DISTANCE LEARNING MATERIAL

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION

MIDNAPORE - 721 102

M. Sc. in Zoology

PART - I

Paper - III - Group : A & B

Module No. - 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 & 36

CONTENT

1.	Module No. 25	Paper – III	Page No. 01-24
2.	Module No. 26	Paper – III	Page No. 25 – 60
3.	Module No. 27	Paper – III	Page No. 61 – 81
4.	Module No. 28	Paper – III	Page No. 82 – 93
5.	Module No. 29	Paper – III	Page No. 94 – 115
6.	Module No. 30	Paper – III	Page No. 116 – 131
7.	Module No. 31	Paper – III	Page No. 132 – 159
8.	Module No. 32	Paper – III	Page No. 160 – 180
9.	Module No. 33	Paper – III	Page No. 181 – 233
10.	Module No. 34	Paper – III	Page No. 234 – 297
11.	Module No. 35	Paper – III	Page No. 298 – 344
12.	Module No. 36	Paper – III	Page No. 345 – 395

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Group-A Paper-III Unit-I
MODULE NO.-25

BIOSPHERE AND ECOSPHERE:

Ecology is the study of the organism in relation to the environment. Organism and environment are inseperably interrelated and interactive entities which exist as an integrated form called 'Biosphere'.

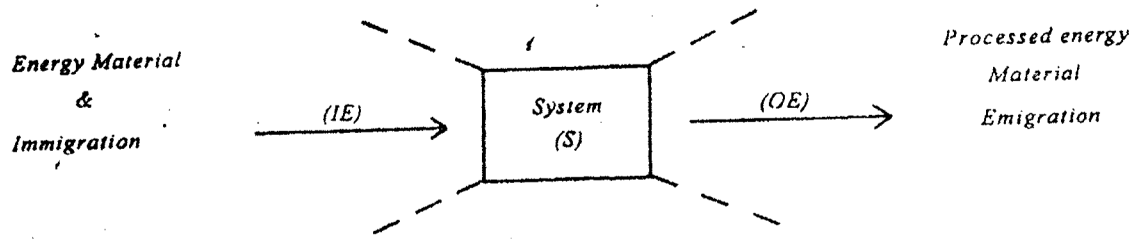
Biosphere is that narrow interface of Earth which can sustain living being alongwith the life present there. This includes hydrosphere, upper portion of lithosphere and lower portion of atmosphere (troposphere). The troposphere does not have any permanent inhabitant although air is temporarily invaded by many kinds of living being.

Sometimes the term biosphere is used in its original restricted sense to denote merely the blanket of life present on the earth and atmosphere, lithosphere, hydrosphere and biosphere and collectively called 'Ecosphere' (R.L. Smith, 1996). Biosphere in its modern sense and Ecosphere are rather synonymous as has been regarded by E.P. Odum (1983).

CONCEPT OF ECOSYSTEM :

Two main components of biosphere are biotic and abiotic components which interact together and constitute a 'Biosystem' (e.g. organismic system, population system, ecosystem etc.). In, Biosphere the 'biotic community' and the physical environment function togather as an ecological system or 'Ecosystem'. Which is a somewhat technical term for nature. The term was coined by A.G. Tansley (1935). Evans (1956) proposed that the ecosystem is the basic unit for ecological studies. It is a conceptual unit, has both structure and function but no size limit. It is an ecological unit through which energy flows and materials circulate. It is an open, self sustained and self regulated system. As far as fundamental definition is concerned ecosystem is synonymous with geobiocoenosis (Hill, 1960).

Since Ecosystem is an open system both input environment and output environment are essential for the ecosystem to function and maintain itself.



IE + S + OE = Ecosystem

Where IE = input environment.
OE = output environment.
S = system.

Structure and function of ecosystem :

Odum (183) recognised following structural components of ecosystem:

(1) Inorganic subtances (C, N, O, CO_2 etc.) which are involved in material cycles.

(2) Organic compounds (protein, carbohydrates, humus etc.) that link biotic and abiotic components.

(3) Air, water, substrate environment, climatic regime and other physical factors (light temparature, pressure etc.) which support life.

(4) Producers, photosynthetic and chemosynthetic autotrophs (auto=self, trophe=nounrishment).

(5) Macroconsumers or phagotrophs (phago=to eat) which include heterotrophic organisms, chiefly animals which ingest food.

(6) Microconsumers which include decomporets or saprotrophs (sapro = to decompose) and osmotrophs (Osmo= to pass through a membrane) e.g. hetrotrophic bacteria & fungi which obtain energy either by breaking down dead tissues or by absorbing dead organic matter.

The first three components are abiotic and remaining three are biotic components. Wiegert and Owen (1971), however, recognised two types of heterotrophs:

(1) biophages - organisms that consume other living organisms, and

(2) saprophages – organisms that feed on dead organic matter.

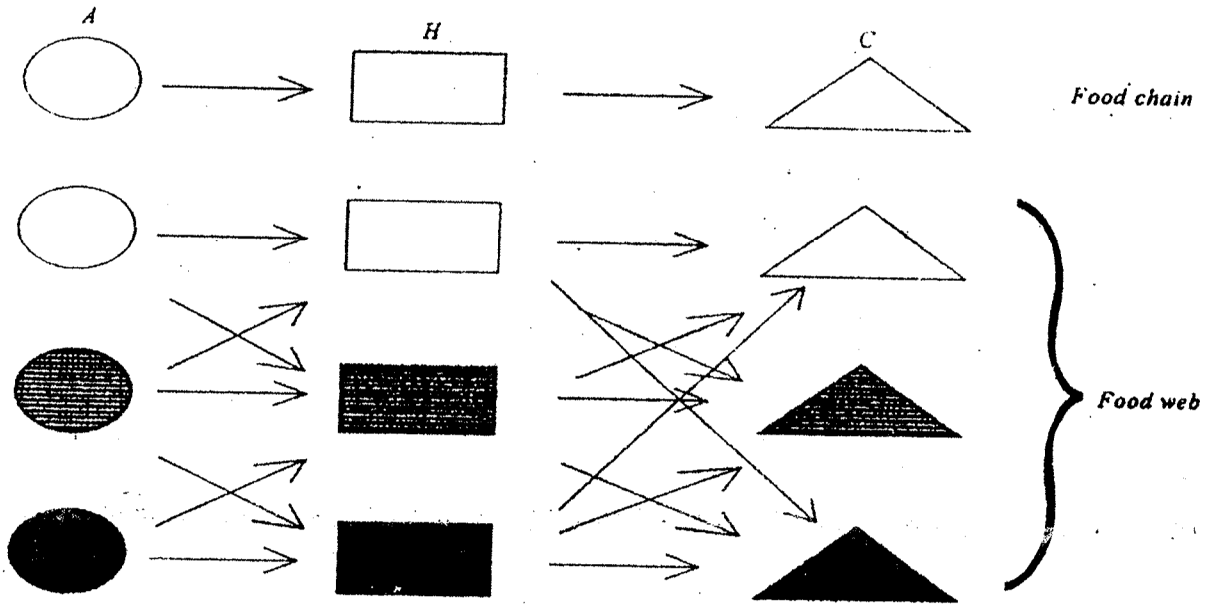
Food chain & food web :

Biotic components of an ecosystem remain linked in a system of 'eat and being eaten up'. In its simplest form it results into a 'food chain'.

Example :

(a) Grass ⟶ Grasshoppher ⟶ Frog ⟶ Snake

But in nature the relationship not so simple since most of the species are enryphagic i.e. they feed on many different kinds of food. As such a complicated network of trophic relation results which is called a 'food web'.



Where A = Autotroph, H = herbivore, C = Carnivore.

Food Chain does not always begin from autotroph. Sometimes food chain starts from dead organic matter. Such food chain are called detritus food chain e.g.

Detritus ⟶ Soil mite ⟶ Pseudoscorpion.

The producers and consumers in the ecosystem can be arranged into several feeding groups,

each Known as a trophic level (e.g. Producers = first trophic level, herbivores = second trophic level etc.). Each trophic level contains at any time a certain amount of living matter (or standing crop) which may be expressed in terms of number per unit area or as biomass : total mass of living matter expressed in fresh weight, dry weight or as energy.

Trophic classification is only a functional approach. A species may occupy more than one trophic level (e.g. Man, bear etc.).

Main functional aspects of ecosystem are :

(1) Production.

(2) Flow of energy.

(3) Biogeochemical cycle.

(4) Biological regulations (Population fluctuation, succession, homeostasis etc.).

Types of Ecosystem :

Different ecosystems can be broadly classified into two categories on the basis of human activity:

(1) Natural (e.g. Lakes, grassland, forest etc.) and

(2) Man made or artificial (Reservoirs, agroecosystems, plantations etc.)

The man made ecosystems have limited life span and are subjected to more allogenic influences.

CYBERNETIC NATURE AND STABILITY OF ECOSYSTEMS :

The science of cybernetics (Gr. Kybernetes = pilot or governor) as founded by Wiener (1948) embraces both inanimate and animate controls. In biological system it is called homeostasis.

The existence of homeostatic mechanisms at different level of biological organisation is well known. This is particularly evident at the physiological level (e.g. maintenance of body temperature in homeotherms). At ecosystem level also homeostasis does exist.

Homeostasis (Homeo = same, stasis = standing) refers to the tendency of a living system to resist change and remain in a state of equilibrium.

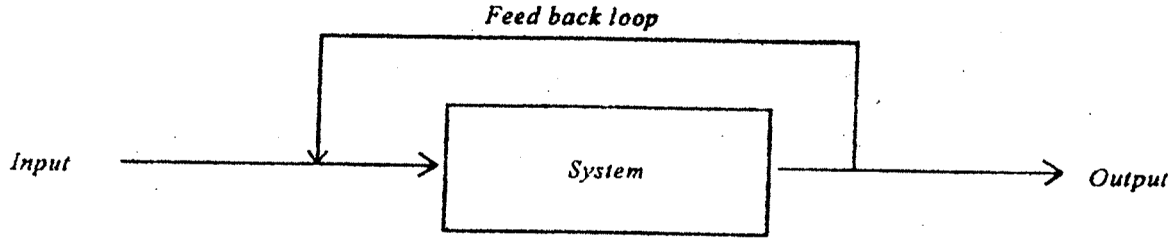
Stability may be achieved in two ways:

(1) Through feedback control.

(2) Through redundancy of components.

Stability through feedback Control :

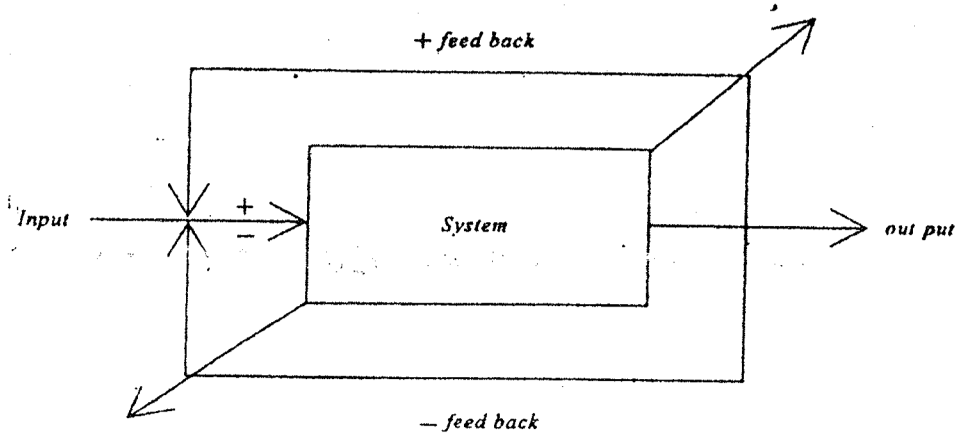
Ecosystem is a cybernetic or homeostatic system. A system consists of regularly interacting and interdependent components forming a unified whole. A cybernetic system is an open system where some of the output is fedback (feed back loop) into the system to control the future output.



In the predator-prey-relationship input of energy in predator (through ingestion) is dependant on and is controlled by the output of energy (chasing of prey). Here the stored energy is fed back to ensure future input of energy.

When the feed back is positive the quantity grows. Positive feed back is 'deviation accelerating'. It is necessary for growth and survival of organism/system. It deviates the system from a set point and may ultimately lead to self destruction of the system if it is not effectively counteracted. (e.g. cancerous growth of cell, population explosion etc.).

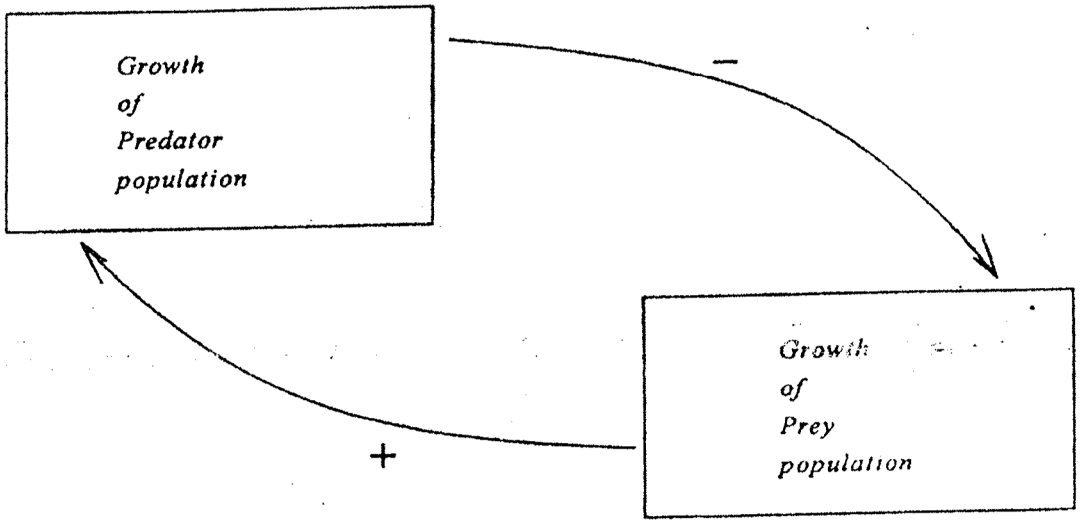
To achieve homeostatic control there must also be a negative feed back which acts as a 'deviation counteracting' input.



Thus homeostatic system is a self regulatory system maintained at a set point. Positive feed back increases the input and deviates it towards higher level increasing the amount. Negative feed back regulates the system and helps it to remain at homeostatic plateau, which has both upper and lower limit. A system, if perturbed, although controlled may not be able to return to the exact condition which prevailed before perturbation took place.

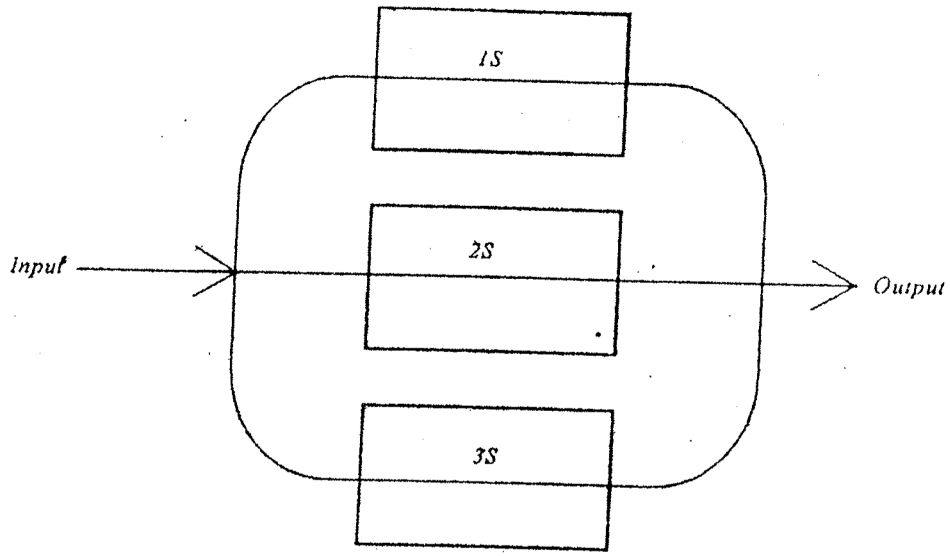
In nature atmosphere has a definite amount of CO_2 (0.03%). Excess of CO_2 is absorbed by the hydrosphere and the set point is maintained. However, CO_2 introduced into the atmosphere as a result of the excessive use of fossil fuel by man is slowly increasing the level of CO_2 in air.

In an ecosystem predetor-prey interaction is a good example of cybernetic stability. Growth of prey population acts as a positive feed back resulting in the growth of predator population which on its part acts as a negative feed back checking the unlimited growth of the prey. Thus an equilibrium of both the population is maintained.



Stability through redundancy of components :

More than one components capable of performing a given function is called redundancy. Redundancy enhances stability. For example diversity of species enhances stability of ecosystem. A food web is more stable than a food chain.



Types of stability :

From ecosystem point of view there are two types of stability:

(1) Resistance stability -

The ability of an ecosystem to resist perturbations and maintain its structure and function intact.

(2) Resilience stability -

The ability to recover when the system is disturbed by a perturbation.

These two types of stability are mutually exclusive and it is difficult to develop both simultaneously. Redwood forest of California has high resistance stability and can not be easily burnt, but has low resilience stability.

Stability may be (a) local or (b) global. The former is the tendency of a system to return to its original state from a small perturbation. Global stability is the tendency of a system to return to its original condition from all possible perturbations.

GAIA HYPOTHESIS :

First proposed by James Lovelock and Lynn Margulis (1973) which holds that organisms have evolved with physical environment to provide an intricate control system - "Gaia".

This control system makes the earth one complex but unified cybernetic system and keeps the earth's condition favourable for existence of life.

According to this hypothesis the prebiotic physical environment of earth was different from the present state of the earth's environment and was more like those of Mars or Venus. Early life on earth was anaerobic deriving energy from CO_2. Gradually organisms have evolved which are more dependant on CO_2. Thus the present physical environment of earth is not the cause of the origin of life but is the consequence of the origin of life in the earth.

PRODUCTIVITY AND FLOW OF ENERGY :

Organisms in an ecosystem are interdependant. Main cause of interdependance is energy.

Primary and secondary production :

The fixing of energy at any trophic level in the ecosystem is called production. Rate of production is productivity. In an ecosystem there are two kinds of production:

(1) Primary production - Production by autotrophs, and

(2) Secondary production - Production by hetrotrophs.

Fixed energy flows through the food web in a unidirectional pathway in accordance with the laws of thermodynamics.

Primary Production :

Source of energy :

In some chemosynthetic autotrophs the low potential energy of inorganic substances act as source of energy. For example in *Beggiatoa* energy is obtained from the oxidation of H_2S of surroundings.

In majority of autotrophs (Photosynthetic) basic source of energy is sun. Sun emits radiation at a constant rate: 2 gcal/cm^2/min ± 15%. This is called solar flux or solar constant. Only about 50% of this reaches the earth's surface. Rest is either reflected back (32%) or absorbed (18%) by cloud, water vapour, CO_2, dust etc. of atmosphere. This, of course, varies with latitude, season and locality.

Only a small fraction of the incident solar energy is utilized by plant for photosynthesis. Ration between the energy utilized for photosynthesis to the incident solar energy is called 'Ptotosynthestic Efficiency'. This varies from species to species.

Net and gross primary production :

Total energy fixed during photosysthensis or total energy retained by the plants as fixed carbon is "gross primary productions" (GPP). The energy remaining after respiration (R) and stored as organic matter in plant tissue is "net primary productions" (NPP).

$$NPP = GPP - R$$

Net production results into growth but it is not always equal to the change in biomass (ΔB) as considerable energy is eliminated (E) by way of grazing, mortality litter fall etc.

$$NPP = \Delta B + E$$

$$\text{Where, } \Delta B = Bt_2 - Bt_1$$

Total annual net primary production of earth is 170×10^9 tons (Whittaker, 1975). The gross primary productivity varies from ecosystem to ecosystem. It is least in desert and deep ocean (<0.5 gms/ m^2/ day). Estuaries, coral reef, tropical forest and perennial plantations are the most productive ecosystems (upto 25 gms/ m^2/ day). Tropical areas are more productive as compared to temperate areas. The mean net primary productivity of a tropical rain forest is 1300 g / m^2/ yr.

