	AG 3	Sephadex AG 50
159	Commerical name of strong anionic exchanger is ……	AG 3	Sephadex AG 50
160	Commerical name of weak anionic exchanger is ……	AG 3	Sephadex AG 50
161	To separate metallic ions ……. exchangers are used.	anionic	cationic
162	Exchangers used for separation of proteins / polysaccharides is	Dowex 50	CM-Sephadex
163	Resin used to prepare deionized water is …….	Mixed-bed resin	QAE-Sephadex
164	Matrix used in affinity chromatography is ……...	QAE-Sephadex	Bio-Rex 70
165	Matrix used in affinity chromatography is ……...	QAE-Sephadex	Bio-Rex 70
166	Matrix used in affinity chromatography is ……...	Bio-Rex 70	Sephacryl S
167	in affinity chromatography, gel is linked with arms called ……..	ligands	matrix
168	For isolation of lipoprotein …… serves as ligand.	NADP	NAD
169	For isolation of immunoglobulins …… serves as ligand.	NADP	NAD
170	For isolation of biotin containing enzymes …… serves as ligand.	heparin	avidin
171	For isolation of coagulation factors …… serves as ligand.	cibacron blue	heparin
172	High Performance Liquid Chromatography is …….	HPTLC	TLC
173	**Unit 3**		
174	Electrophoresis is based on	solubility	molecular mass
175	Electrophoretic movement of particles can be influenced by the following factor:	density	electrical charge
176	Forensic science involves …….	paper chromatography	GLC
177	This cannot be used in gel electrophoresis	starch	agar
178	Electrophoresis involves migration of …… molecules.	neutral	both charged
179	Buffer not used in electrophoresis is	calcium	citrate
180	In electrophoresis use of cellulose acetate paper was introduced in	1959	1960
181	Migration of charge particles is called as	GC	TLC
182	Molecular weight can be determined by ……	immunoelectrophoresis	SDS-PAGE

183	Principle of electrophoresis is based on ……..	charged ions	solar energy
184	Better resolution is obtained in cellulose acetate thatn paper because	less hydrophobic	more hydrophilic
185	Paper used in electrophoresis is made up of	mannose	fucose
186	When serum is subjected to electrophoresis, the fastest moving fraction is	albumin	alpha globulin
187	In PAGE, movement of protein depends on …….. molecule.	charge	size
188	During electrophoresis of proteins in an alkaline medium, they	act as anions and move t	act as cations and mov
189	Electrophoresis using acrylamide gel is known as ……..	TLC	chromatography
190	Polymerization of acrylamide to polyacrylamide is due to addition of	SDS	ammonium persulphate
191	Polyacrylamide is cross-linked with	n-N"methylene bis acryl	agarose
192	Subunits of oligomeric proteins are linked by	hydrogenbonds	van der Waals forces
193	SDS stands for	sodium disulphite	synthetic dihydrogen s
194	In SDS-PAGE, SDS serves as	inititator of polymerizati	an anioinc detergent
195	In SDS-PAGE, the fast moving protein will have	highest charge	low energy
196	Ammonium per sulphate and TEMED initiate ……. Polymerization	agarose	agar
197	The role of mercaptoethanol in electrophoresis is	break hydrogenbonds	pH maintenance
198	Among proteins, ……. carries largest charge and moves faster.	globulin	albumin
199	Staining method for protein electrophoretogram is	silver stain	methylene blue
200	Cross linking agents in PAGE is	ammonium per sulphate	SDS
201	Polymerisation in PAGE is initiated by	bisacrylamide	acrylamide
202	Buffer with a pH of ……. is used for separation of proteins in paper electrophoresis	8.1	8.6
203	SDS-PAGE cannot be used for	enzymes	proteins
204	In electrophoresis, lipoproteins can be detected by staining with	methylene blue	acrydine orange
205	Proteins possesing more than one polypepetide chain are known as	disulphide bridges	oligomeric proteins
206	Subunit of oligomeric proteins can be separated by	solubilizers	glycols
207	In rocket immunoelectrophoresis antibodies are	mixed with buffer	mixed with agar
208	Agarose is produced from	animal oils	metals
209	Nucleic acids are detected by	ninhydrin	ethidium bromide
210	Ampholytes contain	both positive & negative	neutral charges
211	Iso electric focusing separates proteins that differ by	3 charge units	4 charge units
212	Disulphide bonds in proteins are broken by …… in SDS-PAGE.	SDS	Beta-ME
213	pH at which net charge of the protein becomes neutral is called	alkaline pH	acidic pH

