APPLICATIONS OF BIOSTATISTICS AND EVOLUTION

*Editor
Dr. Mandaloju Venkateshwarlu

APPLICATIONS OF BIOSTATISTICS

AND EVOLUTION

***Editor**

Dr. Mandaloju Venkateshwarlu

M.Sc., Ph.D., (Botany) M.Sc (Bio-Tech), FSAB, FHAS.(L.L.B).

Department of Botany, Kakatiya University,

Warangal – 506 009, (TS) India.

Published By:

EJBPS

2021

APPLICATIONS OF EVOLUTION AND BOISTATISTICS

Dr. M.Venkateshwarlu

APPLICATIONS OF BIOSTATISTICS AND EVOLUTION

PREFACE

The book is written in simple language so that the students can easily grasp the matter. Some important terms has been incorporated. So that the students may search the useful related for competitive examinations. In the recent years included in the syllabus of almost all Indian Universities in various subjects of Biology or Life Sciences as an independent evergreen subject. Exponential growth in many areas of basic fundamentals made it necessary in some cases to write several chapters on the same topic which was covered in a single chapter in the earlier book. Similarly, in the present volume, separate new chapters have been written on topics which in the earlier title either did not figure at all or were each covered very briefly as a part of a chapter. In the present book, for instance in separate chapters have been written on new topics. The students of Biology at the post graduate (P.G.) under graduate (U.G.) levels needed to the recent Global changes and developments.

Several of my students in the laboratory helped me either in writing some of the chapters or in preparing the list of references and appendix given at the end of this volume. There may also be errors and omissions of technical nature, since in a vast and fast expanding subject like Botany Biotechnology; one cannot claim to have known everything, despite his best efforts.

I am thankful to Department Head, BOS, Staff and Research Scholars (Botany) my family members, inspiration and cooperation my wife and children's (M. Hamsini) Teachers, Friends, Students and Well-Wishers (Dursheti Sai Charan, M.B.A., Certified Microsoft Office and Windows Specialist). I hope that this book will be useful to students in Life Sciences.

Dr. Mandaloju Venkateshwarlu

CONTENTS

CHAPTER I
EVOLUTION

1. INTRODUCTION

The idea that the present forms of life have arisen from the earlier, simpler ones was far from new when Charles Darwin published his book "*The Origin of Species*" in 1859. The development of the idea of organic evolution through pre-Darwininan, Darwinian, and post-Darwinian eras can be studied under the following periods.

The oldest speculations about evolution are found in the classical mythologies of different centres of ancient cultures, such as India, Chinna, Babylong and Egypt. For instance, the classical Indian (Hindu) mythology has forwarded the idea that universe and the body of all living beings are made up of five elements or "Panch Bhota", namely kshiti (earth), ap (water), teja (fire), marut (air) and vyoum (void orsky). Further, the Padam Puran conceived ten incarnations or "Dashvatarahs" for explaining the evolution of vertebrates. The sequence of "Avatarahs, includes Matsya, fish-like appearance; Kurma, reptile-like; 'Varaha', mammal-like; 'Narasimha', half-man and half-lion; "Vamana", the dwarf man-like appearance (anthropoid); "Parshurama",a Homa saplens of Pleistocene; "Shri Rama", "Balrama", "Shri Krishna" and "Kalpi" – the latter four incarnations depict the formation of clans and communities.

Similar to ancient Indians, the ancient Chinese conceived the world at the beginning as made up of five primary elements – water, wood, fire, soil, and the precious metal, gold. According to their mythology, the positive and negative interactions of these primary five elements were responsible for origin of life and different kinds of adaptations and diversities of organism or "things" were originated from a single source through gradual unfolding and branching.

The Egyptian and the Babylonians gave the idea of cosmic evolution which included the following aspects : (i) a Divine Intelligence and creator, (ii) water as a primitive and boundless substance, (iii) the potentiality of all things in the Divine Intelligence, and (iv) creation as the process of becoming actual by stages.

During this period, because the speculation on evolution idea was much shrouded in mystery and mythology, so, this period has come to be known as the period of obscurity.

2. Period of Ancient Greeks and Romans

The classical writings of certain Greek philosophers contain the vague ideas and speculations about the organic evolution. Thus, Thales (624-548 B.C.) suggested aquatic origin of life, i.e., all life originated from oceans. Anaximander (611-547 B.C.) explained the origin of universe. According to him, living things came from primordial fluid and men were first formed as fishes; eventually they cast off their fish skin and took up life on dry land. Anaximenes (382-524 B.C.) believed that life came from air and he proposed abiogenesis. Xenophanes (576-480 B.C.) is credited with being the first person to recognize that fossils, such as petrified shells embedded in rocks, represent the remains of animals that once lived. He also realized that the presence of fossils of marine animals on what is now dry land indicates that the sea once covered the area. He believed that in ancient times earth was covered with water where life began. Empedocles (495-435 B.C.) is considered as the "Father of evolution idea" by many evolutionists for his ideas which he forwarded in his "Poem of Nature". According to him, life arose spontaneously but gradually and change (evolution) would result when twin forces, the attraction and repulsion, interact with each other. Based on it, he further explained that first living things were cast out in the form of separate parts and by chance these parts were united by attraction force and as a result plants and animals were born. The sex originated later and through reproduction the developing embryo resulted from both paternal and maternal substances and in that fashion distinct species were formed. He believed that when reproduction was established, earth felt no need of special creation of organism. According to Empedocles plants came earlier than animals and perfect forms replaced imperfect types.

………diseardcd Empedocles's views and he observed that organisms could be arranged in graded series from lower to higher. The lowest stage is the inorganic. Organic beings arose from the inorganic by direct metamorphosis. He conceived the organic world to consist of three states: 1. plants, 2. plant-animals, a transitional group in which he included sponges and sea anemones; 3. animals characterized by feeling or sensibility. Within the animal groups he constructed a genetic series leading from the lowest forms up to man placed at the apex. ………is considered as the ………………………from this taxonomic fact, he drew thecorrect inference that one higher group evolved from other lower one in the taxonomic series. However, he had the metaphysical belief that the gradual evolution of living things occurred because nature strives to change from the simple and imperfect to the more complex and perfect. An evolutionary explanation of the origin of plants and animals was given by the Roman poet ………….. (99-55 B.C.) in his poem *De Rerum Natura.*

3. Pre-Darwinian Period

The medieval period is characterized as dark age. During this period, superstition and blind following of ancient thinkers prevailed and Aristotelian reasoning and the spirit of scientific inquiry were died out. Thus, no progress occurred in the idea of evolution during medieval period.

With the Renaissance, interest in the natural sciences quickened and the increasing knowledge of many kinds of animals led more and more scientists to consider the concept of evolution favourably. Thus, before Charles Darwin several possible mechanisms have been proposed to account for the evolution; some of them have been considered here as follows.

(i) Fixity, design and creation. *Nullae speciae novae* (= No species new) was the catch-phrase for the early systematists, including Linnaeus (1707-1778). Species were the units of creation as prescribed in Genesis (The first book of the Bible) and, therefore, immutable. Moreover, the reason they were so well-fitted for the challenges present everyday life was that they had been designed by God for specific functions in nature. John (1627-1705), clergyman, naturalist and early systematist, saw the fitness of species as evidence the existence of a Designer, and this Argument from design was made even more clear ……. William Paley (1743-1805). According to early systematists, animals and plants are wonderful of machinery, more wonderful than any man-made machine, and so they must be the products. Intelligence more wonderful than that of Man. There was a later, more subtle, version of this argue…. Which insisted on the harmony of form rather than utility. These are described as the idealist utilitarian positions respectively.

This creationist concept held certain biological theological and philosophical problems example, many fossilized organisms had been discovered that no longer existed on earth, and yet it unbelievable that the perfect products of an omnipotent designer could ever have become obs….. Georges Cuvier (1769-1832), a very influential French biologist and founder of palaentol… offered catastrophic hypothesis as a way out. A series of catastrophes, fires and floods (e.g., N….. Flood) had removed some of the species initially created by God. Apparently useless characters another worry because inferior design could hardly be attributed to a super-intelligent designer explain certain useless characters (e.g., toes of pig), …………(1707-1788) suggested that the Su………being had created perfectly designed types embodied in the original species, but that new species a from them by a process of hybridization and degeneration. Thus ass was supposed to be a degenerate horse and the ape a degene….. man. But this meant that the assumption of strict fixity of spe…. Had to be relaxed and the concept of *nullae speciae novae* dip eared from the last revised edition of the book – *systema nat*…. of …………………. Though was not an evolutionist, nor wa……. Greatly concerned with explanations of nature, buthe devoted life industriously in arranging the kinds of animals and pl….. known to him in a convenient system. This system he took to actual and fixed; to discover it and use it was the major tas……. Natural history. He broke the tradition in placing man, "the know……. One, Homo sapiens" in the animal kingdom, the class Mamm……… and the order Primates. Of many categories employed today in classification, it is interesting ……………….. used only the kingdom, class, order, genus, and species, but other naturalists soon foun……..

Lamarckism was supported by E. Geoffroy St. Hilaire (1772-1844), Robbert Chambers (1802-1871) and Herbert Spencer (1802-1903), but bitterly opposed by G. Cuvier (1769-1832) and Weismann (1887).

4. Darwinian Period

Charles Darwin (1809-1882) and Alfred Russel Wallace (1823-1913), both spent many years in collecting and exploring South America and certain islands of its sea and both independently drew similar conclusions about the possible mechanism of organic evolution due to natural selection, after studying the easy on population by Malthus (1789) which suggested that human populations continually tend to outstrip the availability of food supply, but are kept in check by natural diseases, and famine, and man-made wars. Darwin-Wallace theory of natural selection was published in the Journal of the Linnaean Society in July, 1858.

Darwin seemed to agree with Lamarck's idea that characters acquired during the life of an organism might be transmissible, in a heredity sense, to its offspring (e.g. Darwin's theory of pangenesis). However, it was on the mechanism of evolution that Darwin parted company, with both Lamarck and his own grandfather.

In his book, Origin of Species (fist published in 1859), Darwin replaced the idea of direted variation (brought about by a sentiment interieur) with random variation that was heritable. Regarding the origin of this variation, he was a little vague – he thought that it might be due to environmental influences on the ……… and he was equally vague about the mechanism of inheritance. What he was perfectly clear about was that no matter how the variation originated, it did so withoutany reference to the needs of the organism. Moreover, for Darwin the variation upon which evolution was based involved very small continuous changes away from the original form. He accepted that large abrupt changes might occur, but claimed these were rare and most led to monstrosities. Finally, instead of Lamarck's mysterious struggling for perfection, Darwin explained evolution in terms of a completely ordinary and non-mysterious process. Darwin observed that animal and plant breeders had, by the selective breeding of variants, been able to affect considerable changes in the form of domestic plants and animals. This was artificial selection. Similarly, he thought that the struggle between organisms in an overcrowded world by finite resources would ensure that only the fittest survive and this would lead to a kind of natural selection of variants.

The gist of dArwin-Wallace theory of natural selection was as follows: 1. Individuals within species show considerable but continuous variation in the form and physiology. 2. This variation arises in a random fashion and is heritable. 3. The potential for increase within populations of animals and plants is considerable. 4. Since resources are limited, so individuals in a population struggle for their own existence and that of their offsprings. 5. hence, only some (what Darwin called the fittest) survive and leave offsprings with the same traits. 6. Through this natural selection of the fittest, species become represented by individuals which are better adapted.

5. Post-Darwinian Period

The history of evolutionary thought subsequent to the publication of the book '*The Origin of Speceis'* may be divided into the following three periods.

(i) **The romantic period.** The romantic period extended from 1860 to about 1903 and was characterized by extreme enthusiasm for Darwinism, together with un uncritical acceptance of whatever data were claimed to support Darwinism. Negative evidence was given little weight, while illogical extremes of interpretation in order to make observed facts fit Darwinian theory were quite common. Leaders of this group in England included T.H. Huxley (1825-1895), Herbert Spencer (1820-1903) and George Rmanes, while in the United States David Starr Jordan and ASA Gray (1810-1888) were the leaders. As a group, they went to interpretative extremes, reading adaptive significance into every organic structure, even on the most imaginary evidence. They often cited excellent anatomical and taxonomic evidence, but experiments to test adaptive values were unusual.

It should not be thought that these were second-rate biologists who were blinded by the brilliance of a great man like Darwin, on the contrary, they were excellent men in their respective fields. Huxley made splendid contributions to the development of invertebrate zoology, taxonomy and vertebrate anatomy. Spencer was one of the leading philosophers of his time. Romanes started his career as an invertebrate neurologist, but he soon became mainly engrossed in evolutionary problems. Jordan was one of the best ichthyologists of all times. Gray was a botanist of great repute. Nor should it be thought that they never ventured to differ from Darwin, for these men were independent thinkers. Yet the atmosphere of approval was extraordinary.

It has been said that evolution was born in England, but found its home in Germany. In fact, the German evolutionists of the Romantic period were more strictly Darwinian than their English and American colleagues in the sense that they were generally more thorough and careful collector of data. The leaders in Germany were Von Baer, Carl Gegenbaur, Ernst haeckel and august Weismann. Von Baer (1792-1876) was a great embryology of different animals. His basic principles of embryologystates that 1. general characters appear before special characters; 2. from the more general characters, less general and finally the special characters develop, 3. an animal during development departs progressively from the forms of other animals; 4. the young stages of animals are like the young (or embryonic) stages of other animals lower in the scale, but not like the adults of those animals. Gegenbaur (1826-1903) was one of the greatest and most influential comparative anatomists. He and his collaborators made exhaustive studies, in complete detail, upon all classes of vertebrates, and used the data so obtained in support of Darwinian theory. Much of the phylogeny (i.e., the evolutionary or anestal history of organisms) of the vertebrates in current textbooks of zoology is taken from the works of Gegenbaur and his collaborators.

Ernst Haeckel (1834-1919) did significant work in anatomy, embryologyand taxonomy. His studies in comparative embryology led him to broaden the principles of Von Baer into the theory recapitulation or biogenetic law, in support of which he published extensively. In one of his writings '*Generelle Morphologie*', Haeckel (1866) defined the biogeneitic law in the following way – an individual organism in its development (……..) tends to recapitualate (or repeat) the stages passed through by its ancestor (phylogeny), on that antogenyrepeats phylogeny.

August Weismann (1834-1914) was a classical geneticist. For explaining the mode of inheritance in living organisms, be rejected outrightly the Lamarckian ideology of inheritance of acquired characters and Darwin's pangenesis hypothesis by demonstrating a well known experiment in which he cut the tails of mice for twenty-two generations, yet each new lot of offspring consisted only of animals with tails. He forwarded an entirelyspeculative and theoretical hypothesis of "the continuity of germplasm" or as most popularly known "germplasm theory". The Main drawbacks of Weismann were that he was unaware of work of his contemporary geneticist Mendel and his speculations rested on some data on chromosome behaviour during mitosis which were often reported to him by his graduate students as he gradually became blind before he completed his major works. He correctly predicted the occurrence of reduction division (or meiosis) and the chemical nature of the hereditary units (Dodson and Dodson, 1976). Weismann imagined that natural selection had operated at three levels: between individuals (Darwin and Wallace);

between parts, organs, tissues, etc., of the same individual (Roux; according to Roux, 1881, the struggle is waged among the molecules, tissues and organs; what is at stake is the intake of food which, according to its quality, ensures greater or determinants (now called genes; Weismann).

Thus, according to Weismann, the most important struggle is between the "vital invisible" particles (genes), whether they reside in the body cell (soma) or the sex cells. These determinants are never identical with one another and their ability to "assimilate" food varies from one another. The strongest take the most food, and their descendants prevail over the others. The result is that by the second generation the parts of the organism represented in the ovum by the strongest determinants, are the most highly developed. This explains the dominance of certain characteristics; the determinants of organs favoured by natural selection are better fed than the rest and are the winners in the competition, hence, the strengthening of change in that the rest and are the winners in the competition, hence, the strengthening of change in that direction (Grasse, 1977).

Also prominent during the romantic period was Karl Pearson and Francis Galton who laid the foundation of the new sciences of statistics and biometry which play a prominent role in modern evolutionary studies.

(ii) The agnostic period. The agnostic period extended from 1903 to 1935 and involved a lot of skepticism and disillusionment. Many factors converged to cause this change. One factor was the rediscovery of Mendel's laws of heredity in 1900. Today Mendelism is the foundation of most studies in evolution, but then the permanence of the gene seemed to raise fearfulobstacles to the origin of new species. In consequence, genetics was regarded as a sort of blind path at the end of which stood the sign: the gene, the dead end. Another factor was the conclusion which was drawn by Johansen from his hereditary experiment of the size in beans, that selection could be effective only in a stock (or organism) with hereditary variability. Variations produced by the environment (including nutrition, sunlight, temperature and moisture) were unimportant for evolution.

Finally, besides natural selection, many additional mechanisms and causes were sought by different evolutionists for explaining the evolution. For instance, Moritz

Wagner (1868) emphasized the effects of different environments on isolated froms. Gulik pointed out the unavoidability of that an inner directive guides the course of evolution mainly independent of the environment.

Hugo de Vries (1848-1935) in his book The Mutation Theory (1909) distinguished between continuous individual variation and discreate, saltational variations. He applied the term mutation to the latter only. De Vries experimented with the plant evening primrose Oenothera lamarckiana. For De Vries it was these mutations that controlled evolutionary change and which were more important than natural selection itself.

The early Mendelians, such as de vries, therefore, played down the ability of natural selection and emphasized the overriding importance of Mendelian ratios and mutations in moulding evolution. Some evolutionists recognized the importance and challenge of Mendelism and set out to reconcile it with the Darwinian theory. Because the results of their efforts went beyond the mechanism that Darwin had proposed, the synthesis that emerged is often referred to as neo-Darwinism (Calow, 1983). Further, interest of some neo-Darwinist shifted from considering the consequences of breeding from pairs of individuals to the consequences of breeding between individuals in populations. The Mendelian ratio then became frequencies of occurrence of genotypes in populations and genetic and evolutionary changes were defined in the terms of changes in these frequencies. In 1908, G.H. Hardy and W. Weinberg independently concluded that Mendelian Mechanisms themselves do not cause changes in gene frequencies and they cannot be an important element in directing evolutionary change. Moreover, two books – J..S. Haldane's Genetical Theory of Evolution (1930) and R.A. Fisher's The Cause of Evolution (1932) – were greatly important in shaping the subsequent development of neo-Darwinism. Both Fisher and Haldane expressed the intensity of selection in terms of changes in gene frequencies. Those alleles that increase in frequency relative to others are said to be fitter, so the change in these relative frequencies measures neo-Darwinian fitness. Fisher's Fundamental Theorem of Natural Selection, i.e., the rate of increase of fitness of any organism at any time is equal to its genetic variance in fitness at that time.

Sewal Wright, another pioneer of neo-Darwinism, in 1931 developed a conceptual framework for considering the interaction between genes and their influence on fitness. This is referred to as an adaptive landscape or fitness space. Wright's emphasis was on populations of genes. However, R.B. Goldschmidt (1878-1958) who had wide experience in the study of geographic variations, taxonomy and physiological genetics, is the principal proponent of the main alternative to the dominant neo-Darwinian theory.

Punctuated equilibrium concept of evolution. This idea was proposed in 1972 by two American palaentologists, Dr.S.J. gould and Dr.N. Eldredge and has been widely accepted as the best way to interpret evolutionary patterns over geological time (see Henry Gee, 1993). This concept contradicted Darwin's idea that species evolve gradually one into other by insensible degree. According to Gould and Eldredge, species evolve by sudden bursts of rapid change that 'punctuate' long periods of static equilibrium. The punctuated equilibrium concept showed that Darwinian gradualism should have been supported by fossil record, i.e., transitional forms should make up the bulk of fossils. But fossil record shows that species seem to persist with little change over many millions of years. Thus, there is no use to try to find evidence of gradual change in fossil record, because it is not there to begin with (Gee, 1993). According to Gould and Eldredge, species evolve very rapidly, within a few thousand years – too short a time for preservation of the fossil record to be likely. But once evolved species remain unchanged for many times, longer than it makes them to evolve. Thus, stasis (stoppage) of species becomes data of evolution.

To support their idea, Gould and Eldredge took help of the idea of mutation. Thus, according to the process of speciation change, in the form of the success of mutations, it tends to happen most efficiently in small peripheral populations of a species rather than in the general mass in which such mutations might be swamped. Likewise, individuals from small populations in transition have even less chance of fossilization than those from large one. Of course, natural selection can be a force for change in small as well as large populations, but a chance and accident assume more significance in smaller populations. Once formed the new species expand into new habitats and settles down to comfortable statis. The result is not one but two populations – exactly the pattern observed in the fossil record in many cases.

(iii) The modern synthesis period. The synthesis of a coherent theory of evolution which takes into account all the pertinent facts of modern biology, has been the work of several biologists during the past fifty years. The first edition of Theodosius Dobzhansky's now classic book, Genetics and the Origin of Species', which appeared in 1937, set the stage, and stimulated biologists in several fields to contribute to the synthetic evolutionary theory. The term modern synthesis was first used by Julian Huxley in 1942. as presented in the famous book – Evolution, The modern Synthesis by Huxley (1942), the modern synthesis or synthetic theory was an amalgamation of the data and concepts of genetics taxonomy, embryology, biogeography, and other disciplines (Stanley, 1979). Likewise books by zoologist Ernst Mayr (1942, 1963) showed how the modern theory could explain the origin of variation patterns in higher animals and G,L.Stebins (1950) attempted to do the same for higher plants. E.B. Babcock's (1947) study of the genus Crepis, carried over a period of more than thirty years, is one of the most thorough and comprehensive studies ever made upon a single genus of plants and it is one of the major supports of the neo-Darwinian theory. Clausen and coworkers in 1951 have published extensive studies on the behaviour of plants when grown in widely different habitats. A leading palaeontologist G.G. Simpson showed in his two books – Tempo and Mode in Evolution (144) and The Major Features of evolution (1953) that the fossil record of higher animals is best explained by assuming that throughout the evolutionary history of living things those same processes took place which were being experimentally demonstrated by many workers in populations of contemporary animals. In 1960, Bernhard Rensch has made a strong case for the belief that the evolution of genera families and higher categories of animals has taken place through an extension into long periods of time of those same processes which at any one time level govern the origin of races and species.

According to Mayr (1970), most of the earlier evolutionary theories were characterized by heavy emphasis on a single factor, as has been tabulated below in Table 2-1.

Theories of evolutionary changes (E.Mayr, 1970).

A. Monistic (single – factor explanations)

1. Ectogenetic: Change directly induced by the environment.

a. Random response (for example radiation effects)

b. Adaptive response (Geoffroyism)

2. Endogenetic: Changes controlled by intrinsic forces

a. Finalistic (orthogenetic)

b. Volitional (genuine Lamarckism)

c. Mutational limitations

d. Epigenetic limitations.

3. Random events ("accidents")

a. Spontaneous mutations

b. Recombination

4. Natural selection

B. Synthetic (multiple-factor explanantions)

1b + 2a + 2b = most "Lamarckian-type" theories

1b + 2b + 2c = some recent "Lamarckian" theories

1b + 3 + 4 = Late Darwinian, most non-mutationists during first three decades of 20^{th} century

3 + 4 = early "modern synthesis"

1a + 2c + 2d + 3 + 4 = recent "modern synthesis"

The modern synthetic theory selected the best aspects of the earlier hypotheses and combined them in a new a original manner. In essence a two-factor theory, it regards the diversity and harmonious adaptation of the organic world as the result of a steady production of variation and of the selective effects of the environment.

Attempting to explain evolution by a single-factor theory was the main weakness of the pre-Darwinian and most nineteenth-century evolutionary theories. For instance, Lamarckism with its internal self-improvement principle, Geoffroyism with its induction of genetic change by the environment, Cuvier's catastrophism, Wagner's isolation, De Vries' mutationism, all tried to explain evolution by a single principle, excluding all others. Even charles dArwin occasionally fell into this error, yet on the whole he was the first to make a serious effort to present evolutionary events as due to a balance of conflicting forces. The current theory of evolution the "modern synthesis", owes more to Darwin than to any other evolutionist and is built around Darwin's essential concepts. Yet it incorporates much that is distinctly post-Darwinian. The concepts of mutation, variation, population, inheritance, isolation, and species, all of which were quite nebulous in Darwin's days, but now are much that is distinctly post-Darwinian. The concepts of mutatin, variation, population inheritance, isolation, and species, all of which were quite nebulous in Darwin's days, but now ae much better defined and understood, have been incorporated in recent "modern synthesis" (Mayr, 1970).

PRESENT STATE OF EVOLUTIN IDEA

The current synthetic theory of evolution recognizes five basic types of processes – Gene mutation, Changes in chromosome number and structure, Genetic recombination, Natural selection, and Reproductive isolation. The first three provide the genetic variability without which change cannot take place; natural selection and reproductive isolation guide populations of organisms into adaptive channels. In addition, three accessory processes affect the working of these five basic processes. Migration of individuals from one population to another as well as Hybridization between races or closely related species both increase the amount of genetic variability available to a population. The effects of chance, acting on small populations, may alter the way in which natural selction guides the course of evolution.

Revision questions

1. Discuss critically the following statement:

'Evolution was born in England, but found its home in Germany'.

2. Discuss the main features of post-Darwinian romantic period and agnostic period.
3. Discuss how did the idea of evolution mature during the 19^{th} century?
4. What is the theory of modern synthesis? Describe its origin growth and essence.
5. Discuss the main points of pre-Darwinian period.
6. Describe the evolutionary ideas of early Greek philosophers.
7. Write short notes on the following:

(i) Father of family trees; (ii) Argument of design

(iii) Lamarck's evolutionary idea; (iv)Natural selection;

(v) Weismann’s determinants.

Theories of organic evolution

Two explanations have been offered for the origin of the different forms of life – special creation and evolution. Many evidences have been presented in previous chapters to support the concept of evolution, and almost without exception modern biologists are convinced of evolution. Uptill now, the following convincing theories have been forwarded to explain the mechanism of evolution. The first of these theories, which is now only of historical singnificance, was that of Jean Haptiste de Lamarck (1744-1829).

Theory of inheritance of acquired characters (Lamarckism)

The theory of inheritance of acquired characters states that modifications which the organism acquires in adaptation to the environments which it meets during its lifetime are automatically handled down to its descendants and so become part of heredity. This theory was propounded by a renowned French…………………………………………………………Revolution. Lamarck spent the early part of his life as a botanist. Then at the age of 50 he turned his attention to zoology, particularly to the study of invertebrates (The terms “invertebrate” and “biology” have been coined by Lamarck). As a result of his systematic studies he became convinced that species were not constant but rather were derived from pre-existing species. This idea was in total conflict with the view of the period – that of fixity of species. As a result Lamarck’s views were challenged by most of the biologists of that time, particularly by Georges Cuvier. In 1809, Lamarck published philosophie Zoologique, which included his theory explaining the changes that occur in the formation of new types. Although, his views on evolutionary mechanism are outmoded now, he still occupies a very important place in the history of evolutionary thought. He was the first evolutionist to conclude that evolution is a general fact covering all forms of life. His evolutionary ideas can be discussed in brief as follows.

1. Internal forces of life tend to increase the size of the organism. New structures appear because of an “inner want” of the organism, i.e., the internal forces of life tend to increase continuously the size of an organism and its component parts.

2. Direct environmental effect over living organisms. The organs of an animal became modified in appropriate fashion in direct response to a changing environment.

3. Use and disuse. The various organs became greatly improved through use or reduced to vestiges through disuse.

4. Inheritance of acquired characteristics. Such bodily modifications, in some manner, could be transferred and impressed on the germ cells to affect future generation. Thus, inheritance was viewed by Lamarck simply as the direct transmission of those superficial bodily changes that arose within the lifetime of the individual owing to use or disuse (Volpe, 1985).