Factors influencing primary productivity :

A number of factors such as nutrient avialability, moisture / precipitation, temperature, day length, grazing, fire etc. influence rate of production. Tropical areas are more productive because of higher temperature and rainfall and greater day length. Season and age are also very important. In Scot's pine plantation maximum production is achieved at the age of 20 years. Gross primary productivity by phytoplankton in lake Erie is nine times more in summer as compared to winter.

Energy subsidy and primary production :

Man is able to increase primary productivity through energy subsidy in the form of fuel energy, man power, irrigation, fertilizer etc.

Secondary production :

Secondary production is also called consumer's production. Heterotrophic animals consume energy by way of ingestion. The fate of gross energy intake (I) can be summarized as :

$$I = A + F + U$$

Where A is the assimilated energy, F is the energy lost through faeces and U is the energy passed out as nitrogenous waste. Since $A = P + R$ where P is secondary production and R is respiratory energy the energy budget can be refined as

$$I = P + R + F + U$$

Although U should be included as part of A because of the difficulty of resparating nitrogeous waste from faecal waste U is not included in A & often U & F are combined together as excreted energy or E. Therefore

$$I = P + R + E \text{ or secondary production is } P = I - R - E.$$

In secondary production there is no gross production. Since $P = A - R$ and $NPP = GPP - R$ therefore analogous to gross primary production is assimilation in secondary production.

Example of secondary production :

Secondary productivity of African Elephant was studied by Petrides and Swank (1966). The data obtained by them as mofied by Krebs (1985) is given below:

Net primary productivity	-	747 Kcal/m^2/yr.
Ingestion (I)	-	71.5 Kcal/m^2/yr.
Faecal energy lost (F)	-	40.2 Kcal/m^2/yr.
Maintenance metabolism (R)	-	31.00 Kcal/m^2/yr.
Growth (P)	-	0.34 Kcal/m^2/yr.

Only a very small fraction of I is actually converted into P.

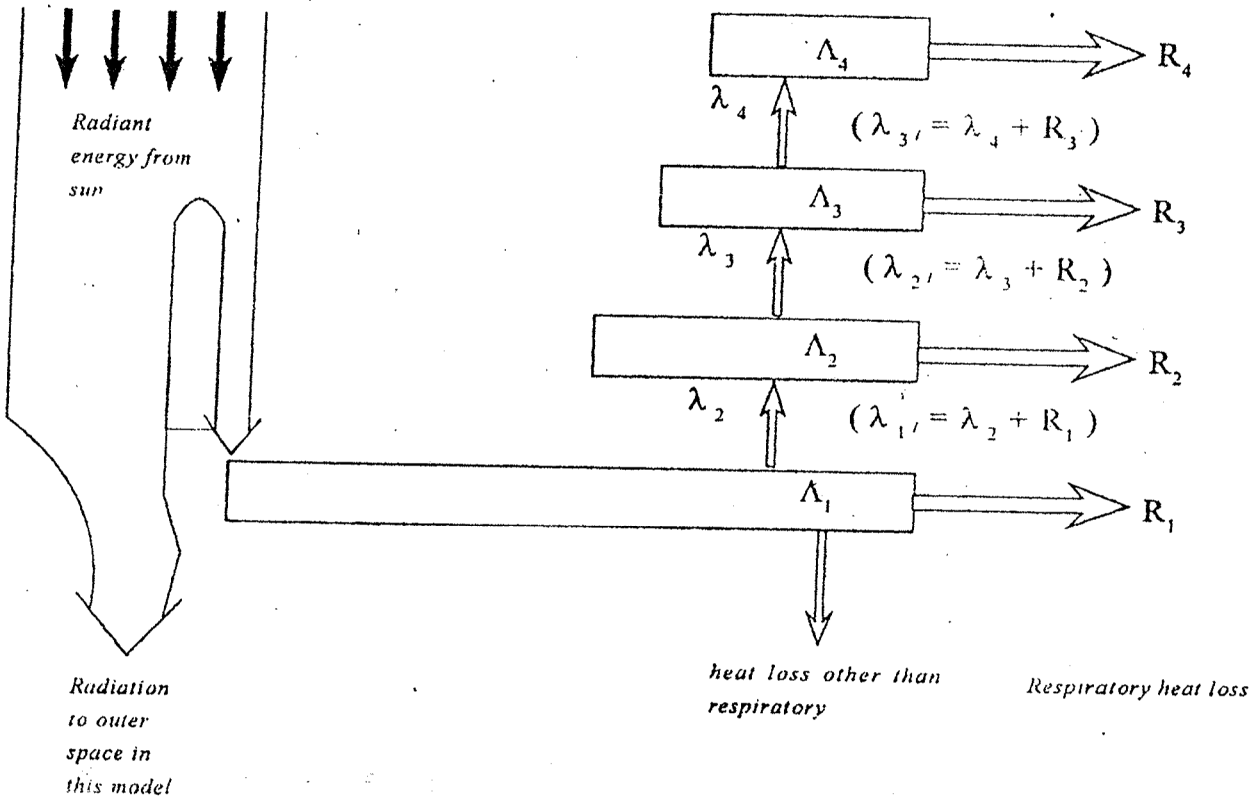
ENERGY FLOW MODELS :

Passage of energy from one trophic level to other is known as flow of energy. The concept has

been explained through various energy flow models. Some important models are given below:

Lindemann's model :

One of the earliest model was put forth by Lindemann (1942) to explain community energetics.



$\Lambda_1 - \Lambda_2$ = Energy content at any trophic level.

$\lambda_1 - \lambda_2$ = Contribution of energy from one trophic level to other.

$\lambda_{1/} - \lambda_{2/}$ = Loss of energy from each trophic level.

$R_1 - R_4$ = Respiratory heat loss.

This model depicts that

$$\lambda_n = \lambda_{n+1} + R_n + \Delta B$$

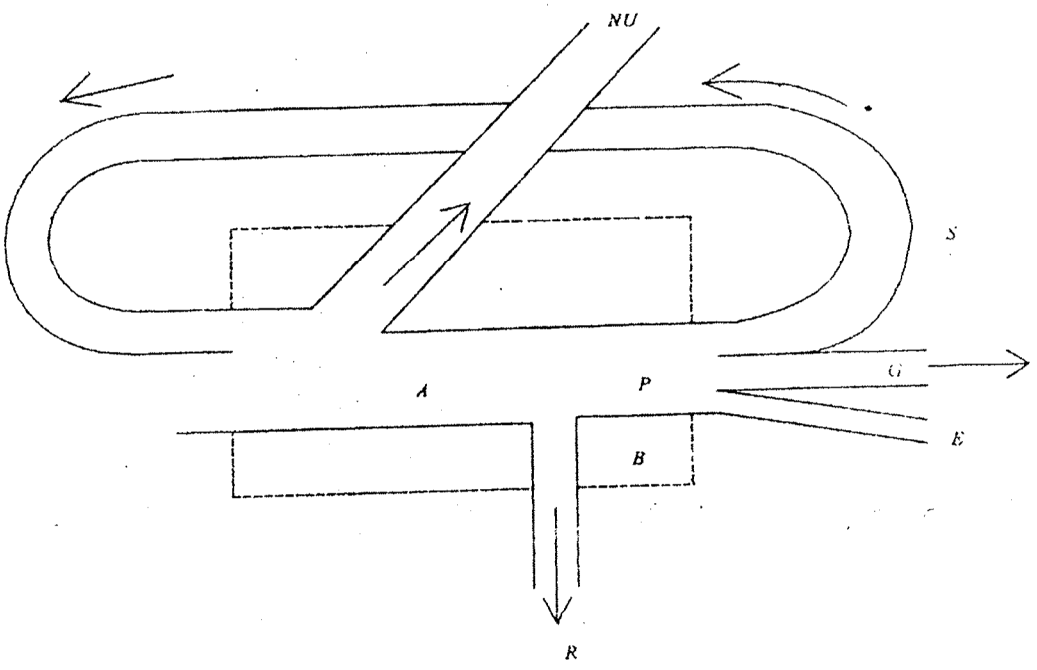
$$\text{or } \lambda_n = \lambda_{n/} + \Delta B$$

$$\therefore \lambda_n \neq \lambda_{n+1}$$

Where ΔB is the gain in biomass as such each trophic level contributes less amount of nergy to the next trophic level as compared to the energy entering into it.

However, this model does not explain energy loss other than heat from the trophic level and flow of energy through decomposers. The first aspect has ben taken care of in the 'Universal model' of E.P. Odum (1968).

Universal Model :



B = Biomass.

I = Input or ingested energy.

NU = Energy not used (Faecal loss).

A = Assimilated energy.

R = Respiration.

P = Production.

S = Stored energy.

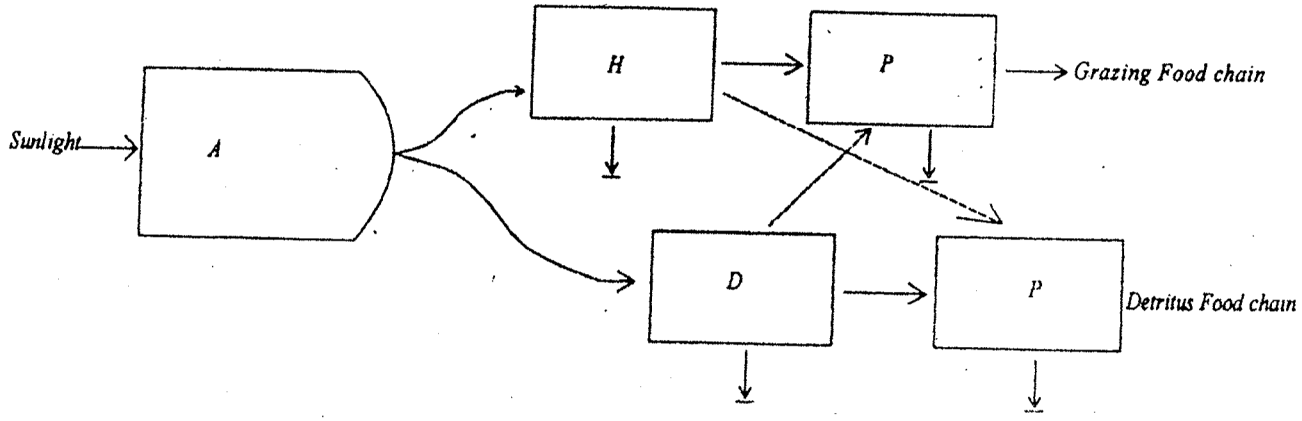
G = Growth.

E = Excreted energy.

The model is called 'universal' because it is applicable to any living component. This model repeated one after another and connected with each other explains flow of energy through food chain (See E.P. Odum, 1971, 1983). The model although explains energy losses other than heat does not explain energy flow through decomposers.

Y shaped or two channel model :

Odum (1963) recognised two distinct pathways of the energy flow model. This is depicted by Odum (1983) is given below:



Where A = Plant.

D = Detritus consumer.

P = Predator.

This model is more realistic as compared to single channel model since it explains utilization of dead organic matter (DOM) also.

The relative importance of two food chains varies from ecosystem to ecosystem. In some cases more than 90% of the autotrophic production flows through the detritus pathway.

Some other important models particularly with reference to energy flow through DOM

compartment and decomposes food chain are those by Wiegert and Owen (1971) and Head and Mac Lean (1975).

ECOLOGICAL EFFICIENCY :

Ratio between energy flow at different points along the food chain expressed in percentage is called ecological efficiency. Since or $\lambda_n = \lambda_{n+1}$ naturally the question arises what is the degree of reduction or in other words what is the magnitude of transfer efficiency. Slobodkin (1962) based on laboratory studies conducted that the ecological efficiency $\left(\frac{\lambda_j \times 100}{\lambda_i}\right)$ was approximately 10%. This is also called Lindemann's efficiency. Subsequent field and laboratory studies revealed that this may vary from 5%-50%.

Types of ecological efficiency :

There are various types. Some important one are given below:

(I) Between trophic level :

(1) Trophic level energy intake efficiency or Lindemann's efficiency = $\frac{I_t \times 100}{I_{t-1}}$

(2) Trophic level assimilation efficiency = $\frac{A_t \times 100}{A_{t-1}}$

(3) Trophic level production efficiency $\frac{P_t \times 100}{P_{t-1}}$

(4) Trophic level utilization efficiency or consumption efficiency = $\frac{I_t \times 100}{P_{t-1}}$

This increases towards higher trophic level and may vary from <1% to 40%.

(II) Within trophic level :

(1) Ecological growth efficiency $\frac{P_t \times 100}{I_t}$

(2) Assimilation efficiency $\frac{A_t \times 100}{I_t}$

It is less than 50% in herbivores detritivores & microbivores and about 80% in carnivores.

(3) Production efficiency $\frac{P_t \times 100}{A_t}$

It is <2% in endothermic vertebrates, approximately 10% in ecothermic vertebrates and 30-40% in invertebrates depending upon cost of maintenance.

In above mentioned expressions.

I = Energy ingested.

A = Energy Assimilated.

P = Production.

T = Trophic level.

Limiting factors and tolerance level :

Based on the type of response abiotic factors may be classified as follows (Fry, 1947):

(1) **Lethal** : Causing death, e.g. extreme heat or cold.

(2) **Masking** : Modifying the effect of some other factor; e.g. low relative humidity increase the ability of warm blooded animals to tolerate high air temperature.

(3) **Directive** : Producing an orienting response in relation to some environmental factor; e.g. some insect move towards the source of light.

(4) **Controlling** : Influencing the rate of activity; e.g. temperature influences rate of development.

(5) **Deficient** : Curtailing an activity if some factor is present in low amount or is absent; e.g. Sitophilus oryzae fails to breed if moisture content of grains falls below 9%.

Often the distinction between controlling factors and deficient factors is not made and these together constitute limiting factors.

Limiting factor :

Any factor that tends to slow down potential growth in an ecosystem is said to be a limiting factor (Odum, 1963).

Blackman (1905) developed the concept when he said five factors control the rate of photosynthesis. This concept is nothing but the extension of the 'law of the minimum' put forth by Liebig 91840) which holds that a minimum quantity of substance or condition is required in soil for development of plants. The law equally holds good for animals. Now it is known that not only there is a lower limit (threshold) for a factor but there exists an upper limit too.

Law of toleration / tolerance :

According to Shelford (1913) for each species there is a range in an environmental factor within which the species functions at or near optimum. Towards each ectreme the functions of the species are first curtailed and then inhibited.

On the basis of above mentioned concept following zones of tolerance may be identified.

(1) Zone of intolerance :

Exists beyond upper and lower limit of tolerance. Organisms do not survive here.

(2) Zone of physiological stress :

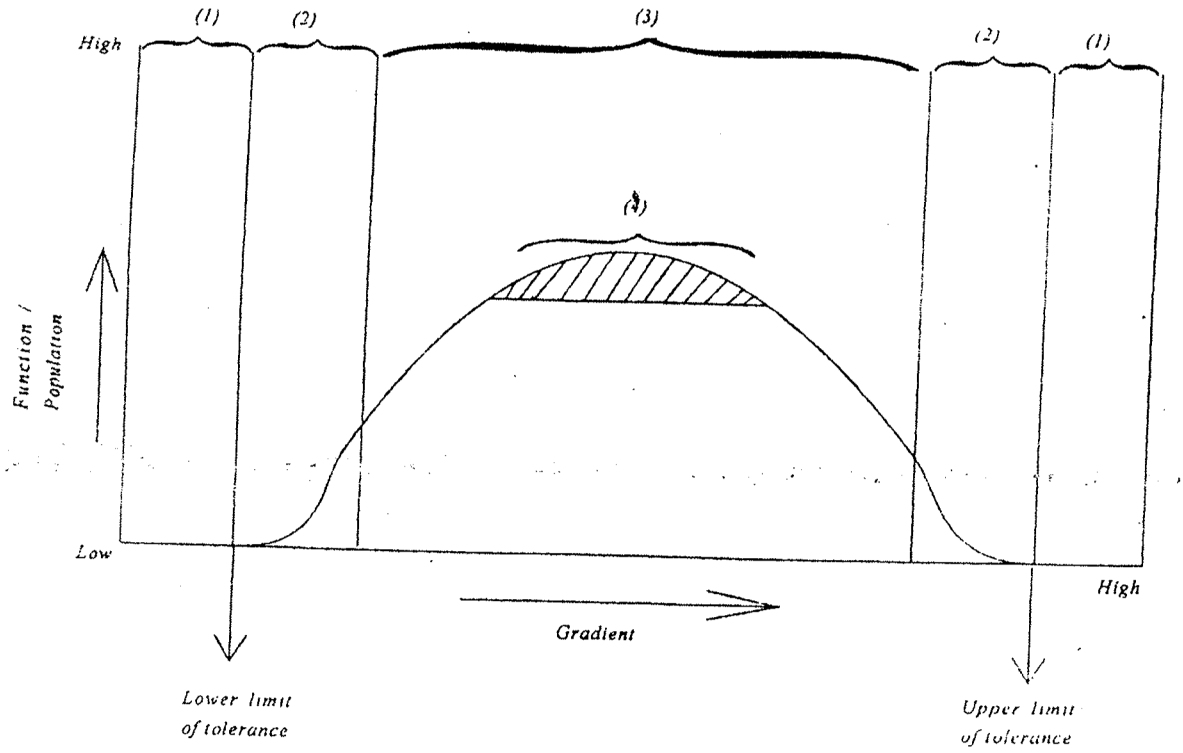
This is situated close to upper and lower limits but within the range of tolerance. Organisms remain in dormant or stressed state.

(3) Opimal Zone :

In between zone of physiological stress. Organisms show normal function.

(4) Zone of preferendum :

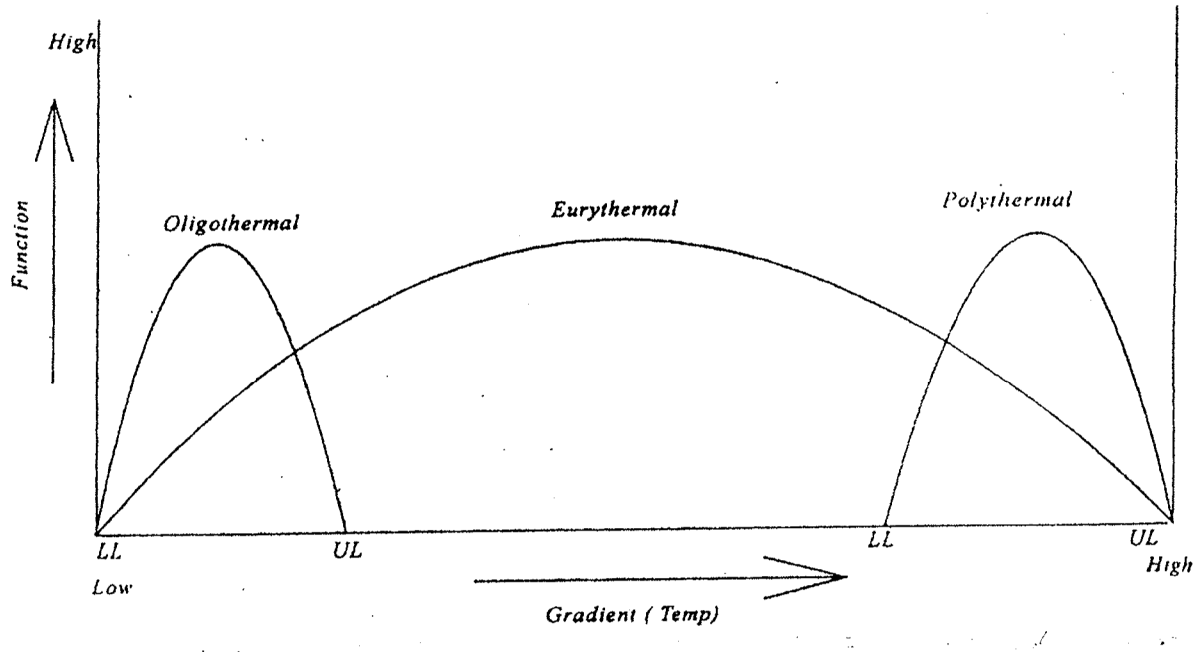
Here conditions are best suited. Organisms prefer to live under such condition.