214	Electrophoresis was first discovered by ……...	Faraday	Michael
215	Agar gel used for immunoelectrophoresis was intorduced by ...	Faraday	Graber & Williams
216	**Unit 4**		
217	Which of these is not a method of data collection.	Interviews	Questionnaires
	Secondary/existing data may include which of the following?	quantitative data	response ser
218			
	An item that directs participants to different follow-up questions depending on their response is called a _____.	Probe	Response set
219			
	Which of the following terms best describes data that were originally collected at an earlier time by a different person for a different purpose?	Secondary data	Primary data
220			
	Researchers use both open-ended and closed-ended questions to collect data. Which of the following statements is true?	Closed-ended questions provide quantitative data in the participant's	Open-ended questions directly provide quantitative data
221			
222	Open-ended questions provide primarily _____ data.	Qualitative data	Confirmatory data
	Which of the following is true concerning observation?	It costs less money than self-report approaches	It takes less time than self-report approaches
223			
	Qualitative observation is usually done for exploratory purposes; it is also called _____ observation.	Naturalistic	Structured
224			
225	The word statistics is used as	Both singular and plural words	neither singular nor plural word
226	In chronological classification data are classified on the basis	attributes	Time
227	Bar diagrams are ---------- dimensional diagrams	three	Two
228	Diagrams and graphs are tools of	presentation	Collection of data
229	Data are generally obtained from	secondary sources	Primary sources
230	In geographical classification data are classified on the basis of	attributes	Area
231	In qualitative classification data are classified on the basis of	attributes	Area
232	In quantitative classification data are classified on the basis of	attributes	Area
233	Number of source of data is	3	2

234	Squares and rectangles are	One dimensional diagram	Two dimensional diagram
235	Data originally collected for an investigation is known as	primary data	Tabulation
236	The heading of a row in a statistical table is known as	caption	stub
237	Range is defined as	S-L	L+S
238	Statistics is also a science of	probabilities	Estimates
239	Which one of the following is a measure of central tendency?	range	Median
240	The total of the values of the items divided by their number of items is known as	Arithmetic mean	Median
241	In the short-cut method of arithmetic mean, the deviation is taken as	A – x	x – A
242	The sum of the deviations of the values from their arithmetic mean is	one	– 1
243	The formula for the weighted arithmetic mean is	$\sum w / \sum wx$	$\sum wx / \sum w$
244	Find the Mean of the following values. 5, 15, 20, 10, 40	18	5
245	Which of the followings represents median?	Third quartile	First quartile
246	Which of the measure of central tendency is not affected by extreme values?	Median	Mode
247	Sum of square of the deviations about mean is	one	Maximum
248	Median is the value of ---------- item when all the items are in order of magnitude.	second	First
249	Find the Median of the following data 160, 180, 175, 179, 164, 178, 171, 164, 176.	175	160
250	The position of the median for an individual series is taken as	(N + 2) / 2	(N + 1) / 2
251	Mode is the value, which has	less frequency density	Average frequency density
252	A frequency distribution having two modes is said to be	bimodal	unimodal
253	Mode has ---------- stable than mean.	more	less
254	Which of the following is not a measure of dispersion?	quartile deviation	Range
255	Which one of the following shows the relation between variance and standard deviation?	S.D = square root of variance	var = square root of S.D
256	Range of the given values is given by	L+S	L- S
257	Which one of the following is relative measure of dispersion?	Q.D	Range
258	Coefficient of variation is defined as	(S.D* 100)/A.M	(AM * 100)/S.D
259	In a symmetrical distribution	A.M >H.M >G.M	A.M =G.M=H.M