Thus, Lamarck believed that organic changes seen in animals were resulted by the influence of environment on the gradual changes of species due to their tendency to become more and more perfoect. According to him, when an animal's environment changes, its needs change, and this leads to special demands on certain organs. Organs used more extensively would enlarge and become more efficient. Conversely, an organ or organs, no longer used, would degenerate and atrophy. He postulated that such changed characteristics (acquired traits) would be transmitted to the offspring.

Examples of Lamarckism

1. The deer-like ancestor of giraffe lived in places (Africa) where the ground was almost invariably parched and without grass. Obliged to browse upon trees, it was continually forced to stretch upwards. This habit maintained over long periods of time by every individual of the race had resulted in the forelimbs becoming longer than the ……. Ones, and neck so elongated that a modern giraffe can raise his head to a height of eighteen fet without taking his forelimbs off the ground.
2. Ducks and other aquatic birds invaded waters from land in search of enough food, because food was scarce on land and these birds did not had power to fly. In water, the duck would stretch its toes apart to give more push during swimming. This new characteristic would be inherited, and the subsequent generation of duck would upon stretching their toes form a more defined web. Each generation would do the same, until the webbed foot seen on ducks today was fully formed. This would then be passed on from generation to generation, essentially unchanged once the perfected state was attained.
3. Falt fishes (deep sea fishes) present at the bottom of sea where there is no sunlight, led an inactive life, lying on one side of the body. The eye of that side

(lying towards bottom) migrated towards upper side and, thus, both eyes are on one side of the body.

4. The whales lost their hindlimbs as the consequence of the inherited effect of disuse.
5. The wading birds (e.g., Jacana) developed its long legs through generations of sustained stretching to keep the body above the water level.
6. Snakes have elongated body accompanied by loss of limbs. The continuous creeping through holes and crevices made limbs continuously useless for locomotion with the result that limbs become completely lost in snakes.
7. Eyes are reduced in moles since they live under ground. In cave animasl also, eyes might become functionless and might even disappear.
8. For plants Lamarck accepted the theory of his compatriot St.Helaire, who developed the notion that plant form is shaped by the combined effects of the environment.

There are other effects such as vestigial organs in living animals due to disuse. Claws in Carnivora, sensitive skin and tactile points on the ventral side of the body and callosities of plam in hard workers, exemplifying Lamarckian theory.

Significance of Lamarckism. Lamarckian theory was simple and it had some appeal, as it provided a way in which changes in organisms could come about. It was the first completely comprehensive mechanistic theory that was offered. Furthermore, it was the theory that lent itself to predictions and, therefore, to testing. Thus, Lamarckian theory enjoyed popular acceptance for near about 70 years, because it was exemplified by many common examples. Most persons know that exercise results in larger muscles.

Critical Analysis of Lamarck's Propositions

Lamarck defended his evolutionary theory vigorously until his death. For it, he suffered both social and scientific ostracism, but he had the courage of his convictions. However, he was criticized for the following reasons by the contemporary scientists during his lifetime and afterwards.

1. The first proposition of Lamarck suggests the tendency to increase in size. While the evolutionary trend in a certain groups of organisms may be associated with the

increase in size, there are many cases, where evolution proceeded not only without any increase in size but rather through a reduction in size. Many plants contradict this Lamarckian principle by showing such a reduction in size during their evolution. Many ferns and conifers which became extinct were gigantic trees and the more highly evolved flowering plants are really much smaller in size.

2. The second Lamarckian principle that new organs result from new needs, is quite manifestly false. In the case of plants, Lamarck believed that the environment acted directly upon the plant causing the production of such new characters as might adapt the plant to its environment. Acted through the nervous systems; in other words, the desire of the animal leads to the formation of new structures. In its crudest form this would mean that the man who mused "*Birds can fly, so why can't I ?* should have sprouted wings and taken to the air.

3. The third Lamarckian principle that organs will develop due to use and degenerate due to disuse, may be correct as far as growth of an organ within the lifetime of an individual is concerned. For example, it is a commonly observed fact that if muscles are put to use these would develop. Hwoever, this principle is meaningful only when it is studied in relation to the following fourth principle.

4. The fourth and final proposition of Lamarck was that the inheritance of characters acquired during the lifetime of the individual. This principle has been tested by many biologists who have devised many types of experiments for it and have found it entirely incorrect. Certain experiments which have discredited it are the following.

1. The noted German scientist August Weismann was the first person who for the first time made a definite distinction between heritable changes and those which cannot be inherited. In 1890, he performed some experiments to test if characters may disappear due to disuse. This he did by cutting the tails (mutilation) of white mice for more than 20 generations to see if this has any effect on tail length. The measuring of tail length of the offsprings of 20 successive generations, revealed that on average, the tails were not shorter. It means that, acquired character (cut tail) was not inherited.

2. Cast and Phillips performed transplantation experiments to show that environment has no effect on heredity. In one of the experiments they transplanted the ovary of a black female guinea pig into the body of white female guinea pig and the recipient female was mated with a white male guinea pig. They found that all the individuals

from this pair were black. This shows that the environment does not affect the heredity as has been suggested by Lamarck.

3. Loeb artificially fertilized the eggs of sea urchin by certain chemical stimuli and these parthenogenetically fertilized eggs produced the generations. The members of which possessed no parthenogenesis.
4. Boaring of ears and nostrils in Indian women has been continued as the tradition from centuries but their offsprings do not show any trace of holes in ears and nostrils.
5. The Chinese bound the feet of their women for many generations, yet this has not resulted in any modification of the feet of present-day Chinese women.
6. Jewish boys and also Muslim boys) have been circumcised for thousands of years, but this has not resulted in a tendency toward the reduction of the prepuce in them.

All the aforesaid cases and experiments clearly showed that Lamarckian theory is not tenable.

Neo-Lamarckism

The evolutionists who support the Lamarckian doctrine of inheritance of acquired characters come under the heading of neo-Lamarkian. Among the neo-Lamarckians notable supporters are Cope (1840-1897), Giard (1846-1908), Packard, Spencer and MeBride tried to modify Lamarckism in order to make it acceptable. These neo-Lamarckians considered that adaptation is universal. It arises as a result of casual relationship of structure, function and environment. Changed environmental conditions alter habits of organisms, hence, in response to new habits, organisms acquire new structures in place of old structures. Consequently, variations among animals have become distinct. These variations have finally become engrained in the heredity of the race. The kind of argument is, in fact, a modified version of Lamarckian principles, because it has omitted Lamarck's view that of a general perfecting tendency in evolution. It stressed direct action of environment on organic structure. For example, according to neo-Lamarckians, the development of fur on skin by animals as protection against cold weather, is the consequence of changed environment from warmer to colder state only. But if the environment reverts to normal state, the fur would also disappear. To account it, neo-Lamarckism included also the effects of use and disuse. Based on this, neo-Lamarckians have rejected natural selection as the sole mechanism of evolution. On the other hand, they believed in the interplay of structure, function and environment as the whole truth of evolution. However, no evolutionist of today is adherent to neo-Lamarckism.

The following are the evidences of transmission of acquired characters.

1. McDougall conducted learning experiments in rats and from the results obtained he tried to suggest that learning as an acquired character can be inherited. During his experiments, the rats were dropped into a tank of water from which there were two exits, one lighted and one dark, but no always the same one. A rat leaving by the lighted exit received an electric shock, while one leaving by the dark exit received no shock. Thus, the number of trails required fro a particular rat to learn always to select the dark exit constituted a measure of the speed of learning. These rats were then bred and their descendants were similarly studied. It appeared that the speed of learning increased from generation to generation and, so, McDougall concluded that learning, an acquired trait par excellence, is inherited.

Some serious criticisms have been raised against McDougall's experiments. It is suggested that genetic constitution of rats was not controlled. Moreover, in the progeny of control rats not subjected to these learning tests, the learning habit was found to change the same way as in the treatments. So, it was said that some unanalyzed changes in the technique may be responsible for the recorded increase in the speed of learning. Lastly, the most strange aspect about McDougall's experiment is, that when the same experiments, when repeated in other laboratories, never gave similar results.

2. F.B. Summer reared the white mice at 20°C to 30°C, because of which they developed longer bodies, tail and feet. When these mice were bred at lower temperatures their offsprings showed their normal proportions of size.

3. Lindsey subjected certain cold-blooded animals, warm-blooded animals and plants to unusual environmental conditions and found that environment affects certain changes in these organisms and these acquired modifications were inherited up to some extent in their offsprings.

4. Guver and Smith induced the hereditary changes in the eyes of fetuses of rabbit by simply destroying lens of living female with a needle *in situ*. The antilens serum has been produced in the blood of these animals.

5. Gr…….. and Detie….. reared rats in cages placed on rotating table for several months. Consequently, they became adapted to the rotating condition to such an extent that when rotation was stopped, they showed signs of ………….(dizziness) and other physiological conditions. This condition was inherited for several generations.

6. Various acquired diseases described by Brown Sequard are inherited from generation to generation. For example, exophthalmia was caused in parents by injuring in brain the restiform body. This disease was, later on, inherited to several generations. The other diseases such ……………. And …………….. are produced by injury to the restiform body near the rib of calamus. Later, these were found to be inherited.

7. Kammerer worked on the tailed amphibian proteus anguinus which lives in complete darkness in the water of underground cave. It was blind and colourless. He brought proteus in daylight, due to which it became coloured (brown and blacm), which passed on to its progeny. Eyes of Proteus also developed normally in daylight. This showed that any change in environmental conditions, induced changes in the animals and the acquired characters were inherited.

Further, Winterbert (1962) has tried, with some success, to give Lamarckism a chemical basis by involving the capacity of everyliving being to react to environmental changes and aggressions. According to Winterbert, nothing is acquired by the living organism that is not the response of internal factors to an outside influence or ethological change. Living means reacting, never undergoing. Above all, it means not waiting around for a fortunate chance occurrence to save the situation (Grasse, 1977).

2. Theory of Natural Selection (Darwinism)

More than a century ago in 1859, Charles Robert Darwin (1809-1822) gave the biological world the master key that unlocked all previous perplexities regarding the mechanism of organic evolution. His natural selection theory can be compared only with such revolutionary ideas as Newton's law of gravitation and Einstein's theory of relativity (Volpe, 1985). The concept of natural selection was explained clearly and convincingly by Darwin in his masterpiece – The Origin of Species (The full title of the book was On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life). This epoch-making book was the fruition of more than twenty years of meticulous accumulation and analysis of facts. The first edition of the book, some 1,250 copies, was sold out on the very day it appeared, November 24, 1859. This book had opened a Pandora's box; it was immediately both acidly attacked and effusively praised.

Charles Robert Darwin was born on February 12, 1809 in Shrewsbury, England. His personality was not the type that one associates with a successful man, let alone a greatone. While his family held a distinguished reputation (his grandfather Erasmus Darwin was a noted poet scientist, his father Robert, was an eminent physician, and his mother was Wedgwood, the family of pottery fame), Charles was a singularly undistinguished individual whose contemporaries viewed him as just a normal, gentle, cautious, morally upright English lad. After completing his early education at Shrewsbury, Charles Darwin was sent at the age of fifteen to study medicine at the University of Edinburg. His medical interests soon dissipated, and he became something of an academic rover, studying a little science, but finding it too formal for his tastes.

His, father then suggested him to become a clergyman, which appealed him. So, he got transferred, after two years, to Christ's college, Cambridge University, to study theology. There too he spent his energy in card-playing and drinking. He was obviously not ready for clerical commitments.

He did manage, however to find some constructive interest in naural history and in collection of natural objects of all kinds. His friendship with Professor John Stevens Henslow, botanist at Cambridge,stimulated an interest in botany, while his exposure to the geologist Sedgwick stimulated an interes in geological explorations. Through the recommendation of Dr.Henslow, Darwin accepted a most inauspicious position as a naturalist without pay aboard the H.M.S. Beagle. The ship was to spend five years in exploration, mostly of South America. Darwin was to provide a collection of material representing the natural composition of the areas visited.

The ship Beagle left Plymouth on December 27, 1831 and visited many islands of the Atlantic ocean, some coasts of South America and some islands of south Pacific, of which Galapagos Islands are the most important. During his five years (from 1831 to 1836) on the Beagle, Darwin recorded almost everything he observed and sent an enormous amount of material back to England.

Upon his return home in 1836, he spent two years writing a book of his experiences, The Voyage of the Beagle. In 1838, he got married and later had two daughters and five sons. From 1838 to 1841. Darwin acted as the secretary of geological society, where he came in contact with an eminent geologist, Charles Lyell. It was not until 1844 that Darwin developed his idea of natural selection in an essay, but not for publication. He showed the manuscript to ……… who encouraged him to prepare a book. Darwin still took no steps toward publishing his views. Perhaps he was well aware of how Lamarck's earlier theory had been received and water all possible supporting evidence before publishing his own new theory. It appears that Darwin might not have prepared his famous volume had not a fellow naturalist in the Dutch East Indies, Alfred Russel Wallae (1823-1913), independently conceived of the idea of natural selection. Wallace had traveled widely in tropical South America and Sourheastern Asia. For studying the flora and fauna of these regions. Walla…. Also inspired by reading Malthus's essay, and the idea of natural selection came to him in a flash of insight during a sudden fit of malarial fever (this happened in February, 1858, when he was workin gin the island of Ternate in Indonesia). In June of 1858, Wallace sent Darwin a short essay "On the tendency of varieties to depart indefinitely from the original type", and asked him if he thought of sufficient interest to present it to the Linnaean Society. We can easily imagine Darwin's amazement upon receiving Wallace's essay. Upon the insistence of the geologist Charles Lvell and the botanist Joseph Hooker, Darwin prepared an abstract of his conclusions for joint publication with Wallace's essay. Wallace's essay and a portion of Darwin's manuscript, each containing remarkably similar views, were read simultaneously before the Linnaean Society in London on July 1, 1858. The joint reading of the papers stirred little interest.

Darwin then laboured for eight months to compress his voluminous notes into a single book, which he modestly called "only an Abstract". Wallace shares with Darwin the honour of establishing the mechanism by which evolution is brought about, but it was the monumental The Origin of Species with its impressive weight of evidence and argument that left its mark on mankind. The fact that this book, written by Darwin alone, exposed the public at large to the theory of natural selection, and that Wallace was still in Indonesia at the time of the controversy and could not champion Darwin's defence by adding his own views, brought the public to associate only Darwin with this theory. Thus the theory is called the Darwinian, rather than the Darwin-Wallace theory of evolution.

Darwin died on April 19, 1882, when he was 73 years old. During his life, he write numerous books, journals, etc. some of them can be listed in Table (5-1).

Table 5-1 Some publications of Charles Darwin.

Title of book	Year of Publication
1. Journal of researches	1839
2. The structure and distribution of coral reefs	1844
3. Geological observations on South America	1846
4. A monograph on Cirripedia (Barancles, living and fossil, 4 volumes)	1851, 1854
5. Origin of species	1859
6. The fertilization of orchids	1862
7. The variation of plants and animals under domestication	1868
8. The descent of man	1871
9. The expression of the emotions in men and animals	1872
10. insectivorous plants	1875
11. The effects of cross and self-fertilization in the vegetable kingdom	1876
12. different forms of flowers on plants of the same species	1877
13. The power of movement in plants	1880
14. The formation of vegetable mould through the action of worms	1881

Facts that influenced Darwin's Thoughts

During the period in which Darwin massed his evidence and developed his natural selection theory, many things affected his thinking, of which three, in particular, deserve mention here.

1. During his voyage, Darwin took only a few books on board. One of them was the newly published first volume of Charles Lyell's Principles of Geology, a parting gift from his Cambridge mentor, John Henslow. ; In this book, Lyell challenged the prevailing belief that the earth has been created by a divine plan merely 6,000 years ago. On the contrary, the earth's mountains, valleys, rivers, and coastlines were shaped not by Noah's Flood but by the ordinary action of the rains, the winds, earth

quakes, volcanoes, and other natural forces. Darwin was impressed by Lyell's emphasis on the great antiquity of the earth's rocks, and gradually came to conclusion that the characteristics of organism as well as the face of the earth could change over a vast span of time.

2. In the late 1830s, Darwin attended the meetings of animal breeders and intently read their publication. Animal breeders were conversant with the variability in their pet animals, and dwelled on the technique of artificial selection. Thus, the breeders selected and perpetuated those variant types that interested them or seemed useful to them. The breeders, however, had only vague notions as to the origin, or inheritance, of the variable traits.

3. Having explained the origin of variation (although incorrectly), Darwin wondered how artificial selection (a term familiar then only to animal breeders) could be carried in nature. There was no breeder in nature to pick and choose. In 1838, Darwin found the solution in Thomas Robert Malthus book 'An Essay on the Principle of Population' in which Malthus asserted that the reproductive capacity of mankind far exceeds the food supply available to nourish an expanding human population. Humans compete among themselves for the necessities of life.

It, thus, occurred to Darwin that competition existed among all living things. Darwin then envisioned that the "Struggle for existence" might be the means by which the well-adapted individuals survive and ill-adjusted are eliminated. Darwin was the first to realize that perpetual selection existed in nature in the form of natural selection. In natural selection, as contrasted to artificial selection, the animal breeder or horticulturist is replaced by the conditions of the environment that prevent the survival and reproduction of certain individuals. Thus, natural selection is a term serving to inform us that some individuals leave more offspring than others. It is not purposeful or guided by a specific aim; it does not seek to attain a specific end.

Theory of Natural Selection

Darwin-Wallace explanation of the way in which evolution occurs may be generalized as follows – the change in species by the survival of an organismal type exhibiting a natural variation that gives it an adaptive advantage in a an environment. Thus, leading to a new environmental equilibrium, is evolution by natural selection".

Thus, natural selection is a continuous process of trial and error on a gigantic scale, for all of living matter is involved. It includes the following elements.

1. The universal occurrence of variation. Variation is the characteristic of every group of animals and plants and there are many ways in which organisms may differ. (Darwin and Wallace did not understand the cause of variation, and assumed it was one of the innate properties of living things. We now know that inherited variations are caused by mutations.)

2. An excessive natural rate of multiplication. Every species, in the absence of environmental checks, tends to increase in a geometrical manner. If a population of a given species doubles in one year and if there are no checks on its increase, it will quadruple the next year, and so on. Such a great reproductive potential of different species may be easily observed in nature. It has been estimated that a common Atlantic coast oyster may shed as many as 80 million eggs in one season. A salmaon produces 28,000,000 eggs in a season. A single pair of English sparrows would be the ancestors of over 275 billion individuals in 10 years if they and their descendants could reproduce at their natural rate without any check. Darwin calculated that even a pair of elephants which are about the slowest breeding animals known, could in the absence of any checks, have 29 million descendants at the end of 800 years.

Thus, more organisms of each kind are born than can possibly obtain food and survive. Since the number of each species remains fairly constant under natural conditions fairly constant under natural conditions, it must be assumed that most of the offsprings in each generation perish. If all the offsprings in each generation perish. If all the offsprings of any species remained alive and reproduced they would soon crowd all other species from the earth.

3. Struggle for existence. Since more individuals are born than can survive there is an intraspecific or interspecific or environmentals are born than can survive there is an intraspecific or interspecific or environmental struggle for survival, a competition for food, mates and space. This contest maybe an active kill-or-be-killed struggle or one less immediately apparent but no less real, such as the struggle of plants or animals to survive drought or cold.

4. The consequent elimination of the unfit and the survival of only those that are satisfactorily adapted. Some of the variations exhibited by living things make it easier for them to survive; others are handicaps which bring about the elimination of their possessors. This idea of "The survival of the fittest" is the core of the theory of natural selection.

5. The inheritance of the mutations or recombinations that make for success in the struggle for existence. The surviving individuals will give rise to the next generation and, in this way, the "successful" variations are transmitted to the succeeding generations. The less fit will tend to be eliminated before they have reproduced.

Successive generations in this way tend to become better adapted to their environment; as the environment changes further adaptations occur. The operation of natural selection over many generations may produce descendants which are quite different from their ancestors, different enough to be separate species. Furthermore, certain members of a population with one group of variations may become adapted in a different way, or become adapted to a different environment. In this way two or more species may arise from a single ancestral stock.

Critical Analysis of Darwinism

With the appearance of Darwin's book in 1859 a veritable storm of controversy broke out. Many people became interested in his presentation; the arguments for and against that developed were often heated. The concept of natural selection was opposed primarily on ethical and religious grounds. But the sheer weight of evidence and the logic of Darwin's presentation finally convinced a majority of the educated people. The concept of evolution affected many aspects of man's thinking and it will continue to do so. However, the idea of "The survival of the fittest" of Darwinism has been erroneously interpreted by many to mean survival only by "Tooth and claw", an interpretation that has led in many cases to rationalizations for the attitude of "Every man for himself" in social and economic affairs. In fact, the process of evolution is not all "Tooth and claw", plants have evolved too, and there is no bloody competition between them. In all forms, plants and animals alike, a part of the selective process involves the action of inanimate nature – factors such as drought, storm, moisture, temperature, etc. Active competition between organisms is a part of the process, but as Walder Allee and others have pointed out, cooperation has, in many cases, been involved in survival Natural selection results in the survival of those forms that are best integrated with the various factors of the environment in which they live.

Certain other demerits of Darwinism. Besides its numerous merits, Darwinism has following demerits.

1. Darwinism or natural selection theory does not account for the beginning of organs, which may appear at first as the veriest rudiments having as yet no selection value. In other words, it remains concerned with the survival of the fittest, but not for the arrival of the fittest. Thus, to give rise to such specializations as elaborate mimicry, or the electric organ of the torpedo, etc., which are of apparent advantage only in the perfected state, natural selection, acting only upon minute gradations toward perfection, seems inadequate. The same is true of so complex a specialization as the eye and its function in the vertebrates or in the insects and crustaceans.
2. Over-specialization in certain cases like extinct Irish deer in which huge antlers outweigh the entire skeleton, or the immense spiral tusks of the Jefferson mammoth, or the minute fidelity of certain mimicking insects such as *Kallima,* or huge dinosaurs of Mesozoic – all cannot be explained on the basis of continuous variations and natural selection. These organs or body structures should not have reached such a harmful stage, if natural selection was operating. However, such cases of overspecializations have been explained by Darwin on the basis of discontinuous variations or "sports" which, according to him, do not play any role in evolution.
3. Natural selection cannot account for degeneracy. To say an organ is no longer useful and, hence, disappears, is to state the effect and not the cause. If under changed conditions a character built up by natural selection becomes a menace, the reversal of selection can accomplish its removal but this will not suffice where the characteristic is an indifferent one.
4. One of the classical objections to natural selection is that new variations would be lost by "dilution" as the individuals possessing them bred with others without them. We now know that although the phenotypic expression of a gene may be altered when it exists in combination with certain other genes, yet the gene itself is not altered and is transmitted to succeeding generations.
5. Darwin indirectly accepted the Lamarckian idea of inheritance of acquired characters in the form of pangenesis hypothesis, which cannot be accepted in the light of knowledge of genetics made available in the present century.

In year 1875, Galton made pangenesis hypothesis untenable by presenting several experimental proofs. He and others at later time made a series of experiments involving blood transfusions and later, transplants of ovaries between black and white varieties of rabbits and chickens. Gametes produced by the transplanted ovaries were consistent with the phenotype of the individual in which the ovary originated and not with the animal currently carrying the ovary. Blood transfusions had no effect on the gametes produced. These experiments readily demonstrated that the pangenesis hypothesis was incorrect.

6. Lastly, some persons have objected natural selection because it is essentially a materialistic doctrine, depending as it does purely on the laws of chance.

Neo-Darwinism

Neo-Darwinism is a modified from of Darwinism. The Neo-Darwinians like T.H. Huxley and Herbert Spencer of England, D.S. Jordan and Asa Gray of United States, and E. Haeckel and A. Weismann of Germany believed that natural selection has accounted everything that is involved in evolution. In their dogmatic belief they went further to secure enough proof towards natural selection.

Certain Neo-Darwinians, such as Weismann and his followers rejected Darwin's theory except its principal element of natural selection. These Neo-darwinians, though distinguished between germplasm and somatoplasm of living organisms in their germplasm theory, yet they could not appreciate the role of mutations in evolution. While Darwin believed that the adaptations result mainly by a single source, i.e., natural selection, Neo-Darwinians thought that adaptations result from multiple forces and natural selections is only one of these many forces. Neo-Darwinians also believed that characters are not inherited as such but there are character determiners, the determinants or biophores, which control only the development. The ultimate charcter would result out due to the interaction of the determiners, activity of the organism and the environment during development. Thus, Neo-Darwinism was incomplete and partly wrong because it lacked present understanding of genetics.

Maturation of Neo-Darwinism into Modern Synthesis

During later half of 19^{th} century and at the beginning of 20^{th} century, many important ideas have been forwarded for explaiing the inheritance of variations from one generation to successive generations and also for explaining evolution. Thus, the rediscovery of Mendel's laws of heredity by Correns, de Vries, and Tschermak in

1900 made it clear that (1) the factors given to the offspring by the parents do not "mix" but are segregated, and (2) in more than one pair of contrasting characters are considered in the same cross, the factors responsible for these are inherited independently. Likewise, mutation theory of evolution of Hugo de Vries (1886, 1887) stated that new species arise by sudden changes or steps called mutations rather than by gradual processes. According to de Vries, it was mutations and not selection that should be considered as the primary factor in evolution. Further, the studies of Wagner (1868) suggested the role of geographic or spatial isolation in the formation of every species, race, or tribe of animals or plants on the earth. Even more, certain population geneticists realized that the actual physical struggle between animals for survival or the competition between plants for space, sun and water is much less important as an evolutionary force than Darwin believed. The evolution of any given kid of organism occurs over many generations during which individuals are born and die, but the population has a certain continuity. Thus, the unit in evolution is recognized not the individual but rather a population of individuals.

A population of similar individuals living within a circumscribed area and interbreeding is termed as a deme or a genetic population. The next larger unit of population in nature is the species, composed of a series of inter-grading demes. Further, in a population, the relative frequencies of the genes will remain constant from one generation to the next (1) if the population is large, (2) if there is no selection for or against any specific gene or allele, i.e., if mating occurs at random, (3) if no mutations occur, and (4) if there is no migration of individuals into or out of the population. The operation of the Hardy-Weinberg Principle will result in maintaining a given gene frequency in a population. Thus, a population undergoing evolution is one in which the gene pool is changing from generation to generation. The gene pool is the sum total of all the allelic genes in a population, and the gene pool of a given population may be changed (1) by mutation, (2) by hybridiation, that is by introduction into the population of gene from some outside population, or(3) by natural selection. Recombination brought about by crossing over and by the assortment of chromosomes in meiosis may also lead to new combinations of genes and phenotypes with some specific advantage or disadvantage for survival that would be reflected in a change in the gene pool.

All these modern understandings in cytology, genetics, cytogenetics, population genetics, and evolution gave away for the formulation of a coherent theory called modern synthesis around 1930s, by S.Wright, H.D. Muller, Th.Dobzhansky, R.B. Goldschmidt, J.S. Huxley, R.A. Fisher, J.B.S. Haldane, Ernst Mayr and G.I. Stebbins.

3. MODERN SYNTHETIC THEORY

The modern synthetic theory of evolution involves five basic processes – mutations, variations, heredity, natural selection and isolation. In addition, three accessory processes affect the working of these five basic processes. Migration of individuals from one population to another as well as hybridization between races or closely related species both increase the amount of genetic variability available to a population. The effect of chance acting on small populations, may alter the way in which natural selection guides the course of evolution (Stebbins, 1971).