Limits of tolerance :

The limits of tolerance with respect to any factor varies from species to species, from one geographical race to other, from individual to individual and from age to age. The prefix 'steno' &

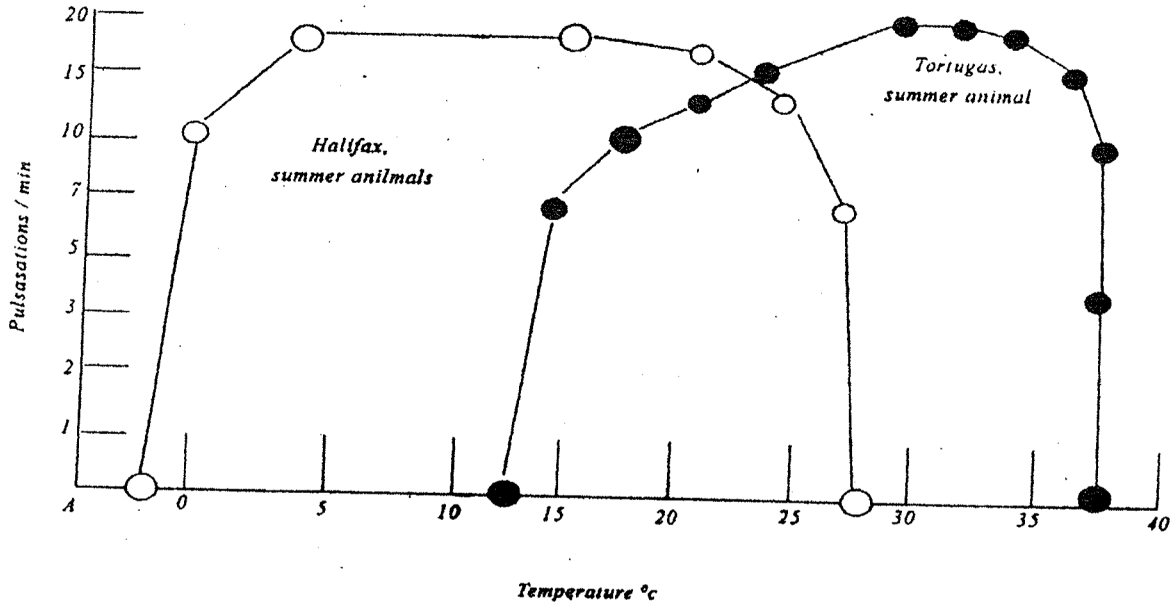
'eury' are used to denote narrow and wide range of tolerance respectively. For instance eurythermal and stenothermal mean braod and narrow range of tolerance with respect to temparature. There may be two kinds of stenothermal creatures oligothermal, narrow towards lower gradiant and polythermal, narrow range towards upper gradient. *Trematomus bernacchi* lives in antarctic sea and is extremely stenothermal (oligothermal) with a range from –2°C to +2°C. Desert pupfish (*Cyprinodon macularius*) can tolerate a range between 10°C and 40°C and thus is a *polythermal* species.



Ecotype

A species with broad range of tolerance is expected to be widely distributed. Such species, however, have locally adapted populations or races which are called 'ecotypes'. The concept of ecotype was first developed by G. Turreson (1920-30).

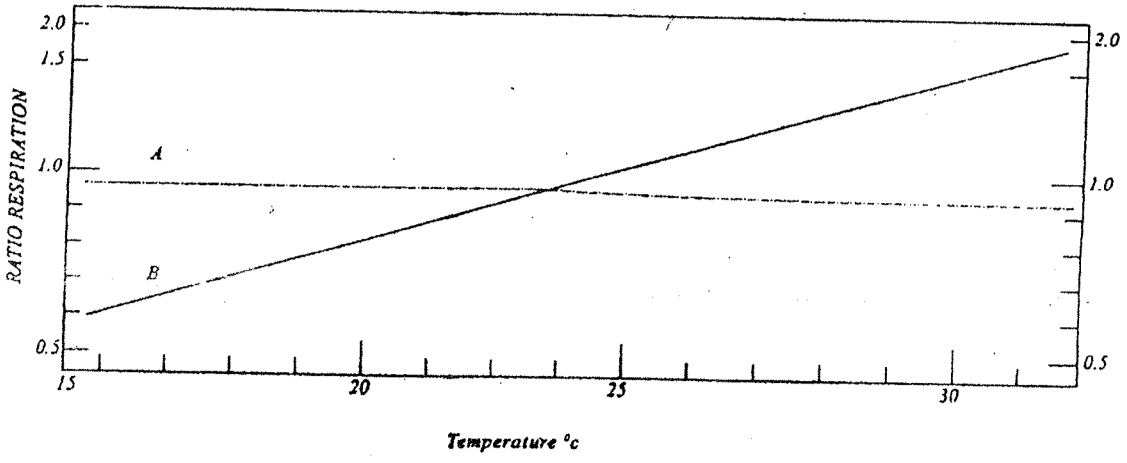
Mean water temperature of halifax sea is 14°C and that of Tortugas sea is 29°C. In Halifax variety of *Aurelia aurita* 10°C the optimum temperature for locomotion but in Tortugas variety this is threshold for locomotion. Similarly 29°C is the upper limit of tolerance for Halifax variety bt it is optimum temperature for tortugas variety. Thus these two varieties of *Aurelia aurita* are ecotypes.



Usually repreoductive individuals and sub adult stages (egg, larvae etc.) have narrow range of tolerance than the adults. Atlantic lobster can live in water with temperature ranging from 0-17°C but eggs & young stages can not thrive below 11°C.

Factor compensation :

Animals are not just 'slaves' to the physical environment. They adapt themselves and modify the effect of physical environment to reduce the limiting effect of abiotic factors. Ability to reduce the limiting effects of abiotic factor is 'factor compensation'. This is particularly effective at the community level but also occurs at species level (ecotype) and at individual level (acdimatization).



The figure shows that the respiration of the whole community in a balanced microcosm (A) is less effected by the change in temperature as compared to the respiratory rate of one species (B). This is an example of factor compensation at the level of community.

Temperature as an ecological factor:

Temperature as an ecological factor has two aspects:

(I) Temperature as a physical environment and

(II) Temperature and living beings.

(I) Temperature as a physical environment :

Both aquatic and terrestrial environment are subjected to fluctuating temperature but such fluctuations are relatively less intense in aquatic environment. Minimum temperature of water in sea is -3^0C and in fresh water ponds it never goes below 0^0C. Water temperature is influenced by aerial temperature and normally it is below 40^0C but in some hot springs water temperature may be over 90^0C. Terrestrial environment are subjected to considerable diel and seasonal variations in temperature. Temperature of soil may vary by 30^0C between sunlight and shade. Temperature also varies with latitude and altitude. With every 150m of increase in altitude temperature decreases by 1^0C. Such changing temperature, therefore, have obvious influences on organisms.

(II) Temperature and animals :

Physiologically animals can be divided into three groups based on the nature of the variation in body temperature:

1. **Poikilotherms :** Body temperature varies with environmental temperature e.g. invertebrates, fish, amphibia, reptiles.
2. **Homeotherms :** Body temperature is fairly constant regardless of environmental temperature e.g. birds & mammals.
3. **Heterotherms :** Sometimes regulate body temperature & sometimes do not e.g. bats, camel.

Based on the mechanism of thermoregulation animals are classified into two groups

1. **Ectoghermic :** Principal source of body heat is external temperature.
2. **Endotherms :** The body temperature is a product of animals own oxidation metabolism.

Some ectoderms (amphibians, reptiles) can raise their body temperature by basking in the sun (heliothermism) or by remaining partially submerged in water.

Range of temperature tolerance :

For animals 52°C is considered as the upper limit of tolerance. Desert insects can be found on hot sand with a temperature of 65°C. Some rotifers can be exposed to a temperature of 150°C for few minutes. For most insects 8°C is considered as a lower limit but many animals can tolerate subzero temperature. Some rotifers, nematodes & tardigrades can be exposed to a temperature of –272°C.

Effect of temperature :

Temperature influences both structure and functional response. Some examples are cited below:

Temperature and Morphology :

1. Animals are dark in colour in warm climates & light in colour in cold climates (Gloger's rule) e.g. Black bear & polar bear.
2. Animals are of bigger size in cold climates as compared to those of warm climate (Bergmann's rule).
3. Body extremitis (tail, ear, legs, snout) are smaller in mammals of cold areas (arctic fox) as compared to warm areas (red fox): Allen's rule.
4. Cyclomorphosis -

Some animals exhibit a cyclic change in morphology with respect to the seasonal variation in temperature - the cyclomorphosis. In Daphnia a distinct helmet like structure is seen in summer which decreases in size during autumn, completely disappears in winter and reappears in spring.

Temperature and sex ratio

Sex ratio in some animals (flea, copepods, rotifers and reptiles) is influenced by environmental temperature. *Xenopsylla cheopis* males outnumber females on rats when the mean temperature of day is between 21°C-25°C but the position is reversed in cooler days.

In red eared slider turtle (*Trachemys siripta*) eggs when incubated at 20°C-20.8°C produce all males and when incubated at 29.4°C-35°C produce all females.

Temperature and function / rate of activity

Temperature influences various aspects of physiological & other activities.

1. Temperature influences speed of development. In *Metriocnemus hirticollis* development is complete in 26 days at 20°C but at 10°C 94 days are required.
2. In *Melanoplus mexicanus* threshold temperature for development of eggs is 8°C. If the egg is given a brief exposure at 0°C and then transferred to the optimum temperature of 27-32°C it develops at a faster rate.
3. Prevailing temperature also influences number of moults. In *M. mexicanus* there are 6 moults when temperature is between 22-27°C and 5 moults when temperature is between 32-37°C.
4. Temperature influences rate of respiration, feeding & other activities. The large mouth bass does not eat at 5°C but it feeds at the rate of 1.6 minnows/ week and 4.1 minnows/week at 12°C and 20°C respectively.

Temperature and behaviour

Temperature influences behaviour of the animals. Although some animals can tolerate extremes of temperature yet mostly these are avoided. Animals orient themselves in such a way so that they may find optimum temperature conditions.

1. Most of the desert animals are nocturnal in nature or they are active during dawn or dusk. High day temperature is avoided by living in burrow or in shades.
2. Desert birds nest on the eastern face of canopy avoiding hot afternoon sun.
3. Amphibians and reptiles bask in sun to raise their body temperature and enter into water to cool down their body temperature.
4. Extremes of temperature also induces hibernation (winter dormancy) and estivation (summer dormancy) which are not confined to ectotherms and may be seen in several endotherms such as bear, bats, arctic ground squirrels etc.

5. Fluctuating temperature causes thermal migration which may be diel or seasonal, vertical or horizontal and may involve distance ranging from few csm. to several hundred kilometers.

Temperature and distribution of animals :

Distribution of a species is largely influenced by temperature both density and diversity of species decline towards higher latitude.

1. Range of distribution may be governed by the critical temperature for reproduction. e.g. *Homarus americanus* is distributed in sea where water is warmer than 11°C which is the minimum temperature required for breeding even though adult can tolerate a temperature between 0-17°C.

2. Distribution is often ogverned by the upper or lower limit of tolerance. *Pecten greenlandicus* is never found in areas where temperature rises above 0°C and *Leptinotarsa decemlineata* is not found in areas where temperature of soil falls below –12°C.

These are only some selected examples. In nature it is quite difficult to isolate effect of one factor from other. Quite often it is the interplay of several environmental factors which influences structure & functions of organisms.

Review questions :

1. Distinguish between :
   - i) Biosphere & Ecosphere.
   - ii) Food chain & food web.
   - iii) Grazing food chain & detritus food chain.
   - iv) Natural & artificial ecosystem.
   - v) Resistance & resilience stability.
   - vi) Net production & gross production.
   - vii) Controlling & deficient factors.
   - viii) Oligothermal & polythermal animals.
   - ix) Ectotherm & Endoderm.

x) Grazing & detritus pathway of the energy flow.

2. **Write notes on :**

i) Gaia hypothesis.

ii) Energy subsidy.

iii) Limiting factors.

iv) Law of tolerance.

v) Ecotype.

vi) Factor compensation.

vii) Cyclomorphosis.

viii) Thermal migration.

ix) Temperature & sex ration.

x) Hetrothermism & Heliothermism.

3. Define Ecosystem. Give a brief account of the structural components of an aquatic ecosystem.

4. What is homeostosis? How stability of an ecosystem is maintained through feed back mechanism & through redundancy of components.

5. Contrast primary & secondary production. Give a comparative account of the Lindemann's model & Odum's two channel models of energy flow.

6. Discuss various types of ecological efficiencies. "Endothermic carnivores are good assimilators but poor producers". Justify the statement.

7. Discuss the effect of temperature on behaviour & distribution of animals.

Suggested readings :

1. Ecology & Field Biology, R.L. Smith. Harper Collins Publications.

2. Basic Ecology, E.P. Odum. Saunders College Publishing.

3. Ecology, M. Begon, J.L. Harpers & C.R. Townsend. Blackwell Scientific Publications.

- - - - -

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Group-A Paper-III Unit-I
Module No. - 26

COMMUNITY STRUCTURE AND ORGANIZATION

Biotic Community :

A biotic community is a naturally occuring, mutually sustaining and interacting assemblage of plants and animals living in the same environment, fixing, utilizing and transferring energy (R.L. Smith, 1980).

There are two opposing views explaining the nature of the community : (1) The organismic concept, which regards community as some sort of super organism, the highest state in the biological organization (F.E. Clements, 1916). (2) The in dividualistic concept on the contrary argues that the community is a collection of species that requires similar environmental conditions (H.L. Gleason, 1917). The individualistic approach was ignored till 1950 but now seems to be more widely accepted of the two.

Major and minor community :

Major communities are of sufficient size and completeness of organization. They need only to receive solar energy from the outside and are relatively independent of inputs and outputs from adjacent communities. For example a forest community.

Minor communities are more or less dependent on the neighbouring communities or on the major communities in which these are situated for inputs and outputs. For example a fallen log, a root spring.

Pioneer and climax community :

Biotic communities tend to maturity. Initial simple stage is called pioneer community and mature final stage is called climax community.

Naming of community :

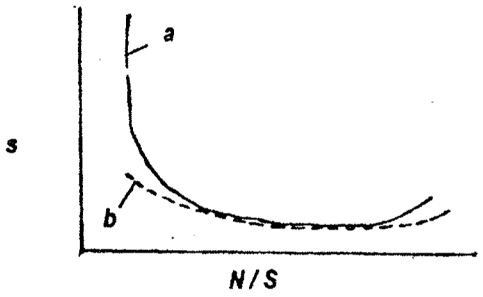
The community is often named after the dominant floral component. For example grassland community or pine forest community.

Structure of the community :

A biotic community consists of individuals belonging to more than one species. Therefore, the two basic parameters in the study / analysis of biotic community are number of species (S) and number of individuals belonging to each species (N). Some communities are richer than the others in the number of species they contain. This does not mean that such communities should have more number of individuals per unit area or per species.

Relationship between species and number :

In an undisturbed, natural environment the general relationship between species and number can be pictured as a concave curve (a) i.e. there are many species with a few individuals and there are a few species which have many individuals. Stress will tend to flatten the curve (b) by eliminating accidental and rare species.



Structure of two communities may differ with respect to the density and diversity of species. A brief account of the basic components are discussed below.

(1) **Abundance :**

Abundance (A) is the commonness of a species :

$$A = \frac{n_i}{x_i}$$

(2) **Density :**

Density (D) is the average number of individuals of a species per sample

$$D = \frac{n_i}{x}$$

Density expressed as per unit area is population density.

(3) **Frequency :**

Frequency (F) is the degree of occurrence expressed as percentage.

$$F = \frac{x_i \times 100}{x}$$

A frequency of 40% means that the species is present in 40% of the area surveyed irrespective of its density.

In above mentioned expressions

n_i = Number of individuals belonging to ith species.

x_i = number of samples in which ith species is present.

x= total number of samples.

NOTE : Sometimes distinction between abundance and density is not made and the two terms are used interchangeably.

Frequency increases with the increase in abundance unless the species is contagious in distribution.

Raunkiaer (1934) established five frequency classes for plants viz., A (1-20%), B (21-40%), C(41-60%), D (61-80%) and E (81-100%). He further opined that $A > B > C \overset{>}{<} D < E$. On the basis of frequency Raunkiaer classified plants as accidental (frequency upto 40%), common (41-80% frequency) and characteristic (>81% frequency) species. Such classification is arbitrary and varies from one group of organisms to other depending upon the choice of authors. For example Wallwork (1970) suggested following scale for soil fauna :

Frequency range	**Class of species**
Upto 5%	Rare
5 – 14%	Occasional
15 – 29%	Frequent
30 – 100%	Abundant

Frequency is also a measure of *constancy* of a species. Skubala (199) suggested following constancy scale:

Frequency	Constancy Class
>75%	Euconstant species
50.1 – 75%	Constant species
25.1 – 50%	Accessorial species
<25	Accidental species

(4) **Relative abundance :**

Abundance of a species in relation to other species in the community expressed as per cent is relative abundance.

$$RA = \frac{n_i \times 100}{N}$$

Where, RA = relative abundance.

n_i = number of individuals of ith species.

N = total number of individuals of all the species.

Relative abundance is an important community parameter particularly when seasonal fluctuations are concerned and also in ascertaining dominance of the species in a community. Some species are strictly *opportunistic* (Mac Arthur, 1960) and occur in great abundance when conditions are highly favourable and become scarce when conditions are unfavourable. Such species show considerable temporal variation in relative abundance. In contrast to these the species which are neither very common nor very rare show little variation in relative abundance. Dominance of a species in the community may be ascertained with the help of relative abundance. S kubala (1999) has proposed following scale for soil mites :

Relative abundance	Dominance status
>10%	Eudominants
5.1 – 10%	Dominants
2.1 – 5%	Sub dominant
1.1 – 2%	Recedents
<1%	Sub-recendents.

Dominance :

In a community some species are common either in number or in biomass i.e. have high 'importance value'. Such species usually occupy more space (have greater frequency) and play important role in community dynamics. These are considered as dominant species. A species may become dominant due to prevailing climatic factors. A species may become dominant as a result of coaction between two or more species. Prairie dog can produce short grass community as these prefer long grass in a mixed prairie (Koford, 1958). Predators control the structure of the community so must be regarded as dominant and *Keystone species.* A species whose activities have a significant role in determining the community structure are called Keystone species (R.L. Smith, 1996). Tiger is a Keystone species.

Dominance index :

Dominance indices based on relative abundance are of particular importance in comparison of the structure of community. One of the earliest such index, index of dominance (C) was developed by Simpson (1949) –

$$C = \Sigma(n_i / N)^2$$

Mc Naughton & Wolf (1970) have shown that the dominance can be related to the 'harshness' of environment. They define dominance index (DI) as the concentration of influence in the two most abundant species. According to them $DI = \frac{100(y_1 + y_2)}{y}$

Where y_1 & y_2 = the abundance of species 1 & 2.

y = the total for all species in the community.

An environment may be harsh for a species but may be equitable for some other species. Structurally less complex habitat show harsher environment and involve greater diel or seasonal temperature and hydrological fluctuations. Temperate environment are regarded harsher than the tropical environment. Dominance index increases with the harshness of the environment and decreases with the vegetational development. For example DI for avifauna (Karr, 1971).