260	If the values of median and mean are 72 and 78 respectively, then find the mode.	60	16
261	If variance is 64, then find S.D.	13	8
262	**Unit 5**		
263	The probability of drawing a card of King from a pack of cards is	1/11	(1/4)
264	In the case of poisson distribution , if the mean is 4, the standard deviation is ,	4	16
265	For a poisson distribution	mean = Variance	mean < Variance
266	In coin, the probability of getting head is	(1/3)	(1/2)
267	The probability that a leap year selected at random contain 53 Sundays is	(2/7)	(1/7)
268	The square of the S.D is	coefficient of variation	variance
269	A bag contains 7 red and 8 black balls. The probability of drawing a red ball is	8/15	7/15
270	For Binomial distribution ,mean is	n	npq
271	The probability of drawing a card of clubs from a pack of 52 cards is	(1/3)	(1/52)
272	The probability of drawing an ace or queen card from a pack of 52 cards is	1/4	1/13
273	The total probability is	2	0.5
274	If binomial distribution is symmetrical if p=q=?	0.4	1
275	Binomial distribution is positively skewed if	p<0.5	p>0.5
276	Binomial distribution is negatively skewed if	p<0.5	p = 0
277	The probability of drawing diamond and a heart card from a pack of 52 cards is	(1/4)	13/102
278	The probability of drawing king and queen card from a pack of 52 cards is	(1/4)	13/102
279	The probability of drawing a card of King from a pack of cards is	1/11	(1/4)
280	In the case of poisson distribution , if the mean is 4, the standard deviation is ,	4	16
281	For a poisson distribution	mean = Variance	mean < Variance
282	In coin, the probability of getting head is	(1/3)	(1/2)
283	The probability that a leap year selected at random contain 53 Sundays is	(2/7)	(1/7)
284	The square of the S.D is	coefficient of variation	variance
285	A bag contains 7 red and 8 black balls. The probability of drawing a red ball is	8/15	7/15
286	For Binomial distribution ,mean is	n	npq
287	The probability of drawing a card of clubs from a pack of 52 cards is	(1/3)	(1/52)
288	The probability of drawing an ace or queen card from a pack of 52 cards is	1/4	1/13
289	The total probability is	2	0.5
290	If binomial distribution is symmetrical if p=q=?	0.4	1

291	Binomial distribution is positively skewed if	p<0.5	p>0.5
292	Binomial distribution is negatively skewed if	p<0.5	p = 0
293	The probability of drawing diamond and a heart card from a pack of 52 cards is	(1/4)	13/102
294	The probability of drawing king and queen card from a pack of 52 cards is	(1/4)	13/102
295	Two coins are tossed five times, find the probability of getting an even number of heads	1	0.25
296	Mean of a Binomial distribution is 24, Standard deviation = 4, n, p, q respectively are :	60, 1/3, 2/3	72, 1/3, 2/3
297	1 out of 10 electrical switches inspected are likely to be defective. The mean and standard deviation of 900 electrical switches inspected are	81, 9	90, 9
298	If the mean of a Poisson distribution's 4, find S.D.	2	0.25
299	If the mean of a binomial distribution is 5 and standard deviation 2 find the number of items in the distribution	25	20
300	In a binomial distribution mean and mode are equal only when	p=0.9	P=0.5

opt3	opt4	Answer
hydrogen bonds	covalent bonds	hydrogen bonds
citrulline	hydroxyl proline	selenocysteine
Threonine	Valine	Valine
Cation exchange	anion exchange	Affinity chromatography
conductivity meter	fluorimeter	fluorimeter
red	yellow	blue
MS	TLC	Flourimeter
intensity of light	width of medium	length of the medium
decreased exponentially	increased exponentia	decreased exponentially
NMR	GCMS	spectrophotometer
both a and b	filter	both a and b
0.5mm	1mm	1cm
fluorimeter	calorimeter	double beam spectroscopy
filter	condensing lens	photomultiplier tube
galvanometer	slit	photomultiplier tube
microwave	infra red	microwave
colourless	black	black
selective absorption	refraction	selective absorption
a (or) b	black	colourless
magnet	sensor	filter
condensing lens	light source	filter
dynodes	anode	dynodes
conductivity meter	spectrophotometer	fluorimeter
400-550nm	380-800nm	380-800nm
UV light	infra red	UV light
condense	straight	parallel