1. Mutation. Alteration in the chemistry of gene (DNA) is able to change its phenotypic effect (i.e., nature of polypeptide) is called point mutation or gene mutation. Mutation can produce drastic changes or can remain in significant. There are equal chances of a gene to mutate back to normal. Most of the mutations are harmful or deleterious and lethal but not all. Most of the mutant genes are receissive to normal gene and these are able to express phenotypically only in homozygous condition. Thus, point mutations tend to produce variations in the offspring.
2. Variation (Recombination). The nature of genetic variations caused by reshuffling of genes during sexual reproduction (recombination) was very little known at the time of Darwin. Recombination – that is, new genotypes from already existing genes – is of several kinds: 1. the production of gene combinations containing in the same individual two different alleles of the same gene, or the production of heterozygous individuals (meiosis); (2) the random mixing of chromosomes from two parents to produce a new individual (sexual reproduction); (3) the mixing of a particular allele with a series of genes not previously associated with it, by an exchange between chromosomal mutations such as polyploidy, deletion, duplication, inversion and translocation also result in variation.
3. Heredity. The transmission of characteristics or variations from parent to offspring, is an important mechanism of evolution. Organisms possessing hereditary characteristics that are helpful, either in the animal’s native environment or in some

other environment that is open to it, are favoured in the struggle for existence. As a result, the offsprings are able to benefit from the advantageous characterstics of their parents.

4. Natural selection. Natural selection brings about evolutionary change by favouring differential reproduction of genes. Differential reproduction of genes produces change in gene frequency from one generation to the next. Natural selection does not produce genetic change, but once genetic change has occurred it acts to encourage some genes over others. Further, natural selection creates new adaptive relations between population and environment, by favouring some gene combinations, rejecting others and constantly moulding and modifying the gene pool.

The workings of natural selection are exceedingly complex because of the range of organizational levels at which it functions. Selection discriminates among available reproducible biotic entities to produce more efficiently adapted units. Natural selection operates upon every stage in the life history of an organism it produces non-random differential reproduction of biological units and may affect any biotic entity from the molecular to the community level. Examples of levels at which natural selection makes differential discrimination are the following : intermolecule, intergene, interchromosome, intergamete, interindividual (Darwinian selection), interdemic, interracial, interspecific and interecommunity. Darwinian selection may result from differential natality among others.

5. Isolation. Isolation of organisms of a species into several populations or groups under psychic, physiological or geographical factors is supposed to be one of the most significant factors responsible for evolution. Geographical isolation includes physical barriers such as high mountains, rivers, oceans and long distances preventing interbreeding between related organisms. Physiological …………….. help in maintaining the individuality of the species, since these isolations do not allow the interbreeding amongst the organisms of different species. This is called reproductive isolation.

Speciation (origin of new species). The populations of a species present in the different environments and are segregated by geographical and physiological barriers, accumulate different genetic differences (variations) due to mutations (both point and chromosomal), recombination, hybridization, genetic drifts and natural selection.

These populations, thus, become different from each other morphologically and genetically, and they become reproductively isolated, forming new species.

4. WEISMANN'S GERM PLASM THEORY

AUGUST Weismann (1834-1914) was the German neo-Darwinian biologist who proposed the germ plasm theory which was published in the book *Das Keimplasma*. He asserted the continuityof germ plasm as the main criterion for inheritance of characters. All the heritable variations have their origin in germ cells and a new type of organisms arise only from changed type of germ cells. Thus, Weismann's germ plasm theory rejected outrightly the Lamarckian concept of inheritance of acquired characters and Darwin's pangenesis hypothesis. While Darwin's pangenesis hypothesisis centripetal in nature, germ plasm theory is centrifugal. The main points of germ plasm theory are the following.

1. Existence of somatoplasm and germ plasm. All substances a of an organism can be divided into two parts, the germ plasm (which is theprotoplasm of germ cells such as sperms and ova) and somatoplasm (which is the protoplasm of somatic or body cells). The germ plasm is thought to be the actual vehicle of heredity; it is not affected by any influence either from the body or the external environment. The somatoplasm is thought not to play any role in heredity.

2. Continuity of germ plasm. In reproduction,a portion of germ plasm is derived from each parent, the paternal sperm and maternal ovum, each of which has an equivalent share in the inheritance. These combine to form the fertilized egg. The contents of egg divide and redivide to produce both body (soma) and the germ plasm of new individual. When the egg divides in the first or subsequent cleavages each daughter cell receives an equal share of germ plasm, and this holds true for all cells which go to form the adult body. None of the somatic cells, however, has any share in subsequent generations but only the germ cells which are derived directly from the original egg. It follows, therefore, that there is a continuous stream of germ plasm from generation to generation, which the somatic cells contribute nothing. Hence, only those mutations which are germinal in their origin can possibly be handed down, and as the hereditarystream of germ plasm is already set apart before the adult body comes in use, one cannot see how any modification impressed upon the body through use or disuse can possibly become a part of the organism's heritage. Thus, germ plasm is immortal since theproduct of the division of previous germ cells. Soma

(somatoplasm) is mortal, it perishes with the death of an organism. Germ plasm can produce the somatoplasm but somatoplasm cannot produce the germ plasm

3. Architecture of germ plasm (Concept of determinants). In 1904, Weismann proposed that every distinct part of an organism is represented in the sex cell by a separate particle – the idioplasm or determinant. Each determinant is supposed to be made up of still smaller units called biophores. The sum total of determinants would represent the parts of the adult organism with all their peculiarities. The complete set of determinants would be handed down from generation to generation, which would account for hereditary transmission of characters. The determinants, according to Weismann, are localized in the chromosomes of the nucleus, just as are the genes of modern genetics.

Further, during the cleavage of the egg, the various determinants become segregated into different cleavage cells or blastomeres. Eventually each blastomere would contain only determinants of one kind and it would be forced to develop or differentiate in a specific way in accordance with the determinants present in it. Only the cells having the sex cells among their descendants tend to preserve the complete set of determinants, since these would be necessary for directing the development of next generation.

4. Paralled induction. In apposition of Lamarck's inheritance of acquired characters, Weismann had introduced the idea of paralled induction. According to this concept, the stimulus affects simultaneously the germ plasm and the body (soma). He proposed the occurrence of an internal stimulus which affects germ cells and results in heritable variations. The stimulus, according to Weismann, is the nourishment which is necessary for determinants and biophores (since they are living units). Those determinants or particles which obtain better nourishment are fast-growing and stronger than those getting less nourishment. Correspondingly these particles tend to produce either strong or weaker part or organ in organisms. Thus, Weismann assumed a struggle for existence between better nourished determinants, and there lies the causes of appearance and disappearance of variations.

Objections to Weismann's Germ plasm Theory

Germ plasm theory is criticized mainly for its speculativeness (i.e. it lacks any experimental support), and also for its idea of determinants and their segregation during cleavage and for its failure for explaining causes of regeneration and asexual reproduction.

Significance of Weismann's Germ plasm Theory

1. It made ground for the understanding of the concept of particulate inheritance of Mendel.
2. It provides some clue about genes (determinants) which reside in chromosomes and represent some part of animal body.
3. According to the embryologists, the greatest contribution of this theory is that it proposes the division of germ plasm and somatoplasm during cleavage of the zygote during the embryogenesis (e.g., cleavage of Ascaris).
4. The idea of continuity and immortality of germ plasm prepares the ground for the continuity of chromosome or DNA from one to next generation.

5. MUTATION THEORY

Darwin recognized two types of variations in nature, viz., 1. Minor and continuous and 2. major and discontinuous ones. He considered minor and continuous variations important in the origin of new species. Darwin considered major variations quite insignificant and for them he coined the terms sports or salutatory (L. saltare = to leap) variations. In England, William Bateson collected a large number of species showing discontinuous or salutatory variations. In Holland, Hugo de Vries (1901) gave much importance to these variations and proposed that new species arise not by the accumulation of minor and continuous variations through natural selection, but by the suddenly appearing salutatory variations for which he coined the term mutation (L.*mutare* = to change). De Vrie's mutation theory states that species have not arisen through gradual selection accumulated for hundreds or thousands of years but have appeared by sudden jumps (saltations) and transformation.

Dutch botanist Hugo de Vries (1848-1935) was born at Haarlem; he studied at Leiden, Heidelberg and also Wurgberg. He was a university lecturer at Amsterdam and also held the professorship of Plant Physiology. His book entitled "*Die Mutation Theorie*" was published in 1901. In this book, de Vries proposed the mutation theory in order to explain the mechanism of evolution. This theory was based on his observations on evening primrose, Oenothera lamarckiana grown in a field near Amsterdam. He studied this plant in wild form for many years continuously and observed certain spontaneous changes in some of these wild plants. These plants differed considerably in stem height, flower colour andleave's shapes. He observed that these changes were heritable and ultimately led to several new varieties. He succeeded in cultivating all these new varieties and named them as mutant varieties. In fact, he selected for his breeding experiments two mutant varieties –*Oenothera laevifolia*, characterized by smooth leaves and *O. brevistylis* characterized by short styles. And he observed that these features were breeding true and so, he regarded these mutant strains as the distinct species. Thus, de Vries recognized thefollowing seven distinct species of *Oeonothera.*

1. O.gigas. Giant and stout plants with large flowers and deep green leaves.
2. O.rubrinervis. Fruit red veined, leaves pale green and stem slender and brittle.
3. O.oblonga. dwarf and weak plants with oblong leaves.
4. O. albida. Weak plants with whitish-pale leaves.
5. O.laevifolia. Leaves narrow and smooth and pale flowers.
6. O.brevistylis. Round leaves, flowers with very short style and flattened stigma.
7. O. vanilla. Very short or dwarf variety (20-30 cm); leaves sessile.

Of these varieties which arose suddenly from the parental variety *O. lamarckiana*, the *O.gigas* and *O.rubrinervis* were progressive varieties; the *O.laevifolia*, *O.brevistylis* and *O. vanniella* were retrograde varieties and *O. oblonga* and *O. albida* were weak and inconstant forms and could be grown only under strict protection.

Characteristics of Mutation theory

De Vries's mutation theory has the following characteristics.

1. Mutations appear from time to time among the organisms of a naturally breeding species or populations. The organisms with mutations are called mutants. These mutants are clearly distinct from their parents.
2. Mutations are heritable and form new species. They do not disappear by crossing.
3. Mutations arise suddenly in one step, i.e., new species arise suddenly in one step and not gradually.
4. Mutations occur in all possible directions and may be advantageous or disadvantageous.
5. Unsuitable mutants are destroyed by natural selection.

6. Mutations appear full-fledged and, hence, there is no question of incipient stags in the development of an organ.

Types of Mutation

Mutation in a broad sense include all those heritable changes, which alter the phenotypeof an individual. In genetics, much investigations have been done in the phenomenon of mutation by Morgan, ……………………………………….etc. according to their work; following two types of mutations have been recognized.

1. Gene……………………………… Any chemical change in a particular gene leading to the appearance of different alleles and affecting the viability or phenotype of the organism (for details see Chapter 16 in Part II – Genetics).
2. Chromo…………………. Any change in the number and structure of the chromosome or chromosome sets, leading to …………(e.g., polyploidy, trisomy) and chromosomal aberrations (e.g., translocation) respectively (for details see Chapter 14 and 15 in Part II-Genetics).

Advantages of Mutation Theory

The advantages of de Vrie's mutation theory are the following:

1. The mutation theory describes the importance of mutation in selective value of organisms.
2. The mutation theory explains the occurrence of evolutionary changes within short period in contrast to natural selection (which describes slow and continuous variations).
3. Mutation theory explains the absence of connection links as no criteria against evolution but its possibility exist.
4. Occurrence of mutations large and divergent direction removes the possibility of species disappearance by crossing.
5. Since mutations appear fully formed from the beginning, there is no difficulty in explaining the incipient stages in the development of organs.
6. Mutation is of great service to breeders in developing new useful varieties.

Objections to Mutation Theory

1. The mutation theory was unable to explain the presence of flightless birds on oceanic islands.

2. It could not explain the existence of discontinuity in distribution among individuals.
3. Many mutations, described by de Vries in *Oenothera lamarckiana*, are now known to be due to certain numerical and structural changes in the chromosomes. For instance, "*gigas*" mutant of *O. lamarckiana* was later found to be due to polyploidy.
4. Mutation theory alone could not explain evolution. It, however, provided raw material for other forces to act upon it and bring about evolutionary changes.

Revision questions

1. What is Lamarckism? How does it differ from Darwinism?
2. Define and describe the hypothesis of "Inheritance of acquired characters".
3. Explain the modern concept of evolution and discuss how does it support Darwinism.
4. Give an account of Lamarckism. Give various evidences in support and criticism of it.
5. Write an essay on the theory of natural selection.
6. Describe Darwinism and explain the postulates of Neo-Darwinism.
7. What is modern synthetic theory? Explain it.
8. Describe those circumstances which helped Darwin in the formulation of natural selection theory.
9. Write short notes on the following:

a. Neo-Darwinism; and
b. Germ plasm theory
c. Modern synthetic theory.
d. De Vries's mutation theory;

SELECTION IN ACTION

Darwin clearly saw the importance of selection as a prime evolutionary force in the natural world. He cites countless examples of adaptation through selection, and thousands of new cases are recognized each year. But one problem in discussing selection has usually been the correlation of experimental studies under controlled laboratory conditions with what actually happens in nature. During the last forty years a series of studies combining experiment and observation in the field have been carried out on the problem of industrial melanism in moths; they provide an exciting insight into the operation of selection under natural conditions. These brilliant investigations were undertaken originally by R.A. Fisher (1929) and E.B. Ford (1940, 1964, 1971) and more recently by H.B.D. Kettlewell (1956, 1961, 1973), all of Britain.

1. Melanism in Moths or Industrial Melanism

With increasing industrialization due to industrial Revolution in Britain, entomologists began to record the replacement of light or grey colours in many different species of moths with dark or black ones (called melanics). A particularly good example is the peppered moth, *Biston betularia*, where the dark form in referred to as carbonaria. The following facts were exceptions, all corals require water of maximum purity and salinity. They are never found near the mouths of large rivers such as the Amazon (South America), the Orinoco (Venezuela) and the Mississippi (USA), where the salinity is greatly changed. Great barrier Reef on the cast coast of Australia acts as barrier. It hinders the dispersal of animals.

Further, like the salinity the depth of ocean acts as an effective barrier. Thus, animals living on the surface and mid-water cannot move to deeper waters because their bodies would not be able to withstand the increase in atmospheric pressure and decrease in light. Similarly, animals from depth cannot survive in surface waters because their bodies will not be able to withstand the decrease in pressure and increase in light intensity.

6. Biological Barriers

the biological barriers include preferential type of food. Certain animals depend on particular type of plants for their food, shelter and breeding and if those plants are not present or removed, then that particular species would gradually perish. However, if such plants are available in some nearby area they would migrate. Similarly, certain animals act as parasites and others as predators. Of these animals availability of the host animals and the prey respectively, is most essential for the survival. If the host species or prey are destroyed or removed from an area, the parasites and predators would either disperse or become extinct.

Revision Questions

1. Describe the various barriers, which restrict the zoogeographical distribution of animals.
2. What are the barriers? How do they affect the distribution of animals in space?
3. Discuss the role of barriers in the dispersal of animals.

ORIGIN OF LIFE

The kind of evolution which includes evolution from atoms and molecules to simple and then complex sub stances and from these to still more complex ones capable of self-duplication, has been termed as inorganic molecular or chemical evolution. The most significant out comes of chemical evolution are the origin of biologically important macromolecules such as proteins and nucleic acids (RNA and DNA), origin of life and an environment to sustain life on the planet earth.

Historical and Theories

The question of origin of life on our planet, the earth,has remained a most complicated and puzzling problem for thinkers, philosophers and naturalists of modern as well as ancient times. Different views have been forwarded concerning the origin of life on earth by thinkers of different ages. Most of them are quite bizarre and merely having a historical significance. Certain most important hypotheses and theories concerning the phenomenon of origin of life are the following:

1. Special Creation Theory

According to the special creation theory, the life is mysterious force or *vital spirit* that set living things completely apart from the non-livings and it is originated on our planet due to some supernatural event, which cannot be studied scientifically but must be accepted on faith. Most religions of different human cultures accepted the concept of special creation in one way or other. The followers of this theory argue that all living beings on the earth were originally created by a super natural power – God. According to Bible, for instance, world was created within six natural days. Plants were created on third day, fish and fowl on the fifth day, and animals on the sixth day. Lastly human beings were created, first man then women. Adam, the first man, was moulded by God from inanimate matter – clay, which he furnished with a soul, thus, breathing life into him.

Objections to special creation theory. The special creation theory lacks sound logic and scientifically sound evidences, therefore, it could not convince the scientists. Charles Darwin condemned the special creation theory by saying that earth has not always been inhabited by the plants and animals as well know now.

2. Hindu Concept of Origin of Life

Different schools of Indian philosophy and Hindu religion have elaborated the problem of origin of life in sufficient details, but that is beyond the scope of this book, however, it will not be out of context to discuss here some ideas of Upanishad-philosophy. For instance, the cosmology elaborated in the "Taittiriya Upanishad" describes the origin of life in the following way – from Brahman sprang ether, from ether air, form air fire, from fire water, from water earth, from earth herbs, from herbs food, from food seed and from seed man". Likewise, according to the "Mundaka Upanishad", life arose from food and from food arose the mind. Most of the early Upanishadic philosophers considered Brahman as not being something transcendental, superhuman, supernaturalor otherworldly. They did not recognize any God as the creator, controller and designer of the world. To them the Brahman was not a god, but the infinite totality of existence (for details see, K. Damodaran, 1970 – "Man and Society in Indian Society", People's Publishing House, New Delhi, India).

3. Theories of Spontaneous Generation or Abiogenesis

For centuries, the concept that life arose spontaneously from inanimiate material was a principal doctrine of how life originated. The spontaneous generation or abiogenesis of life was visualized as beginning with either inorganic materials or with putrefying organic matter. In China, for example, even from the earliest times, there was belief in the spontaneous generation of aphids or other insects under the influence of heat and moisture. In sacred books of India there are indication of the sudden emergence of various parasites, flies, and beetles from sweat and manure. It has been deciphered from Babylonian cuneiform texts that worms and other creatures were formed from the mud of canals. In ancient Egypt the conviction prevailed that the layer of humus deposited by the Nile in its flood stage could give birth to living creatures. When it was warmed by the sun…. frogs, toads, snakes and mice were formed thus.

Apparently, before the 1700s in Europe, it was common among laymen and scientists alike to believe that living matter could spontaneously develop from non-living materials. Infestations of worms, flies, fungi and other organisms appeared regularly in food and organic matter without any apparent explanation. People were unaware of the life history of most of these organisms and could not know how easily eggs and spores were transferred from place to place.

Disproof of the spontaneous generation theory. During the late seventeenth century, an Italian doctor Francesco Redi (1626-1698) performed what are now considered classical experiments to test the validity of the theory of spontaneous generation and had the hunour of being the first to come forward with experimental refutation of spontaneous generation so firmly established for many centuries. In his paper "*Esperienze intorno alla generazione degli netti* (1668), he described a series of experiments in which he placed meat or fish feel) under clean muslin coverings and demonstrated that while flies laid eggs on muslin, maggots or larvae appeared only when those eggs were transferred to the meat and allowed to hatch. Eh concluded that maggots came only from pre-existing flies and were not spontaneously generated by any other form of material.

While Redi's experiment proved that even the simple forms of life known in the seventeenth century did not arise by spontaneo us generation, the discovery of microbes by Van Leeuwenhoek (1632-1723) brought the questin into the forefront of biological thought once again. After all, maggots are far more complex than those little spherical, rod, or corkserew-shaped bacterial forms observed under the microscope. So it was considered that microorganisms arose by spontaneous generation. This idea was very authoritatively maintained by the noted German philosopher Leibiz (1646-1716). French naturalist Buffon (1708-1788) and Englishman John Nedham (1713-1781).

The Decline and Fall of the Theory of Spontaneous Generation

An Italian scholar Abbe Spallanzani in 1765 tested the theory of spontaneous generation of microbes. Eh prepared flasks of meat broth which were boiled for several hours and then sealed. The broth remained clear for months, and when the seals were broken and the broth tested, it was shown to be free of microbes.

Spallanzani's experiments were neither conclusive nor satisfying to many of his contemporary scientists. Some claimed that by boiling he had driven out a "vital force" necessary for spontaneous generation. When Priestly proclaimed oxygen as a necessary factor for life and experiments of Gay Lussac demonstrated that such boiled and sealed flasks contained no oxygen, than these were sized upon as evidence that the vital principle of life required oxygen. Many of Spallanzani's successors modified the experimental procedure in attempt to quell the criticisms, but each new technique gave rise to new criticisms. The death blow to the theory of spontaneous generation of microbes was finally dealt with Louis Pasteur (1822-1895) in the nineteenth century, some two hundreds years after the initial experiments of Spallanzani.

The period of Pasteur. Pasteur devised several experimental means by which the spontaneous generation of microbes was disproved. The simplest and most sophisticated one was with the use of a swan-neck flask. He prepared a meat broth in this flask and boiled it for several hours. He then left the flask unsealed on a laboratory bench. Previous experiments have shown that the mere boiling of the broth did not diminish its generative capacity, so the experiment could not be criticized on these grounds. The flask was not sealed, and there was a free exchange of air with the environment, so the system did not lack oxygen. Still, the swan-neck remained free of microbial contamination for months, because their swan necks were shaped so to trap viable microbial particles and to allow only air to enter the flask. After several months when he broke the neck of one of these flasks, contamination by air and proliferation of microorganisms in the fluid ensued. This simple experiment, thus, altogether disproved the concept of spontaneous generation.

4. Hypothesis of Panspermia

some scientists of nineteenth century assumed that this planet was 'seeded' from space. One of such a theory called cosmozoic theory or hypothesis of panspermia was developed by Richter (1865) and then supported by Thomson, Helmholtz (1884), Van Tieghem (1891) and others. According to this hypothesis, meteorites travelling through the earth's atmosphere are strongly incandescent only on the surface, while the interior remains cold. Thus, the embryos of organisms inhabiting meteorites or the planets from which they were formed were preserved alive in the interior of meteorites, finding fertile soil on earth, and grew and evolved to produce all the species now present.

Another hypothesis of panspermia which was developed by Arrhenius (1911) holds that spores of life together with particles of cosmic dust could be transformed from one heavenly body to another under the pressure of stellar rays. But it is doubtful that whether such spore remained viable because as especially serious danger for living embryos are short wavelength ultraviolet radiations which penetrate inter-planetary and interstellar space and are destructive to all life.

From 1870 to 1940, most areas of science progressed toward a level of maturity where their firmly established principles collectively created a unified picture of the world. Biology had brought life down to the chemical level through physiology and biochemistry. Chemistry had raised the elementary structure of matter to large synthetic molecules. ……………………….. considerations on which reactions were based. Geology had measured the age of earth as a few billion years and that of the universe as close to ten billion years. The stage was set for the idea of spontaneous generation to be given a more rigorous inspection. The first form of life was not to be a maggot or bacterium, or even a virus but a collection of biologically important molecules, capable of producing organized forms of life through processes of natural selection. The laboratory was the entire surface of earth, the time was in terms of several billion years, and the experimental procedure was natural selection.

5. Theory of Chemical Evolution and Spontaneous Origin of Life at Molecular Level

Since Pasteur' time, countless generations of students have been taught not to believe in spontaneous generation. However, Pasteur's experiments revealed only that life cannot arise spontaneously under conditions that exist on earth today. In fact, a Russian biochemist Alexander I. Oparin in 1924 has proposed that conditions on the primeval earth billions of years ago were definitely different from present conditions, and the first form of life, or self-duplicating particles, did arise spontaneously from chemical inanimate orabiotic substances. Thus, Oparin was the first to suggest that a long evolution of chemical substances occurred before life actually originated. In his book "*The Origin of life on Earth*", in 1924 he provided a biochemical explanation of origin of life. According to him, origin of life occurred along the orign and evolution of earth and its atmosphere. The primordial atmosphere of earth had water, methane and ammonia from which as a resultof a series of changes, a colloidal body called coacervate and containing a mixture of biologically important macromolecules like proteins, lipids, nucleic aids, etc., were formed. These coacervates though were not living, but they behaved in a manner similar to biological systems by being subjected to natural selection, by being chemically directive and by having the capacity of reproduction by fragmentation.

Synthesis of coacervate droplets. The coacervate droplets have been made typically by combining solutions of oppositely chared colloids, e.g., gelatin + gum Arabic, or, serum albumin + gum; or clupein +gum, or histone + gelatin, histone + RNA, or gelatin + gum + DNA. Each solution is uniform, but when the two are brought together they interact to form ‘swarms’ or ‘clusters’. When these clusters reach a certain size they separate from the solution in the form of coacervate droplets.

Characters of coacervate droplets. Oparin’s coacervate droplets can take up materials from the external medium and some of them show enzymatic activities. For example, Oparin and his associates have made coacervate droplets from gum Arabic and histone at pH 6.2 and have included phosphorylase enzyme durng the formation of droplets. When glucose-1-phosphate is dissolved in the equilibrium liquid, the droplets are found to store starch. In another experiment, coacervate droplets of some sort have been prepared with phosphorylase and β-amylase. In this case, the starch that formed from the entering glucose-1-phosphate then decomposed into maltose which diffused into the external medium where it was detected.

Similarly, poly……cleotides were synthesized enzymatically within coacervate droplets. These experiments with coacervate droplets, thus, show some sense of the special benefit of encapsulation of enzymes in living entities. Oparin proposed that although the coacervates which might be originated during chemical evolution, were not living, they behaved in a manner similar to biological systems (1) by being subjected to natural selection (in terms of a positive selection for stability as expressed through persistent size and constant chemical properties, (2) by being chemically directive (selective accumulation of material) and (3) by reproduction (by fragmentation).

……………………………………………………………standing and they were synthesized from polymers obtained from living systems, they no longer served as the models of protocells.

Laderberg has recognized the following three steps in the Oparin's thesis: (1) Chemogeny (Chemical evolution); (2) Biogeny (formation of self-reproducing biological units) and (3) Cognogeny (Evolution of mechanisms of expressions, communication, perception and feeding). Oparin like observations were made by an English biologist J.B.S. Haldane (1929) about the origin of life. Both of these scientists got the ideas from the studies of biochemistry made by Sir F.G. Hopkins, and both were of the opinion that the evolution is purely chemical, the gradual transmutation of inorganic compounds into organic ones.

Experimental Support of Oparin's Hypothesis

Oparin's ideas created a rash of excitement, but it was not until 1953 that some experimental evidence was gathered to support Oparin's hypothesis of spontaneous generation of life at molecular level.

Miller's Experiment

In 1953, Stanley Miller then at the University of Chicago and a student of Nobel Laureate Harold Urey, synthesized organic compounds under conditions resembling the primitive atmosphere of the earth. At that time it was believed that theprimitive atmosphere had probably been composed chiefly of water vapour, methane, hydrogen and ammonia (latter three in the ratio of 1:2:2). Such a gas mixture was circulated through a closed apparatus by steam from boiling water and subjected to an electric spark discharge (75,000 volts) between tungsten electrodes (Fig. 12.4). The electric discharge duplicated the effects of violent electrical storms in the primitive universe. The gas mixture was then condensed and added to the boiling water for recirculation; any non-volatile substance that appeared would accumulate in the water. This apparatus was permitted to run for a week. The results were spectacular. Several amino acids were synthesized (e.g., alanine, glycine, glutamic acid, aspartic acid, valine and leucine; in both D- and L- forms) and a number of other organic compounds, such as hydrogen cyanide (HCN), aldehyde and cyano compound (-C≡N), appeared; about 15 per cent of the carbon in the gas mixture became incorporated into these compounds.

Miller's instructive experiment has been successfully repeated by a number of investigators; amino acids can also be generated by irradiating a similar gas mixture with ultraviolet light. Both purines and pyrimidines can be formed when HCN and cyanoacetylene (HC≡C-C≡N) are present ……………………………………… ………………………………………… purines (mainly adenine) can be regarded as condensation products of five HCN molecules, although the route of synthesis is still not clear (Miller and Orgel, 1973). Polymerization of HCN also gives rise to arotic acid, which can be photochemically decarboxylated by sunlight to pyrimidine (uracil). Even ATP can be formed under prebiotic conditions, especially in the presence of the common mineral apatite (calcium phosphate).