Study area	**DI for avifauna**
Temperate	50
Bare ground	
Shrub	17 – 20
Tropical	42 – 62
Grassland	
Shrub	12 - 16

Species diversity :

Number of species vary from one community to other. Some communities have fewer species as compared to others but this does not mean that such communities are poor in species diversity. Species diversity is ratio between number of species (S) and number of individuals (N) in a community.

α – β - γ diversity :

Diversity of species within a habitat is called 'α' diversity. Diversity of species between two habitats in the same geographical area is called 'β'- diversity and diversity of species between different geographical areas is called γ- diversity.

Species diversity indices :

A number of diversity indices are available in ecological literatures which are useful for the analysis of the structure of community, particularly for the purpose of community comparison. Some commonly used indices are mentioned below :

(1) Richness or variety indices (d) :-

i) Margalef's index (1958) : $d = \frac{S - 1}{\log N}$

ii) Odum, Cantlon & Kornickar's index (1960) :

d = 'S' per thousand 'N'.

iii) Menhinick's index (1964) :

$$d = \frac{S}{\sqrt{N}}$$

These indices though can be easily calculated are not much useful in the study of biotic communities because these change with the size of the sample. For example if all the species of a habitat are already represented in a collection, further collection would reduce the proportion of species to individuals.

(2) Index of general diversity $(\overline{H})$ or Shannon -- Wiener diversity index (1964)

This index (often wrongly referred to as Shannon – Weaver index) is considered best for making comparison because it is reasonably independent of the sample size. It can be computed by the following

expression :

$$\overline{H} = -\Sigma\left(\frac{n_i}{N}\right)\log\left(\frac{n_i}{N}\right)$$

$$\text{or } \overline{H} = -\Sigma P_i \log P_i$$

Where n_i=number of individuals of ith species or importance value.

N = Total for all the species.

P_i= Importance probability of each species.

General diversity index has been used by some authors as a biocriteria for adjudging the level of pollution in water bodies. For example according to Wilhm and Dorris (1968) water may be considered as clean if $\overline{H}>3$, as moderately polluted if $\overline{H}$ is between 1 & 3 and as heavily polluted if $\overline{H}<1$.

Shannon's index has some limitations. It does not indicate the homogenity or heterogenity of the community or about the equitability of the community which may be analysed by the 'evenness index' or 'equitability index'.

Evenness index or equitability index :

These indices are further informative and reflect the evenness of allotment of individuals among species. For example two communities each containing 10 species and 100 individuals may have the same S/N index but in these communities individuals may not be evenly packed among species. One community may have *minimum evenness* at one extreme with 91+1 +1 +1 +1 +1 +1 +1 +1 +1 individuals and the other community may have *perfect evenness* with all the 10 species having 10 individuals at the other extremes. These two communities, therefore, would have different evenness index. Some commonly used such indices are –

(1) Evenness index (e) of Pielou (1966) :

$$e = \frac{\overline{H}}{\log S}$$

(2) Equitability index (J) of Sheldon (1969) :

$$J' = \frac{H'}{H'_{max}}$$

where $\overline{H}$ or H' = Shannon's index.

S= Number of species.

H'max=maximum diversity possible for the community if all species were equally abundant (log S).

Evenness index may provide other information's too. For example Khotko and Artemieva (1995) opined that evenness index for soil biota decreases with the increase in stress in the environment.

Some important ecological principles associated with the concept of species diversity :

(1) Diversity of species reflects complexity, stability and maturity of the ecosystem and depends on the availability of ecological niche. Diversity is more in complex, stable and mature ecosystem.

(2) Species diversity tends to be low in physically controlled ecosystem and high in biologically controlled ecosystem.

(3) Diversity increases with decrease in the R/B ratio (total community respiration to total biomass ratio) or 'ecological turnover'. When the R/B ratio is low more of the community energy can go into diversity.

(4) Diversity is also influenced by trophic relationship such as amount of grazing & predation.

(5) Diversity decreases with the increase of stress in the environment.

Some low diversity environments are :

(1) New environment e.g. newly exposed sand dunes.

(2) Severe environment e.g. Desert.

(3) Unpredictable environment e.g. Polar regions.

(4) Stressed environment e.g. Polluted river.

Community boundry :

Community boundry is difficult to define. Although some communities may have a clearcut boundry between themselves (e.g. grassland & pond, Old field & a burnt field) often two adjacent communities meet and blend into each other forming an 'edge' or 'ecotone' a concept first stated by Leopold (1933). Although previously regarded as similar the concept of 'ecotone' and 'edge' are now regarded as different (Smith, 1996).

(1) **Edges :**

When community boundries are sharp or abrupt.

(2) **Ecotone :**

When community boundry is diffused blending into each other forming a transient zone.

Types of edges:

(1) **Inherent edge :**

These are natural such as between terrestrial and aquatic systems formed due to changes in topography and soil types.

(2) **Induced edge :**

Arising due to natural disturbances like fire, floods or man made disturbances like agriculture, clearing, harvesting etc.

Types of ecotone :

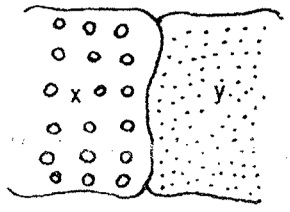
Ecotone may be classified as 'simple' ecotone and 'multiple' ecotone as well as 'narrow' and 'wide' ecotone.

(1) **Simple & multiple ecotone :**

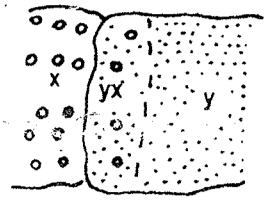
When ecotone is formed by the blending of two communities then it is called simple ecotone. When formed by the advancement of more than two communities then known as multiple ecotone.

(2) **Narrow and wide ecotone :**

Narrow ecotone is formed due to advancement of one community into other. Wide ecotone is formed when both the communities blend into each other.

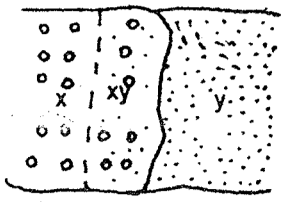


Edge between x & y communities

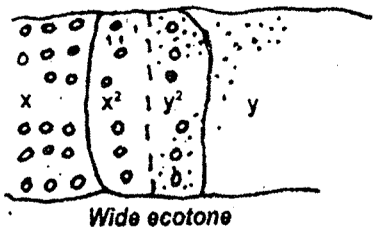


Narrow & simple ecotone

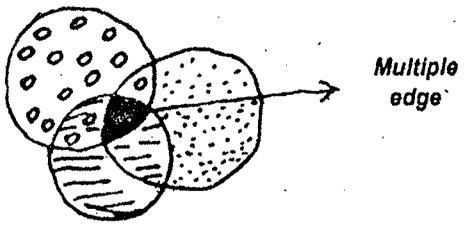
Community 'y' extending into community 'x'.



community 'x' extending into community 'y'.



Wide ecotone
Community 'x' & community 'y' both blending into each other.



Multiple edge

Edge index (EI) :

Patton (1975) developed this index which is the ratio between edge to area. This can be calculated by the following expression :

$$EI = \frac{TP}{2A\pi}$$

Where TP = the total perimeter around area plus any line or edge within the area.

A = Area.

π= 3.14

Edge index is lowest (i.e. 1) for a circle.

Edge effect and edge species :

In ecotone there is an overlapping of flora and fauna of the either communities. In addition to this there may be species which are characteristic and often restricted to the ecotone. Thus both number of species as well as density of species are greater in the ecotone than in the communities flanking it. The tendency of increased diversity and density of species at ecotone is called 'edge effect'. The ecotone is characterised by steep gradient of environmental factors (e.g. wind flow, turbulance, temperature, moisture, light etc.). Those species which are restricted to ecotone only are known as edge species.

POPULATION ECOLOGY

LIFE TABLE

Mortality is one of the key parameters that cause population change. It is important to know how mortality is occurring in population. Is mortality higher in juvenile forms? Do older individuals have higher mortality? What is the average expectation of life? Such questions can be answered by constructing a 'life table'. A life table is a convenient format for describing the mortality schedule of a population. It is an age specific summary of death rates operating in a population.

Life tables were developed by human demographers particularly those working for insurance companies as they are interested in knowing how long a man can be expected to live so that they may fix the premium accordingly. The concept was introduced in ecology by Raymond Pearl in 1921.

Construction of 'Life table' :

To construct a life table following basic columns are required to be set up.

x = age in terval.

n_x = number of survivors at start of age interval x.

l_x = proportion of organisms surviving at the start of age interval x.

d_x = number dying during the age interval x to x+1.

q_x = mortality during age interval x to x+1.

L_x = the average number of individuals alive in each age interval. This is also called life table age structure.

T_x = cumulative value of L_x.

e_x = mean expectation of life for individuals alive at start of age x.

To set up a life table it is important to decide on the age interval to be taken up. For humans age interval may be 5 years for deer 1 year, for field mice it may be 1 month, for insects and mites it may be 1 day and for protozoa it may be 1 hour.

In fact all the columns can be calculated from the data collected for the second column i.e., n_x. For example

$$l_x = \frac{n_x}{n_0}$$

$$d_x = n_x - n_{x+1}$$

$$q_x = \frac{d_x}{n_x}$$

$$L_x = \frac{n_x + n_x + 1}{2}$$

T_x can be calculated by summing up L_x values cumulatively from the bottom of the life table $\left(T_x = \sum_x^{\infty} L_x\right)$ For example $T_6 = L_9 + L_8 + L_7 + L_6$

Finally e_x is calculated as follows :

$$e_x = \frac{T_x}{n_x}$$

Life table for *Balanus glandula* as constructed by Connell (1970) is given below :

Age (year) x x	**Observed number alive** n_x	l_x	d_x	q_x	L_x	T_x	e_x
0	142	1.000	80	0.563	102	224	1.58
1	62	0.437	28	0.452	48	122	1.97
2	34	0.239	14	0.412	27	74	2.18
3	20	0.141	(4.5)	0.225	17.75	47	2.35
4	(15.5)*	0.109	(4.5)	0.290	13.25	29.25	1.89
5	11	0.077	(4.5)	0.409	8.75	16	1.45
6	(6.5)*	0.046	(4.5)	0.692	4.25	7.25	1.12
7	2	0.014	0	0.000	2	3	1.50
8	2	0.014	2	1.000	1	1	0.50
9	0	0	-	-	0	0	0

In the above table any individual of 8 year age is expected to survive only ½ year more so there will be no individual alive next year.

FERTILITY SCHEDULE/ FERTILITY TABLE :

Like life table, fertility table can also be constructed. It is a convenient format for describing the fertility or fecundity (fecundity schedule) of a population. It is an age specific summary of birth in a population. This table gives an account of the number of females produced per unit of time per female aged x (b_x) or as individual fecundity i.e. the mean number of eggs produced per surviving female (m_x). A hypothetical example is given below :

* Estimated number alive. Actual observation in field could not be made.

Age x	l_x	**No. of births/ eggs(m_x or b_x)**	V_x **(Product of l_x & m_x)** $l_x . m_x$
0	1.00	0	0
1	0.80	0	0
2	0.60	1	0.6
3	0.40	2	0.8
4	0.20	2	0.4
5	0.10	2	0.4
6	0.05	0	0
7	0.00	0	0

From this table 'Net reproductive rate' or 'Basic reproductive rate (R_0) can be calculated as –

$$R_0 = \sum_0^\infty l_x . m_x = \sum_0^\infty V_x$$

In this hypothetical example R_o = 1.9

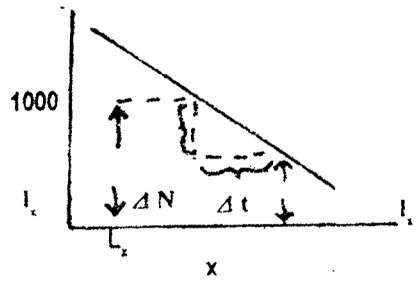
Net reproductive rate is the multiplication rate per generation. It can also be defined as number of female offsprings that replace each female of previous generation. In a stable population R_o would be '1'.

SURVIVORSHIP CURVES

Crude death rate gives very little information. For example mortality among old adults will affect a population very differently from mortality among juveniles. Consequently an age specific death rate is a more specific statistics than the crude death rate.

Ecologists are primarily interested in organisms that survive rather than those die. As such mortality is commonly expressed in the form of survivorship curve. These curves indicate the number of survivors out of a 'cohort' after a specific time interval. A group of individuals of the same age is called a cohort.

Diminishment of a cohort through time can be shown by a curve known as survivorship curve. The number in the cohort remaining alive (l) at any particular age (x) is symbolized as l_x. Thus survivorship curves are also called l_x curves. These may be constructed by plotting l_x values of the life table in y axis and age in the x axis.



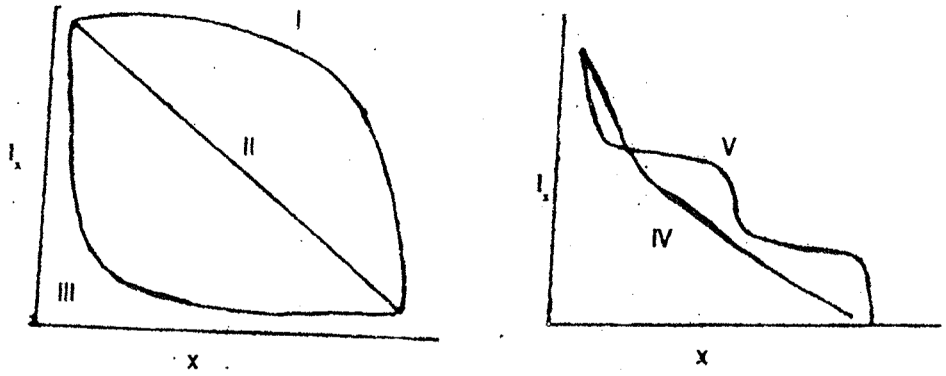
Where ΔN = number of animals dying.

x = age.

Such curves may be plotted either in arithmetic scale or in logarithmic scale.

Types of survivorship curves :

There may be infinite variety of possible survivorship curves. Three general types (type I, II, III) of curves were recognised by Raymond Pearl (1928). Two more types, type IV and type V were recognised by Slobodkin (1962) and Ito (1961) respectively. These are the five main types of survivorship curves.



(1) **Type I, Right angle Curve or convex type :**

Individuals survive for almost equal length of time. Juvenile mortality is little. Individuals die at old age. Rarely encountered in natural population.

Examples : Laboratory culture of starved fruit flies. Human population of industrialized society. Natural population of Dall mountain sheep (*Ovis dalli dalli*) and some rotifers (*Porales decepiens*).

(2) **Type II, Theoretical curve or Linear Type :**

Mortality occurs at a constant rate throughout the period of life span of individuals in the population. Rather uncommon in natural population.

Example : *Hydra fusca*, Many birds (British Robin), some rodents.

(3) **Type III or concave type :**

Mortality is very high during the young or juvenile stages, particularly in eggs and free swimming larval stages. Survivorship is high for remaining individuals. This curve is typical for organisms producing large number of young with low parental investment.

Example : Many marine invertebrate population such as oyster. Most teleost. Most parasites.

(4) **Type IV or slightly sigmoid type :**

Mortality is high among young, low and almost constant in the adults. When plotted in logarithmic scale a straight line is formed.

Example : Seen in many birds (e.g. *Melospiza melodia*), mice and rabbits.

(5) **Type V or stairstep type :**

Survival rates sharply differ from one life history stage to other. Alternate high and low survivorship forms a stairstep like l_x curve.

Example : Holometabolous insects (Butter flies).

Remark :

Many intermediate types of curves are also seen. For example wild *Drosophila* shows a curve intermediate between type I and type II.

Factor affecting shape of survivorship curve

(1) **Parental care :**

Ito (1959) pointed out that the shape of survivorship curve is related to the degree of parental care or protection given to the young. Thus survivorship curves are much less concave in those organisms which protect their young (e.g. honey bee, robin etc.) than those which do not protect their young (e.g grass hopper, sardine fish etc.).

(2) **Density :**

The shape of the curve may vary with the density of the population. Taber and Dasman (1957) found that in "black tailed deer' the curve of the denser population is more concave.

(3) **Sex :**

Survivorship may differ for males and females of the same species (e.g. red deer). In human beings females have a greater convex curve as compared to males.

(4) **Race and mutants :**

Shape may also vary from one race to another as well as among different mutants. In *Drosophila* 'speck' variety has type I curve and 'vestigeal wing' mutant has type II curve whereas the wild type has an intermediate curve between type I and type II (Bodenheimer, 1958).

POPULATION REGULATION:

No animal population increases indefinitely Eventually it arrives at some state of equilibrium about which it fluctuates. Population regulation is under the influence of various kinds of factors.

Factors regulating population density :

1. **Density independent influences :**

Intensity of action is independent of the population size. Effect is constant regardless of the population density e.g. climatic factors.

2. **Density dependent factors :-**

The intensity of action varies with the density of population. Intensity of effect increases with the increase in population density and *vice –versa* e.g. competition for food.

Insect ecologist considered density dependent influences to be more important but vertebrate ecologist thought that density dependent factors play more important role. In nature it is difficult, if not impossible, to separate the effect of one from another. Now ecologist prefer to classify the factors as extrinsic & intrinsic influences.

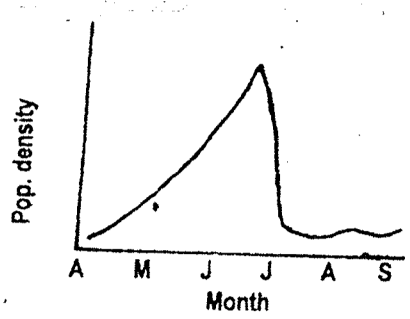
(I) **Extrinsic influences :**

Weather, resource availability, predation, disease & human activities are important extrinsic factors regulating population size.

Weather :

1. Most powerful, particularly extremes of temperature and moisture.
2. Influence population by affecting growth, maturation, reproduction, survival, movement dispersal & by elimination.
3. In general the influence tends to be irregular & unpredictable.

Example :



i) In California when temperature exceeds 100°F for several days aphid population in walnut orchard decline sharply to form 'J' shaped curve. The population fails to resurge after temperature declines (Sluss, 1967).

ii) Severity of winter and heavy snowfall result in sharp decline in Adirondack mountain deer (in USA) due to high mortality, particularly of fawns 5 out of 6 fawns die. Decline in reproduction rate also occurs.

iii) Rainfall affects reproductive success in *Dipodomys merriami* which becomes reproductively active during January and February when plants grow as a result of rainfall. These plants provide nutrients and vitamin for pregnant & lactating females. If there is no rainfall then there will be no re production.

2. **Resource availability :**

Abundance or otherwise of resource (food, space etc.) often regulate population size largely through intra specific competition.

Nicholson (1954) performed a series of experiments with sheep blowfly (*Lucilia cuprina*) and found that when unlimited supply of food was given scramble for food and wastage of food took place and population oscillated violently. When limited supply of food was given then there was contest type of competition population fluctuated less violently and mean population increased by four times.

In nature situation may be quite different. Poor nutrition causes repression in growth, delayed sexual maturity, lower conception rate and increased intrauterine and fown mortality. Consequently there is slowing down of the growth rate. High quantity & quality of food on the other hand cause increased fertility, low mortality and increases growth rate. Shortage of food forces herbivores to migrate frome one place to other. Herbivores are followed by carnivores.

Shortage of food also results into delayed reproduction & death due to starvation in carnivores. Rusch (1972) found that when the population of snow shoe hare (*Lepus americanus*) was high 100% of the owl (great horned owl, *Bubo virginiana*) nested but when population of prey was low only 20% the owl nested.

Faced with same shortage of prey canada lynx (*Lynx canadensis*) in the same area continued reproduction but the young died of starvation.

4. **Disease & parasites :**

Disease becomes an important regulating mechanism and reaches epidemic level when the density of population is high. Increase in population density results into decline in defense mechanism (antibody formation declines) and susceptibility to disease increases.

Disease thus may be the consequence of the high population density rather than the cause of population decline.

High density – High mortality due to disease.

e.g. Ranikhat disease in poultry. Rabies in wild mammals.