hydrogen	mercury	mercury
secondary filter	photo cell	secondary filter
perpendicular light	refraction	linear resolution
different	condensed	single
absorbance	path length	absorbance
prism	absorbance meter	spectrophotometer
condensing lens	absorption coefficient	diffraction
polychromatic light	spectra of light	monocromatic light
double beam spectrophotomet	NMR	double beam spectrophotometer
prism	monochromator	reversed prism
monochromatic	spectrum	spectrum
white	green	yellow
gamma rays	UV rays	gamma rays
absorbance	prism	fluorescence
single spectrum	central spectrum	polychromatic
decreases exponentially	increases exponential	decreases exponentially
electric coefficient	absorption coefficient	extinction coefficient

muscular force	mechanical force	centripetal force
applied centrifugal force	type of rotor	applied centrifugal force
towards bottom	towards top	radialy outwards
rotation per second	cm/min	rpm
981cm sec ^{2}	981m sec ^{2}	981cm sec ^{2}
number of times 'g'	time	number of times 'g'
radius and angle of rotation	high speed	rpm and radius of rotation
RCF=(1.118x10)(rev. min ^{2}	RCF=(1.118x10)(re	RCF=(1.118x10)(rev. /min)^{2}r
$\omega=2\pi$ rev. rev/120	$\omega=2\pi$ rev. min^{2}/60	($\omega=2\pi$ rev. min ^{1})/60
minimum	higher	zero
stable	maximum	increases
constant	becomes maximum	decreases
Sedimentation Unit	Svedberg Unit	Svedberg Unit

20˚C	15˚C	20˚C
1x10 ^{13}sec	1x10 sec	1x10 ^{13}sec
constant	higher	faster
40S to 10,000S	40S to 1,000S	40S to 1,000S
35,000S	10,000S	40,000S
30x10^{3}S to 70x10^{3}S	20x10^{3}S to 60x10^{3}S	20x10^{3}S to 70x10^{3}S
Amount of solvent eluted + ar	Amount of solvent elu	Amount of solute eluted / amount of impurity eluted
four	six	two
6,000 rev. min	3,000 rev. Min ^{1}	3,000 rev. Min ^{1}
upside down	parellel	vertically
diagnolly opposite	vertical	diagnolly opposite
6,500g -	5,000g -	6,500g
6,500 rev. min ^{1}	5,000 rev. min ^{1}	6,000 rev. min ^{1}
50,000 rev. min ^{1}	60,000 rev. min ^{1}	25,000rev. min ^{1}
25,000g	50,000g	60,000g
Continious flow-centrifuge	hand centrifuge-	Continious flow-centrifuge
Four types	Six types	Two types
50,000 rev. min ^{1}	80,000 rev. min ^{1}	80,000 rev. min ^{1}
60,000g	45,000g	60,000g
a glass chamber	a backlite meterial	Heavy Armour platting
Air flow centrifuge	Air fuge -	Air fuge
3.7cm	3.5cm	3.7cm
2,00,000 rev. min ^{1}	1,50,000 rev. min ^{1}	1,00,000 rev. /min
1,70,000g	1,60,000g	1,60,000g
clinical centrifuge	airfuge	High speed refrigerated centrifuge
Airfuge	Clinical Centrifuge	Analytical Ultra Centrifuge
Thermostat	Thermoinductor	Thermistor
Condensing and concave	condensation	Condensing and Collimating
Four cells	Six cells	Two Cells
Five	Four	Three

Rayleigh interference system	Electro magnetic ligh	Rayleigh interference system
Rayleigh interference	Schlieren optical syst	Schlieren optical system
Photo emissive plate	Photo deviating plate	Photographic plate
Titanium	Plutonium	Titanium
fixed zero angle rotor	above 1	fixed zero angle rotor
4˚ to 14˚	30˚ to 40˚	14˚ to 40˚
Swinging bucket rotor	Zonal rotor	Swinging bucket rotor
Axis of rotation	axis of deviation	Axis of rotation
Angle of rotation	Angular force	Centrifugal force
Four types	Six types	Two types
Corrosive effect	Wall effect	Wall effect
Swinging bucket Rotors	Fixed angle rotors	Elutriator Rotors
Four types	Five types	Two types
Mechanical Centrifugation	cooling centrifuge	Analytical Ultra Centrifugation
Sedimentation constant metho	Central coefficient	Sedimentation equilibrium method
Cellular organisation	sedimentation	Sub cellular fractionation
Pellet	cell waste	Homogenate
Supernautent	Homogenate	Marker
Glucose 6 Phosphotase	malate	5' Nucleotidase
Glucose 6 Phosphotase	Lactate Dehydrogena	DNA
Nucleotidase	Lactate Dehydrogena	Glutamate dehydrogenase
Acid Phosphotase	Lactate Dehydrogena	Acid Phosphotase
Nucleotidase	Glutamate dehydroge	Glucose 6 phosphate
Nucleotidase	Glucose 6 Phosphate	Lactate dehydrogenase
size of the particle	time	buoyant density of the particle
Centrifugal elutriation	Centrigugal constant	Rate zonal centrifugation technique
Sodium bromide	Ficoll	Sodium bromide
Caesium sulphate	Sodium bromide	Ficoll
Dextron	Ficoll	Glycerol
Bonding of DNA and RNA	Purification of Protei	Separation of whole cell
Sucrose	Ficoll	Caesium sulphate