Protenoid Microspheres

The synthesis of monomers (e.g., amino acids) is only a small step toward the synthesis of a living cell. In 1964, Sidney W.Fox, then at Florida State University and later at the University of Miami, reasoned that proteins were synthesized from amino acids in the primitive earth by thermal energy or heat. Fox accordintly heated a mixture of 18 amino acids to temperatures of 160-200°C for varying periods of time. He obtained stable, protein-like macromolecules, which he termed protenoids. These thermally produced proteinoids are similar to natural proteinoids are similar to natural proteins in many respects. For instance, bacteria can actually utilize the protenoids in a culture medium, degrading them enzymatically into individual amino acids. Equally important, when the protenoids material was cooled and examined under a microscope, Fox observed small, spherical, cell-like units that has arisen from aggregations of the protenoids. These molecular aggregates wre called protenoid microspheres.

Physical properties of protenoid microsphees. The protenoid microspheres were spherical, mircroscopic and of uniform diameter. Their size remained in between the range of 0.5-0.7mμ diameter. They had a number of properties in common with coccoid bacteria and microfossils found in ancient geological strata (Fox, 1969), in containing uniformity in size, stability in centrifugation, electronmicroscopy, extreme pH variations, etc.

Structural properties of protenoid microspheres. Electron microscopic studies of protenoid microspheres have revealed concentric double layered boundaries around them through which diffusion of material from the interior to the exterior occurs. Further, protenoid microspheres has the ability of motility, growth, binary fission into two particles, and a capacity of reproduction by budding and fragmentation. Further, some of them are stainable by gram-positive stains others by gram-negative stain.

Enzyme-like activities of protenoid microspheres. They are found to have catalytic activity such as degradation of glucose and this enzymatic activity of them is partiallylost during heating.

The model of protenoid microsphere for protocells is widely accepted because of its following significances (1) Such protenoid microspheres arise from monomers, rather than from polymers obtained from organisms already in biota, as is true for the usual experiments with coacervate droplets. (2) This model suggests that protenoids are informational and it shows the origin of communication which may be intercellular or intergenerational communication. In fact, it suggests the simultaneous origin of both types of communication.

Since both proteins and nucleic acids (plus other simpler substances) are required to develop and reproduce organisms living today, an obvious question is which of these substances arose first. No clear answer is available. To date protein-like materials have proven easiest to synthesize prebiotically; ………………………………… ……………………………………………………………… with early proteins upon which selection could act. This replicating material could have been a protein itself. It is difficult to say that this is impossible, but no self-replicating proteins are known (J.W. Valentine, 1976). Nucleic acids seem difficultto form, but they do replicate; perhaps the earliest gene was a simple nucleic acid (i.e. RNA). The trouble with nucleic acids is that they do not do anything themselves. This paradox has been recently resolved with the discovery of a catalytic RNA by Cech (1982-1986). At this stage, let us analyse the following models regarding the genetic substance.

1. Cairns-Smith’s model. The model of primitive genetic system proposed by Cairns-Smith (1971) postulates that the first genes were minerals, minute crystals associated with proteins as concentration sites and to some extent as templates. Certain minerals, e.g., clay or other layered silicates, were in very large supply as products of

early rock weathered. They might have favoured the formation of classes of polymers for which they provided good templates. Natural selection would favour a polymer whose activities preserved or even increased the supply of its mineral template, perhaps by helping to weather a parent material to produce the clay gene. The association of two (and eventually many) different polymers might have aided in mineral formation, while others provide adhesion for the polymer colony, and still others cleaved prebiotic organic compounds to provide building block materials for the colony, and so forth.

The polymer in any such early colonies would have proteins but eventually a nucleic acid was added to a colony fitness, any process that would replicate it *in situ* would be considerably favoured. Thus, a nucleic acid (probably a simple RNA at first) would replace minerals as genetic material. Presumably, the nucleic acid gradually took over the task of coding for all the proteins employed in the colony, including those associated with its own replication. Eventually DNA evolved (presumably coded by RNA originally) and an essentially modern nucleic acids – protein system of gene – enzyme was present. At some stage membranes were developed around the colonies and those membrane-bound units became cells. Thus, according to this model genes evolved in the following way.

Clay genes → RNA genes → DNA genes
Having proteins

Cairn-Smith scenario is highly speculative. Evidence are lacking which may prove that life following such a protoadaptive pathway from prebiotic materials to living cells (see Valentine, 1976).

2. RNA first model. In recent years, evidences are in favour of RNA to be material of first formed gene (Woese, 1967; Crick, 1968; Origel, 1973, 1986; Watson *et al.*, 1986; Darnell *et al.*, 1986). Thus, a number of protein-free (i.e., non-enzymatic) reactions that could have been decisive in the development of the first biologically important polymers have been discovered by molecular biologists. These reactions include (i) formation of RNA oligonucleiotide from activated monomeric precursors; (ii) template-directed RNA synthesis; (iii) site-specific RNA cleavage (with and without the aid of second RNA molecule), and (iv) RNA-RNA splicing. This, RNA may have been the first polymer and some form of reverse transcription may have

given rise later to DNA. How the coordination between the RNA oligonucleotides and amino acids or peptides occurred remain unknown, but the result was a single universal three-letter code that is still used by all cells today.

Further, comparison between the rRNAs of many different cell types as well as comparison between bacterial rRNAs and rRNAs of organelles (mitochondria and chloroplasts) have suggested two important conclusions about evolution: (i) the overall two-subunit structure of ribosomes is universal; (2) secondary (stem-loop) structure is extremely similar in all rRNAs. Thus, ribosomes can be considered as highly preserved but fairly complicated early cellular structure.

Catalytic RNA. The precursor of rRNA of the ciliate protozoon *Tetrahymena* was found to …………………………………………………………. Itself from the rRNA precursor without the help of a protein enzyme (Cech and coworkers, 1982, 1983, 1986). The ability of rRNA intron to function as enzyme implied that the very first living molecule might have been an RNA replicase that catalysedits own replication without the help from a protein (see Watson *et al.*, 1986).

The discovery in 1983 of a second catalytic RNA species (377 nucleotides long RNA component in RNP or ribonucleoprotein) of *E.coli* lent strong support to this notion. In streak virusoid (*Lucerne transient*) too RNA enzymes work like protein enzymes (Pace and Marsh, 1985). Catalytic RNA are now considered as molecular fossils whose history dates back to the earliest life forms.

Thus, the discovery of intervening sequences, the introns or junk DNA (mainly in the eukaryotic genomes) has given rise to much speculation about the nature of the earliest genes. Because protein free RNA reactions include cutting and splicing, and because spliced gene structures are relatively well preserved over enormous evolutionary intervals (i.e., more than a billion years), it is possible that the first gene had non-contiguous coding regions and that splicing was a prominent feature of the most primitive genomes. According to this view, cells lacking introns may have lost them due to selective pressures associated with rapid growth, whereas eukaryotic precursors were slower growers that did not lose their introns. The antiquity of the eukaryotic lineage is supported by the rRNA lineage studies (see Darnell *et al.*, 1986).

Why RNA and not DNA was the first living molecule ?

New examples of RNA molecules with enzymatic activity are constantly being discovered, but no enzymatic activity has ever been attributed to DNA. Further, ribose is much more readily synthesized than deoxyribose under simulated prebiotic conditions.

PROCESS OF ORIGIN OF LIFE

With such a historical background about the speculated modes of protobiogenesis, now we are in better position to discuss the present (current) understanding of molecular evolution and its culmination into the origin of life. The pre-requisite for the treatment of the problem of protobiogenesis is the origin of the raw materials (matter, elements, stars, planets, etc) for the molecular evolution, viz., a brief introduction of cosmology.

1. Structure of Cosmos

The universe or cosmos is made up of countless aggregation of stars organized into systems of various shapes and types, called galaxies. These galaxies, each of which includes hundreds billions of stars with their accompanying planets and satellites, are distributed into the vast reaches of space. Milky was is a galaxy which is composed of 100 billion stars besides sun and its planets (viz., solar system). The sun is a star being five billion years old and remains in a rapid state of motion. The planets of solar system are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune and Plato. All the planets of solar system rotate around their axes and revolve around the sun in elliptical paths determined by the gravitational force.

2. Primitive Earth

It is a tenable hypothesis that the cosmos as we know it originated in gigantic explosion (big bank hypothesis of Abbe Lemaitre, 1950) which blow matter from a primordial fireball, scattering outwards the materials of the universe in an ever-expanding volume. Clouds of dust and gases, chiefly helium and hydrogen, spread through the space, thinned in some regions and dense in others. Stars can be formed from the condensation ofsuch clouds; our own sun probably formed from a rotating mass of dust and gas that was flattened by the spin and which collapsed upon a central region due to the force of gravity. As the gaseous disc condensed inwards, the spin of the central region must have increased in order to conserve momentum. At some point the spin could have become so great that material would be ejected owing to centrifugal force. However, as such a point was approached, and the centrifugal force increased, matter in the outer portions of the condensing disc would tend to be more attracted to local dense patches than do the central region. Thus, at some distances from the proto-sun, matter would be condensed into separate bodies. These would be the planets, which lie in the plane of the sun's rotation. The solar satellites, including the earth and presumably the other planets and asteroids and the earth's moon as well, appear to date from about 4.6 billion years ago (see Valentine, 1976).

3. Prebiotic Synthesis

The core of the primitive earth melted from heat of condensation, but the outer crust cooled quickly except where convection or volcanic eruptions brought heat and mass from the molten core to the surface. The primitive atmosphere was formed as various gases escaped from the interior of the earth (outgassing). The early gas cloud was especially rich in hydrogen (H_2). The hydrogen of the primitive earth chemically united with carbon to form methane (CH_4), with nitrogen to form ammonia (NH_3), and with oxygen to form water vapour (H_2O). Thus, the early atmosphere had a strongly reducing (non-oxygenic) character, containing primarily hydrogen, methane, ammonia and water. Darnell *et al.,* (1986) has, however, argued that these gases (or compounds) may not have existed free in the atmosphere in large quantities for extended times. Carbon dioxide (CO_2) and hydrogen sulphide (H_2S) were probably more permanent constituents of the early atmosphere.

The crust and atmosphere of the primitive earth cooled slowly, and the first shallow seas were formed as the less volatile gases including water, rained back down to earth. The oceans formed as a result of millions of years of torrential rains. The rains eroded the rocks and washed minerals (such as chlorides and phosphates) into the seas. The stage was set for the combination of the varied chemical elements. Chemicals from the atmosphere mixed and reacted with those in the waters to form a wealth of hydrocarbons (that is, compounds of hydrogen and carbon). Water, hydrocarbons and ammonia are the raw materials of amino acids which, in turn, are the building blocks for the larger protein molecules. Thus, in the primitive seas, amino acids accumulated in considerable quantities and became linked together to form proteins or oligopeptides (prebiotic synthesis).

Energy sources for prebiotic synthesis. Complex carbon compounds, such as amino acids and proteins, are termed as organic because they are made by living organisms. Our present-day green plants use the energy sources in the primitive earth, and how was synthesis of orgaic compounds effected in the absence of living things? It is generally held that ultraviolet (UV) rays from the intense solar radiations penetrating through the thin ozone-free atmosphere, electric discharges such as lighting from rainstorms and intense dry heat from volcanic activity furnished the energy for prebiotic synthesis such as joining the simple carbon compounds and nitrogenous substances into amino acids (Orgel,1973). Superheated hot springs were also formed along rifts in the sea floor. Alternating cycles of cold nights and hot days too might have affected the prebiotic synthesis (see Watson, *et al.*, 1986).

Evidence that oxygen was absent in primitive atmosphere. The sedimentary rocks that remain from 3.8 to about 1.9 billion years ago, which appear to have been deposited chiefly in shallow epicontinental seas, generally suggest an anoxic atmosphere of the primitive earth: the minerals are mostly in low states of oxidation relative to today (although more highly oxidized layers do appear locally). Themost famous examples are the banded iron formations which contain vast amounts of ferrous iron. This indicates that free oxygen was not available in the atmosphere; otherwise the iron would be in a ferric state (Cloud, 1968).

There are certain other considerations which show why the primitive atmosphere lacked free oxygen? Free oxygen is available in quantity from only two sources. One is photodissociation of water vapour, which occurs high in atmosphere. Hydrogen, freed from water molecules, leaks out into space, while the heavier oxygen is retained. The extent of this process in the past, and the rate at which photodissociated oxygen could accumulate, are much debated. Oxygen is also produced from water via photosynthesis. It is commonly believed that the bulk of free oxygen was produced by plants. Before the advent of life, and indeed until plant biomass became large, this source of oxygen was absent or unimportant.

D

R

A

W

I

N

G

One final evidence that oxygen was not ordinarily present in the early atmosphere and, therefore, was not dissolved in early oceans, is that organic compounds are stable in the presence of hydrogen (that is, under reducing conditions) but are quickly destroyed in the presence of oxygen (Haldane, 1933). This indicates that compounds required as precursors or building blocks for living systems must have originated and accumulated in a reducing environment.

Primitive sea served as broth for the prebiotic synthesis. Classical experiment of Miller (1953) has established that organic compounds can be formed without the intervention of living organisms. Thus, it appears likely that the sea of the primitive earth spontaneously accumulated a rich mixture of organic molecules. In the absence of living organisms and oxygen, the organic compounds would have been stable and would persist for countless years. The sea became a sort of dilute organic soup or broth (or aquatic Garden of Eden of Volpe, 1985 or chemical cauldron of Watson *et al.*, 1986). In the organic soup of cauldron the molecules collided and reacted to form new molecules of increasing levels of complexity. Proteins capable of catalysis or enzymatic activity, had to evolve, and nucleic acid molecules capable of self-replication must also have developed.

4. Evolution of Progenote

The precellular evolution involved the following three main steps.

(i) Origin and evolution of RNA world. The living cell is an orderly system of chemical reactions (which are directed by enzymes) that is self-duplicating nucleic acids) evolved before the development of the metabolic (enzymatic) machinery of the cell. It is likely that self-replicating polynucleotides (nucleic acids) slowly became established in the primordial earth some 3.5 billion years ago. In the absence of proteinaceous enzymes in the prebiotic environment, les efficient catalysts in the form of minerals or metal ions would be sufficient to promote the reactivity of the nucleotide precursors. It is postulated that errors continually occurred in the copying process and that sequence of nucleotides in the original polynucleotides molecule became altered on many occasions. Large numbers of polynucleotide variations undoubtedly maladapted. Just as organisms today compete for available resources, molecules with different nucleotide sequences competed for the available nucleotide precursors in promoting copies of themselves. Any new mutant sequence with a replication rate higher than the antecedent sequence could be “fittest” and prevail. Natural selection (or differential reproduction), thus, operated on population of molecules as it did later on population of organism.

The highlights of RNA would include (i) distinction between genomic and functional RNA molecules, (ii) recruitment of exons into transcription; (iii) inclusion of introns in most genes; (iv) no regular cell growth and (v) formation of membranes.

(ii) Origin and evolution of ribnucleoprotein (RNP) world. Certain polynucleotides not only specified their own sequences, but apparently directed the synthesis of macromolecules of an entirely different type-namely, polypeptides (proteins). In present-day organisms, RNA guides protein synthesis, which suggests that the nucleotides of RNA (not DNA) were the first carriers of genetic information. Thus, RNA established a primitive genetic code for ordering amino acids into proteins. Evidently, the genetic code – the translation of nucleotide sequences into amino acid sequences – became established at a very early stage of organic evolution. The universality of the genetic code in present-day organisms attests to the early origin of genetic code.

The highlights of RNP world include (i) origin of primitive rRNA, rRNA and aminoacyl-rTNA synthetase enzyme; (ii) evolution of template-dependent translation apparatus (true mRNA) and transcription and replication of segmented double standed RNA genomes.

(iii) Origin of plasma me…………………………………………………. Directs the synthesis of a protein that accelerates the replication of that particular RNA (i.e. RNA polymerase). However, a protein specified by a particular variant of RNA could not foster the reproduction of that kind of RNA unless it is retained in the immediate vicinity of the RNA. The free diffusion of proteins could be prevented if some form of a compartment evolved to enclose or circumscribe the specific protein made by a particular RNA. All present-day cells have a limiting plasma membrane composed principally of phospholipids and proteins. Most probably the first cell was formed when polarized films of phospholipids and proteins. Most probably the first cell was formed when polarized films of phospholipids formed soap-like bubbles enclosing aggregations of complex macromolecules. Specifically, a membranous structure arose that encircled a self-replicating aggregation of RNA and protein molecules. Once bounded by a limiting membrane, a given RNA molecule could be assured of propagating its own special protein molecules.

(iv) DNA world, At some later stage in the evolutionary process, DNA took the place of RNA as the repository of the genetic information. This step may have occurred due to evolution of an enzyme called inverse transcriptase. In other words, the genetic code was stored in DNA rather than RNA. Today, RNA molecules serve essentially their primal function ; directing protein synthesis. Thus, in modern cells, genetic information is stored in DNA, transcribed into RNA, and translated into protein. An endless permutation of nucleic acids and proteins has fostered an enormous richness of life.

Thus, evolutionary highlights of DNA world include (i) evolution of multiple pieces of DNA, (ii) stable storage of information in DNA; (iii) sure existence of plasma membrane; and (iv) growth and division.

(v) Origin of progenote. The precellular stage of organization has been termed progenote. The nature of progenote is uncertain. It might have been a functional, slow replicating, independently evolving organism, or it might have been an amorphous group of primitive transcription units plus a primitive transcription and translation apparatus from which many different cell lineages evolved; three of which have survived to the present day (i.e., archaebacteria, urkaryote and eubacteria). The progenote contained DNA, most genes with introns, a slow growth and heterophic nutrition.

The first living system (the progenote) drew upon the wealth of organic materials in the sea broth. Organisms that are nutritionally dependent on their environment for ready-made organic substances are called heterotrophs. The primitive one-celled heterotroph probably had little more than a few genes, a few proteins and a limiting plasma membrane. The heterotrophs multiplied rapidly in an environment with a copious supply of dissolved organic substances. However, the ancient heterotrophs could survive only as long as the existing store of organic molecules lasted. Eventually, living systems evolved the ability to synthesize their own organic requirements from simple inorganic substances. Thus, in the course of time, autotrophs arose, which were able to manufacture organic nutrients from simpler molecules.

However, primitive autotrophs would require a whole array of enzymes to direct a multistep chain of metabolic reactions involved in the synthesis of a protein. Indeed, it would be too much to expect that all the necessary enzymes evolved at the same time. Norman H. Horowitz (1959) has proposed that the chain of steps in a complicated chemical pathway evolved backward. Let us analyse his concept in the following way:

Retrograde evolution. According to Horowitz an organism might acquire, by successive mutations, the enzymes required in a biosynthetic pathway in the reverse order from that in which they normally appear in a sequence. In other words, evolution began with the end product of the pathway and worked backward, one step at a time, toward the beginning of the reaction chain. The total pathway became constructed when the organism had tapped from its surroundings all of the intermediate precursors of the reaction.

For instance, let us suppose that an organic compound *O* is synthesized through the following steps, where *A* through *D* represent precursors.

………………………………………………………………………………………

The first primitive heterotroph, lacking synthetic ability, would require the presence of *O* in its environment. This essential organic compound, continually being used by the heterotroph an ability to catalyze the reaction from *D* to *O*, then the organism would no longer remain dependent on the availability of *O* in the environment. Indeed, in an environment where *O* had become depleted, the new mutant would have a survival or selective advantage over the ancestral type and replace it. As the new mutant reproduced, *D* in turn would become scarce. Another mutation might occur in the mutant, converting it to a still newer form capable of catalyzing the reaction *C* to *D*. At this juncture, both the first mutant and the first two-gene systems.

The intimate two-gene combination would be able to survive and reproduce in the presence of *C* and *D.* as other compounds in synthetic pathway become progressively rarer, additional mutations for their synthesis would be favourably selected. Ultimately, a multigenic system would evolve capable of directing the synthesis of *O* from inorganic substance A by way of intermediate products *B, C, D.* At first glance, it might appear that an unreasonable number of mutational events has been involved. But it should be recalled that mutations occur normally and continually in living organism. In fact, the capacity of mutation should be regarded as an essential property of life.

(vi) Adaptive radiation of progenote. The first simple autotroph arose in an anaerobic world, one in which little free oxygen was available. The primitive autotrophs obtained their energy from the relatively inefficient process of fermentation (the breakdown of organic compounds in the absence could also be developed to provide reducing power in the form of NADH that could be used to derive biosynthetic reaction.

(i) Glucose + ADP → Lactate + ATP (e.g., *Lactobacillus*)

(ii) Glucose + ADP → Ethanol + Lactate + CO_2 + ATP (e.g., *Leuconostoc*)

(iii)Glucose + ADP → Ethanol + CO_2 + ATP (e.g., *Zymomonas*)

(iv)Lactate + ADP → Propionate + Acetate + ATP + CO_2 (e.g., *Clostridium)*

Thus, the early fermentative autotrophs were much like our present-day anaerobic bacteria (*Lactobacillus,* etc.) and yeast. The metabolic processes of the anaerobic autotrophs resulted in the liberation of large amount of carbon dioxide (CO_2) into the atmosphere. Once this occurred, the way was paved for the evolution of organisms that could in synthesizing organic compounds and could use sunlight as the only source of energy. Such organisms would be the photosynthetic cells.

Early photosynthetic cells probably split hydrogen containing compound, such as hydrogen sulphide (H_2S). In other words, as is still performed today by anaerobic green/purple sulphur bacteria, hydrogen sulphide is cleaved into hydrogen and sulphur. The hydrogen is used by the cell to synthesize organic compounds, and the sulphur is released as a waste product (as evidenced by the earth's great sulphur deposits). In time, the process of photosynthesis was refined so that water (H_2O) served as the source of hydrogen. The result was the release of oxygen (O_2) as a waste product. At this stage, free oxygen became established for the first time in the atmosphere.

The first organisms to use water as the source of hydrogen in photosynthesis were the blue-green algae (cyanobacteria). Since blue-green algae were active photosynthesizers, atmospheric oxygen accumulated in increasing amounts converting reducing primitive atmosphere into oxidizing atmosphere. Oxygen would be toxic to anaerobic bacteria. Many primitive anaerobic bacteria, incapable of adapting to free oxygen, remained in portions of environment that were anaerobic, such as sulphur springs and oxygen-free muds. Today, there are bacterial types that are anaerobic as well as aerobic.

Geologic evidence indicates that free oxygen began to accumulate in the atmosphere about two billion years ago. The rising levels of atmospheric oxygen set the stage for the appearance of one-celled eukaryotic organisms, which arose at least one billion years ago. Then, within the comparatively short span of last 600 million years, the one-celled eukaryotes evolved in various directions to give rise wealth of multicellular life forms inhabiting the earth.

Evolution of Eukaryotes

The simplest cells in nature – the bacteria and blue green algae (cyanobacteria) – have been classified as prokaryotic cells. These cells have no nuclear membrane by which the hereditary materials (DNA) are set apart from the cytoplasm, and lack specialized cytoplasmic bodies (organelles) such as mitochondria and chloroplasts. In a bacterial cell, the DNA molecule, which does not contain introns but does have operons, forms a simple closed loop that is attached to the inside of the plasma membrane. The replication of a bacterial cell occurs rapidly by the simple division of the cell into two.

To recall, phylogenetically, the progenote evolved into three types of cells: 1. Archaebacteria, 2. Eubacteria and 3. Urkaryote. Archaebacteria are the prokaryotes which appear to have diverged from the true bacteria (or eubacteria) very early in evolution. The strange class of archaebacteria consists of bacteria that are tolerant of acid (thermophiles or thermoacidophiles), bacteria that thrive in salt (halophiles) and bacteria that generate methane (methanogens). Recent developments in molecular biology reveal that these bacteria share some striking biochemical features (such as 16

S rRNA sequence) that are absent from other bacteria. Eubacteria includes cyanobactgeria (aerobic, H_2O photosynthetic); chloroxybacteria (aerobic, H_2O photosynthetic); paracoccus (aerobic, heterotrophic), nor-sulphur bacteria (aerobic, H_2S photosynthetic), sulphur bacteria (anaerobic, H_2S photosynthetic), spirochaetes (anaerobic, heterotrophic). *Clostridia* (anaerobic, heterotrophic) and *Desulphovibrio* (sulphate respiration) see Sleigh, 1986). The urkaryote is the ancestor of modern eukaryote.

In contrast to the prokaryotes, the cells of the morphologically more complex plants and animals, or eukaryotic cells, have a distinct nuclear membrane that encloses strands of DNA, or discrete chromosomes. Eukaryotic cells also have an elaborate system of membrane-bound cytoplasmic organelles. The enzymes that are responsible for releasing energy from complex organic molecules are packed inside sausage-shaped organelles called mitochondria. In the eukaryotic plant cell, a prominent organelle is the chloroplast, the photosynthesizing particle that converts light energy into chemical energy. The following two hypotheses have been proposed regarding the origin of semiautonomous organelles of eukaryotes, i.e., mitochondria and chloroplasts.

D
R
A
W
I
N
G

A Endosymbiotic hypothesis. One of the most intriguing speculations is that mitochondria and chloroplasts arose not by a gradual evolutionary process but abruptly in an unusually striking manner, i.e., symbiosis. In fact, mitochondria have a number of interesting properties that advocate that they were once free-living or independent, bacteria-like (i.e., Prokaryotic) organisms. Mitochondria have small amounts of their own DNA; this DNA exists as a loop-shaped molecule like the DNA of bacteria. Mitochondria also possess the genetic capacity to incorporate amino acids into proteins. These properties have led, in 1970, Lynn Margulis of Boston

University to propose the serial symbiotic hypothesis or endosymbiotic hypothesis. He hypothesis or endosymbiotic hypothesis. He hypothesized that mitochondria may have been derived form primitive aerobic bacteria (i.e., *Paracoccus*) that were engulfed by predatory organisms, probably fermentative bacteria, destined to become eukaryotic. The predatory hosts became dependent on their enslaved mitochondria and the latter, in turn, became dependent of their hosts. Thus, the association was of mutual advantage to the predatory cell and the engulfed prey. Such a close association or partnership is called symbiosis. In essence, the mitochondria (formerly oxygen-respiring bacteria) established permanent residence within the hosts. These predatory hosts became the first eukaryotic cells (i.e., animals cells).

Eukaryotic plant cells contain both mitochondria and chloroplasts, just as aerobic bacterial cells can be equated with mitochondria, blue-green algae cells are basically equivalent to the chlorophyll containng chloroplasts. Chloroplasts, like the mitochondria, have their own unique DNA and the associated protein-synthesizing machinery. In this case too it as speculated that those predatory cells (i.e., fermentative bacteria) that engulfed but did not digest the photosynthetic blue green algal cells and the aerobic bacteria evolved into eukaryotic plant cells. Thus, the engulfed prey became permanent symbiotic residents – either as mitochondria or chloroplasts – within the predatory cell.

B. Invagination of surface membrane hypothesis. Many authors are not convinced with the hypothesis suggesting the relatively abrupt origin of the organelles of eukaryotes by a process of symbiosis. According to them, the evolution of living organisms normally proceeds through a series of gradual changes. Thus, continual modification of the surface membrane (plasma membrane) may have been the basic evolutionary mechanism in the differentiation of eukaryotic cell from prokaryotic cell. The organelles of eukaryotic cells may have evolved by the invagination (or drawing inwad) of the surface membrane of a primitive cell. The postulated mechanism for the origin of a mitochondrion from the cell surface membrane. In support of this scheme, it is interesting that the enzymes for the breakdown of carbohydrates for energy production in bacteria are incorporated in the structure of the cell surface membrane. In similar manner other specialized internal structures of eukaryotes, including the membrane enclosed nucleus, may have differentiated from invaginated cell surface.