Disease may have a density independent influence when it is introduced into a population lacking any resistance (e.g. Bird malaria in Hawaii, USD of fish) or where it results from environmental changes e.g. radisotopes, arsenic, heavy metals etc.

Parasites may have a density depended effect. Varley's (1947) study on Knapweed gall fly reveated that *Eurytoma curta* killed more gall fly when population density was high. Thus action was density dependant.

Year/ density	Initial larval population	Larvae killed by parasite	% of population killed
1934, Low	43/m^2	6	14
1935, High	148/m^2	66	45

(5) Predation :

Predator population keeps a check on the prey population and on its part is regulated by the prey population. Both the populations maintain an equilibrium. (For further informations see Module No. 27). Example : Deer & tiger population.

(6) Human activites :

In the modern era human activity has become an important regulatory influence on the density and diversity of animal populations. What ever man does to environment has some effect or other on the animal population whether intentional or non intentional. Some of such important anthropogenic activities are deforestation, afforestation, cultivation, war, industrialization, urbanization and hunting.

i) Destruction of forest causes disappearance of forest animals and birds (e.g Woodpecker) and increase in number of open field species (e.g. sparrow).

ii) Plantation has just the converse effect.

iii) Cultivation increases desired plant species and dependent animal species such as pests and their predators. Use of pesticide not only causes mortality of target species but also eliminates non target species through biological magnification of pesticide and non availability of food in case of predators.

iv) War is another human interference which adversely affects wild population. Marine biota of persian gulf was severely affected during the gulf war.

v) Urbanization and industrialization cause decrease in the density and diversity of most animal groups but some forms are favourably affected. For example rat and crow population increase due to urbanization. Likewise *Tubifex* and *Chironomus* larvae population increase in polluted water.

vi) Ruthless hunting by man has caused decline as well as total extermination of many species. For example extermination of Caribou from Labrador due to hunting has caused disappearance of wolves and Red Indians. Similarly passenger pigeon with an estimated population of three thousand million in 1850 completely disappeared by 1913 due to ruthless hunting.

II. **Intrinsic influences :**

Some important intrinisic population regulating mechanisms are – crowding, stress, behaviour, dispersal and genetic one. Some species can also consiously self regulate their own population.

1. **Crowding, stress & behaviours**

Crowding causes shrinkage of resource & space resulting into social stress and influences behaviours. Social stresses act on the individuals through physiological feed back involving endocrine system particularly adreno pituitary system. Stress affects physiology, growth & behaviour which ultimately affects population growth. Increased population of mice in laboratory results into :

(i) Curtailment of reproductive functions.

(ii) Delay in sexual maturity in both sexes.

(iii) In some cases no females reached normal sexual maturity.

(iv) Increased foetal mortality particularly in sub-ordinate females.

(v) Such effects may continue upto the next generation.

(vi) Ultimate effect is lower birth rate and lower population growth rate.

Myers et al. (1971) studied rabbits (*Oryctologus cuniculus*) by confining them in different densities in natural habitats and found that it causes :

i) Increased aggression.

ii) Decrease in reporduction and in rate of ovulation.

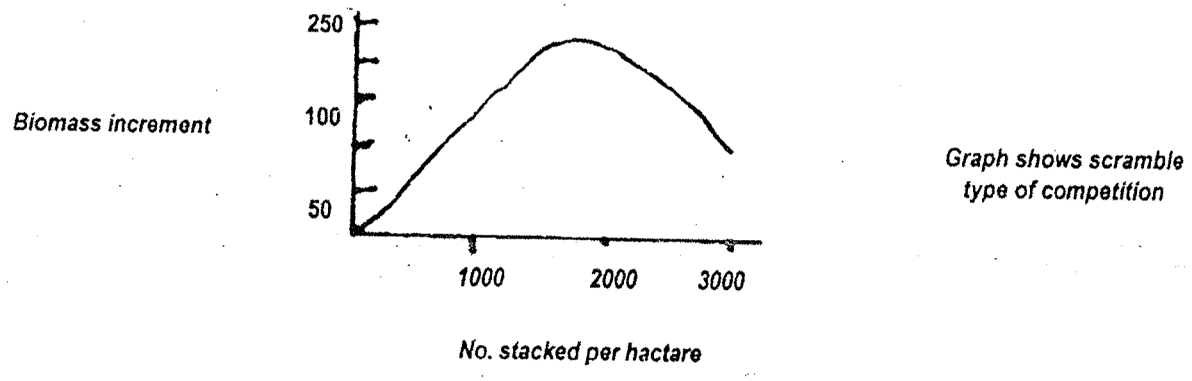
iii) Histological signs of stress were evident in individuals such as

a) Enlargement of adrenal.

b) Weight of liver & spleen decreased.

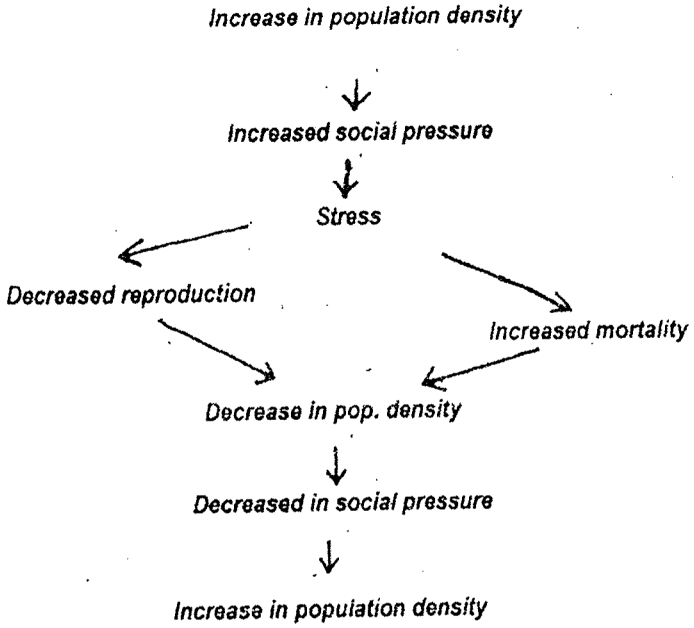
c) Inflammation of kidney.

Decline in growth due to crowding has also been observed in fish population (Backiel & LeCren, 1967).



Stress hypothesis of population regulation (Christian's, 1963):

The hypothesis states that the stress acts as a self regulatory mechanisms of population regulation as shown below :



Social behaviours of animals regulates number of individuals in an area through (I) territoriality & (ii) social dominance by increasing social stress & mortality. (See Module No. 29 for further informations).

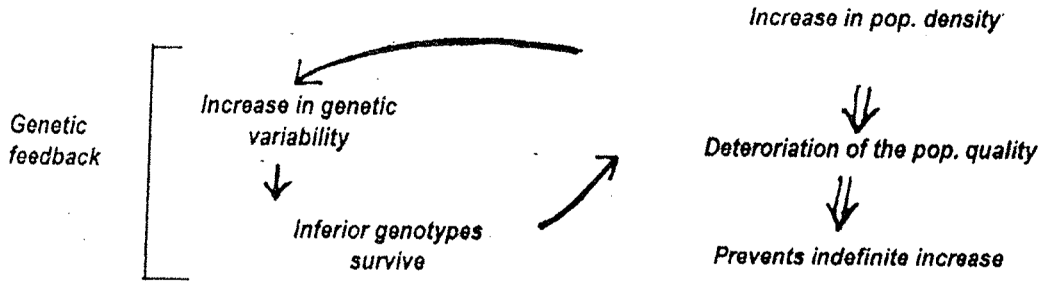
Dispersal

From ecological point of view dispersal may be classified as density dependent (*saturation dispersal,* after reaching carrying capacity sub-ordinates & juveniles move away) or density independent (*pre saturation dispersal,* healthy population move away before the carrying capacity is reached). This is more important in population regulation. *Social hierarchy & territoriality* can also force low ranking individuals to disperse. Dispersal introduces new genotype and increases reproductive potential in invaded areas.

Genetic cause :

Many animals die to unknown causes even though the environment appears to be favourable. Chitty (1960) working on voles (mice) and Wellington (1960, '64) working on forest tent catterpiltar (*Malacosoma pluviale*) concluded that when population increases in density, quality of population deteriorates which prevents indefinite increase through a *genetic feed back* involving genetic change within the population. Population increase brings about increase in genetic variability in the population and many inferior genotypes survive.

Wellington (1960, 1964) found that when population of *Malacosoma* increases, active adult females leave the area to lay eggs leaving behind sluggish females which are easily subjected to ill effects of environment. In an experiment Carson (1968) introduced a wild type male into a population of inbred



Drosophila female. This small change in the gene pool resulted into a population explosion which increased by 3 times and ultimately crashed abruptly. The cause seems to be a genetic one.

Self regulation :

Human population is regulated by reducing reproductively active phase of females by late marriage. After marriage by adopting various birth control measures the population growth is further self regulated. Self regulation of population is also common in animal kingdom. For example Great horned owl delays reproduction when there is shortage of prey (snow shoe hare).

Review questions:

1. Distinguish between:
   a) Major and minor community.
   b) Diversity index and evenness index.
   c) Ecotone and edge.
   d) Crude death rate and age specific death rate.
   e) Extrinsic and intrinsic population regulatory influences.
2. Write short notes on :
   a) Relative abundance.
   b) Dominance index.
   c) Keystone species.

d) Species diversity.

e) Edge species.

f) Fertility schedule.

g) Stress as a population regulatory factor.

h) Man as regulatory influences of the population density of animals.

i) Biotic community.

j) Relation between 'S' & 'N'.

k) Low diversity areas.

3. What is life table? Discuss various steps in the construction of a life table. Add a note on the significance of life table.

4. What is species diversity? Explain the advantage of evenness index over diversity index. Discuss important ecological principles associated with species diversity concept.

5. Define survivorship curve. Explain different types of survivorship curves with the help of suitable example. Add a note on the major factors affecting the shape of survivorship curve.

6. What do you understand by population regulation? Give a brief account of the population regulation under the influence of extrinsic/ intrinsic influences.

7. What do you understand by dominance of a species? How dominance may be ascertained in a biotic community? Discuss relationship between dominance index and quality of environment.

8. Give a brief account of the types of edges and ecotones. Explain edge effect. What is edge index?

Suggested Reading :

1. Ecology by C.J. Krebs. Harper International Edition. Harper and Row Publishers, New York.

2. Ecology & Field Biology by R.L. Smith 5th Ed. 1996 Harper Collins College Publishers, New York.

3. Basic Ecology by E.P.Odum. Saunders College Publishing, New York.

4. Concepts in Ecology by E.J. Kormondy. Prentice – Hall of India.

.....

SPECIES DIVERSITY HYPOTHESIS

1. *Introduction :*

Why should some localities have more species than other localities ? Why should tropical areas have more diversity than other areas and why should there be variation in species diversity in relation to latitudes and altitudes ? These are some of the questions which ecologists are trying to answer. In doing so they have generated a number of hypothesis. Some important hypothesis relating to species diversity are mentioned below :

2. *Various hypothesis :*

2.1. *Evolutionary time hypothesis (Fischer, 1960, Simpson, 1964)*

According to this hypothesis species diversity is related to the age of the community. Older communities (in an evolutionary sense) hold greater diversity than younger communities. Thus the diversity increases with the age of the community. Tropical communities are older as such have greater diversity as compared to the temperate communities.

2.2. *Ecological time hypothesis :*

According to this hypothesis new habitats are invaded by few species due to less time available for invading, dispersal and settlement by species. Enough time has not passed since glacier periods in temparate regions to be invaded and saturated by species from tropical areas.

Remarks :

Above mentioned both the hypothesis are related to time, first one in a longer time scale and second one in a shorter time scale.

2.3. *Spatial heterogeneity hypothesis (Simpson, 1964).:*

This hypothesis holds that more complex and heterogenous is the physical environment more complex will be its fauna and flora. Thus more types of habitats the area has (i.e. more is the β diversity) more types of species will be there (α diversity).

2.4. *Climatic stability hypothesis (Fischer, 1960, Connell and Orias, 1964)*

According to this hypothesis diversity is related to the climatic stability. There are more types of organisms in stable and constant environment with little seasonal or diel fluctuations in aboitic

conditions present a favourable environment to species as compared to an unstable environment. Tropical regions are more equitable as compared to temperate regions and as such have greater diversity.

2.5. *Climatic predictability hypothesis :*

This states that species diversity could be related to the precdictability of the climate. In precditable climate species diversity is greater in unpredictable variations in climate it is lower. In tropical regions climate is more predictable in comparison to temperate and polar regions. Species also respond to predictable variations (such as bird migration) and cause temporal variation in diversity.

Remarks :

Above mentional two hypothesis (2.4 and 2.5) are in relation to climate. Sanders (1968) modified climatic stability hypothesis to stability time hypothesis.

2.6. *Stability time hypothesis*

This hypothesis assumes that two cantrasting type of community exist — the physically controlled and the biologically controlled. In the former density independant physical factors like climate, temperature, rainfall etc. control population which are subjected to greater stress. In biologically controlled communities density dependant factor play more important role in population regulation (e.g. predation, parasitism, competition for etc.) since physical factors are comparatively stable and predictable. As a consequence species diversity is more in biologically controlled communities than in physically controlled communities.

2.7. *Energy hypothesis (Brown, 1981, Curie, 1991) :*

This hypothesis holds that species diversity is related to energy flux per unit area. Species diversity will be greater in regions receiving more energy per unit area. Primary production is more in areas with higher energy flux. Thus latitudinal variations in energy flux influences species diversity. Higher latitudes have lower diversity.

2.8. *Productivity hypothesis (Connell and Orias 1964)*

This hypothesis states that the level of diversity is determined by the amount of energy flow entering in production. When more energy is available for production and less is required for maintenance more will be diversity. Thus diversity is related to R/B ratio or 'ecological turnover'. When R/B ratio is > 1 diversity will decrease when R/B ratio is < 1 diversity will increase. Since

ecological turnover is dependant on climatic stability this hypothesis is also related to climatic stability.

Remarks :

The two aforesaid hypothesis (2.7 and 2.8) are related to energy.

2.9. Competetion hypothesis (Dobshansky, 1951, Williams, 1964)

Under competitive pressure species occupy narrow ecological niche. They search new ecological niche which permit enhancement of diversity. More there will be ecological niche for occupation more will be species diversity.

2.10. Predation hypothesis (Paine, 1966)

Predation reduces number of prey individuals which reduces interspecific competetion and permits existence of more types of prey species by eliminating competitive exclusion. It has been experimentally shown that when a keystone predator is removed from an area then number of prey species decreases. Thus predation increases species diversity.

2.11. Dynamic equilibrium hypothesis (Huston, 1979)

Disturbance in community decreases diversity. In the absence of disturbance competitive species grow and achieve dynamic equilibrium which increases diversity. For example forest fire or any other type of perturbance reduces species diversity.

3. Conclusion :

Although many hypothesis have been put forth these are difficult to test experimentally. Each hypothesis has its own merit and demerit and no single hypothesis can provide a definitive answer to the problem the solution of which largely remains elusive.

II. SPECIES ABUNDANCE HYPOTHESIS :

1. Introduction :

In most of the communities few species are common and most of the species are rare. Species diversity indices tell us little about species abundance and dominance. To determine species abundance we need to plot the proportion present into abundance classes. Theoretically, abundance classes represent the manner in which species divide the niche space or the environment. On a semilog graph paper species are arranged on the 'x' axis in sequence from most abundant to least abundant. On 'y'

axis relative importance of species are plotted. A trend line connecting the points results into the dominance-diversity curve as proposed by Whittaker (1965). Three statistical models or hypotheses are used to describe species abundance; the random niche boundary model, the niche preemption model and the log normal hypothesis (Fig. 1).

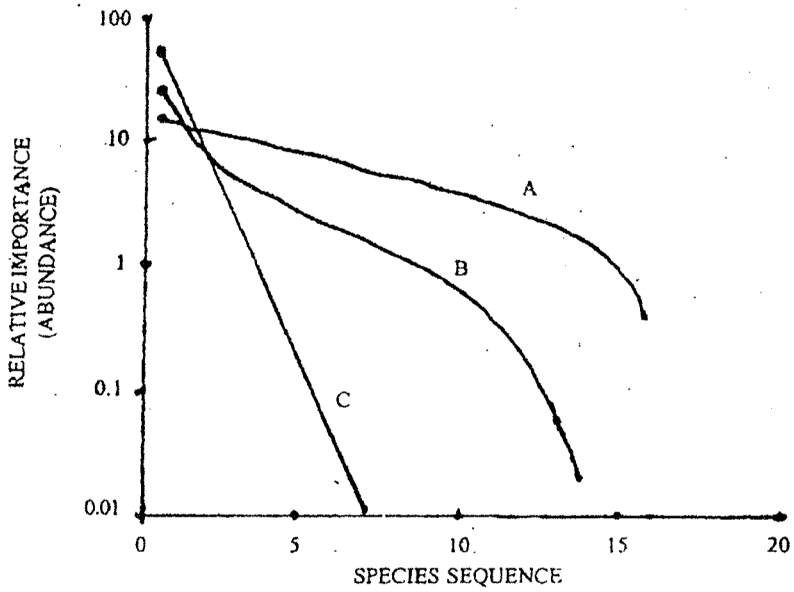


Figure 1 : Dominance - Diversity curve

2.1. *The random niche hypothesis or Broken stick model of Mac Arthur (1960) :*

This hypothesis views abundance as a random partitioning of resources distributed along a continuum. According to this model niche space is divided into random, contiguous, nonoverlapping segments. Length of each segment represents abundance of a species. It results into a broken stick type of curve (curve A). It produces the highest evenness of the three models. This type is rarely seen. Groups exhibiting interspecific competition and territorial behaviour of homogenous community with least stress show this type of curve. For example nesting bird in a forest.

2.2. *The niche preemption hypothesis or geometric distribution hypothesis*

According to this hypothesis the most abundant species is twice as numerous as the next most abundant one which in turn has twice the abundance of the third one. From this one might assume that the most dominant species preempts half the available niche space and second species half of the

remaining space (e.i. 25%) and so on. There is no overlapping of space. This results into a straight line (curve C) and represents lowest evenness. Such distribution is achieved in communities found in harsh and severe environment such as deserts.

2.3. *The log normal hypothesis (Preston, 1962) :*

The above mentioned two curves represent extreme situations. In most natural populations an intermediate sigmoid type of curve (curve B) is seen indicating a more complex pattern of niche differentiation and niche overlap with partial competition. In such case species distribution is determined by large number of variables such as food, space, microhabitates etc. This is seen in relatively undistributed habitat.

A comparison of the three models are given in table.

Table : Comparison of three abundance hypotheses

Attribute	Random niche	Niche preemption	Log normal
1. Type of curve	Broken stick	Straight line	Sigmoid
2. Niche overlap	No overlap	No overlap	Overlapping
3. Niche partitioning	Random partitioning	Niche preemption	Niche differentiation
4. DiversityHighest	least	intermediate	
5. Evenness	Highest	least	intermediate
6. Dominance	Least dominate	One or few species	intermediate
7. Stress operating	Least	Maximum	intermediate
8. Competition	Intense competition competition	Most intense	Partial competition
9. Occurrence	Rare in nature Environment	In stressed and severe environment	Common in undisturbed

3. *Conclusion :*

From the three abundance model / hypothesis some generalizations can be arrived at :

1. Higher and flatter the curve, the higher is the overall diversity. Thus diversity of A > diversity of B > diversity of C.

2. The steeper the curve the lower will be the diversity and dominance of one species over other. Dominance A < B < C.
3. Dominance — diversity curve can be used to assess the effects of perturbation on species structure.
4. Stress, natural (severe weather) or anthropogenic (pollution) tends to steepen the curve.

III. REVIEW QUESTION :

1. Explain (α, β and γ diversity. Briefly discuss various species diversity hypothesis.
2. Briefly discuss niche preemption hypothesis. How does it differ from log normal and random niche hypothesis?
3. Write notes on :
   - i) Dominance - Diversity curve.
   - ii) Diversity hypothesis based on climate.
   - iii) Diversity hypothesis based on energy.
   - iv) Diversity hypothesis based on time.
   - v) Predation hypothesis.