Dextran	Sucrose	Dextran
Cytosol	Mitochondrion	Mitochondrion
paper	thin layer	affinity
elution volume	retention time	chromatogram
less than 1	less than 2	less than 1
HPLC	TLC	TLC
agarose	polyacrylamide	silica gel
polyacrylamide	agarose	filter paper
disulphide bridges	hydrogenbonds	capillary forces
Kieselguhr G	polyacrylamide	Kieselguhr G
hexane, water	butanol, acetic acid, w	butanol, acetic acid, water
equal to RF	RF-1	greater than Rf
column	thin layer	partition
0.01 N HCl	0.001N HCl	0.001N HCl
methanol	ethanol	ninhydrin
gel matrix	axis	gel matrix
β-1,4 linked arabinose	β-1,4 linked glucose	β-1,4 linked glucose
α-1,6 linked fuc ose	α-1,6 linked arabinose	α-1,6 linked gluc ose
L-fucose	L-arabinose	D- galact ose
orthosilicic acid	hydrochloric acid	orthosilicic acid
ECG	chromatogram	chromatogram
only positive	only negative	oppositely charged particles
matrix without charges	void volume	cation exchanger
polystyrene	starch	polystyrene
retention	elution	elution
retention	elution	void volume
retention time	elution	retention time
retention time	isocratic elution	isocratic elution
4	5	4
Richard	Edwin	D.T. Day
plant pigments	animal cells	plant pigments

Richard	Edwin	M.S.Tswett
tar	coconut	coconut
gel matrix	salt	minerals
calcium phosphate	sodium chloride	calcium phosphate
calcium	geometrical isomers	geometrical isomers
Bio-Rex 70	AG 1	Sephadex AG 50
Bio-Rex 70	AG 1	Bio-Rex 70
Bio-Rex 70	QAE-Sephadex	QAE-Sephadex
Bio-Rex 70	QAE-Sephadex	AG 3
resin	rexin	resin
both 1 & 2	cellulose	cellulose
AG 1	Bio-Rex 70	Mixed-bed resin
Bio-Gel P	Dowex 50	Bio-Gel P
AG 1	Sepharose	Sepharose
QAE-Sephadex	Dowex 50	Sephacryl S
gel	compounds	ligands
heparin	avidin	heparin
heparin	Protein A & G	Protein A & G
NADPH	Protein A & G	avidin
NADPH	Protein A & G	cibacron blue
GLC	HPLC	HPLC
absoption	filtration	molecular mass
magnetic filed	TCA	electrical charge
immunoelectrophoresis	TLC	immunoelectrophoresis
polyacrylamidealbumin	albumin	albumin
negatively charged	positively	both charge
formate	phosphate	calcium
1958	1957	1958
centrifugation	electrophoresis	electrophoresis
agarose gel electophoresis	Cetrifugation	SDS-PAGE