Finally, one of the new twists contributed by modern molecular biology is the possibility that today's prokaryotes are very highly evolved cells whose evolution has made them the most flexible and efficient cells on the planet.

MOLECULAR EVOLUTION

Evolution is a process of change. At the molecular level, this process involves the insertion, deletion, or substitution of nucleotides in the DNA. If the DNA encodes a polypeptide, these events may cause a change in the amino acid sequence. Over time, such changes can accumulate, leading to a molecule that bears little resemblance to its ancestor. Comparisons of the amino acid compositions of proteins in present-day species, the more changes will be evident in the amino acid sequences of the proteins of the two contemporary species. Viewed another way, the number of amino acid modifications in the lines of descent can be used as a measure of time since the divergence of the two species from a common ancestor. In principle, protein molecules incorporate a record of their evolutionary history that can be just as informative as the fossil record. Thus, proteins can be considered as chemical fingerprints of evolutionary history, because they bear amino acid sequences that have changed as a result of genetic changes.

The Evolution of Proteins

A few thousand kinds of proteins are required for the functioning of an *E.coli* cell and about a million kinds are required by human beings. They offer a preferable field for evolutionary study than other macromolecules such as nucleic acids because proteins are more heterogeneous both structurally and functionally, and are easy to isolate for the analysis. Recently a number of proteins have been characterized by the method of sequence analysis. The common procedure for establishing the amino acid sequence of a polypeptide requires, first, breaking the protein into small fragments of peptides, and then determining the amino acid sequence in each peptide. The polypeptide is broken into fragments with enzymes (e.g, trypsin and chymotrypsin) that hydrolyze the bonds between contiguous amino acids at specific sites. The resulting peptides are separated from each other using such procedures as column chromatography and two-dimensional paper chromatography. The amino acid sequence in each peptide is ascertained usually with the help of an apparatus known as a "protein sequencer" or "sequinator" (see F.J. Ayala, 1976).

Sequence data are now available from diverse biological groups ranging from microorganisms to mammals. Comparison of sequences allows study of interrelationship between structure and function, and to deduce how proteins have evolved. The laws governing evolution of proteins are similar to those for heritable traits, and interrelatedness of different organisms points toward their descent from a common ancestor. In general the following features have been noted.

1. Identical proteins are not found among different living species. But homologous proteins with some similarities occur in diverse organisms.
2. Different positions in the amino acid sequence vary with respect to the number of amino acid substitutions that can take place without impairing the function of the polypeptide chain. Thus, some positions allow more substitutions than others.
3. Due to the tightly packed three dimensional structure of the protein, a change undergone by an amino acid situated in the interior of the chain is likely to affect residues in neighbouring positions. Such evolutionary changes, therefore, occur in pairs.
4. Enzymes performing similar functions have similar structures.

Examples of Protein Evolution

1. Insulin. The first protein to be sequenced was insulin (Insulin is protein hormone produced by the beta cells of the islets of Langerhans of pancreas which participates in carbohydrate and fat metabolism) which composed of only 51 amino acids (Sanger and Thompson, 1953). Comparison of the amino acid sequence of insulin in a variety of mammals, including cattle, pigs, horses, sperm whales the amino acids, the protein is identical in each of these species.

2. Haemoglobin. In comparison between haemoglobin of vertebrates, E.Zuckerkandl (1965), calculated an average of 22 differences between human haemoglobin chains (α and β) and the similar haemoglobin of horse, pig, cattle and rabbit. It was, therefore, concluded that human haemoglobin diverged from those of animals by about one amino acid change per 7 million years. The rate of change of amino acids is, however, different for different proteins.

3. Cytochrome c. Cytochrome c is an ancient, evolutionarily conservative molecule that serves as an essential enzyme in respiration. It is relatively small protein of a chain length slightly over 100 amino acid residues, which apparently exists in all eukaryotic organisms. This ubiquitous protein has been extracted, purified and analyzed in more than 80 eukaryotic species (Fitch and Margoliash, 1967; Diekerson, 1971). Cytochrome c has 104 amino acids in vertebrates and a few more in certain species lower in the phylogenetic tree. Since cytochrome c performs similar vital functions in all organisms, fairly rigid evolutionary limitations have been placed on the acceptance of chance alterations (or mutations) of the primary amino acid sequence (i.e. an evolutionary conservatism is maintained).

Some amino acid residues of cytochrome c have not varied at all throughout time. In particular, the same amino acids have been found in 20 positions in all organisms tested, from moulds to humans. Obviously, the 20 amino acids at specific positions are essential, and any substitutions at these sites are likely to interfere with the function of the enzyme. In fact, there is one region of cytochrome c formed by a sequence of 11 amino acids (positions 70 through 80) that has remained intact in all organisms.

In Table 12…. The number of amino acid replacements in cytochrome c of several species have been compared (Dickerson, 1971). Firstly, it is evident that the chromosomes of closely related vertebrates differ in only a few residues, or not at all. For example, there is no difference at all in the composition of the amino acid residues between humans and the chimpanzee. Secondly, the greater the phylogenetic differences, the greater the likelihood that the cytochrome composition differ. This becomes clear when one compares a vertebrate with an insect (i.e. tuna and moth) or an insect with yeast.

Neutral Theory of Protein Evolution

Evolution of proteins have also been studied in certain other proteins such as myoglobin, α-globin, histone, carbonic anhydrases, fibrinopeptides, etc. analyses of the amino acid sequences of polypeptides suggest that some evolutionary changes may be random. This fact led to the formulation of a theory assumes that for any gene, a large proportion of all possible mutants are harmful to their carriers; these mutants are eliminated or kept at very low frequencies by natural selection. A large fraction of mutations, however, are assumed to be adaptively equivalent (i.e. the rate of evolution is equal to the rate of mutation).

Evolution of cytochrome c (Source: Volpe, 1985)

A. comparison of cytochrome c between different organisms

Organisms	Number of differences of amino acid residues
1. Cow and sheep	0
2. Cow and whale	2
3. Horse and cow	3
4. Rabbit and pig	4
5. Horse and rabbit	5
6. Whale and Kangaroo	6
7. Rabbit and pigeon	7
8. Shark and tuna	19
9. Tuna and fruitfly	21
10. tuna and moth	28
11. yeast and mould	38
12. Wheat germ and yeast	40
13. Moth and yeast	44

B. Comparison with human cytochrome c	
Organisms	Number of differences of amino acid residues
1. Chimpanzee	0
2. Rhesus monkey	1
3. Rabbit	9
4. Cow	10
5. Kangaroo	10
6. Duck	11
7. Pigeon	12
8. Rattle snake	14
9. Bull frog	18
10. Tuna	20
11. Fruitfly	24
12. Moth	26
13. Wheat germ	37
14. Mould (*Neurospora*)	40
15. Baker's yeast	42

Although the neutral theory does admit a role for natural selection as a purifying agent, that is, as a force that eliminates harmful mutations, it has little or no place for selection in the positive Darwinian sense. In the neutral theory species do not get better by fixing beneficial mutations; they simply do not get worse by fixing only neutral or nearly neutral mutations. For this reason, this theory is also called non-Darwinian theory of evolution (see Gardner *et al.*, 1991).

Revision questions

1. What is chemical evolution and origin of life? Discuss.
2. Describe Urey and Miller experiment. How does this experiment support Oparin's hypothesis?
3. Describe different theories for the origin of life on earth. Discuss in detail the Oparin's theory regarding the origin of life.
4. Describe how life originated on earth.
5. Write short notes on the following.

(i) Coacervates; (ii) Microspheres; (iii) Origin of eukaryotic cells; (iv) Endosymbiotic hypothesis; and (v) Current postulated role of RNA in origin of life.

CHAPTER II
BIOSTATISTICS

INTRODUCTION

Biostatistics has been defined as the application of statistical methods to biological sciences. However, the use of statistical methods is constantly increasing in biological sciences which are toughing newer horizons every day. The development of biological theories is closely associated with statistical methods. The doctrine of heredity rests on statistical basis. Therefore, a good understanding of biostatistics is essential for the students of biological sciences, as the methods of biostatistics are indispensable tools for the design and analysis of data and in the interpretation of experimental results for dependable conclusions. A preliminary acquaintance will help not only in applying bio-statistical methods but also in a better appreciation of their potential value. This book is primarily designed to outline the logical basis of the statistical approach to experimental problems and also to give the main features of those statistical methods commonly used in biological agricultural and medical experimentations.

Biostatistics has a venerable history. Considerable progress was made in the development of biostatistics during the period of Sir Francis Galton (1822-1911) who is regarded as “the father of biometry and eugenics”. His major contribution to biology was the application of statistical methods to the analysis of biological variation and correlation and regression. However, his methodology has been recognized as the foundation for the application of statistics to biology. W.F.R. Weldom (1860-1906) is credited with the coming of the term “biometry” Karl pearson (1857-1936) who is regarded as “the father of modern statistics”, was greatly motivated by the researches of Francis Galton. He established a laboratory and started the issue of “Biometrika Journal”. His major contribution was correlation and regression. Of the several mathematical methods of measuring correlation, the Karl Pearson’s method, popularly known as Pearson’s coefficient of correlation, is the most widely used in practice. William S. Gosset (1876-1937), a student of Karl Pearson, propounded a number of statistical formulae under the name of “student”. His contribution is also worth praising. Ronald A. Fisher (1890-1962) is another great figure in the history of biostatistics of the present century. Fisher is the real giant in the development of biostatistics. He propounded many concepts and conducted tests. Fisher must be credited with at least half of the essential and important developments as the theory now stands.

Let us first define some basic terms that are necessary for an understanding of biological and agricultural data. The data in statistics are generally based on individual observations. The data collected for statistical analysis donot consist of observations that are identical, since there would belittle reason to study such a variation. The data counted or measured for analysis purposes will represent the varying values of variable, i.e., a characteristic that shows variation. Any quantity or quality liable to show variation from one individual to the next in the same population is known as variable. An individual observation of any variable is known as variate. Some examples of variables include the plant height, heights of adult males, the weights of preschool children and the ages of patients seen in a dental clinic. To start with, we have to distinguish between qualitative and quantitative characters. In the case of qualitative characters, the individuals comprising the material under consideration are distinguished by some quality. Petal colour of the flower and smooth or wrinkled nature of surface in a seeds are instances of qualitative characters. We refer to variables of this kind as qualitative variables. On the other hand, quantitative variable is one where the individuals are distinguished by a measurement. For instance, quantitative variables where the individuals are distinguished by a measurement. For instance, quantitative variable is one where the individuals are distinguished by a measurement. For instance, quantitative variables are further divided into two kinds; continuous variables and discontinuous or discrete variables. Continuous variables are those which can take any value within the certain range exhibited by the population. Many of the variables studied in biology are continuous variables. Quantitative characters more commonly belong to this group; such as weight, height, volume, time and percentages etc. The discontinuous or discrete variables are those variables which have only certain fixed numerical values with no intermediate values possible in between. Examples are the number of organisms per unit area. The variable distributed according to the normal curve is a continuous one, whereas in binomial and poisson distributions, the variable is discrete.

It has been already started that the data in biostatistical studies are generally basedon individual observations. The selection of a part of a population to represent the whole population is known as sampling the part selected is known as sample. The object of sampling is to get information regarding the population from which the sample is obtained. In statistics, population always means the total number of individual observations from which inferences are to be made at a particular time. A sample represents the small collection of the population which has actually been observed. For example, all plants in a wheat field represent a population. Whereas individual observations on ten high yielding plants from this population refer to a sample. Populations may be finite or infinite. If a population of values consists of a fixed number of these values, the population is said to be finite. For example, the number of plants in a quadrat, the number of wheat plants in a plot and the number of patients in a hospital etc. On the other hand, population in which it is theoretically impossible to observe all the values, is an infinite one. An infinite population is unlimited in size. For example, the number of phytoplanktons in a pond, the number of RBCs in the human body and so on.

In all statistical studies, sampling is an important factor for obtaining the best possible values of the statistical parameter of the population. This aim can be achieved successfully if the sample studies are made in such a way that they disclose in relationship between the values of distribution. For example, if it is found that a particular frequency distribution of a sample study confirms to binomial, poisson and normal distribution. A random selection is one in which each unit of the population has the same chance of being included in the sample. In random sampling, the selection of units is done in such a way that the chance of selection of each unit of the population is the same. However, the selection of units depends entirely on chance and one does not know before hand which unit will actually constitute the sample. This method is also known as chance selection. Random sampling is done in different ways to eliminate human bias. That is why selections are actually made with the help of machines or lots. The most common example of random sampling is that the quadrats are made in the field at random where environmental conditions are uniform. It is also important to take into consideration the fact that the size of the sample has a relation to the degree of accuracy. Greater would be the accuracy if the size of the sample is large as it covers all types of variations. If the population is large, the greater should be the size of the sample. Therefore, the accuracy in random sampling has more or less a relation to its size. In brief "sample studies can give better results only when the sample is drawn systematically, their size is adequate and an appropriate sample design is used".

Samples and populations can be described by using measures like Mean, Mode, Median and Standard Deviation. These measures describing the characteristics of a sample, are called statistics. However, measures when they describe the characteristics of a population, are called parameters. Hence, statistics is a characteristic of a sample and a parameter is a characteristic of a population. For example, the mean pod length of the early maturing mungbean varieties growing in a field is 60 mm. is the characteristic of the population of all mungbean varieties and can be called as population parameter. On the other hand, the mean pod length of the early maturing mungbean variety, G-65 is 60mm; hence, this 60 mm. pod length is used to describe a characteristic of the sample "G-65". In this case, 60 mm. would be a sample statistics.

Statisticians always maintain the distinction between population parameters and sample statistics through the use of different symbols. For example, Greek letters are usually used to denote parameters while Roman letters denote sample statistics. The population mean is represented by μ while the sample mean is indicated by x. The standard deviation of the population is expressed by σ whereas the sample standard deviation is delineated by s. The population percentage is designated by π while sample percentage is symbolized by p.

The term standard error is used in biometrical analysis because it conveys a specific meaning. For example, if we take the number of samples and calculate the mean pod length for each sample, all the sample means will not be the same. We expect some variability in our observed means. This variability in the sample statistics results from sampling error due to chance, i.e. there are differences between each sample and the population and among several samples.

Table 1: Differences between a population and a sample.

Distination	Population	Sample
Definition	It is a defined as a collection of the items under consideration	It is defined as a part of the population selected for study
Characteristics	Parameter	Statistics
Symbols	N= population μ = population mean σ = population standard deviation π = population percentage	n= sample size x = sample mean s = sample standard deviation p = sample percentage

In all fields of research, there is one feature in common, i.e. the variability of the experimental material. When there is considerable variation from observation to observation in the same material. The experimenter is forced to use an experimental design which allows for unbiased estimates of the treatment differences. However, replication and randomization are two required conditions for an experimental design. It means that replication and randomization are necessary to obtain a valid estimate of the error variation. The object of comparison in the experimental trials is termed treatments. The repetition of the treatments under investigation is known as replication. The value of replication in experimental design is easily understood. Replication increases the accuracy of the scope of the experiment and it enables us to determine the magnitude of the uncontrolled variation that is usually referred to as an error. In all experiments, any one treatment is represented by a sample which is made up of one unit in each replication. Therefore, in general, the accuracy of the experiment increases in proportion to the square root of the number of replications. However, the process of randomization will ensure that the various soil conditions represented in the field will have an equal chance of being used in the experiment. The benefits to be obtained from the arrangement of treatments in replications where in each replication contains one of each of the treatment are well known to the experimenter. It controls the experimental error. The comparison between treatments will only hold good, provided that the treatments are allocated randomly to various plots. For example, five treatments are randomized with three replication in the experimental plot. Since all the treatments appear in each replication randomly, the influence of any fertility gradients will be eliminated by such an arrangement. This arrangement for five treatments in three replications gives the following design.

Replication I	T_1	T_2	T_3	T_4	T_5
Replication II	T_4	T_5	T_1	T_2	T_3
Replication III	T_2	T_3	T_4	T_5	T_1

The knowledge of biostatistics is becoming an important ally to the understanding and interpretation of the facts and findings of the biological, agricultural and medical research. In this book the approach is widened to highlight the principles of biostatistical methods to students and researchers. The governing principle behind this approach is to ensure that they appreciate the utility and usefulness of the subject.

Measures of Central Tendency

Although frequency distributions serve useful purposes, there are many situations that require other types of data summarization. Measures of central tendency and sometimes needed to make meaningful interpretation of the data. Generally, it is found that in any distribution, values of the variable tend to congregate around a central value of the distribution. This tendency of the distribution is known as its central tendency and the measures devised to consider this tendency are known as measures of central tendency. One of the most important objectives of this statistical analysis is to get a single value that describes the characteristic of the entire mass of data. Such a value is called an “average”. Since an average is somewhere within the range of the data, it is also called the measure of central value. It reduces the complexity of the data and compares them. According to R.A. Fisher, “The inherent inability of the human mind to grasp in its entirely a large body of numerical data compels us to seek relatively few constants that will adequately describe the data”. The human mind is incapable of remembering the entire mass of unwieldy data. Therefore, the data need to be summarized into a few summary measures which can reveal their basic features.

An average is a method of condensing a mass of data to a single figure, which is typical and fully representative of the entire data. It represents the whole group. The functions of an average are.

(i) To present the salient features of a mass of complex data.

(ii) To facilitate comparison

(iii)To know about the population from a sample.

(iv)To help in decision making.

However, some desirable properties of a good measure of central tendency are as follows.

(i) It should be rigidly defined.

(ii) It should be easily comprehensible and easy to calculate.

(iii) It should be capable of further mathematical treatment.

(iv) It should be based on all the observations.

(v) It should not be affected by fluctuations of random sampling

(vi) It should not be unduly affected by extreme observations.

(vii) It should be easy to understand.

(viii) It should have sampling stability.

When we discuss the distribution of data in biology, agriculture or in medicine, there are usually five basic measures of the central tendency. These are the arithmetic mean, median mode, geometric mean and harmonic mean.

Conventional symbols

A sample of a population consists of *n* observations with a mean of x (read *x*bar). A measure computed from the data of a sample is called a statistic. However, the mean of a population is symbolized by μ which is pronounced as mu. A measure computed from the data of a population is called a parameter. The number of observations in a population is denoted by the capital italic letter *N*. Generally in biostatistics, we use Roman Letters to symbolize sample information and Greek letters to symbolize population information. It is impossible and impractical to get all observations of the population, thus μ remains unknown and the sample mean, x serves as an estimate of μ in bioexperimental studies.

$$\text{Samplemean } (x) = \frac{\Sigma x}{n}$$

$$\text{Population mean } (\mu) = \frac{\Sigma x}{N}$$

Where Σx = sum of values of all observations

n or *N* = number of observations in the sample or population, respectively.

Arithmetic mean

The most familiar and widely used measure of central tendency is the arithmetic mean. It represents the entire data by one value which is obtained by adding together all the values and dividing this total by the number of observations. In other words, the arithmetic mean is the sum of all observations divided by the total number of observations. However, each independent value plays an equal part in the determination of the mean.

Calculation of arithmetic mean

Calculation of arithmetic mean in a series of individual observations

The arithmetic mean of individual series can be calculated either by direct method or by short-cut method. However, the results of both the methods will be the same.

Direct method: the arithmetic mean is computed by summing up the observations and dividing the sum by the total number of observations. Symbolically;

Equ

Where x = arithmetic mean

Σx = sum of all values of the variable, x

e.e. x_1, x_2, x_3, ----- x_n

n = number of observationsS

Example 1: Calculate the arithmetic mean of the following set of observations.

7,6 8, 10, 13, 14

Solution

Equ

Example 2: The following data represents the number of productive tillers per plant of a wheat variety. Calculate the mean number of tillers per plant.

Number of productive tillers = 10, 11, 10, 11, 9, 7, 9, 11, 12, 10

Solution

Equ

Short-cut method: The short-cut method is used where the number of observations are large to avoid the considerable difficult in the calculation of arithmetic mean. This method is based on an important property that the algebraic sum of the deviations of a series of individual observations from their mean is equal to zero. Four example, the arithmetic mean of 7, 6, 8, 10 is 8. the deviations from the mean are -1, -2, +1, 0, +2. Thus the total is zero. Therefore, we can assume an arbitrary mean to find out the deviations from this assumed mean. Here, the total of deviations will not be zero. The total is divided by the number of observations and added to the assumed mean which will be the actual arithmetic mean. Symbolically.

Equ

Where x = arithmetic mean

A = assumed mean

D = deviation of items from the assumed mean (x-*A*)

Σd = sum of deviations

Example 3: The following example shows the calculation of the arithmetic mean from the number of the spikelets per spikein wheat.

Number of spikelets per spike = 18, 20, 21, 19, 28, 22, 29, 30, 31, 35

Solution: Assumed mean = 24

Spikelets per spike X	**Deviations from the assumed mean *d*=(x-*A*)**
18	-6
19	-5
20	-4
21	-3
22	-2
28	4
29	5
30	6
31	7
35	11
n=10	Σd = -20+33 = 13

Equ

Calculation of arithmetic mean in a discrete series

For computation of arithmetic mean in a discrete series, the values of the variable are multiplied by their respective frequencies. However, the number of observations are equal to the total of the frequencies. The formulae for computing the arithmetic mean by direct and short-cut methods are as follows.

Direct method

In a discrete series, the frequency of each row is multiplied by its respective value. The total is then divided by the total number of frequencies to find out the arithmetic mean. Symbolically.

Equ

Where x = arithmetic mean

Σf = sum of frequencies

Σfx = sum of values of the variable and their corresponding frequencies

Example 4: The data recorded on the number of chlorophyll deficient plants in a lentil population are given below. Calculate the arithmetic mean.

Number of chlorophyll deficient plants	Number of plants
0	34
1	14
2	20
3	24
4	25
5	33

Solution

Number of chlorophyll deficient plants *X*	Number of plants *f*	*fx*
0	34	0
1	14	14
2	20	40
3	24	72
4	25	100
5	33	165
	$\Sigma f = 150$	$\Sigma fx = 391$

Equ

x = 2.61

Short-cut method: In this method, the deviations of the items from an assumed mean are multiplied by their respective frequencies and summed up. This total is then divided by the total frequencies and added to the assumed mean. The resulting figure is the actual arithmetic mean. The arithmetic mean is calculated on the basis of this given formula.

Equ

Where x = arithmetic mean

A = assumed mean

Σfd= sum of the deviations from the assumed mean and the respective frequencies

Σf = sum of the frequencies

Example 5: Calculate the arithmetic mean from the data given in the table.

Number of leaves	Number of plants
5	6
10	5
15	9
20	4
25	3
30	6
35	4
40	3

Solution : Assumed mean = 25

Number of leaves ***X***	Number of plants ***F***	***D* = (x-25)**	***Fd***
5	6	-20	-120
10	5	-15	-75
15	9	-10	-90
20	4	-5	-20
25	3	0	0
30	6	5	30
35	4	10	40
40	3	15	45
	Σf = 40		Σfd=-190

Equ

x = 20.25

Calculation of arithmetic mean in a continuous series

In a continuous series, the arithmetic mean may be calculated after taking into consideration the mid-point of various classes. However, the method will be the same for both inclusive class-intervals as well as for exclusive class-intervals. However, the mean can be calculated by any of the following methods. The arithmetic mean will be same in all these methods.

Direct method

The following formula can be used for calculation of arithmetic mean in a continuous series by direct method.

Equ

Where x = arithmetic mean

Σfm = sum of values of mix-points multiplied by the respective frequency of each class

Σf = sum of frequency

f = frequency of each class

m = mid-point of various classes.

$$\text{Mid-point } (m) = \frac{\text{Lower limit + Upper limit}}{2}$$

Example 6: Compute the arithmetic mean from the following data.

Plant height (cms)	0-10	10-20	20-30	30-40	40-50	50-60
Number of varieties	5	10	25	30	20	10

Solution

Plant height (cms) Classes	**Number of varieties *F***	**Mid-points *M***	***Fm***
0-10	5	5	25
10-20	10	15	150
20-30	25	25	625
30-40	30	35	1050
40-50	20	45	900
50-60	10	55	550
	$\Sigma f = 100$		$\Sigma fm = 3300$

Equ

x = 33

Short-cut method: The following formula is used for calculating arithmetic mean in a continuous series by a short-cut method.

Equ

Where x = arithmetic mean

A= Assumed mean

d = deviations of mix-points from assumed mean (m-*A*)

Σ*f* = sum of frequency

f = frequency of each class

Example 7: Calculate the arithmetic mean from the data of example 6.

Solution: Assumed mean = 35.

Plant height (cms) Classes	Mid-point *m*	Number of varieties *f*	*d*=(m-*A*)	*fd*
0-10	5	5	-30	-150
10-20	15	10	-20	-200
20-30	25	25	-10	-250
30-40	35	30	0	0
40-50	45	20	10	200
50-60	55	10	20	200
		Σ*f* = 100		Σ*fm* = -200

Equ

x = 33

Step deviation method

In this method, we take class-intervals into consideration to simplify the calculations. This is true for the equal class-intervals. However, in the case of unequal class-intervals, we can further simplify the calculations by taking a common factor instead of the class-intervals. Symbolically;

Equ

Where x = arithmetic mean

A= Assumed mean

D` = deviations of mix-points from assumed mean (m-*A*)/i or c

Σ*f* = sum of frequency

i = class-invervals

c = common factor

d= deviations from assumed mean

Example 8: Calculate the arithmetic mean from the data of example 6.

Solution: Assumed mean = 35

Plant height (cms) Classes	**Mid-points *M***	**Number of varieties *f***	***d*=(m-35)**	***d`*=(m-35)/i**	***fd`***
0-10	5	5	-30	-3	-15
10-20	15	10	-20	-2	-20
20-30	25	25	-10	-1	-25
30-40	35	30	0	0	0
40-50	45	20	10	1	20
50-60	55	10	20	2	20
		$\Sigma f = 100$			$\Sigma fd` = -20$

Equ

Calculation of arithmetic mean in series having open-end classes

In some cases, the lower limit of the first class-interval and the upper limit of the last class-interval are not known. Such class-intervals are called open-end classes. Arithmetic mean cannot be calculated in open-end classes unless we know the lower and upper limits of the classes. In such cases, an assumption should be made for unknown class-limit. It depends upon the class intervals.

Example 9

Number of pods	**Number of plants**
Below 10	4
10-20	6
20-30	20
30-40	12
40-50	10
50-60	5
Above 60	4

It is clear from the above example that the class-interval is uniform, therefore, the lower limit of the first class would be zero and the last limit would be 70. The first class will be 0-10 and the last 60-70.

Example 10

Number of pods	**Number of plants**
Below 10	4
10-30	6
30-60	20
60-100	12
100-150	10
Above 150	5

However, in this case the class-interval is increasing by 10. Therefore, the appropriate assumption would be that the first class 0-10 and the last class is 150-210.

Merits and demerits of arithmetic mean

Merits: Arithmetic mean is the simplest measurement of central tendency of a series. It is widely used because.

(i) It is easy to understand and easy to calculate.

(ii) It is rigidly defined.

(iii) It is based on all the observations.

(iv) It provides a good basis for comparison.

(v) It is amenable to further mathematical treatment.

(vi) It is the most reliable mean to use when sample data are being used to make inferences about population.

(vii) It is not affected by fluctuations of sampling.

Demerits

(i) The mean is unduly affected by the extreme items.

(ii) It cannot be a accurately determined even if one of the values is not known.

(iii) The arithmetic mean cannot be calculated in open-end classes without making assumptions about the size of the class-intervals.