IV. SUGGESTED READINGS :

1. *Ecology and Field Biology* : R. L. Smith. 5th Edition, 1996. Harper Collins, New York.
2. *Essentials of Ecology:* C. R. Twonsend, M. Begon, J. L. Harper. 2nd Edition, 2003. Blackwell Publishing. Malden, USA.
3. *Basic Ecology : E. P.* Odum, 1983. Saunders College Publishing. Philadelphia.

ORGANISMIC AND INDIVIDUALISTIC CONCEPT OF BIOTIC COMMUNITY

1. *Introduction :*

A biotic community is a naturally occurring, mutually sustaining and interacting assemblage of plants and animals living in the same environment, fixing, utilizing and transferring energy (R. L. Smith, 1980). Thus the species which make up a community are not merely occurring together but they do interact with each other. Such groups are natural and since found in the same area are related by geography.

There are two opposing views explaining the nature of the community : the *organismic concept* and the *individualistic concept.*

1.1. *The Organismic Concept :*

This concept was first put forth by the American plant ecologist F. E. Clements (1916). He viewed community as a *'superorganism'*. He regarded a community as a discrete, discontinuous, discernible and hence describable entity. According to this view community is the highest stage in the organization of the living world from cell to tissue, organ, organism and community. As such the community is comparable to a super organism, an integrated natural unit that has definable boundary — a definite entity Clement regarded final stage of succession, climax as a super organism.

This view was supported by earlier ecologists like S. A. Forbes, W. C. Allee, J. Braun — Blanquet and E. P. Odum among others. The organismic school dominated the ecological thoughts during the first half of the Twentieth century.

1.2. *The Individualistic Concept :*

The organismic concept was challenged by another American ecologist H. A. Gleason (1917) and subsequently supported by A. G. Tansley and R. H. Whittaker. Gleason argued that a community is far from being a distinct natural unit. It is a mere collection or association of species which require similar conditions for existence. Their adaptations to similar conditions enable them to live in a particular place. As far as interactions are concerned they do interact for survival but the presence or absence of any one species in a community is independent of all other species. In his view community is an association of convenience. It does not represent a natural unit and it has

no functional significance as a community. According to the individualistic concept as each species is distinct or individualistic similarly each community is also distinct from all others — it is an *individualistic community.* Species not the community is the highest grade of living organization. Community is not an integrated unit it is a loose collection of population not always with well definable boundary.

Clements believed that the species belong to a community were closely associated with one another and the ecological limits of distribution of each species coincided with the distribution of the community as a whole. This concept of community organization is referred to as a *closed community.* Gleason, on the other hand, believed that each species was distributed independently of others in a community. This concept of community organization is referred to as *open community.*

Thus while the organismic concept emphasises that community is a definite entity the individualistic concept considers it as an abstraction or idea. Although the individualistic concept was ignored till 1950 but now it is more universally accepted even though some modern ecologist still support organismic concept and the organismic / individualistic debate is far from settled (Kormondy, 1996).

A comparative idea of the two concepts are given is table 1.

Table 1 : A comparison of the two concepts of community

	Organismic Concept	Individualistic Concept
1.	Community is a discrete entity.	Community is not a distinct entity it is an abstraction.
2.	Community is an integrated natural unit with definable boundary.	It is an assemblage of population requiring similar conditions.
3.	Community is the highest stage in living organization.	Not so. Species is the highest state in living organization.
4.	Community is a super organism.	Community is an individualistic entity — an association of convenience.

II. GUILDS :

1. Introduction :

R. Root (1967) introduced the concept of *'Guild'* in ecology. He gave the following definition. "A Guild is defined as a group of species that exploit the same class of environmental resources in a similar way." In other words "Guilds are group of species that forage or feed similarly in a given habitat" (R. L. Smith, 1996).

A community can be viewed as a *complex assembly of guilds* containing one or more species which may be taxonomically related or not related and which may have little likeness of appearance. For example Kangaroo and sheep in Australia belong to the same guild although they have little likeness in appearance. Rabbit and bandicoot, on the other hand, although are similar in appearance belong to different guilds. Nectar feeding birds in a forest exploiting different flowers belong to a common guild.

Thus a community is a complex assembly of guilds containing one or more species. Therefore guilds are ecological units of great taxonomic diversity.

2. ***Relationship among the various terms used to describe group of organisms :***

At this stage it would not be out of place to discuss relationship of various terms used to describe group of organisms like Taxa, Community and Guild. These three sets of group are respectively related to by phylogeny, geography and resource (Fig. 1) :

i) *Taxa* are groups of organisms that are phylogenetically related.

ii) *Communities* are groups of organisms found in the same area i.e. are related by geography. When these populations are taxonomically related in the same geographic area these are called *assemblages.*

iii) *Guilds* are groups of similar resource use patterns i.e. these are related by resource. Members of a guild of same geographic area are called *'local guilds'*. When these are taxonomically related are regarded as *ensembles.*

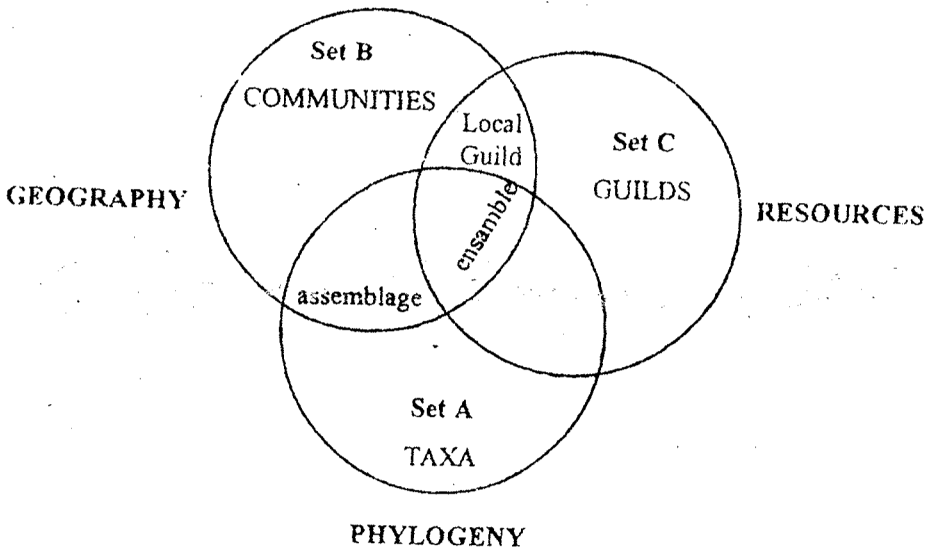


Figure 1 : Relationship between different types of groups

Guild denotes a functional attribute and thus may be confused with the term niche. While niche relates to the functional role of an individual species the guild relates to the functional role of a group of species. Both the

terms apply for the potential competitors or partitioners of same resource. The members of guild, however, exploit the resource together in a way that minimizes niche overlap and competition.

3. ***Some examples:***

Root in a collard stand *(Brassica oleracea)* found three herbivore guilds of insects.

i) Pit feeders :

Which rasp small pits on the leaf surface and have rasping type of mouth parts. 19 species of insects belonged to this group.

ii) Strip feeders

Insects with mandibulate mouth parts and feed by clearing holes in the leaf. 16 species of insects belonged to this guild.

iii) Sap feeders

Insects with piercing and sucking type of mouth parts which feed and sap. 58 species of insects belonged to this guild.

The species composition, of course, may change from year to year. For example cabbage aphid was the most abundant aphid in 1966 and 1968 but was absent in 1967, when other aphids increased in abundance. The implication is that within some guilds, species can replace one another and carry on the same functional role.

In a tropical forest there may be different kinds of guilds of insect eating birds. For example guild of bark insect feeders, feeding on insects present in barks. There may be a 'ant following guild' in which birds hunt insects fleeing the march of army ants. There may be another guild feeding on the insects present on leaves.

4. ***Conclusion :***

The concept of guild is new and little understood but three hypothesis can be set which require testing.

1. A community always has a group of guilds but the group may be packed with different number of species.
2. Many species form interchangeable members of a guild. These species are functional equivalents.
3. The number of guilds within a community is small in relation to the number of species.

The entire concept can be diagramatically shown as in Figure 2.

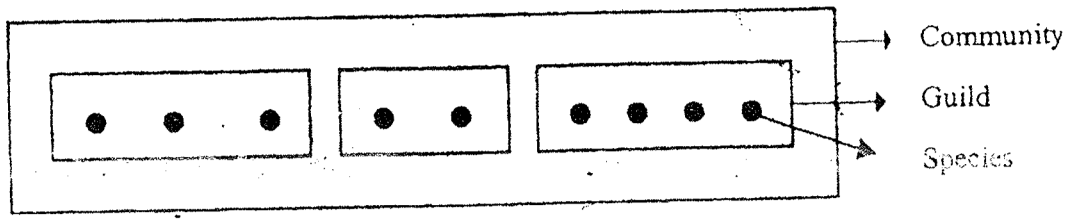


Figure 2 : Scehatic of Guild structure.

III. ECOLOGICAL EQUIVALENT:

Two species can not permanently occupy same or identical niche in the same community. However, two species can occupy same or similar ecological niche in different geographical areas. Such species are called ecological equivalents but these may not be taxonomically closely related. Similar ecosystems tend to develop in widely separated geographical regions which have similar climatic conditions. A grassland will develop wherever there is a grassland climate, but the species of grasses and grazers may be different. Some examples of mammalian ecological equivalents of grassland ecosystem are given in table 2.

Table 2 : Mammalian ecological equivalents

Ecological niche	North America	South America	Africa	Asia	Australia
1. Saltatorial herbivore	Jack rabbit	—		Jerboa	Red Kangaroo
2. Fossorial herbivores	Prairei dog		Golden Mole	Mole	Marsupial mole
3. Cursorial gregarious herbivores	Bison	Pampas deer	Zebra	Wild ass	
4. Cursorial predators	Coyote	Red wolf	Cheetah	Jackal	Tasmanion wolf

Among Cursorial herbivorous birds Ostrich of Africa, Rhea of South America and Emu of Australia are ecological equivalents.

IV. REVIEW QUESTIONS :

1. What do you understand by ecological guild ? Briefly discuss the concept of guild citing examples.
2. Define biotic community. Explain organismic and individualistic concepts of biotic community.
3. Write short notes :
   - i) Ecological equivalents
   - ii) Local guild
   - iii) Ensemble
   - iv) Clements idea about biotic community.

V. SUGGESTED READINGS :

1. *Concept of Ecological :* E. J. Kormondy, Fourth Edition, 1996. Prentice Hall of India Pvt. Ltd. New Delhi.

2. *Ecology :* R. E. Ricklefs and G. L. Miller. Fourth Edition, 200. W. H. Freeman and Company, New York.

3. *Ecology : S. C.* Kendeigh. Prentice Hall of India Pvt. Ltd. New Delhi, 1975.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Group-A Paper-III Unit-I
Module No. - 27

POPULATION INTERACTION

Two populations inhabiting any area does not mean that these are merely existing there. Either these require similar environmental conditions or/and there exists some interaction between themselves. Some possible types of population interactions are summerized in the following table —

Types of Interaction:

	Type of Interaction	Effect of Interaction		Result of Interaction
		On Population A	On Population B	
1.	Neutralism	0	0	Nil, neither population is effected.
2.	Amensalism	–	0	Unfavourable for one population which is inhibited, no effect on other
3.	Competition	–	–	Unfavourable for both the populations
4.	Parasitism	+	–	Favourable for A (parasite) unfavourable for B (host)
5.	Predation	+	–	Favourable for A (predator) unfavourable for B (prey)
6.	Commensalism	+	0	Favourable for A (commensal) but B is unaffected
7.	Protocooperation	+	+	Favourable for both but not obligatory
8.	Mutualism	+	+	Favourable for both and obligatory

0 = No significant interaction, + = Interaction is beneficial resulting into growth, – = Interaction is detrimental resulting inhibition.

1. Neutralism:

There is no apparent interaction between man and butterfly in an area. This is neutralism.

2. Ammensalism:

Chlorella vulgaris produces 'Chlorellin' which is toxic to other species of algae *Gymnodinium veneficium*, a dinoflagellate secretes a toxic substance which affects muscular and nervous system of fish and causes death of fish.

Inhabition of the growth of a plant species by another species of plant by secreting some chemical substance is called 'allelopathy'.

3. Competition:

In competition both the populations are adversely affected even though eventually one of the competiting population wins. Completion may be for some common resource. ***Example***: deer and antelope complete for food.

4. Parasitism:

It is an association between two species in which one species (parasite) utilizes other species as its habitat and source of food. Parasite is benefitted and host suffers.

5. Predation:

Predation is a step in the transfer of energy. Here one species (predator) kills and feed on the other species (prey). Obviously the predator is benefitted at the expense of prey. ***Example***: Tiger & deer.

6. Commensalism:

It is a one sided relationship between two populations in which one is benefitted (commensal) but the other one is neither benefitted nor harmed. Barnacles on the back of whales & king crab get a place for attachment and also obtain a free ride. *Remora* attaching to the bellies of sharks in addition to these feeds on the leftovers of shark's prey. Whale, king crab or shark remain unaffected.

7. Protocooperation:

Bird sitting on the back of rhino or deer and feeding on ectoparasites are examples of protocoperation since both the populations are benefitted but the interaction is not obligatory.

8. Mutualism:

It is an obligatory association between two species. Such relations are mutually beneficial to both the species involved. The algae living within the gastrodermal cells of hydra utilises CO_2 released by hydra for photosynthesis. 10% of the sugar is released to hydra for its use. *Trychonympha* living in the alimentary canal of termite digests cellulose for it.

COMPETITION

Competition can be defined as interaction between two organisms for some common resource usually in short supply. It is a negative type of interaction in which both the interacting organisms are harmfully affected.

Types of Competition:

Competition may be of two types —

1. Intraspecific competition:

Competition between individuals of the same species which limits the density of a species in a given area.

2. Interspecific competition:

Competition between members of different species. It can reduce the densities of competitive species as well as can cause exclusion of a species from an area.

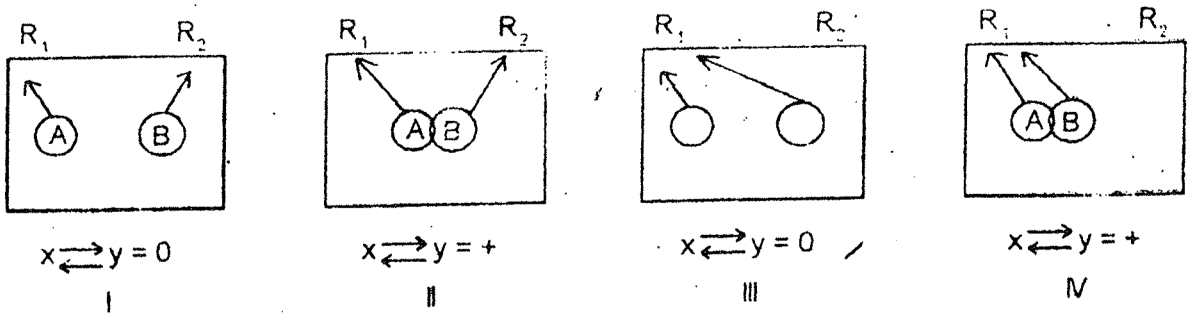
Competition may also be classified as —

1. Scramble type, in which all the individuals have access to the resource, scramble for the same and get a small share of it. The resource is divided among the competitors.

2. Contest type, in which there is a contest for the resource and the winnter takes the lion's share and losers are denied access to it.

Exploitative and interference types of competition:

Park (1962) suggested that the specific factors responsible for the elimination of species fall into two categories — **exploitation and interferance**. Exploitation refers to the utilization of common resources, of course with different efficiencies. Interference, on the other hand, refers to direct detrimental effects of one species on the growth, survival and erproduction of the other. Competition thus can also be classified as exploitative and interference types. Competitive interactions are most intense when both the processes are involved.



R_1 & R_2 = Resource

X & Y = Interaction between two species

O & + = No interference and presence of interference

A & B = Two species

In the above mentioned hypothetical examples

I — There is no competition since there is no exploitation of common resource and no interference.

II — Competition exists because of the presence of interference even in the absence of the exploitation of common resource.

III — Competition exists because of the exploitation of the common resource.

IV — Competition is intense because of both interference and exploitation of common resource.

MATHEMATICAL MODEL / LOTKA-VOLTERRA MODEL FOR INTERSPECIFIC COMPETITION

The idea of interspecific competition was first emphasized by Darwin (1859), Joseph Grinnell (1917) pointed out that two species can not occupy for long identical ecological niche. Volterra (1931) and Lotka (1934) developed mathematical model to explain interspecific competition and G.F. Gause (1934) conducted the first controlled experiments on interspecific competition.

Lotka-Volterra model:

Let us assume that there are two species '1' & '2' which are limited, by the resource availability, at the levels K_1 and K_2 respectively. Growth of the independent non competitive populations of these species may be represented by the logistic equations:

$$\frac{dN_1}{dt} = \gamma_1 N_1 \left(\frac{K_1 - N_1}{K_1} \right) \quad \text{..... (1) for species 1}$$

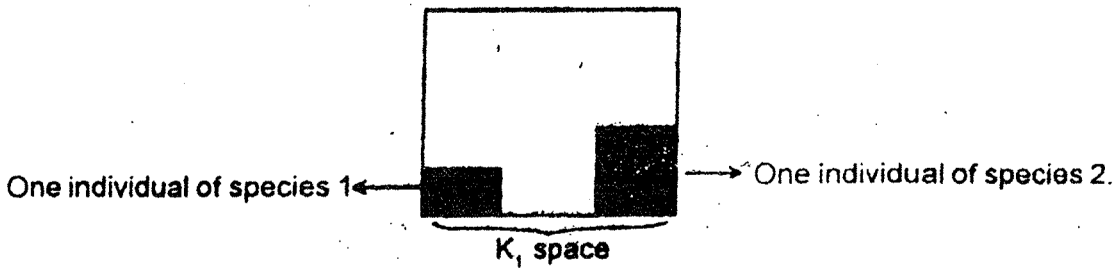
$$\frac{dN_2}{dt} = \gamma_2 N_2 \left(\frac{K_2 - N_2}{K_2} \right) \quad \text{..... (2) for species 2}$$

Where N_1 & N_2 are population size of species 1 & 2 respectively.

γ_1 & γ_2 are the intrinsic rate of growth for the two species.

Now if these two species occur together and show some degree of competitive interaction, the degree of realization of potential growth $\frac{K-N}{K}$ by one species (N_1) will be influenced not only by its own population density but also by that of its competitor (N_2). This influence would reduce the degree of realization of potential growth by a factor related to the population size of the competitor.

Since the individuals of both the species differ in body size and food intake ability therefore these will also differ in their competitive ability and thus their 'K' level will also be different.



Since space occupied by one individual species 1 is not equal to the same for species 2, more time and individuals would be required by species 1 to reach carrying capacity as compared to species 2. As N_1 is not equal to N_2 in its competitive ability we need a conversion factor (α) to convert the population densities of competitors into equivalent values:

$$N_1 = \alpha N_2$$

Therefore realization of growth rate for species 1 will be

$$\frac{dN_1}{dt} = \gamma_1 N_1 \left(\frac{K_1 - N_1 - \alpha N_2}{K_1} \right) \quad \text{..... (3)}$$

The same line of argument can be applied to species 2.

$$N_2 = \beta N_1$$

Therefore the logistic equation for the second species would be

$$\frac{dN_2}{dt} = \gamma_2 N_2 \left(\frac{K_2 - N_2 - \alpha N_1}{K_2} \right) \quad \text{..... (4)}$$

In these equation α & β are 'competitive coefficient' which serve to convert population densities of competiting species into equivalent values. α represents the inhibitory effect of one N_2 individual on the population growth of species 1 and β is the inhibitory effect of one N_1 individual on the population growth of species 2.