colour	UV	charged ions
less hydrophilic	aal	less hydrophilic
galactose	cellulose	cellulose
beta globulin	gamma globulin	alpha globulin
size & charge	weight	size & charge
do not move	disappear	act as anions and move towards anode
PAGE	gel electrophoresis	PAGE
beta mecaptoethanol	urea	ammonium persulphate
styrene	TCA	n-N"methylene bis acrylamide
disulphide bridges	carbons atoms	disulphide bridges
sodium dihydrogen phosphate	sodium dodecyl sulph	sodium dodecyl sulphate
cationic detergent	neutralizing agent	an anioinc detergent
lowest charge	no charge	highest charge
acrylamide	agarobiose	acrylamide
impart negative charge to prot	break S-S bonds	break S-S bonds
keratin	hemoglobin	albumin
Ponceau-S	ethidium bromide	silver stain
bisacrylamide	TEMED	bisacrylamide
TEMED	CBB	TEMED
6	7.5	8.6
vitamin A	nucleic acids	vitamin A
vinyl orange	Sudan black	Sudan black
alpha chain	beta chain	oligomeric proteins
detergents	emulsifiers	solubilizers
applied in well	sprayed on gel plate	applied in well
plants	algae	algae
ninhydrin	Coomassie blue	ethidium bromide
positive groups	negative groups	both positive & negative groups
2 charge units	one charge unit	one charge unit
APS	Coomassie blue	Beta-ME
isoelectric pH	nutral pH	isoelectric pH

DuBois	Alexander Reuss	Alexander Reuss
DuBois	Alexander Reuss	Graber & Williams
Experiments	Observations	Experiments
Archived research data	Structured	Archived research data
Semantic differential	Contingency question	Contingency question
Experimental data	Field notes	Secondary data
Open-ended questions provide qualitative data in the participant's own words	Closed-ended questions directly provide qualitative	Open-ended questions provide qualitative data in the
Predictive data	Quantitative only	Qualitative data
It is often not possible to determine exactly why the people behave as they do	It costs more money than self-report approaches	It is often not possible to determine exactly why the
Complete	Probed	Naturalistic
plural word	singularwords	Both singular and plural words
class interval	location	Time
one	multi	one
analysis	summarization	presentation
both primary&secondary	neither primary nor secondary	both primary&secondary
time	location	Area
time	location	attributes
time	magnitude	magnitude
4	1	2

three dimensional diagram	Multi dimensional diagram	Two dimensional diagram
secondary data	published data	primary data
title	heading	stub
L+S	L-S	L-S
both estimates nor probabilities	neither estimates nor probabilities	both estimates nor probabilities
variation	correlation	Median
mode	range	Arithmetic mean
(x – A) / c	(A – x) / c	x – A
two	zero	zero
∑x / n	∑x / ∑f	∑wx / ∑w
41	20	18
Second quartile	Q.D	Second quartile
sixth deciles	Mean	Median
zero	Minimum	Minimum
Middle most	last	Middle most
176	180	175
N/2	N/4	(N + 1) / 2
greatest frequency density	graetest frequency	greatest frequency density
trimodal	modal	bimodal
same	most	less
standard deviation	median	median
variance = S.D	variance / S.D = 1	S.D = square root of variance
S+L	LS	L- S
S.D	coefficient of variation	coefficient of variation
S.D/A.M	(1/S.D)*100	(S.D* 100)/A.M
H.M >G.M >A.M	A.M <H.M <G.M	A.M =G.M=H.M

70	76	60
14	11	8
1/12	1/13	1/13
2	1	2
mean > Variance	mean < Variance	mean = Variance
2	0	(1/2)
(3/7)	(1/53)	(2/7)
square of variance	square of coefficient	variance
1/15	14/15	1/15
p	np	np
¼	1/13	1/13
2/13	1/52	1/13
1	0	1
0	0.5	0.5
p=0.5	p = 0	p<0.5
p=0.5	p>0.5	p>0.5
2/13	(7/16)	13/102
(2/13)	8/663	8/663
1/12	1/13	1/13
2	1	2
mean > Variance	mean < Variance	mean = Variance
2	0	(1/2)
(3/7)	(1/53)	(2/7)
square of variance	square of coefficient	variance
1/15	14/15	1/15
p	np	np
¼	1/13	1/13
2/13	1/52	1/13
1	0	1
0	0.5	0.5

p=0.5	p = 0	p<0.5
p=0.5	p>0.5	p>0.5
2/13	(7/16)	13/102
(2/13)	8/663	8/663
0.4	0.25	0.25
87, ¼, 3/4	90, 1/5, 4/5	72, 1/3, 2/3
88, 10	91, 11	81, 9
3.25	4	4
16	9	25
q=0.1	p=1	P=0.5