Median

The "Median" is another important and widely used measure of central tendency. The median is usually defined as that value which divides a distribution so that an equal number of items occur on either side of it. In other words, 50 per cent of the observations will be smaller than the median. If the number of values is odd, the median will be the middle value when all values have been arranged either in ascending or descending order of magnitude. Let us now obtain the median of the sample consisting of values 70, 84, 86, 99 and 103 of the total number of spikes per panicle in rice. The value of median is 86, since there are two items below this values, and two items above it. When the number of observation is even, there are two middle values. In this case, the median is taken to be the mean of these two middle values when data are arranged in order of their magnitude. Let us illustrate this by finding the median of the values 70, 84, 86, 99, 103, 107, 108 and 109 of the total number of spikes per panicle in rice. In this instance, the median would be (99+103)/2 = 101.

The definition is, therefore, recast thus: "the median is that value of the variable which divides the group into two equal part, one part comprising all values greater and the other, all values less than the median"

Calculation of median

The data are arranged in ascending order of magnitude to find out the value of the median. If the data set contain an odd number of values, the middle of the array is the median. If there is an even number of items, the median is the average of the middle items.

Calculation of median in a series of individual obseravations

For calculating median in a series of individual observations, the following steps are taken.

(i) Arrange data in ascending or descending order. The answer will be the same both ways.

(ii) Median is located by finding the size of n+1/2 th item.

Symbolically

M = size of the n+1/2 th item

Where *M* = median

n = number of observations

Example 11: Find out the median from the data recorded on the number of clusters per plant in a pulse crop.

Number of clusters = 10, 18, 17, 19, 10, 15, 11, 17, 12

Solution: Arrange the data either in ascending or descending order.

S.No.	Data arranged in ascending order
1	10
2	10
3	11
4	12
5	15
6	17
7	17
8	18
9	19

Median = Size of n+1/z th item

Median = size of 9+1/2 th item

Median = Zise of 5th item + 15

Median = 15

Example 12: Data recorded on the production of butterfat during 10 consecutive days in cows is presented below. Calculate the median.

Butterfat (kgs) = 4.0, 5.7, 3.9, 4.2, 6.6, 7.0, 7.9, 8.0, 9.0, 10.0

Solution: Data arranged in ascending order

S.No.	Butterfat (kgs.)
1	3.9
2	4.0
3	4.2
4	……
5	6.6
7	7.9
8	8.0
9	9.0
10	10.0

Median = Size of (n+1)/2th item

Median = Size of (10+1)/2th item of 5.5th item

Size of 5.5th item = Size of 5th item + Size of 6th item/2

Size of 5.5th item = 6.6 + 7.0 / 2 = 13.6/2 = 6.8

Median = 6.8

Calculation of median in a discrete series

For the computation of the median in discrete series, one should take into consideration the following steps:

(i) Data should be arranged in ascending or descending order of magnitude.

(ii) Find out the cumulative frequencies.

(iii)Median = Size of (n+1)/2th item

(iv) Find out the value of (n+1)/2th item. It can be found by first locating the cumulative frequency which is equal to (n+1)/2 or the next higher to this and then determine the value corresponding to it. This will be the value of median.

Example 13: Find out the value of the median from the following data.

Number of angular seeded plants	12	8	17	10	11	16	18	14	6	7
Number of plants	39	33	42	40	47	42	60	50	22	25

Solution

Angular seeded plants in ascending order	**Number of plants** ***f***	**Cumulative frequency** ***cf***
6	22	22
7	25	47
8	33	80
10	40	120
11	47	167
12	39	206
14	50	256
16	42	340
18	60	400

Median = Size of (n+1)/2th item or

Median = Size of (400+1)/2 or 200.5th item

Median = Size of 200.5th item = 12

Median = 12

Calculation of median in a continuous series

While computing the value of the median in a continuous series, first determine the particular class in which the value of the median lies. Use *n*/2 as the rank of the median and not (n+1)/2. Hence, it is *n*/2 which will divide the area of the curve into two parts. The following formula used for determining the exact value of the median.

Equ

Where L = lower limit of the median class

cf = cumulative frequency of the class preceding the median class

f = frequency of the median class

i = class-interval

However, it should be remembered that instead of the lower limit of the median class, one can take the upper limit of the median class. In this case, class-intervals should be in descending order and frequencies are to be cumulative from top to bottom.

Example 14: Calculate the value of the median from the data recorded on the number of grains per earhead on 300 wheat earheads.

Number of grains per earhead	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-50
Frequency	2	27	52	118	57	27	13	4

Solution

Number of grains per earhead classes	**Frequency *f***	**Cumulative frequency *Cf***
5-10	2	2
10-15	27	29
15-20	52	81
20-25	118	199
25-30	57	256
30-35	27	283
35-40	13	296
40-45	4	300

Median = Size of n/2th item =300/2 = 150

Median lies in the class = 20-25

L=20; n/2 = 150; *cf* = 81; i = 5; *f* = 118

Equ

Median = 22.92

Calculation of median in unequal class-intervals

In unequal class-intervals, frequencies need not be adjusted to make the class intervals equal. The formula given in continuous series can be used here.

Equ

Example 15: Calculate the median from the following data.

Classes	01-10	10-30	30-60	60-80	80-90	90-100
Frequency	5	16	30	12	6	1

Solution

Classes	**Frequency** ***f***	**Cumulative frequency** ***cf***
0-10	5	5
10-30	16	21
30-60	30	51
60-80	12	63
80-90	6	69
90-100	1	70

Equ

Median = Size of n/2 = 70/2 = 35th item

Median lies in the class = 30-60

L = 30; n/2 = 35; cf = 21; f = 30; i = 30

Equ

Median = 44

Check: If we make the class-intervals equal, the same answer should be obtained

Classes	**Frequency** ***f***	**Cumulative frequency** **(*cf*)**
0-10	5	5
10-20	8	13
20-30	8	21
30-40	10	31
40-50	10	41
50-60	10	51
60-70	6	57
70-80	6	63
80-90	6	69
90-100	1	70

Equ

Median = 44

Calculation of median in open-end classes

Since the median is not affected by the values of extreme ends, we are not concerned with extreme values for calculation of the median in open-end classes.

Example 16: Calculate the median in open-end series.

Value	Less than 10	10-20	20-30	30-40	40 and above
Frequency	4	8	14	6	4

Size of item classes	**Frequency** ***f***	**Cumulative frequency** ***cf***
Less than 10	4	4
10-20	8	12
20-30	14	26
30-40	6	32
40 and above	4	36

Median = size of n/2 = 36/2 = 18^{th} item

Median lies in the class = 20-30

Equ

L =20; n/2 = 18; cf = 12; f = 14; i = 10

Equ

Median = 24.29

Graphic location of median

Median can be located graphically by applying any one of the following methods:

(i) Draw only one ogive by "less than" or "more than" method.

(ii) Draw two ogives – one by "less than" method and the other by "more than" method.

Example 17: Determine the value of median graphically.

Variable	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**	**60-70**	**70-80**
Frequency	4	6	10	15	11	7	3

Solution

Less than ogive		More than ogive	
Variable *x*	**Cumulative frequency** *cf*	**Variable** *x*	**Cumulative frequency** *cf*
20	4	10	56
30	10	20	52
40	20	30	46
50	35	40	36
60	46	50	21
70	53	60	10
80	56	70	3
		80	0

Median = Size of *n*/2th item = 56/2 = 28^{th} item.

Median = 28^{th} items.

Method 1: if we draw one ogive by “less than” method, we can determine the value of the median from it. Take 28 on the y-axis and draw a perpendicular on the ogive. From the point where it meets the ogive, draw another perpendicular on the x-axis. This will give the value of the median (Figure 4.1a).

Method II: Draw twoogives – one by “less than” method and the other by “more than” method. From the point where these two curves intersect, draw a perpendicular on the x-axis. This point on the x-axis will give the value of the median (Figure 4.1b).

However, with both these methods, the value of the median will be the same.

Check Median = Size of n/2 th item

Median = 56/2 = 28^{th} item.

Figure 4.1(a): Graphic location of median

Figure 4.1(b) : Graphic location of median

Equ

Median = 40 + 5.33

Median = 45.33

It is clear from these methods as well as from the calculations, that the median value is 45.33. This indicates that the median value can be determined graphically also.

Merits and demerits of median

Merits

(i) It is easy to define and easy to understand.

(ii) It is especially useful in the case of open-end classes. Since the median value is located in the median class-interval and that interval is virtually certain of not being open-ended, the median may be determined.

(iii) The median is also recommended in unequal class distributions.

(iv) It is frequently used in skewed distributions. The median will not be affected by the size of values of extreme items.

(v) The value of the median can be determined graphically. However, the value of the mean cannot be graphically ascertained.

Demerits

(i) It is not based on all the observations since it is a positional average.

(ii) Median is affected more by sampling fluctuations, than by the value of the mean.

(iii) If the number of observations is even, we cannot calculate the median. In this case, the mean of two middle values will be the estimate of the median.

(iv) Median would be unsuitable in case of large and small items in a series.

MODE

The "Mode" is another measure of central tendency which is conceptually very useful. It is derived from the French word "La mode" which means fashion. Mode is the most fashionable or typical value of a distribution becauseit is repeated the highest number of times in the series. The modeis, by definition, the most commonly occurring value. However, according to Croxton and Cowden, "The mdoe of a distribution is the value at the point around which the items tend to be most heavily concentrated". It may be regarded as the most typical of the series of values. The main feature of mode is that it is the size of that item which has the maximum frequency. A set of data may have a single mode, in which case it is said to be unimodal. When concentration of data occurs at two or more points, such a series is called *bimodal* or *multimodal*, depending on the number of such points. For example, the data recorded on the number of seeds per pod are given in three sets to find out the most commonly occurring value.

(i) 8, 9, 11, 10, 10, 7, 6, 12

(ii) 8, 10, 10, 9, 12, 7, 11, 11

(iii) 6, 7, 9, 12, 11, 10, 8, 5

It is clear from the above given data that the mode in the first series would be 10, since this value occurs more frequently than any other value. It has a single mode 10. In the second case, it has 10 and 11 as modes because both occur two times. However, the third set of data has no mode.

The following diagram shows the mode value. The value of the variable at which the curve reaches the maximum is called the mode. Mode would be the value on x-axis which corresponds to the maximum frequency on y-axis (Figure 4.2).

Figure 4.2: Diagram representing mode

Mode in symmetrical and skewed distributions

In a perfectly symmetrical distribution, there is only one mode, the three measures of central tendency – the mode, median and mean-coincide with the highest point. If the data are such that a skewed distribution, the mode is at the highest point and the median lies to the right of this point, but the mean falls to the.

Figure 4.3 (a) : Mode median and mean in symmetrical distribution

Figure 4.3 (a) : Mode median and mean in positively skewed distribution

Figure 4.3 (a) : Mode median and mean in negatively skewed distribution

mean falls to the right of the median. However, in negatively skewed distribution, the mode will have the largest value, the median lies to the left of the mode and the mean falls to the left of the median. In the negatively skewed distribution, the value of mean is lower than those of the other two measures. Whereas, in positively skewed distribution, the opposite is true. In brief, whatever may be the shape of the curve, the mode is always located at the highest point (Figure 4.3 a,b,c).

Calculation of mode

Calculation of mode in a series of individual observations

Mode can be ascertained by mere inspection in case of individual observations. The values occurring the maximum number of times are the modal values.

Example 18: Find the mode of the following data relating to the weights of a sample of 10 experimental animals:

S.No.	1	2	3	4	5	6	7	8	9	10
Weight (kgs.)	10	11	10	12	12	11	9	8	11	11

Solution

Weight (kgs)	Number of animals
8	1
9	1
10	2
11	4
12	1
13	1

Since the item 11 occurs the maximum number of times, i.e times, the model values is 11. If two values have the maximum frequency, the series is bimodal.

Calculation of mode in a discrete series

Mode can be determined by looking to that value of the variable around which the items are most heavily concentrated.

Example 19: Find the value of the mode from the following data.

Value	6	8	10	12	14	16	18	20	22	24
Frequency	20	30	40	40	55	60	55	20	15	25

Solution

It is clear from the above given example that the modal value is 16, because it occurs the maximum number of times, i.e. 60 times. However, the values on either side of 16 have also large frequencies. In such cases, it is desirable to prepare a grouping table and an analysis table.

Grouping method: A grouping table has size columns. Frequencies can be added in two's or in three's. Grouping is done as given below.

(i) In column 1, maximum frequency is marked.

(ii) In column 2, requencies are added in two's (1 and 2; 3 and 4.5 and 6 and soon)

(iii) In column 3, frequencies are added in two's leaving the first frequency (2 and 3; 4 and 5; 6 and 7 and so on).

(iv) In column 4, frequencies are added in three's (1,2 and 3; 4,5 and 6;7,8 and 9 and so on)

(v) In column 5, frequencies are added in three's leaving the first frequency (2,3 and 4;5, 6 and 7; 8,9 and 10 and so on).

(vi) In column 6, frequencies are added in three's leaving the first two frequencies (3,4 and 5; 6,7 and 8; 9, 10 and 11 and so on)

In all the processes, mark down the maximum frequencies by bold letters or by a circle. After preparing a grouping table, prepare an analysis table to find out the exact value of the modal class.

Variable (x)	2	4	6	8	10	12	14	16	18	20
Frequency (f)	15	20	25	27	30	20	15	21	10	11

Solution

Grouping table

Frequency (*f*)

Variable (x)	**Col.1**	**Col.2**	**Col.3**	**Col.4**	**Col.5**	**Col.6**
2	15					
		35				
4	20					
			45			
6	25					
		5		60	72	
			57			82
10	30					
		50		77	65	
12	20					
			35			47
14	15					
		27		37	33	
16	12					
			22			
18	10					
		21				
20	1					

Analysis table

Variable (x)

Column number	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**
1					1					
2			1	1						
3				1	1					
4				1	1	1				
5		1	1	1						
6			1	1	1					
Total		1	3	5	4	1				

In this example, the model values is 10, because it shows the maximum frequency. However, the concentrations of the items is more on either side of this value. But it was disclosed after the groping analysis that the modal value is 8 and not 10. Therefore, one should not rely on the inspection of data to locate the mode at the point of maximum frequency. According to Elhance and Elhance, "Mode is affected by the frequencies of the neighbouring items also and, therefore grouping is essential as it reveals the true point of maximum concentration".

Calculation of mode in a continuous series

In continuous series, the modal class should be a certained first which is defined as the class having maximum frequency. In an other way by preparing the grouping table and the analysis table, mode is calculated by applying the following formula:

Equ

Where L = the lower limit of the modal class

Δ_1 = the difference between the frequency of the modal class and the frequency of the preceding class (ignoring signs)

Δ_2 = the difference between the frequency of the modal class and the frequency of the succeeding class (ignoring signs)

i = the interval of the modal class

The above given formula can be applied for calculation of mode, if the class-intervals are uniform throughout. It is already mentioned that the mode is affected by the frequencies of the neighbouring classes. If, in case the frequency of the preceding class is greater than the frequency of the succeeding class, the mode would be nearer to the lower limit of the class-interval. In the other case, the situation will be the reverse when the frequency of the succeeding class is more than the preceding class, the mode would be nearer to the upper limit of the class-interval. The value of the mode cannot be determined by applying this formula in bimodal distributions. If, it is used when the size of the class-intervals and the frequencies are very small, the distribution becomes irregular.

Example 21: From the following data of the weight of 125 students, determine the model weight.

Weight (lbs.)	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**	**90-100**	**100-110**	**110-120**
Number of students	8	12	24	31	32	11	5	2

Solution

It is difficult to determine the modal class by inspection. Therefore, it is necessary to prepare the grouping table and an analysis table.

Grouping table

Frequency (*f*)

Weight (lbs)	**Col.1**	**Col.2**	**Col.3**	**Col.4**	**Col.5**	**Col.6**
40-50	8					
		20				
50-60	12					
			36			
60-70	24					
		55		44	67	
70-80	31					
			63			87
80-90	32					
		43		74	48	
90-100	11					
			16			18
100-110	5					
		7				
110-120	2					

Analysis table

Variable (x)

Column number	50-60	60-70	70-80	80-90	90-100
1				**1**	
2		**1**	**1**		
3			**1**	**1**	
4			**1**	**1**	**1**
5	**1**	**1**	**1**		
6		**1**	**1**	**1**	
Total	**1**	**3**	**5**	**4**	**1**

It is clear from this table that the modal class is 70-80 and not 80-90 as observed by inspection. Therefore,

Equ

Mode = 78.75

Calculation of mode in unequal class-intervas

The above given formula can be applied only in problems where the class-intervals are equal. Before, we compute the mode in unequal class-intervals, the class-intervals should be made equal and frequencies should be adjusted accordingly.

Example 22: Calculate the mode of the following distribution.

Number of grains	100-110	110-130	130-140	140-160	160-170	170-180
Number of plants	11	40	27	34	12	6

Solution: In this problem, the class-intervals are unequal, therefore, we must adjust the frequencies and make the class-intervals equal.

Number of grains per panicle	Number of plants
100-110	11
110-120	20
120-130	20
130-140	27
140-150	17
150-160	17
160-170	12
170-180	6

Mode lies in the class = 130-140

Equ

L = 130; Δ_1 = (27-20) = 7; Δ_2 = (27-17) = 10; i =10

Equ

Mode = 134.12

Graphic location of mode

In a frequency distribution, the value of mode can be determined graphically.

For this we should draw the histogram of the data first. Since, the modal class is the tallest rectangle, we can locate the mode in it by.

(i) Drawing the line from the top right corner of the tallest rectangle to the top right corner of the rectangle to its immediate left.

(ii) Drawing a second line from the top left corner of the tallest rectangle to the top left corner of the rectangle to its immediate right.

(iii) Then a perpendicular is drawn from the intersection of these lines to the x-axis which gives the value of mode.

Example 23: The data on the number of pods per plant are distributed as follows. Determine the modal value of the given distribution graphically and verify the results by direct calculation.

Number of pods per plant	**01-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**
Number of plants	1	5	11	24	30	20	15	7	3

Solution

Number of pods per plant	**Number of plants**
0-10	1
10-20	5
20-30	11
30-40	24
40-50	30
50-60	20
60-70	15
70-80	7
80-90	3

Equ

Mode lies in the class = 40-50

L = 40; Δ_1 = (30-24) = 6; Δ_2 = (30-20) = 10; i =10

Equ

Mode = 43.75

Graphic location: Since the modal line touches the x-axis at 43.75, the value of the mode is 43.75 (Figure 4.4)

Figure 4.4: Graphic location of the mode

Merits and demerits of mode

Merits

(i) Mode is easy to calculate and can be determined by a mere observation of the data.

(ii) The mode is not unduly affected by extreme items.

(iii) It is simple and precise.

(iv) Mode is that point where there is more concentration of frequencies. Hence, it is the best representative of the data.

(v) It can be located in open-end distributions.

Demerits

(i) The mode is not based on all the observations.

(ii) The value of the mode cannot be determined in bimodal distribution.

(iii) It is not a rigidly defined measure. Sometimes the exact value of the modal class cannot be known by inspection of the data. Therefore, it is necessary to prepare the grouping table and then an analysis table to find out the modal class.

Relationship among Mean, Median and Mode

In a perfectly symmetrical distribution, the value of mean, median and mode coincide (i.e. mean = median = mode). However, in asymmetrical distribution, these values are not equal. When distribution is positive skewed, the mean is with the highest value ; followed by the median, and then the mode. In the case of negatively skewed distribution, the mode has the highest value; followed by the median, and then the mean. An important relationship exists among mean, median and mode which is given by the following formula.

Mean-mode = 3 (mean-Median)

Mode = 3 Median – 2mean

Example 24: In a moderately skewed distribution, the values of mode and median are 11 and 14, respectively. Find out the value of mean.

Solution

Mode = 3 median – 2 mean

Mode = 11; Median = 14

Mode = 3 Median – 2 Mean

11 = 3 x 14 – 2 Mean

11 = 42-2 Mean

2 Mean = 11-42 = 31

Mean = 31/2

Mean = 15.5

Example 25: Find out the value of median if mean = 16 and mode = 21.

Solution

Mode = 3 Median – 2 Mean

Mean = 16; Mode = 21

Mode = 3 Median – 2 Mean

21 = 3 Median – 2 x 16

21 = 3 Median - 32

21 + 32 = 3 median

3 Median = 53

Median = 53/3

Median = 17.67

Example 26: The median of a series is 44 and mean is 36. What will be the value of mode?

Solution

Mode = 3 Median – 2 Mean

Median = 44; Mean = 36

Mode = 3 Median – 2 Mean

Mode = 3 x 44-2 x 36

Mode = 132-72

Mode = 60

Geometric Mean

The most important measures of central tendency are mean, median and mode because of their wide usefulness, simplicity and general applicability. Under certain conditions, other measures of central tendency may be useful, and we shall, therefore, consider the geometric mean and the harmonic mean.

The geometric mean is defined as the nth root of the product of n items of a series. If the geometric mean of items 5 and 20 is to be calculated, we apply the square root. Similarly, the geometric mean of items 10,20 and 30 would be the cube root of the product of these figures. If there are four items, it would be the 4^{th} root of the product of these four figures and so on.

Symbolically;

Equ

Where *G.M.i* = geometric mean

n = number of items

x = values of the variable

Thus, the geometric mean for four values 5, 10 , 12 and 15 would be:

Equ

It is interesting to note that the arithmetic mean of these four items is 10.5. The value of the geometric mean is always less than the value of the arithmetic mean. However, it is difficult to calculate the geometric mean if the number of items are more. Therefore, in such cases, calculations have to be donewith the help of logarithms.

Graphic representation of data

Graphic methods enable the statisticians to present quantitative data in a simple, clear and effective manner. However, the important step in statistical analysis is to prepare a frequency distribution table. But the graphic representation of data in the frequency distribution table reveals the relationship that might be overlooked in a table. The frequency table of most biological variables develops a distribution which can be compared with the standard distributions, such as normal, binomial and poisson.

A graph is a visual form of representation of statistical data. Comparison can be made between two or more phenomena very easily with the help of a graph. According to I.R. Vesselo, "The simplest and commonest aid to the numerical reading is the graph. This gives a picture of numbers in such a way that the relations between them strike the eye. It is of the greatest value in making numbers clear". The frequency distribution can be represented graphically in any of the following ways:

(i) Histogram
(ii) Frequency polygon
(iii) Frequency curve
(iv) Cumulative frequency curve

The data collected on the number of grains per spike of a wheat variety, "Bansi" are given in table.

Table: Number of grains per spike in wheat.

31	27	31	30	27	32	17	34	23	24	30	37	26
30	29	33	26	21	36	30	32	22	34	25	37	31
24	36	27	38	32	28	29	29	22	23	28	30	19
18	31	41	226									

Table: Rearranged data in the ascending order of magnitude of table.

Table 3.6 provides the raw data for the number of grains per spike for 130 observations. The frequency depends o the umber of plats carrying the definite number of grains per spike. Data are arranged in the ascending order of the magnitude (Table 3.7). The difference between the largest and the smallest values is called the range of the distribution of data. In the present example, the rage is 43-17 = 26 which can be divided into 9 classes by taking class intervals of 3 as shown in the following page.

Equ

Where *L* = largest value = 43
S = smallest value = 17
c = 1 + 3.322 *log N*

N = number of observations

Equ

The mid-point of each class is taken as the class value ad the frequency of a class consists of the number of observations falling within the limits of that class. Normally, there should be 7 to 15 classes which are recommended in the frequency distribution. For instance, the class includes observations from 17 to 19 and values from 20 to 22 fall in the second class, and similarly values from 23 to 25 fall into the third class. If the number of observations are large, we form more classes; if the number of observations are small, we form fewer classes. The lower class boundaries of 9 classes are 17, 20, 23, 26, 29, 32, 35, 38 and 41 and the upper classes boundaries are 19, 22, 25, 28, 31, 34, 37, 40 and 43, respectively (Table 3.8). Therefore, the class length (class width) of any class is the difference between the upper and the lower class boundaries.

Construction of a graph: It will be convenient in discussing frequency graphsto use the conventional mathematical terms, ordinate and abscissa. In the construction of frequency graphs, the values of the variable are measured on the x-axis and the corresponding frequencies on the y-axis. For construction of graphs, two lines are drawn which cut each other at right angles. The horizontal line is called abscissa and the vertical line is called ordinate. The point where these lines cut each other is called the point of origin or O. Together, the ordinate and abscissa are called co-ordinates to the point.

Table 3.8: Frequency distribution of number of grains per spike as given in table 3.6.

Class intervals	Class value	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency
17-19	18	8	8	0.062	0.062
20-22	21	15	23	0.115	0.177
23-25	24	18	41	0.138	0.315
26-28	27	21	62	0.162	0.477
29-31	30	26	88	0.200	0.677
32-34	33	19	107	0.146	0.823
35-37	36	12	119	0.092	0.915
38-40	39	7	126	0.054	0.969
41-43	42	4	130	0.031	1.000
Total		130			1.000

To plot a point on a graph, two values are needed, 'independent variable' or 'x-variable' and 'dependable variable' or 'y-variable'. In graph, independent variables are shown along x-axis and dependable variables along y-axis. However, two values are related in such a way that the changes in one variable depend upon the changes in the other. For instance, with the increase is number of pods per plant, there will be a concomitant increase in plant yield.

For construction of a graph, the graph paper is divided into four parts, called 'quadrants'. However, both the x and y values are positive in quadrant I. in quadrant II, the x's are negative and y's are positive. In quadrant III, both x and y are negative. In quadrant IV, x values are positive and y values are negative. The following details are given in figure 3.8.

Figure 3.8: Construction of graph.

Histogram

Graphs of frequency distributions are useful because they emphasize and clarify the trends that are not so readily discernible in the frequency tables. There are certain methods of presenting the frequency distribution graphically; however, histogram is the most important method for displaying the frequency distribution. Histogram is a set of vertical bars whose areas are proportional to the frequencies represented.

In constructing the histogram, the variable, number of grains per spike should be taken on the horizontal axis (x-axis) and the frequencies depending on it on the vertical axis (y-axis). Each class is represented by a distance which is always proportional to its class interval. When all the classes are of equal lengths, the heights of the rectangles will be proportional to the frequencies of the respective classes. In this way, there are a number of rectangles each with a class interval distance as its width, and the frequency distance as its height. Histogram is two-dimensional where both the length as the width are important. Whereas a bar diagram is one-dimensional, i.e. only the length of the bar is important, and not the width.

The histogram can be constructed in two ways depending upon the class – intervals:

(i) For distributions that have equal class intervals.

(ii) For distributions that have unequal class intervals.

In the first case, if the class intervals are equal, the variable number of grains per spike is taken on the x-axis and the number of platns on the y-axis. In such a case, the height of the rectangles will be proportional tot eh frequency (Figure 3.9 and table 3.1).

Table 3.9: Draw a histogram from the data presented in table 3.8.

Number of grains	Number of plants
17-19	8
20-22	15
23-25	18
26-28	21
29-31	26
32-34	19
35-37	12
38-40	7
41-43	4

In case of unequal class intervals, a correction must be made. One has to take the lowest class interval into consideration for making suitable adjustments in the frequencies of other classes. For example, if one class interval is twice as compared to the lowest class interval, the height of the rectangle is divided by two. On the other hand, if it is three times more, we divide the height of its rectangle by three and so on. The following example will illustrate frequency distribution that has unequal class intervals (Figure 3.10 and table 3.10).

Table 3.10: Represent the data by means of a histogram (unequal class intervals).

Number of pods	Number of plants
10-20	16
20-30	24
30-40	39
40-50	25
50-70	20
70-110	20
110-150	12

Figure 3.9: Histogram representing the distribution of number of grains per spike as given in table 3.9.

Solution: If the class intervals are unequal, "frequency density" has to be calculated. After this adjustment, histogram can be drawn on the basis of the data given below.