$\because N_1 = \alpha N_2$ and $N_2 = \beta N_2$

$\because N_1 = \alpha K_2$ and $K_2 = \beta K_1$

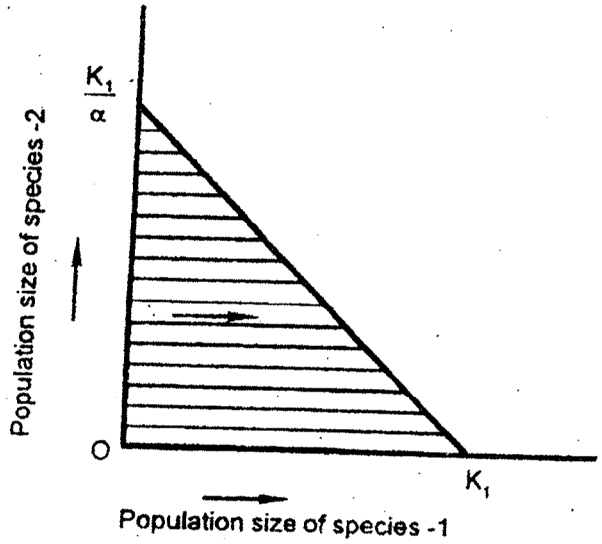
$\because \alpha = \frac{K_1}{K_2}$ and $\beta = \frac{K_2}{K_1}$

Thus, if species '1' had a K value of 200 and species '2' had a K value of 100 than α would be equal to 2 and β would be equal to 0.5. In effect this means that one individual of species 2 is equal to 2 individuals of species 1 in competitive influence.

When two species are competiting depending upon the competitive ability of the species both the species will be stabilized at one point and then

$$\frac{dN_1}{dt} = 0 = \frac{dN_2}{dt}$$

The situation can be depicted with the help following diagram.



All the species for species 1 will be used when there are K_1 individuals of species 1 (in the absence of species 2) or when there are K_1/α (or K_2) individuals of species 2 (in the absence of species 1). When both the species are present then the populations will increase in size until reach an equilibrium along the diagonal line, somewhere between K_1 and K_1/α.

According to the precise values of α and β the outcome of competition differs. The different outcomes depend upon the relative intensity of intraspecific and interspecific competition for each species. Interspecific competition is more intense when $\alpha > \frac{K_1}{K_2}$ and intraspecific competition is more intense when $\alpha > \frac{K_1}{K_2}$ for species 1. Based on this assumption four possible cases may be postulated:

1. $\alpha > \frac{K_1}{K_2}, \beta > \frac{K_2}{K_1}$ Interspecific competition more intense than intraspecific competition for both the species.

2. $\alpha < \frac{K_1}{K_2}, \beta < \frac{K_2}{K_1}$ Intraspecific competition more intense than interspecific competition for both the species.

3. $\alpha > \frac{K_1}{K_2}, \beta < \frac{K_2}{K_1}$ Intraspecific competition more intense for species 1 and interspecific competition more intense for species 2.

4. $\alpha < \frac{K_1}{K_2}, \beta > \frac{K_2}{K_1}$ Intraspecific competition more intense for species 1 and interspecific competition more intense for species 2.

Other factors influencing fate of competition:

Several other factors such as abiotic factors, parasitism, predation etc. influence the outcome of interspecific competition. The basic model has been further elaborated by many authors. For example Slobodkin (1963) has shown how nonselective predation may influence a competitive relationship. If a predator takes fraction 'm' of individuals, regardless of species per unit time. Such an action can be incorporated into the competitive model as—

$$\frac{dN_1}{dt} = \gamma_1 N_1 \left(\frac{K_1 - N_1 - \alpha N_2}{K_1} \right) - mN_1$$

$$\frac{dN_2}{dt} = \gamma_2 N_2 \left(\frac{K_2 - N_2 - \beta N_1}{K_2} \right) - mN_2$$

PREDATION:

Predation is a step in transfer of energy. In restricted sense it means members of one species (predator) eating those of other species (prey) and thus involves killing of the prey. But predation in its broadeast sense may be defined as one organism feeding on another living organism and thus include herbivory, carnivory, cannibalism and parasitoid.

LOTKA-VOLTERRA MODEL FOR PREDATION:

Lotka (1925) and Volterra (1926) independently developed mathematical model to express the relationship between predator and prey. The model involves two equations one for the predator and another for the prey. The equations are based on two suppositions — (1) Number of prey will increase in the absence of predator, and (2) Number of predator will decrease in the absence of prey.

In an unlimited environment population growth may be expressed as

$$\frac{dN}{dt} = \gamma N \qquad \text{..... (1)}$$

In the absence of predator (p) the prey population (N) is expected to increase

$$\frac{dN}{dt} = \gamma_1 N \qquad \text{..... (2)}$$

Where, N = prey density and

γ_1 = intrinsic rate of increase of prey in the absence of predator.

Similarly the predator population is expected to decrease in the absence of prey.

$$\frac{dP}{dt} = -\gamma_2 P \qquad \text{..... (3)}$$

Where, P = density of predator

γ_2 = instantaneous specific death rate of predator in the absence of prey

If the predator and prey are put together in the limited space the rate of increase of prey will be slowed down by a factor that depends on the density of predator.

$$\frac{dN}{dt} = (r_1 - \varepsilon P) N \quad \ldots\ldots\ldots (4)$$

Where ε is constant measuring ability of prey to escape predator.

Thus ' $-\varepsilon P$ ' denotes decrease in the prey density per individual of predator present which is a negative phenomenon.

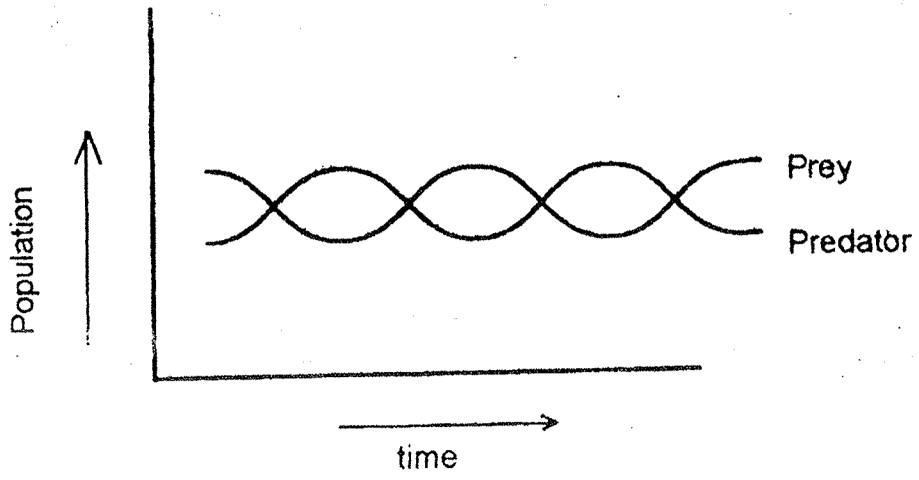
Similarly, in such a situation, the predator population will increase as a rate that depends on the density of the prey population.

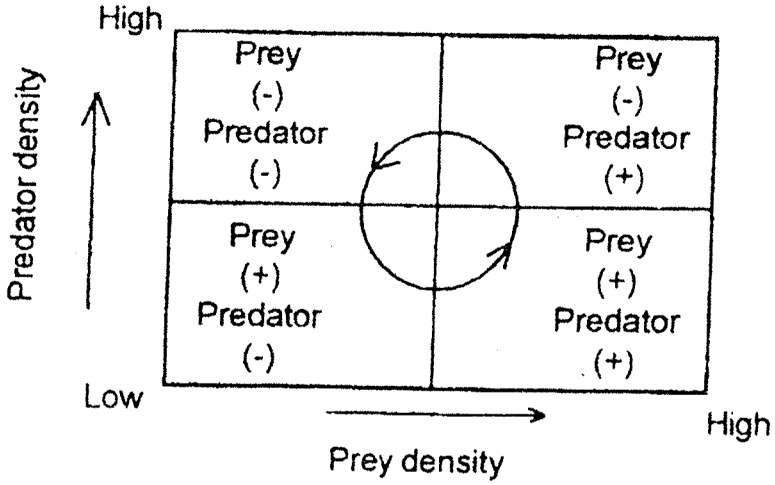
$$\frac{dP}{dt} = (-\gamma_2 + \theta N) P \quad \ldots\ldots\ldots (5)$$

Where θ is a measure of the skill of the predator in catching prey.

Thus θN is the increase in the predator population per individual of prey present.

The Lotka-Volterra model shows that as the perdator population increases the prey population decreases to a point which reversed the influence. The prey neither quite destroyed and the perdator never completely exterminates. Thus both the populations oscillate through time.





Limitations of the model:

The equations are based on certain false assumptions —

1. Both population have unlimited rate of growth. Neither prey not the predator population inhibits its own growth.
2. Environment is completely closed and homogeneous i.e., prey is randomly distributed and predator also moves randomly in the prey population.
3. Every prey has equal probability of being attacked and every encounter of a predator with prey results into capture of prey. There is no wasted attempt.
4. All responses are instantaneous with no time lag for handling or ingestion prey.
5. Age structure of both the population is unimportant. Young and old individuals are equally capable of capturing prey or escaping predator.
6. Density dependent mortality of predator or prey is unimportant.
7. Interaction of prey with its food has not been taken into consideration.

Responses of predators to the changing density of prey:

Predator prey interaction equations suggest two distinct responses of predators to changes in prey density

(1) Functional response:

As the prey density increases each predator may take more prey or take more frequently.

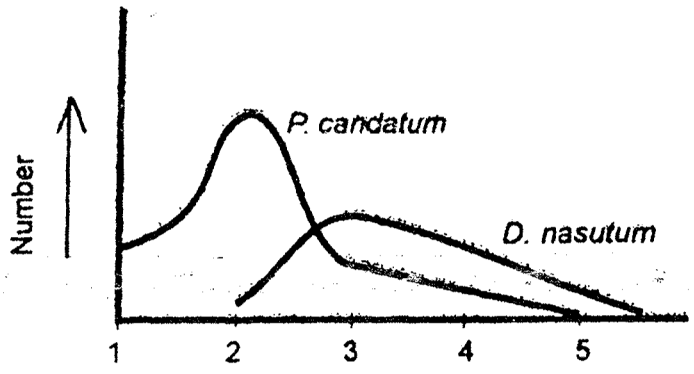
(2) Numerical response:

Predators may become more numerous through increased reproduction or immigration

Laboratory evidence in favour of predator prey interaction:

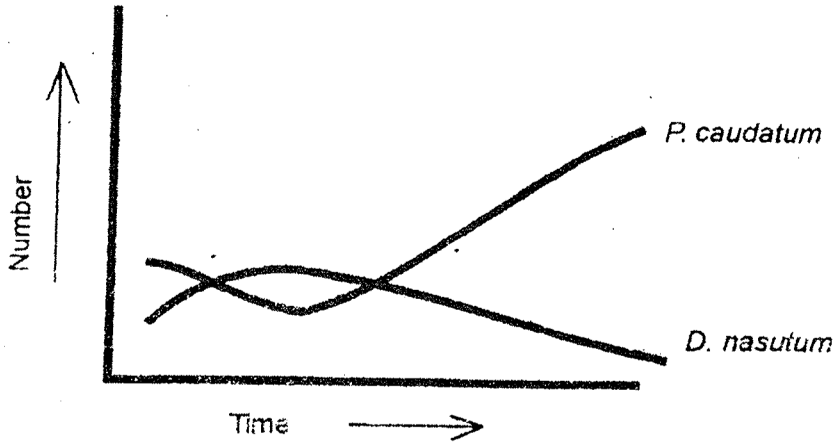
Gause (1934) performed a series of experiment involving *Paramecium caudatum* and its predator *Didinium nasutum*.

(1) In a culture medium containing oat meal and bacteria 5 *Paramecium* were introduced on first day and 3 *Didinium* were introduced on the second day. Both the populations exterminated after some days.



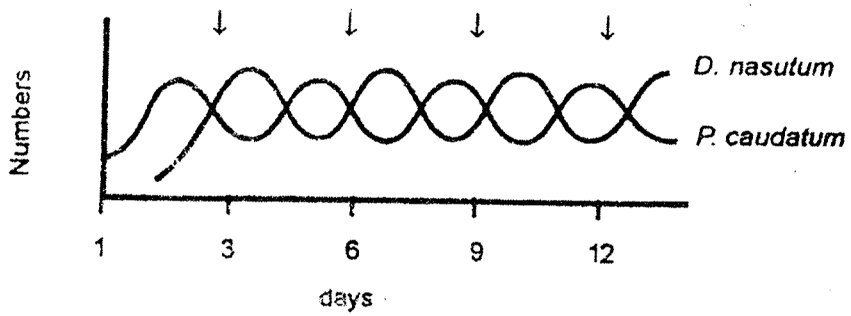
Oat meal without sediment

(2) When sediments were added to the culture, which provided refuge for *Paramecium*, *Didinium* was exterminated as it could not follow *Paramecium* which took refuge underneath the sediments due to its large size. Thus only *Paramecium* survived.



Oat meal with sediment

(3) However, when every third day one *Paramecium* and one Didinium were added to the culture, both the population showed desired oscillation.



Oat meal culture without sediment but with immigration

Thus Gause conduded that the Lotka-Volterra model operates only in the presence of immigration. Subequent experiments by Utida (1957), Huffaker (1958), Pimentai et al (1963) etc. have shown that oscillation may be obtained even without immigration and that a minimum level of prey population is required for a self sustained predator-prey relationship.

Evidence from nature:

Canada Lynx (*Lynx canadensis*) and snow showe hare (*Lepus americanus*) have peaked every nine to ten years for over past 200 years (Elton and Nicholson (1942). This is, however, debatable since the hare shows such peak even in the absence of lynx. Therefore, some other factors may also be involved in the process. Keith (1983) showed that food shortage in winter can also cause decline in hare population.

Coevolution of predator prey system:

As predators evolve more efficient ways of capturing prey, the prey also evolves ways of avoiding or countracting predation. Among mammals prey have evolved sharp sense of hearing and protective colouration and predators have evolved sharp sense of smell and offensive mechanisms.

ECOLOGICAL NICHE

Different species occupy distinct "ecological niche" in an ecosystem. Ecological niche is quite distinct from habitat". The habitat of an organism is the place where it lives. The ecological niche on the other hand includes not only the physical space occupied by the organism but also its functional role in the community and its position in environmental gradients of temperature, moisture, pH, soil and other conditions of existence. The species may be found in the same habitat but normally they occupy distinct ecological niche. Odum (1963) compared habitat with the 'address' and ecological niche with the 'profession'. The ecological niche is a particular combination of physical factors and biotic relations requierd by a species for the normal course of its life activities (Kendeigh, 1961).

Spatial and trophic niche:

Joseph Grinnell (1917) used the word ecological niche for the first time but he thought of the niche mostly in terms of the 'microhabitat' or what is now called as the **spatial niche** or **habitat niche** — the physical space occupied by the species. Charles Elton (1927) on the other hand used the term niche in the sense of the functional status of a species in its community. Since he placed emphasis on the energy relation (trophic relation) his version of the niche concept might be considered as **trophic niche**.

Multidimensional/hypervolume niche:

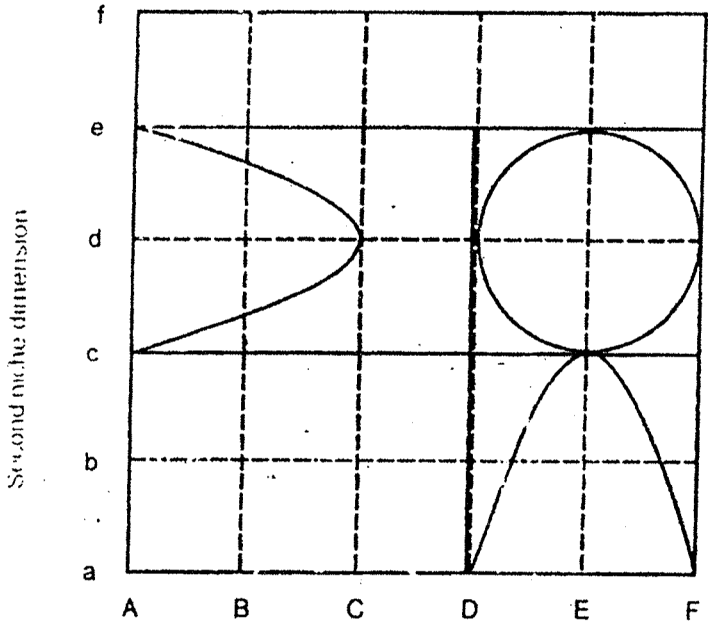
In 1927 G.E. Hutchinson suggested that the niche could be visualized as hypervolume within which the environment permits an organism to survive. This is the multidimentional niche concept which refers to the population of species in the environmental gradients.

Fundamental niche and realized niche:

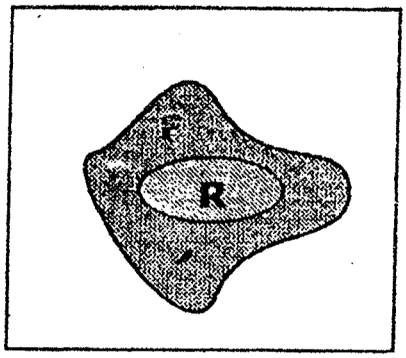
Hutchinson further made distinction between the **fundamental** and **realized** niche. Former is the maximum "abstractly inhabited hypervolume" when the species is not constrained by competition with the other species and

the latter is a smaler hypervolume occupied under biotic constraints. Fundamental niche is determined by environment gradient (climate) and biotic relations such as presence of food/prey, predator disease etc.

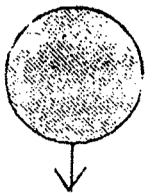
Consequently ecological niche of a species depends not only on where it lives but also on what it does (how it transforms energy, responds to or modifies the physical and biotic environment and also on how it is regulated by other species.



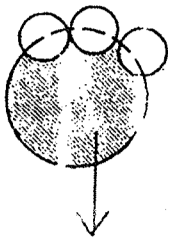
Niche space based on two gradients



Fundamental (F) and Realized (R) niche



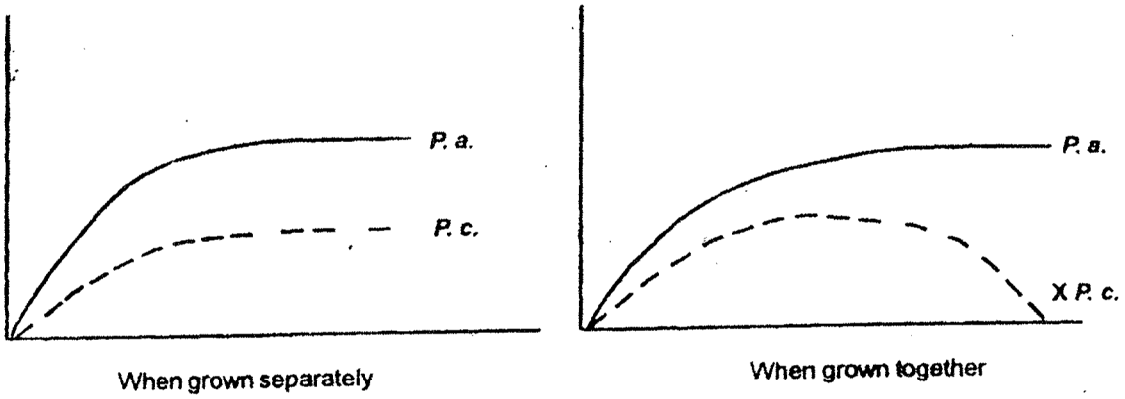
Fundamental niche of species 'A'



Realized niche of spaces 'A'

COMPETITIVE EXCLUSION:

Two species can not permanently occupy identical ecological niche. In doing so one species competitively excludes another species from that niche. This was first experimentally demonstrated by G.F. Gause (1934). Working on two species of *Paramacium, P. aurelia* and *P. caudatum* he showed that when both the species were grown separately, but under identical condition both the species survived *P. aurelia*, however, grew at a faster rate than *P. caudatum*. When these were cultured together they only *P. aurelia* survived and *P. candatum* was exterminated.