Number of pods	Number of plants	Adjustment (10=1)	Frequency density
10-20	16	1	16
20-30	24	1	24
30-40	39	1	39
40-50	25	1	25
50-70	20	2	10
70-110	20	4	5
110-150	12	4	3

Histogram for unequal class intervals is drawn in figure 3.10.

Figure 3.10: Histogram representing the distribution of the number of pods per plant as given in table 3.10

Frequency polygon

Frequency distribution can be portrayed graphically by means of a frequency polygon. To construct a frequency polygon, we mark the frequencies. (number of plants) on the vertical axis and the values of the variable (number of grains per spike) on the horizontal axis as in the case of histogram. A dot is placed above the mid-point of each class and the height of a given dot corresponds to the frequency of the relevant class interval. By connecting the dots by a straight line, the frequency polygon is prepared. Maximum frequency is observed at 26 number of grains per spike and the class frequencies diminish towards both the sides. Figure 3.11 is the frequency polygon for the data presented in table 3.8.

How can a frequency polygon be turned into a histogram ? A frequency polygon is simply a line graph that connects the mdi-points of the all the bars in a histogram. We may draw a histogram of the data given in table 3.8 and then join with a straight line the mid-points of the upper horizontal side of each rectangle with the adjacent ones. The frequency so formed is called the frequency polygon. Figure 3.12 shows the frequency polygon. One may notice that figure 3.11 is superimposed on the histogram (figure 3.9). This figure (3.12) shows the relationship between the two graphic forms.

Figure 3.11: Frequency polygon representing the distribution of number of grains per spike as given in table 3.8

Frequency curve

The frequency polygon or histogram will approach more and more the form of a smooth curve. Such a curve is obtained in normal distribution of individuals in a large sampleor in a population. This sample frequency curve serves as the population frequency curve in almost all statistical analyses. In a majority of the biological characters, the frequency distributions approximate to a symmetrical bell-shaped curve known as the normal curve. The frequency curve is drawn freehand to eliminate as far as possible, the accidental variations that might be present in the biological, agricultural and other data. The total area under the curve should be equal to the area under the original histogram or polygon (Figure 3.13 and table 3.8).

Figure 3.12: Frequency polygon and histogram representing the distribution of number of grains per spike as given in table 3.8.

Figure 3.13 : Frequency curve representing the distribution of number of grains per spike as given in table 3.8

Cumulative frequency curve or ogive

In the case of this graphic representation of data, it is desirable to determine the number of observations that fall above or below a certain value rather than within a given interval. In such cases, the regular frequency distribution may be converted to a cumulative frequency distribution. A graph of cumulative frequency distribution is called the ogive (pronounced "oh-jive"). There are two methods of constructing ogive, namely:

(i) The "less than" method

(ii) The "more than" method

In the "less than" method, we start with the upper limit of the classes and go on adding the frequencies. However, in the case of "more than" method, we start with the lower limit of classes. The first method gives a rising curve, whereas the second method shows a declining curve (figure 3.14 and table 3.11).

MEASURES OF DISPERSION, SKEWNESS AND KURTOSIS

Introduction

In the preceding chapter the various measures of central tendency gives an estimate of the central value around which individual observations are distributed. A measure of central tendency tells something about the general level of magnitudes of the distribution, but it fails to give its complete description. Thus, it may be such that in several series, the average may be the same but the variables may highly differ in magnitudes and, therefore, the central tendency calculated from such variables may not be the true representative. According to Prof. Neiswanger "Two distributions of statistical data may be symmetrical and have common means, medians and modes and identical frequencies in the modal class. Yet with these points in common they may differ widely in their values about the measures of central tendency". Thus, to know the extent of spread about these averages or the variations of items, we have to resort to one such measure called dispersion. The average, discovers the representative value of a series, and the dispersion finds how individual values are dispersed away from the average.

The measures of central tendency and dispersion do not indicate whether a distribution is symmetrical or not. Skewness is another measure that refers to the amount of symmetry or asymmetry of a distribution. This statistical measure describes the shape of the distribution. In symmetrical distribution, the observations show a perfect balance on either side of the central tendency. However, in asymmetrical distribution, the balance is thrown to one side. But the measure of kurtosis describes the peakedness of the distribution.

Dispersion

The different measures of the central tendency, gives us an idea of the central value, whereas measures of dispersion are the measures of spread about an average. The central value may be the same in two or more distributions, but may differ in respect to dispersion. The measures of dispersion help us in studying the important characteristics of eh distribution. It measures the extent to which there are differences between individual observations and the central value. Dispersion may be defined as the measure of variation of the variable. Spiegel defined dispersion as, "the degree to which numerical data tend to spread about an average value". But according to Brooks and Dick, "dispersion or spread is the degree of the scatter or variation of the variable about a central value". In the words of L.R. Connor, "Dispersion is a measure of the extent to which the individual items vary". According to John I. Griffin, "A measure of variation or dispersion describes the degree of scatter shown by observations and is usually measured as an average deviation about some central value". It is clear from these definitions that dispersion is to the degree of variation of the variable but not the direction.

Dispersion is important not only as merely supplementary to the average, but because of the scatter distribution. According to Spurr and Bonini, "In matters of health, variations in body temperature, pulse beat and blood pressure are basic guides to diagnosis. Prescribed treatment is designed to control their variation". In industrial production, efficient operation requires control of quality variation, the causes of which are sought through inspection and quality control programmes.

The meaning of dispersion will be clear from the three distributions shown in the following page.

The arithmetic mean in all the three distributions is 16. It indicates that these distributions are alike in nature. However, these distributions differ from one another. In a all the values are the same as the arithmetic mean; there is no dispersion. In B, only one value is represented by the arithmetic mean and the other values show a slight deviation. In C, values deviate very widely from one another. Therefore, the dispersion in C is greater as compared to the B distribution. The amount of dispersion maybe small when the values, though different, are close. Figure 5.1 shows that the three curves have equal mean but are different in dispersion (5.1a). The curve C shows more variation or dispersion than the curve B. however, it was observed in (5.1b) that the mean values are different with the same dispersion. Some desirable properties of good measure of dispersion are given in the following page:

Distributions

A	B	C
16	14	3
16	15	8
16	16	17
16	17	24
16	18	28
80	80	80
16	16	16

Figure 5.1: Figure shows same mean but different dispersions (5.1a) and same dispersion but different mean value (5.1b)

(i) It should be rigidly defined.

(ii) It should be easy to understand and calculate.

(iii) It should be based on all the observations of the series.

(iv) It should have sampling stability.

(v) It should not be affected much by extreme values.

Methods of measuring dispersion

For ascertaining the deviations from the central value there are certain specific measures of dispersion. The important measures of dispersion are.

(i) Range

(ii) Mean deviation

(iii) Variance and the standard deviation

The first measure of dispersion is a positional measure whereas the other two methods are calculated measures of dispersion. However, the most popular method of studying dispersion is the standard deviation method. These measures are absolute measures of dispersion.

Absolute and relative measures of dispersion: Measures of dispersion may be either absolute or relative. The absolute measures of dispersion can be compared with another, only if the two belong to the same units of the variables in which the original data are given. For example, they are expressed in cms., kgs., tones etc. depending upon the unit of measurement of the original data. The absolute measures of dispersion do not help us in cases where the variables are expressed indifferent units. For instance, when we measure the height and the weight of the students. However, relative measure of dispersion is called coefficient of dispersion. The relative measure of dispersion is the ratio of the absolute measure of dispersion to the mean and is expressed as a percentage.

$$\text{Relative measure of the dispersion} = \frac{\text{Absolute measure of dispersion}}{\text{Mean}} \times 100$$

The relative measures of dispersion are.

$$\text{(i) Coefficient of range} = \frac{L - S}{L + S}$$

$$\text{(ii) Coefficient of the mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}}$$

$$\text{(iii) Coefficient of the standard deviation} = \frac{\text{Standard deviation}}{\text{Mean}}$$

Range

Dispersion may be measured in terms of the difference between two values of a series. The range is the simplest method of studying dispersion. It is defined as the difference between the highest and the lowest values of the variable in a distribution. For example, the number of seeds per plant in cowpea are: 70, 94, 110, 126, 130, 145, 160 and 175; the values of the range would be 175-50 = 105. This figure indicated the variability in the number of seeds per plant. The range gives a comprehensive values for the data in that it includes the limits within which all the items occurred. Symbolically;

Range (R) = $L – S$

Where L = largest value

S = smallest value

Characteristics of range

The range is very easy to understand, but its usefulness as a measure of dispersion is limited. However, the range considers only the highest and the lowest values of a distribution and it does not take account of many other intermediate observations. Therefore, it is heavily influenced by extreme values and it ignores the nature of variation of all the other observations. The range is likely to change drastically from sample to the next in a given population. In case of grouped frequency distribution, the range is calculated as the difference between the upper limit of the highest class and the lower limit of the smallest class. The open-end distributions have no range because the highest and the lowest values do not exist in open-end classes.

Coefficient of range

The coefficient of the range or the relative measure of the range is obtained by applying the following formula.

$$\text{Coefficient of range} = \frac{L - S}{L + S}$$

When the data are grouped into frequency distribution, the range can be calculated by any one of the following methods.

(i) Calculate the difference by subtracting the lowest limit of the lower class from the upper limit of the highest class.

(ii) Calculate the difference by subtracting the mid-point of the lowest class from the mid-point of the highest class.

Example: Data on number of secondary branches in a pulse are given below.
Find out the range and its coefficient.

No. of secondary branches = 8, 10, 15, 18, 17, 26, 22, 20

Solution: The highest value is 26 value the lowest value is 8.

Range = $L - S$ = 26-8 = 18

$$\text{Coefficient of range} = \frac{L - S}{L + S} = \frac{26-8}{26+8} = \frac{8}{34} = 0.53$$

Example 2: Calculate the range and coefficient of range from the following data recorded on number of clusters per plant in blackgram.

No. of clusters	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of plants	6	10	12	15	11	7	4

Solution: The largest value is the upper limit of the last calss and the smallest value is the lower limit of first class.

$L = 80$; $S = 10$

Range = $L - S = 80 - 10 = 70$

Coefficient of range: It can be calculated by both the methods.

Method I: Calculation based on upper limit and lower limit of two classes.

$$\text{Coefficient of range} = \frac{L - S}{L + S} = \frac{80-10}{80+10} = \frac{70}{90} = 0.78$$

Method II: Calculation based on mid-point of last and the first class.

$L = 75$; $S = 15$

$$\text{Coefficient of range} = \frac{L - S}{L + S} = \frac{75-15}{75+15} = \frac{60}{90} = 0.67$$

Merits and demerits of range

Merits

(i) IN all the measures of dispersion, the range is the easiest to calculate and simplest to understand.

(ii) It gives a rough but quick answer.

Demerits

(i) The range is not based on all observations.

(ii) It changes from one sample to the next in a population.

(iii) It cannot be calculated in open-end distribution.

(iv) It is very much affected by sampling fluctuations.

Mean deviation or average deviation

The range is a positional average which is not based on all observations. Therefore, it does not show any scatter of the observations from an average. However, the mean deviation is a measure of dispersion based on all items ina distribution.

The mean deviation is also known as the average deviation. It is calculated either from the mean or the median. Arithmetic mean is more commonly used in calculating the value of the mean deviation, therefore, it is more frequently called the mean deviation. The advantage in calculating the mean deviation from the median is that the sum of the deviations of items from the median is less than the sum of the deviations from the mean. The value of the mean deviation from the median is always less than the value calculated from the mean.

In computing the mean deviation, the algebraic signs (+ or -) of the deviations are ignored and all the deviations are taken as positive. To indicate the absolute deviations, sign of deviation "*d*" is sandwitched between two vertical lines called "modulus", i.e. /d/.

The mean deviation can be calculated from the following formula.

Symbolically;

Equ

Where dx = deviations from mean

n = number of observatiosn

x = observations

x = sample mean

equ

where *dM* = deviations from emdian

Coefficient of the mean deviation

The mean deviations would be an absolute measure of dispersion as it is expressed in the same units of original data. The coefficient of the mean deviations is obtained by dividingthe mean deviation by the mean or the median. For example.

(ii) Coefficient of the mean deviation $= \dfrac{\text{Mean deviation}}{\text{Mean}}$

(iii) Coefficient of the mean deviation $= \dfrac{\text{Mean deviation}}{\text{Median}}$

Calculation of the mean deviation

a) Calculation of the mean deviation in a series of individual observations: The following steps will explain the procedure to find out the mean deviation.

(i) Calculate themean and take the deviations of items from the mean ignoring + and – signs and denote these deviations by /d/.

(ii) Compute the total of these deviations, i.e. Σ/d/

(iii) Mean deviation is calculated by dividing thesum of deviations by the number of items. Symbolically;

Equ

Example 3: Calculate the mean deviation and its coefficient from the following data:
Variable = 30, 40, 70, 30, 60, 50, 10

Solution: Arrange the data in an ascending order

Variable x	***d* = (x-x)**	**/dx/**
10	-30	30
20	-20	20
30	-10	10
40	0	0
50	10	10
60	20	20
70	30	30
Σx = 280		Σ/dx/ = 120

Equ

Mean deviation = 17.14

$$\text{Coefficient of the mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}} = \frac{17.14}{10} = 0.43$$

Example 4: Data recorded on length of carrots (cms). Calculate the mean deviation and its coefficient.
Length (cms) = 9.2, 9.6, 10.0, 11.0, 12.0, 9.8, 10.2, 9.7, 12.7, 10.6

Solution: Arrange the data in an ascending order

X	d=			

/

/

/

Calculation of the mean deviation in a discrete series: the following formula should be used while calculating the mean deviation in a discrete series.

Equ

Steps

(i) Calculate the mean of the series.

(ii) Take the deviations of the items from the mean ignoring signs /dx/.

(iii) Multiply the deviations with the respective frequencies Σf /dx/

(iv) Divide the total by the number of observations. This will be the value of the mean deviation.

Example 5: Calculate themean deviation from the following data:

x	10	11	12	13	14	15	16
y	3	8	12	14	7	6	4

Solution

Equ

/

/

/

Calculation of mean deviation in a continuous series : In continuous series, the mean deviation from mean can be calculated in the same way as was done in the discrete series. The only difference is that the mid-points of various classes should be obtained first and then the deviations of these points is taken from the mean. Symbolically;

Mean deviation from the mean = ----------------------

Example 6: The following frequency distribution gives the number of chillies per plant. Calculate the mean deviation from the mean for the number of chillies per plant

No. of chillies per plant	10-16	17-23	24-30	31-37	38-44	45-51
No. of plants	8	10	23	29	18	12

Table is given in the following page

Solution

Equ

Merits and demerits of the mean deviation

Merits

(i) It is readily understood and is easy to calculate.

(ii) It is based on all the observations.

(iii) It is less affected by extreme items.

(iv) It is a better measure for comparison.

(v) It is flexible because it can be calculated from any measure of central tendency.

Demerits

(i) It is not amenable to further mathematical treatment because it ignores the algebraic signs of the deviations.

(ii) It cannot be calculated in open-end classes.

(iii) This method may not give accurate results when the mean deviation is taken from the median which is not a satisfactory measure when the degree of variability in the series is very high.

Variance and standard deviation

Variance

Variance is also called Mean Square Deviation. This term was first introduced by R.A. Fisher in 1913. The term “Variance” is used to describe the square of the standard deviation. This is an important measure in the quantitative analysis of variance helps us in isolating the effects of various factors. The variance is very important statistic because of some mathematical properties that can be used in developing statistical theories. To calculate the variance, the deviations of the variables from the mean are squared and then added. The sum of the squares of deviations is divided by the number of observations to get the variance of the sample. This measure has an advantage over the mean deviation because the sum of squares of deviations is always positive. The variance is defined as the mean of squares of deviations. Symbolically;

Equ

Check: The results of the above mentioned formula can be checked up with the help of the following formula:

Equ

In the case of frequency distribution, the variance can be calculated by the following formula.

Equ

Where Σf = total frequency

$(dx)^2$ = mean square deviation

When the size of the sample is small, the difference will be relatively high. However, in the samples of large size, the variance will show little difference. Therefore, while computing variance when the sample size is less than 30, one should divide $\Sigma(dx)^2$ by n-1 to obtain an estimate or variance. On the other hand, the variance is not included in the measures of absolute dispersion because the variance is not expressed in the same statistical unit in which original data are given as cms., kgs. Or grams etc. as it relates to squared deviations, the variance of the yield of the plant measured in grams, should have been expressed in squared grams. Therefore, the practical difficulty can be overcome by taking the square root of the variance value which is called standard deviation.

Standard deviation

Standard deviation is the most commonly used absolute measure of dispersion. The concept of standard deviation was first introduced by Karl Pearson in 1893. It is expressed in the same units of measurement as the observations, whereas, variance is expressed in squared units. The term "standard" is assigned to this measure of variation is probably because it is the most commonly used and is the most flexible in terms of variety of applications of all the measures of dispersion. It is clear that standard deviation is a measure of the spread in a set of observations. In this method, the drawback of ignoring the algebraic sign as in mean deviation is overcome by taking the square of deviations, thereby, making all the deviations positive. Standard deviation is defined as the square root of the variance. Thus, the sample standard deviation (s) would be.

Equ

As it measures the dispersion or variability of a distribution, the larger the standard deviation mean, the greater is the value of var ability. On the other hand, there will be homogeneity in a series when the standard deviation is small.

Calculation of standard deviation

a) Calculation of standard deviation in a series of individual observations : Standard deviation can be calculated by applying any of the following two methods:

Method I: This method can be applied only by taking deviations from the actual mean symbolically;

Equ

Steps

(i) Calculate the arithmetic mean of the observation (x)

(ii) Find the deviations of individual observations from themean and denote it by dx.

(iii) Square these deviations and total them as $\Sigma(dx)^2$

(iv) Divide the sum $\Sigma(dx)^2$ by the total number of observations and then extract the root. This is the value of the standard deviation.

Method II: The following formula can be used to calculate the standard deviation:

Equ

Steps

(i) Find the square of all the values of the individual observations and total them (Σx^2).

(ii) Find the square of the total of all the values and divide it by the number of observations $(\Sigma x^2)/n$.

(iii) Subtract $(\Sigma x^2)/n$ from Σx^2 and divide it by number of observations.

Equ

And then find ouz the square root of this

Equ

This will be the value of standard deviation.

Example 7: Calculate the standard deviation from the following data.

Variable (x) = 10, 13, 17, 22, 27, 30, 31, 32.

Solution : Method I

Equ

Method II

Equ

Note: Whatever may be the way of calculation, the answer will be the same.

b) Calculation of the standard deviation in a discrete series: In this case, standard deviation can be calculated after applying one of the two methods. In discrete series, only frequencies are taken into consideration. Symbolically;

equ

Example 8: Data on waxy endospermic plants were recorded in maize. Calculate the standard deviation from the following data:

Waxy endospermic plants	7	8	9	10	11	12
No. of plants	13	13	18	17	15	14

Solution

Equ

c) Calculation of the standard deviation in a continuous series : For calculation of the standard deviation in a continuous series, any of the methods can be applied which are used in discrete frequency distribution. However, in this case, mid-points are taken into consideration.

Equ

Example 9: Calculate the standard deviation from the data recorded on the number of pods per plant in mothmean.

No.of pods per plant	15-17	18-20	21-23	24-26	27-29	30-32	33-35	36-38	39-41	42-44
No. of plants	2	2	4	5	7	9	6	4	3	2

Solution: Method I.

Solution

Equ

Method II:

Step-deviation method: This method can also be used here. ; here "i" stands for the common factor or calss-interval. Symbolically;

Equ

Where i = class-interval

Equ

Coefficient of variation

The standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original data are collected. The standard deviation of the height of plants cannot be compared with the standard deviation of the weight of plants because they are expressed in different units, i.e. heights (cms.) and weight (gms.). Therefore, the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure of dispersion is known as the coefficient of variation.

Coefficient of variation is of great practical significance and is the best measure of comparing the variability or consistency of two or more samples. The smaller the coefficient of variation, the greater is its consistency. On the other hand, if the coefficient of variation is greater, the sample is said to be more variable or less homogeneous.

Symbolically;

Equ

Where *s* = standard deviation

x = mean

Karl Pearson suggests this measure of coefficient of variation as the most commonly used measure of relative variation. In all the biological problems, standard deviation is always used as the measure of dispersion and the arithmetic mean of the average. Therefore, the resulting percentage of this ratio is known as coefficient of variation.

Example 10: From the following data, compute the values of coefficient of variation. Compare the two series for the data consistency.

Equ

This shows that the data of series "A" are more variable and less consistent as compared to the data of series 'B' which are more consistent.

Example 11: From the following data recorded on the height of plants of varieties, G-65 and PS-16 of mungbean; find out which variety is more consistent.

Plant height (cms.)

Variety G-65	25	50	45	30	70	42	36	48	34	60
Variety PS-16	10	70	50	20	95	55	42	60	48	80

Solution

Equ

The mean of the variety PS-16 is higher than the variety, G-65. However, the C.V. is also higher for the variety PS-16 than G-65. it indicates that the variety, G-65 is more consistent or homogeneous than variety, PS-16 because the C.V. is less in variety, G-65.

Merits and demerits of the standard deviation

Merits

(i) It is rigidly defined

(ii) It is based on all the observations.

(iii) It is less affected by sampling fluctuations.

(iv) For comparing two or more samples or varieties or distributions, the coefficient of variation is useful as it is based on the standard deviation and arithmetic mean. These important points indicate that the standard deviation is the best measure of dispersion. It is the most extensively used measure in statistical analysis of the agricultural as well as the biological and the medical sciences. In normal distribution, the standard deviation helps us in finding the number of items that fall within the specific ranges, i.e.

Mean ± 1s covers 68.27% of the values

Mean ± 2s covers 95.45% of the values

Mean ± 1s covers 99.73% of the values

It is given in figure 5.2

Demerits

(i) It is difficult to calculate as compared to the other measures of dispersion.

(ii) It gives more weightage to extreme values and less to the values which are near to the mean.

Figure

PROBABILITY AND PROBABILITY DISTRIBUTIONS

INTRODUCTION

The physicist Helmholtz (1821-1894) once wrote that “all science is measurement”. This statement is essentially true if we also agree that the method of measurement may differ from one science to another. In genetics, much of what is measured concerns the ratios of different phenotypes and genotyes. These genetic ratios arise from probability relationships, the chance of segregation and assortment of genes into gametes and their chance combination into zygotes. Since these are chance events, exact predictions cannot be made for any particular event. Just as, when one tosses a coin, there is no guarantee that heads must always follow tails, there is no guarantee that any particular genetic event (such as a particular genotype) will occur when a number of different types of events are possible. All we can say in advance of any chance event in genetics is, that it has a certain probability of occurring.

Probability is a part of our everyday lives. We live in a world in which we are unable to forecast the future with complete certainty. Our need to cope with uncertainty leads us to the study and use of probability theory. Probability refers to the chance of happening or nor happening of an event. In Mendelian genetics, the F_2 generation segregates in a particular ratio. The number of individuals in each ratio results from chance segregation of genes during gamete formation and their chance combinations to form zygotes. Since these are the chance events, accurate predictions about the results cannot be made. But what we can say is that there is a certain probability of occurrence of a given genetic event.

The word chance indicates that there is an element of uncertainty about some of the common statements such as chances of one drug being better than the other; the probability that this year the wheat production will be quite high and probability that a mother will give birth to a boy in the first pregnancy. In all the above statements, the event is not certain to take place. This uncertainty is numerically expressed as probability. It measures the relative frequency of a particular event happening by chance.

The theory of probability has it sorigin in the games of chance related to gambling. It was in the first half of the sixteenth century that serious thought was given to the problem of probability by eminent mathematicians. An Italian mathematician, Jerome Cardon was the first man to write a book titled “Book on games of chance” in 1663. However, the classical treatment of probability dates back to the seventeenth century after the work of two mathematicians, Blaise Paseal (1623-1662) and Pierre de Fermat (1601-1665). Much of this theory developed out of attempts to solve problems related to games of chance such as rolling of dice. Jacob Bernoulli (1654-1705) also extended the application of the theory of probability much beyond the restricted field of games of chance. In the nineteenth century, Pierre Simon de Laplace (1749-1827) compiled the first general theory of probability. R.A. Fisher and Von Mises introduced the empirical approach to probability. The modern theory of probability was developed by Russian Mathematicians like Chebychev (1821-1894), A. Markov (1856-1922) and A.N. Kolmogorov (1933). The theory of probability has been used very extensively in all disciplines of sciences. Now it becomes difficult to discuss biostatistics without an understanding of the meaning of probability. A knowledge of probability theory makes it possible to interpret statistical results and we can express numerically the inevitable uncertainties in the resulting conclusions.

Probability

The concept of probability is basic to many principles of classical genetics, e.g. segregation independent assortment etc. However, the principles of population genetics are based on the consideration of laws of probability. Simply defined, probability is the likelihood of occurrence of an event. In daily life, each one of us is often faced with a decision about the likelihood that an event will occur or not. For example, before the commencement of examinations, each student is anxious about the likelihood of his doing well in the examinations. The likelihood of obtaining good marks in an examination may be influenced by a number of factors, such as, the result of an examination depends on the preparation by the student the type of question paper, the attitude of examiner etc. however, in contrast to this there are a number of events whose occurrence depends entirely on chance and not on any factor. For instance, when a coin is tossed, either the head or the tail turns up. Similarly, a new-born baby is either a male or a female. In such cases, the events occur purely because of chance.

Example 1

Sex determination in bisexual organism is a chance event. For a normal fertilization of an ovum, there are only two equally likely ways. The zygote may develop into a male or female depending on the type of gamete the male parent provided in animals, man, Drosophila etc. but in poultry it depends on the type of gamete produced by the heterogametic sex may or may not contain the X-chromosome and these are equally likely too. So the chance that a fertilized egg will develop into a particular sex is ½.

Figure

Probability may be defined as the chance or likelihood that an event will occur. The value of probability ranges between 0 to 1. If an event is certain to happen its probability would be 1 (p=1). On the other hand, if it is certain that the event would not take palace then the probability of its happening is 0(p=0). In agricultural and biological sciences, such occurrences, are very rare like in social sciences.

Statistically, the term probability can be explained in the following way. If an event can happen in 'a' ways, and fail to happen in 'b' ways, then the probability of its happening 'p' can be written as:

Equ

Similarly, the probability of the failure of the event to happen is denoted by 'q'. Therefore,

Equ

For example, if twins are bron once in 80 different pregnancies then p for birth of twins = 1/80 and the probability for single birth will be q=1-1/80=79/80. If probability of being Rh^- is 1/10 then of being Rh^+ will be 1-1/10=9/10.

Example 2

If a surgeon transplants a kidney in 400 cases and succeeds in 160 cases, calculate the probability of survival after operation.

Solution

Equ

Before discussion the theory of probability, let us have an understing of the following terms.

Random experiment or trial: Experiments can be repeated under the same conditions but the results (outcome) cannot be predicted in any repetition; any one of the possible ways is called **random experiment or trial.** Tossing of a coin is a noteworthy example of a random experiment. When you toss, it falls head up or tail up, but exact prediction is not possible in any toss. Similarly, the production of a commodity on different days or price of a commodity in different months may be taken as the outcome of a random experiment.

Event: Any possible outcome of random experiment is called an event. Performing an experiment is called trial and the outcome is termed as event. In simple terms, an event is the occurrence of something. The occurrence of a head or a tail is an event. Similarly, a new-born baby being a boy or a girl is also an event. The events that occur purely due to chance are grouped into two categories. These are: (i) mutually exclusive and (ii) independent events.

Mutually exclusive events: Events that are so related among themselves are said to be mutually exclusive, if the occurrence of one excludes the possibility of the occurrence of the other or in other words, two events are mutually exclusive if both cannot occur simultaneously. All simple events are mutually exclusive. For example, if a coin is tossed, either the head or the tail can occur but not both. Similarly, a new born can be either a boy or a girl. Therefore, in a trial, the occurrence of the head excludes that of the tail and the occurrence of a boy excludes that of a girl and vice-versa. Thus, the probability of getting a head (or a tail) is ½ and the probability that a new born will be a boy (or girl) will also be 1/2. If a coin is tossed a large number of times, one-half of the times it will turn up “head” and the remaining one-half times it will be “tail”. Thus, the probability of an event is the average frequency of that event. Conversely, the average frequency of an event is its probability.

Independent events: A set of events are said to be independent if the occurrence of any event does not affect the chance of the occurrence of any other event of the set. For instance, the sex of a calf to be born is independent of the sex of the calf in the previous calfing by the same dam. Likewise, when two coins (A and B) are tossed together or separately, the occurrence of the head or the tail in the case of the second coin (B). Therefore, they are known as independent events. This is clear from the table given.

Table 1: Probability of different combinations of the head (H) and the tail (T) obtained on tossing of two coins (A and B) simultaneously or separately.

Similarly, when three coins are tossed, there are eight possible combinations of H and T. Each of these combinations will have the probability of 1/8 = ½ x ½ x ½ (Table 2).

Table 2: Probability of different combinations of H and T obtained after tossing of three coins (A, B and C).

These combinations are summarized below according to the order of occurrence of H and T in them.

HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

Therefore, the probabilities of getting 3H, 2H, 1T, 1H, 2T and 3T on tossing three coins together will be 1/8, 3/8, 3/8 and 1/8, respectively.

Permutation and combination : Before explaining the theory of probability, it is necessary to have an understanding of permutation and combination because these terms are used in the calculation of probability. The word permutation refers to arrangement and the word combination refers to group. Permutation refers to different arrangement of objects in a set, where all elements are different and distinguishable.

These arrangements are without repetition of any individual object. If some objects are similar, the permutations will be affected. Permutations are designed to find out the total number of ways in which an event can be performed. For example, we have three varieties of what and can arrangement them in six ways as follows.
ABC, ACB, BAC, BCA, CAB, CBA

Now we are interested in the total number of different ways in which varieties can be arranged. This can also be shown through a tree diagram.

Figure

It may be noted that none of the three varieties has appeared more than once in an arrangement. Thus, we may define that the different ways in which a number of objects can be arrangedin a definite order are called *p*.

Let ${}^{n}p_{r}$ denote the number of permutations of 'n' different objects taken 'r' at a time. It is difficult to write down permutations. Therefore, a formula has been derived symbolically;

Equ

Combination: The number of observations which can be made from a group of things, irrespective of the order, is called combination. The combination differs from the permutation in that one combination such as ABC may be stated in the form of several permutations by simply rearranging the order in which the things occur in the series such as ABC, ACB, BAC, CAB, CBA. All of thee are considered to be one combination but they represent six permutations. Any combination can give rise to a set of permutations, the number of permutations is always greater than the number of possible combinations in any given situation. It will be clear from the given table.

Table 3: Relationship between combination and permutation.

Number of objects (*n*)	Taken at a time (*r*)	Combination (${}^{N}C_{r}$)	Permutation (${}^{N}P_{r}$)
AB	Two	AB	AB, BA
ABC	Two	AB, AC, BC AC, BC, CB	AB, BA, CA
ABC	Three	ABC	ABC, ACB, BCA, BAC, CAB, CBA

Thus, it is clear that there is only one combination for a set

$^{n}C_{n}$

a combination of ‘n’ different objects taken ‘r’ at a timeis denoted by $^{n}C_{r}$. It is a selection of only ‘r’ objects out of ‘n’ objects and we do not consider the order of arrangement. Symbolically;

equ

Theorems of probability

It is very important to have a clear concept of probability as it provides the basis for all the tests of significance. It is estimated usually on the basis of the following two basic rules of chance.

(i) Addition rule

(ii) Multiplication rule

The first has to do with the probability, that one or the other two events will occur, and the second with the probability that they both will occur.

Addition rule

Addition rule for mutually exclusive events

Two events, A and B are said to be mutually exclusive, if the occurrence of one event precludes the occurrence of the other event. For example, the tossing of a coin will result in either the head or the tail, birth of a male excludes birth of a female, Rh birth excludes birth of a Rh baby and so on. Therefore, these two possible outcomes are mutually exclusive.

When two events are mutually exclusive, the probability of the occurrence of either A or B is the sum of their individual probabilities. This addition rule may be stated as follows.

P (A or B) = p(A) + p(B)

The same rule can be extended to three or ore mutually exclusive events.

P(A or B or C) = p(A) + p(B) + p(C)

Example 3

If you roll a single die, the probability of getting either a 1 or 2 would be computed as follows.

Solution

P (A or B) = p(A) + p(B)

P (1 or 2) = p(1) + p(2)

= 1/6 + 1/6

= 2/6 = 1/3

Example 4

From a pack of 52 cards, one card is drawn at random. What is the probability that it is either a king or a queen?

Solution: The events are mutually exclusive. There are 4 kings and 4 queens in a pack of 52 cards.

The probability of drawing a king is 4/52 and similarly the probability of drawing a queen is 4/52.

The probability that a card is either a king or a queen

P (A or B) = p(A) + p(B)

4/52 + 4/52 = 8/52 = 2/13

Addition rule when events are not mutually exclusive

It two events, A and B are not mutually exclusive; it is possible for both events to occur. When the events are not mutually exclusive, the addition rule must be modified.

P (A or B) = p(A) + p(B) – p(A and B)

For example, what is the probability of drawing either an ace or a heart from a pack of 52 cards? The events, ace and heart can occur together because we can draw the ace of hearts. Therefore, ace and heart are not mutually exclusive events.

Example 5

What is the probability of getting either an ace or a spade from a pack of 52 cards?

Solution

P (ace or spade) = p(ace) + p(spade) – p(ace and spade)

Equ

Example 6: A card is drawn at random from a well shuffled pack of cards. What is the probability that the card is a spade or a queen?

Solution: A denotes a spade card

B denotes a queen

Equ

Multiplication rule

Multiplication rule for independent events

This rule states that if the two events, A and B are independent, the probability of joint occurrence is given by the product of their separate probabilities. Symbolically;

P (A and B) = p(A) x p(B)

The same can be extended to three or more events.

P (A, B and C) = p(A) x p(B) x p(C)

Example 7

What is the probability of the heads on two or three successive tosses?

Solution

(i) p(A) = The probability of the head in first toss = ½ or 0.5

(ii) p(B) = The probability of the head in second toss = ½ or 0.5

combined probability

P (A and B) = p(A) x p(B)

= ½ x ½ = ¼ = 0.25

(iii) p(A) = ½ or 0.5

p(B) = ½ or 0.5

p(C) = ½ or 0.5

Combined probability

P (A, B and C) = p(A) x p(B) x p(C)

= ½ x ½ x ½ = 0.5 x 0.5 x 0.5 = 0.125

Multiplication rule for dependent events

This rule states that if two events, A and B are dependent, the probability of occurrence of one event is dependent on the occurrence of the other event.

P (A and B) = *p*(A) x *p*(A/B)

Or

P (A, B and C) = *p*(A) x *p*(A/B) x *p*(C/AB)

Example 8

A bag contains 7 red and 3 black balls. Two balls are drawn at random one after the other without replacement. What will be the probability that both the balls drawn are black?

Solution

Probability of drawing a black ball

Equ

Probability applications

The laws of probability maybe applied to any subject which involves chance or random happenings. The study of probability is closely related to genetics because it involves a statistical analysis of the ratios of various characteristics among the offspring of parents of a known phenotype. Multiplication law of probability is useful to find out the results of more than one generation also, probability applications in genetics will be more clear if we analyse the results of first, second and third generations of human families. We can make the probabilities of events event more explicit using a probability tree of all three generations. Figure 6.1 shows the possible outcome of the first generation. Figure 6.2 will give four possible outcome of the second generation. The complete probability tree is shown in figure 6.2 will give four possible outcome of the second generation. The complete probability tree is shown in figure 6.3 will give four possible outcome of the second generation. The complete probability tree is shown in figure 6.3. However, it is clear that no event is affected by the events preceding or following it.

Figure 6.1: Probability tree of first generation.

Analysis of results of Figure 6.1

(i) Probability of getting a male = ½ or 0.5

(ii) Probability of getting a female = ½ or 0.5

Figure 6.2: Probability tree of the first generation and the second generation

Analysis of results of Figure 6.2

(i) Probability of getting two males = ¼ = 25%

(ii) Probability of getting two females = ¼ = 25%

(iii) Probability of getting one of either sex = ¼ + ¼ = 1/2 = 50%

Figure 6.3: A complete probability tree

Table 4: Possible outcome and their probabilities.

First generation		Second generation		Third generation	
Possible outcome	**Probabilities**	**Possible outcome**	**Probabilities**	**Possible outcome**	**Probabilities**
M	0.5	MM	0.25	MMM	0.125
F	0.5	MF	0.25	MMF	0.125
	1.0	FM	0.25	MFM	0.125
		FF	0.25	MFF	0.125
			1.00	FMM	0.125
				FMF	0.125
				FFM	0.125
				FFF	0.125
					1.000

Complete analysis of results

(i) Probability of getting 3 males

= ½ x ½ x ½ = 1/8 or 0.125 or 12.5%

(ii) Probability of getting 3 females

= ½ x ½ x ½ = 1/8 or 0.125 or 12.5%

(iii) Probability of getting 2 males 1 female

= 1/8 + 1/8 + 1/8 = 3/8 or 0.375 or 37.5%

(iv) Probability of getting 2 females 1 male

= 1/8 + 1/8 + 1/8 = 3/8 or 0.375 or 37.5%

Total probability = = 1/8 + 1/8 + 3/8 + 3/8 = 8/8 = 1 or 100%

(v) Probability of getting MFM a the end of third generation
P(MFM) = *p*(M) x *p*(F) x *p*(M) = 0.5 x 0.5 x 0.5 = 0.125 or 12.5%

- We can understand the law of probability more clearly when it may be applied to genetic problems. In human beings, albinism is a recessive genetic trait, the birth of an affected child of normal looking parents indicates that the parents are heterozygous (carriers) for that trait. Parents are normal and healthy because of recessive nature of gene. Since, the trait is monogenic recessive they can expect affected children and normal children in the ratio of 1:3. The probability of having an affected child is thus ¼ at each birth and so also in all future pregnancies. In other words, the probability for normal child at each birth is ¾. If, parents decide to have three more children and want to know the chance that all the three will be albinos. The answer would be: ¼ x ¼ x ¼ = 1/64; not a very great chance. One the other hand, if they want to know the probability of all, three being normal the answer would be: ¾ x ¾ x ¾ = 27.64. One has to bear in mind that the occurrence of an albino in the first pregnancy will not influence the heredity of future children for this trait. By applying the laws of probability one can know that he/she may be carrying some serious hereditary defect.
- One may often note the working of the laws of probability for predicting the ratio of boys and girls born in a family. The probability for a boy at any birth is ½ and for a girl also ½ since, the human male produces an equal number of X and Y sperms. What is the probability that the first two children born in a family will be males? For this, one has to determine the product of separate probabilities at each conception; ½ x ½ = ¼. What is the probability that the third child in the family in which the first two are males will also be a male? The probability of third child is also ½ because the sex of any child is independent of the sex of the other children.
- The principles of probability can be applied in solving the Mendel's problems of heredity. Probability can be used to obtain the phenotypic ratios to be expected from dihybrid, trihybrid or polyhybrid crosses etc. For example, in pea plant when homozygous tall plant is crossed with homozygous dwarf plant, the expected ratio is three tall and one dwarf in the second generation. Likewise when a cross is made in

between a variety of pea which bore red flowers and another variety with white flowers, the appearances in the second generation will be three red flowered plants and one white flowered plant. When both the crosses are conducted simultaneously the number of tall red would be.

¾ tall x ¾ red = 9/16 tall red

Similarly to find the number of tall white to be expected; multiply the chance of each trait appearing independently.

¾ tall x /4 white = 3/16 tall white

Complete dihybrid results.

Equ

However, the results of trihybrid ratio can also be obtained if we multiply the phenotypic ratio of third cross with the above results.

The same method can also be used in the case of the interaction of genes. In epistasis when we use the duplicate recessive epistasis involving two separate genes which can cause deaf mutism in man, the ratio is as follows.

Equ

- Practical applications of the laws of probability are made by human geneticists and in some instances by animal breeders in analyzing pedigrees. The first step in such an analysis is to determine whether the trait in question is behaving as a dominant or recessive. For instance, deafness may be have as dominant in some families and recessive in others. Recessive genes are difficult to follow because they may remain hidden by their dominant alleles generation after generation. Carriers in the population usually cannot be identified until an affected child is born. Recessives are expressed more frequently in families in which the father and the mother are more closely related than parents in the general population.

However, in the absence of data to indicate which individuals are carriers, the geneticists may resort to probability as the best available tool for determining the likelihood of expression of given recessive gene in a certain family. Probability is then based on the family history which may be recorded in a pedigree chart.

The use of probability in human pedigree analysis is illustrated in Figure 6.4. The trait, adherent ear lobes dependent on a recessive gene appeared only once in the known history of family as indicated by the single darkened circle in the figure. The woman (11-3) in whom the trait is expressed must be homozygous (aa) for the recessive gene. From the parent cross (Aa x Aa), the probability for the occurrence of Aa in any child with free ear lobes is 2/3 and the probability for occurrence of AA is 1/3. the probability that the parents with genotypes Aa having an aa child (Aa x Aa = 1AA, 2Aa, 1aa) will be ¼. Therefore, probability for an expression of the trait in the child of the individuals will be.

Equ

Figure 6.4: Pedigree chart to illustrate the probability in pedigree analysis. In this family group, a trait (adherent ear lobes), dependent on a recessive gene, has appeared in one individual (identified by the darkened circle). Three basic steps maybe followed in all pedigree analysis: (1) What is the chance that oneparent is a carrier of the gene in question ? (2) What is the chance that the other parent is a carrier for the same gene? (3) What is the chance that a child of these two parents (genotypes) could express the trait involved? The product of the separate probabilities is the chance that a particular future child will express the trait.

(Sources: Principles of Genetics by Gardner, pp. 34)

Probability distributions

We have been dealing with the observed frequency distribution in the chapter on diagrammatic and graphic representation of data; measures of central tendency and also measures of dispersion skewness and kurtosis. Such distributions are based on observations and experimentation where frequencies are arrived at by collection, tabulationand classification of data. For example, frequency distribution of the number of pods per plant, plant height and the number of tillers etc. However, sometimes it is possible to devise mathematically the frequency distribution of certain population. Such distributions are called "*Theoretical distributions*" or "*Probability distributions*". Probability distributions are not obtained by actual observation or experiments but are mathematically deduced on certain assumptions. These distributions are based on laws of probability and are very useful for statistical analysis.

There are three main types of probability distributions which are widely used in different sciences. These distributions are discrete as well as continuous. T hey are as follows.

(i) Binomial distribution

Discrete probability distributions

(ii) Poisson distribution

Continuous probability distribution

(iii) Normal distribution

Binomial distribution

The binomial distribution is one of the mot widely used probability distributions of random discrete variable. This distribution is also known as *Bernoulli distribution* since it was introduced by a Swiss Mathematician, J.Bernoulli (1654-1705). The distribution is derived from a process known as a Bernoulli trial or as a Bernoulli process. Bernoulli process is one where an experiment can result in only one or two mutually exclusive outcome such as success or failure, dead or alive and male or female; the trial is called a *Bernoulli trial*. Therefore, the binomial distribution describes the distribution of probabilities where there are only two possible outcome for each trait of an experiment, for example, if a coin is tossed once, there are two possible ways of outcome, the head or the tail. The probability of obtaining a head is $p = ½$ and tail is $q = ½$.

Thus $(p + q) = 1$ and binomial is $(p + q)^n$. similarly, if two coins are tossed simultaneously, there will be four possible outcome.

<table>
<tr><th>First coin</th><th>Second coin</th><th>Probability</th></tr>
<tr><td>H</td><td>H</td><td>pp=p²</td></tr>
<tr><td>H</td><td>T</td><td rowspan="2">pq
= 2pq
qp</td></tr>
<tr><td>T</td><td>H</td></tr>
<tr><td>T</td><td>T</td><td>qq =q²</td></tr>
</table>

Binomial expansion is $(p + q)^2 = p^2 + 2pq + q^2$

Suppose we toss three coins and record the number of the heads, how often should be expect to get 3 heads, 2 heads, 1 head and no head ?

The possible outcome will be.

First coin	Second coin	Third coin	Probability
H	H	H	ppp= p^3
H	H	T	p^2q $\}$ $3p^2q$
H	T	H	p^2q
T	H	H	p^2q
H	T	T	pq^2 $\}$ $3pq^2$
T	H	T	pq^2
T	T	H	pq^2
T	T	T	qqq=q^3

Binomial expansion is $(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$

The expansion $(a + b)^n$ or $(p + q)^2$ is known as binomial expansion. Here, p stands for the probability of the head while q stands for the probability of the tail and n stands for number of coins. According to the different values of n, following expansion of $(p + q)^n$ can be received.

$(p+q)^1$	$p + q$
$(p+q)^2$	$p^2 + 2pq + q^2$
$(p+q)^3$	$p^3 + 3p^2q + 3pq^2 + q^3$
$(p+q)^4$	$p^4 + 4p^3q + 6p^2q^2 + 4qp^3 + q^4$
$(p+q)^5$	$p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$
$(p+q)^6$	$p^6 + 6p^5q + 15p^4q^2 + 20p^3q^3 + 15p^2q^4 + 6pq^5 + q^6$

The numerical coefficients in this expansion can beobtained without actual multiplication. They would be respectively, ${}^nC_0$, ${}^nC_1$, ${}^nC_2$, ${}^nC_3$… ${}^nC_n$. They can also be obtained from Pascal's triangle as given below.

Pascal's Triangle

This triangle was first constructed by Pascal and is actually a set of figures written in a trianglular shape. It is relatively a simple method for determining the coefficients of binomial expansion. Note that each number in the triangle is derived by adding the numbers in the row above which lie on either side of it. For example, in row fourth, 6 is obtained by adding 3 and 3 in the third row; in row seventh 35 is obtained by adding 20 and 15, and 252 is obtained by adding 126 and 126. In this manner, one may write the set of numbers in as many rows as one wishes. The set of numbers thus obtained form a triangular shape and constitute the Pascal's triangle. These set of numbers in this triangle have several properties.

(i) The sum of numbers in a row is twice the sum of those in the preceding row but only half of the sum of those in the following row.

(ii) The set of numbers on one side of the midpoint are mirror images of those on the other side. For example, in the sixth rowof the triangle numbers on the left of the midpoint are 1, 6, 15, 20 while those of the right side of the midpoint are 20, 15, 6, 1.

(iii) The second number of a row is the same as the index of binomial. For example, the second number of the sixth row of the triangle is 6. the set of numbers in this row (1, 6, 15, 20, 15, 6, 1), therefore, are coefficients for the terms of binomial $(p + q)^6$.

(iv) The sequence of numbers in a row is the same as that of terms of the binomial expanded in the manner explained earlier.

PASCAL TRIANGLE

Index of Binomial	Coefficients in the binomial expansion $(p + q)^2$	Total
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256
9	1 9 36 84 126 126 84 36 9 1	512
10	1 10 45 120 210 252 210 120 45 10 1	1024

Assumptions of binomial distribution

A sequence of Bernoulli trials forms a Bernoulli process. We can describe this process which is al follows.

(i) Each trial has only two possible outcome “success” or “failure”.

(ii) The probability of success denoted by ‘*p*’ remains constant from trial to trial. Similarly the probability of failure denoted by ‘*q*’ should also be constant for all trials.

All trials must be independent of each other. The outcome of any particular trial is not affected by the outcome of any other trial.

Formula for the determination of probability in a binomial expansion

In 'n' trials, the total number of possible ways of obtaining 'r' success and failure (n-r) is.

Probability (r successes out of n trials)

Equ

Where *p* is the probability of success with a single trial. The symbol ! means factorial which s computed as follows.

5! is 5 x 4 x 3 x 2 x 1 = 120

4! is 4 x 3 x 2 x 1 = 24

3! is 3 x 2 x 1 = 6

2! is 2 x 1 = 2

1! is 1=1

0! is 1 because, mathematically it is defined as equal to 1.

Example 9

What is the probability of getting 0, 1, 2, 3 and 4 heads when a coin is tossed four times ?

Solution

(a) No. of trials, $n = 4$

Probability of success, $r=0$

Equ

Poisson distribution

Poisson distribution is a discrete probability distribution and is used very widely in physical and social sciences. It was derived by a Frenchman, S.D. Poisson (1781-1840) in 1837. Binomial probability distribution gives an idea of the probability of getting a specific number of occurrences in a given number of trials. However, this distribution studies the probabilities of rare events which are common in science. For example, the emission of particles from a radioactive source, persons dying due to a rare disease, the number of defective articles produced by a high quality machine and the number of mistakes committed by a good typist etc are rare events, in the sense that the probability of their happening is very rare. Poisson distribution is made use of in such cases where '*p*' is very small. Therefore, this is the limiting distribution of binomial distribution when the number of trials '*n*' is very large and the probability, '*p*' of success for every trial is very small; the '*np*' is the fixed number which is called '*Poisson distribution*". Poisson distribution may be obtained as a limiting case of binomial probability distribution under the following conditions.

(i) The poisson distribution is a discrete distribution.

(ii) It has a single parameter which is the mean of the distribution and is denoted by *m*

(iii) *Np*= *m* which remains constant

(iv) Formula used is as follows:

Equ

Where p = probability

r = 0, 1, 2, 3…. n successes

e = 2.7183 (constant)

Table 5: Value of *m* with corresponding value of e^{-m} (e = 2.7183)

m	e^{-m}	m	e^{-m}	m	e^{-m}

Example 10: An average of 5 cars arrive at a tollbooth every minute. Assuming this to be a poisson distribution, what is the probability that exactly 0, 1, 2, 3, and 4 cars arrive in a one minute period?

Equ

Normal Distribution

The binomial and poisson distributions are useful probability distributions for discrete variables. The most important distribution dealing with a continuous variables is the *normal distribution.* Normal distribution is also called normal probability distribution which is extremely useful in the analysis of the agricultural and the biological data. The normal distribution was first discovered by De Moivre (1667-1754) in 1733. Laplace and Gauss also derived this distribution quite independently of De Moivre. The normal distribution is a limiting form of the binomial probability distribution.

Normal distribution plays an important role in making inferences regarding the value of the population mean from the sample mean. It also gives an idea about the distribution of items in a sample of the population. This technique of drawing inferences about the population from the sample is one of the major tools in data analysis. In most of the biological analyses, values are often distributed in accordance with the normal distribution. In large sample size, analysis of data will show a peak of the curve with symmetrical distributed items on both the sides of the peak. Such a curve with important statistical properties is called the normal distribution curve. This type of curve denotes a normally distributed population.

Importance of the normal distribution

The importance of normal distribution would be clear from the following points.

(i) In most of the biological analyses, values are often distributed in accordance with the normal distribution. As the sample size increases, the distribution of mean of a random sample approaches a normal distribution. On the other hand, it the samples of the large size are drawn from a population which is not normally distributed, the successive sample means will form normal distribution. A small sample shows an irregular distribution due to obvious error in the sample. However, large sample size will reduce the error and make the cure smooth (Figure 6.5a).

(ii) When the size of the samples becomes large, the normal distribution serves as a good approximation of discrete distribution such as Binomial and Poisson.

Properties of the normal distribution

The normal distribution curve has the following properties.

(i) The normal curve is “bell-shaped” and is symmetrical in appearance. It has a single peak, thus it is unimodal.

(ii) The mean of a normally distributed population lies at the centre of its normal curve. The height of the normal curve is at its maximum at the mean. This ordinate divides the curve in two equal parts.

(iii) The mean, median and mode are all equal in normal distribution.

(iv) The height of the curve declines on either side of the peak which occurs at the mean. The area on both the sides of the peak is equal to each other.

(v) The cure is asymptotic to the base on either side. The two tails never touch the base line.

Figure

Normal distribution curve

Different figures of normal distribution curves will give an idea about the shape of the cure when values of means and standard deviation are taken into consideration. They are illustrated in figures (6.5 a, b, c, d).

It is clear from the figure (6.5a) that the normal curve is symmetrical and is a bell-shaped curve. Figure (6.5b) shows the normal probability distribution curves with the same mean, but with different standard deviation values. Figure (6.5c) illustrate three normal curve with the same standard deviation, but with different mean values. Figure (6.5 d) shows three different normal probability distribution curves, each with a different mean and a different standard deviation. It can be concluded from these curves that a normal curve can describe a large number of populations differentiated only by the mean and the standard deviation.

Area under normal distribution

The area under normal curve is distributed as follows.

(i) It is approximately 68 per cent of all the values in a normally distributed population that lie within 1 standard deviation (plus and minus) from the mean (Figure 6.6a).

$x \pm 1s$ covers 68.27% area; 34.14% area will lie on either side of the mean.

(ii) It is approximately 95.45 per cent of all the values in a normally distributed population that lie within 2 standard deviation (plus and minus) from the mean (Figure 6.6b).

$x \pm 2s$ covers 95.45% area; 47.73% area will lie on either side of the mean.

(iii) It is approximately 99.73 per cent of all the values in a normally distributed population lie within 3 standard deviations (plus and minus) from the mean (Figure 6.6c).

x±3s covers 99.73% area; 49.87% area will lie on either side of the mean.

Figure

Standard normal curve

The curve with zero mean and unit standard deviation is known as the standard normal cure. However, a curve with mean and standard deviation can be converted into a standard normal curve as shown in figure. For this, a change in scale and origin is essential, because with different values of mean and standard deviation, it become difficult to find out the area between two ordinates. In the original scale (x-scale), the mean and standard deviations are x and s. however, in new scale (z-scale) the mean is zero and standard deviation is unity. The process of changing the x-scale into z-scale is called z-transformation.

Figure

Equ

Where z = number of standard deviations

x = value of random variable

x = mean

s = standard deviation

note: the formula given for normal distribution is on sample basis and not on population basis.

Equ

Example 11: Find out the area under normal curve when z = 2.55

Solution

From the table of the standard normal distribution, the area between z = 0 and z = 2.55 is 0.4946. This area is shown in graph.

Figure

Computing normal curve probabilities

There are two important points to be remembered here

(i) z is the number of standard deviations from the mean. The interval from z = 1 and z = 2 is one standard deviation wide. However, it will not cover the same area between the mean and on z value.

(ii) The total area under the curve is 1 because normal curve is a probability distribution. The area covered by one-half curve is 0.5

The following example would illustrate this point.

“The life expectancy of light bulbs whose life times are normally distributed with a mean life of 750 hours and with a standard deviation of 80 hours”.

Dr. Mandaloju. Venkateshwarlu

Dr. Mandaloju Venkateshwarlu obtained his M.Sc. (1995) and Ph.D. Degree in Botany from Kakatiya University (1999). Best NSS Valentor (1991-1992) (For Social Work and Plantation) (Nick name of Chetla Venkanna), nearly above 4-5 lakhs plantation (1989-1999), Haritha Mithra, KU (1999), Editor in chief (AJBS) Senir Scientist & Honorary fellow FHAS, Life Member in CBEES, Advanced in Plant Sicences, Asian Journal of Bio Science, (Editorial Member) Natural and Environmental Pollution Technology, Indian Botanical Society, Member in IJSET, IJSEAT, IJSR, IOSR & Asia Pacific European Biochemical Society (Senior Member). He has contributed more than 90 papers in National and International Journal of Reoute Published 13 books. His Areas of interest are Plant Tissue Culture and Plant Biotechnology, Plant Mutation Breeding. Cytology and Cytogenetics, Molecular Biology and Crop Improvement. He attended National (70) and International (14) Conferences and many Workshops.

Published by

EJBPS

ISSN:2349-8870