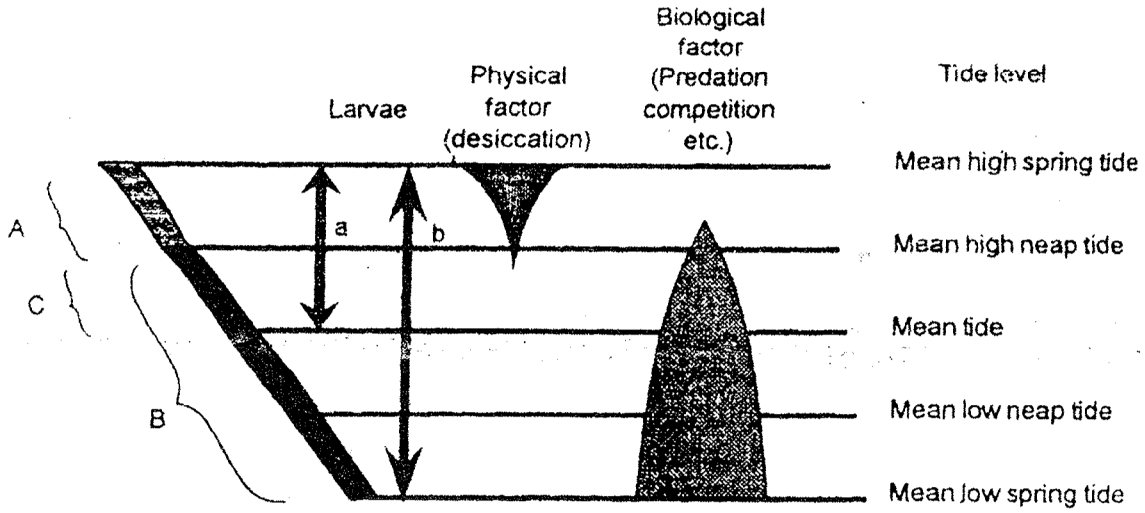


In a subsequent experiment he found when *P. caudatum* and *P. bursaria* were grown together both the species survived because of niche differentiation. *P. caudalum* consumed suspended yeast in the culture while *P. bursaria* restricted its feeding to those settled in the bottom facilitating coexistence. On the basis of these experiments Gause concluded that two species can not occupy exactly same ecological niche for long time. This is known as **Gause's principle** or **competitive exclusion principle** which may be stated as 'two complete competitors can not coexist (Hardin, 1960). Competitive exclusion could be avoided if environmental heterogeneity was provided. Many subsequent experimental evidences proved this principle. Evidences collected on the basis of field studies also provide supporting evidences in this regard.

David Luck (1945) observed that two related species of fish eating sea birds in Britain Cormorant (*Phalacrocorax carbo*) and Shag (*P. aristotelis*) apparently coexisted since both nest in the same cliff and take food from same water. A detailed probe, however, revealed that both occupied distinct ecological niche. Cormorant

nests high on cliff, fishes out on deep sea and is a bottom feeder feeding on mixed diet but no eels. Flounder and shrimp comprised 59% of the diet and clupeoid fishand eel comprised only 1% of the meal. Shag on the contrary nests low in cliff, fishes in shallow estuaries and is a surface feeder with a specialized diet eating mostly eels and clupeoid fish which constituted 82% of its diet whereas flounder & shrimps made only 3% of the meal.

Connell (1961) working in the intertidal zone in Scotland observed that *Chthamalus stellatus* inhabits upper part of intertidal zone and another barnacle *Balanus balanoides* inhabits the lower part of intertidal zone i.e. they occupied distinct ecological niche. *Balanus* grows a bit faster and taking an edge over *Chthamalus* can grow in the lower zone. Although *Chthamalus* can grow in the lower zone it is competitively excluded from that area. Larvae of both the species have a wider distribution. In the absence of *Balanus*, *Chthamalus* grew in the upper portion of 'Balanus zone' but *Balanus* did not grow in the upper portion of 'Chthamalus zone' even in absence of it. Physical factors limit upward distribution of *Balanus* and biological factors downward distribution of *Chthamalus*.



A = Chthamalus zone, adult *Chthamulus* found.

B = Balanus zone, adult Balanus found.

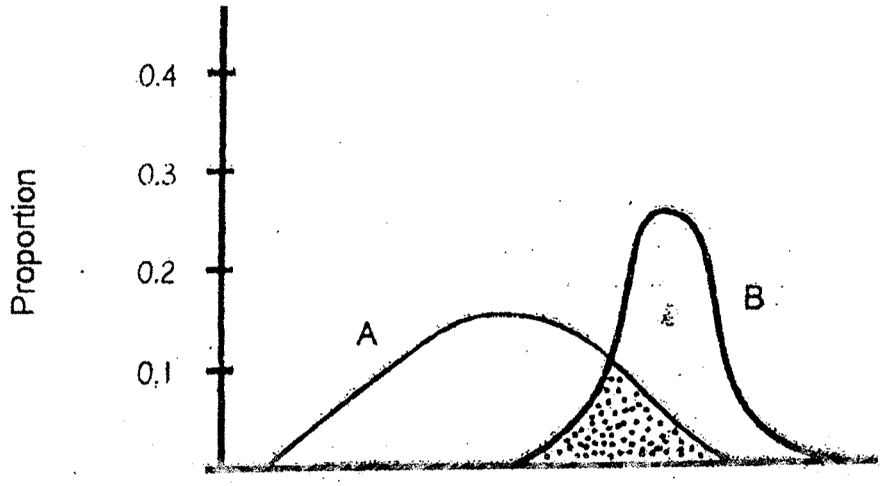
C = Chthamalus zone from where adult *Chthamalus* is competitively excluded.

a = Distribution of larvae of *Chthamalus*.

b = Distribution of larvae of *Balanus*.

NICHE WIDTH / NICHE BREADTH / NICHE SIZE:

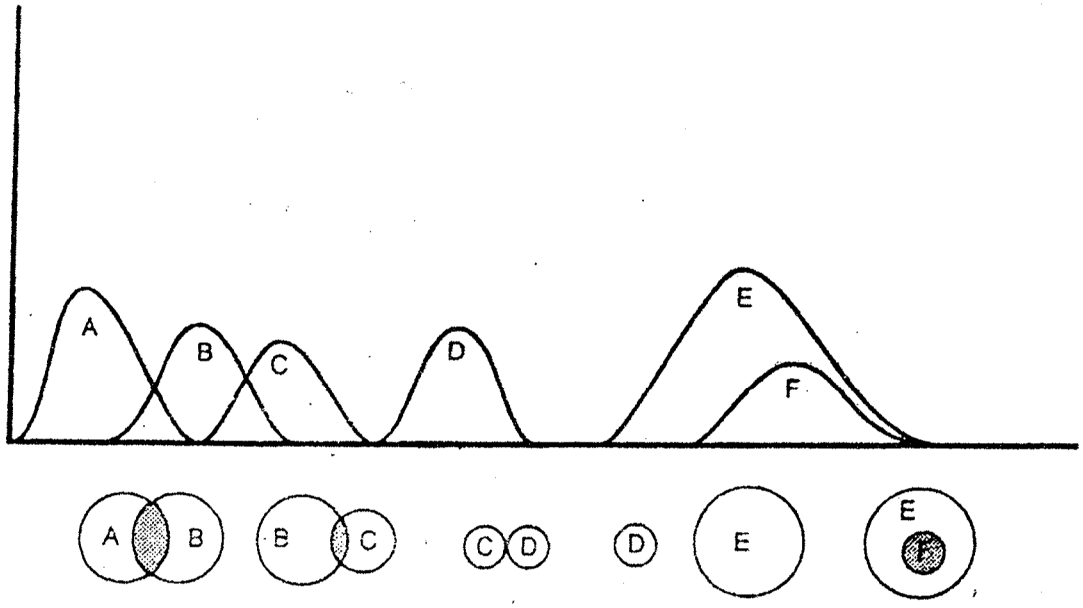
Niche width is the extent of the hypervolume representing the realized niche plotted in a resource gradient. Niche width is generally described as narrow and broad. This definition is derived from plots of niches on a resource axis. The width is measured by the length of the axis intercepted by the curve. Measurements of niche width usually involves some morphological trait such as bill size or some ecological variable as food size or habitat space or type of food (e.g. small fruit to large fruit). More precisely niche width is the sum total of variety of different resources exploited by an organism (Pianka, 1975). The wider the niche more generalized the species is considered, the narrower the niche, the more specialized is the species. Consequently a species may be classified as eurycoeious (wide range of ecological niche) and stenocoeious (narrow range of niche).



Hypothetical distribution of species with a broad niche (A) and a species with a narrow niche (B) on resource gradient. The niches overlap (shaded area). Species A overlaps species B more than species B overlaps species A.

NICHE OVERLAP:

Fundamental niches of the two species may overlap to some extent. Some niche spce may be shared. Niche may overlap with respect to one or few resources. Even niche of one species may be contained in that of another. Rarely do two species have exactly same niche requirement with total overlap. The following diagram shows different types of niche overlap.



Different types of niche relationships visualized as graphs on a resource gradient. Species A and B have overlapping niches of equal breadth but are competitive at opposite ends of the resource gradient. B and C overlapping niches of unequal breadth. Species C shres a greater proportion of its niche with B than B does with C. (In this example, however, B shares its niche also with A at the other end of its resource niche). C and D occupy adjacent niches with little possibility of competition. D and E occupy disjunct niches with no overlap and no competition. Species E has a niche contained within the niche of E. If F is superior to E competitively, it persists and E shares that part of its niche with E.

Review questions:

1. Deduce Lotka-Volterra equation for competition. What would be the types of competition if $\alpha > K_1 / K_2$ and $\beta < K_2 / K_1$? How would you indicate effect of non selective predtion in competitive interaction?

2. "Complete competitors cannot coexist". Elaborate the statement citing one experimental and one field evidence. Under what situations two species may coexist.

3. Differentiate between habitat and niche Explain multidimensional niche concept.

4. Explain Lotka-Volterra equation for predator-prey interaction. Discuss limitations of this model. Cite an experimental proof in favour of predator-prey equilibrium.

5. What is niche width? Can niche width affect the success of a species? What is niche overlap? Mention different possible types of niche overlap.

6. Discuss different types of population interactions with species reference to negative interactions.

7. Distinguish between

   (i) Protocooperation and mutualism.

   (ii) Parasitism and predation.

   (iii) Spatical niche and trophic niche.

   (iv) Fundamental niche and realized.

   (v) Exploitative and interference competition.

   (vi) Scramble and contest type of competition.

   (vii) Functional and numerical response of predator.

8. Write notes on

(i) Competitive coefficient.

(ii) Amensalism.

(iii) Commensalism.

(iv) Competitive exclusion principle.

(v) Coevolution of predator & prey.

Suggested reading:

In addition to books listed in module 25 and 26 following book may be consulted —

Ecology, R.E. Ricklefs and G.L. Miller 4th Ed. 2000. W.H. Freeman & Company, New York.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Group-A Paper-III Unit-II
Module No. - 28

Ethology

Basic Consideration :

Branch of Zoology dealing with the study of animal behaviour is called Ethology (Gr. Ethos = custom). The term was coined by Hilaire (1859). Prehistoric man was a keen observer of animal behaviour. They needed this to protect themselves from predators and to hunt their prey. Aristotle (4th cent. B.C.) and Pliny (1st Cent. A.D.) made important observations on animal behaviour yet the modern study of animal behaviour did not start until Charles Darwin (1872) wrote his famous book "Expression of Emotions in Man and Animals" where he argued that the difference in emotion between man and animals was one of degree and not of kind. This was followed by Animal Intelligence written by George Romanes (1882). Tremendous development of the subject took place in the Twentieth century. In 1973 Konrad Lorenz, Karl von Frisch and Niko Tinbergen got Nobel Prize for their work on animal behaviour.

Approach in the behavioural study :

Behaviour of an animal may be studied and analysed in three ways

1. **Vitalistic approach or observational approach :**

   This is the classical approach in which behavioural activities are observed in nature through field observations and are related to environment.

2. **Mechanistic approach or experimental approach**

   Behaviour is studied in laboratory under controlled conditions. Animals are regarded as some sort of

machines with set rules.

3. Ethological approach :

This is the modern approach which involves and combines both the above mentioned techniques.

In interpreting behaviour care must be taken to avoid 'anthropomorphism' (interpreting what has been observed in terms of human experience) or anthropocentrism (putting oneself in the place of animals).

The first step in the study of animal behaviour is the compilation of an "ethogram" — an inventory of the behaviour performed by a species.

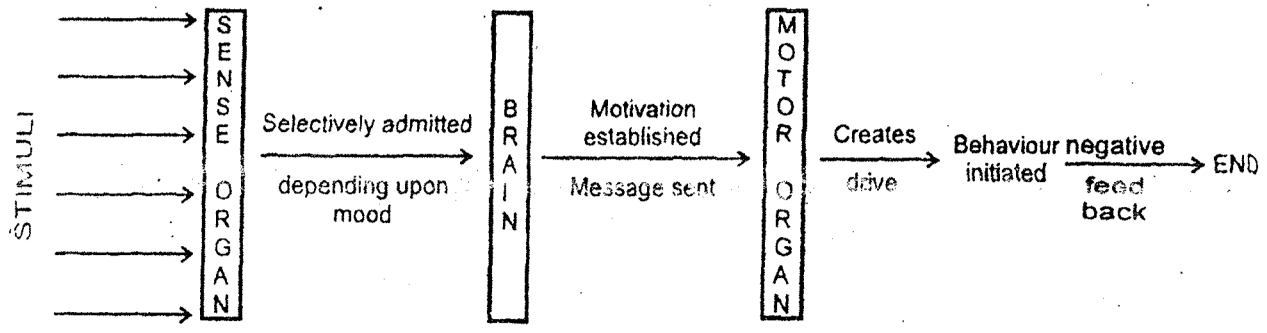
Definition of behaviour :

In popular language behaviour may be defined simply as what an animal does. More technically it is external manifestation of a stimulus response phenomenon in which response is performed voluntarily and consciously. Reflex is the simplest form of reaction to stimulus. Since it is automatic and involuntary it is not regarded as behaviour.

Most primitive type of behaviour are *taxes* and *kineses*,seen in invertebrates.

Mechanism of behaviour :

Behaviour is essentially a stimulus — response phenomenon in which external stimuli are selectively admitted depending upon the internal state of animals (mood and motivation). Much of the sensory inputs are rejected between sense organ and brain. Sensory data are processed and integrated through neural mechanism and drive is established. Message is sent to the motor organs for appropriate actions and behaviour is initiated. Once a behaviour is initiated some internal mechanism appears which acts as a negative feedback and prevents the animal from carrying the behaviour indefinitely resulting into the end of the behaviour.



What is stimulus ?

Anything which elicits response may be called *stimulus, cue* or *signal.* This may be an object, a part of it or any property of it or a situation.

Types of stimulus :

Depending upon the sensory organ involved in the perception of stimuli these may be classified as : Visual, olfactory, auditory, gustatory or tuctile stimuli..

Based on origin and function stimuli may be classified as :

1. **Motivational stimuli :**

   Which may be external (increasing day length causing reproductive behaviour in birds) or internal (depleting food reserve in hibernating animals causing awakening & food seeking) and provides drive.

2. **Releasing stimuli**

   Always external, produced by a member evoking response in other member e.g. Mustache of male flicker bird.

3. **Terminating stimuli**

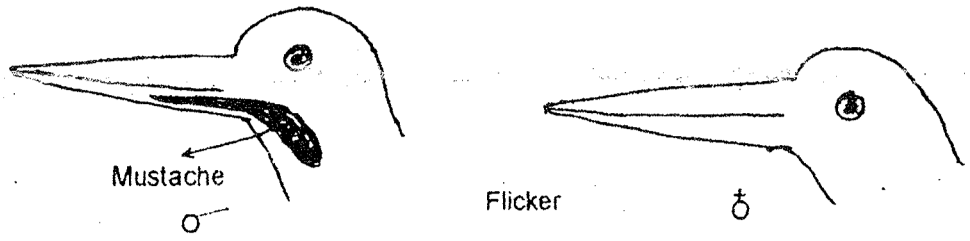
   Stimuli which terminate a behavioural response. May be external (e.g. sight of completed nest) or internal (e.g. full stomach, satiation).

Same stimulus may establish different motivation, may evoke different response in same or different animal e.g. sight of food will evoke different response in starved & well fed animal. Even starved animal will refrain from taking food if it is associated with danger.

Sign stimulus and releaser :

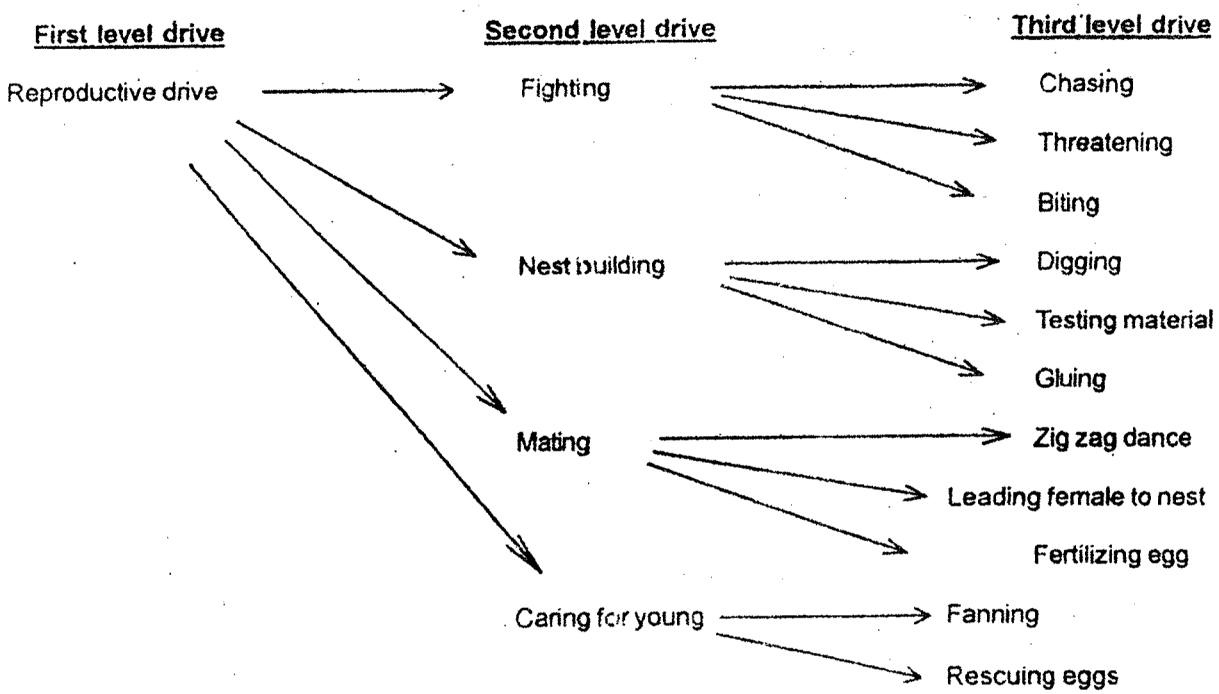
Specific external stimilus that triggers stereotype response from conspesifics is called a sign stimulus (Drickamer, Vessey and Meikle, 1996).

A structure of an animal that serves exclusively to send out sign stimuli and results into a high degree of response is called releaser e.g. Mustache of made flicker is a releaser which evokes reproductive behavi our in females and agonistic behaviour in other males.



Hierarchy of drives :

Tinbergen (1954) opined that different levels of drives and action exist in animals. In three spined stickle back male it may be depicted as :



Types of behaviour :

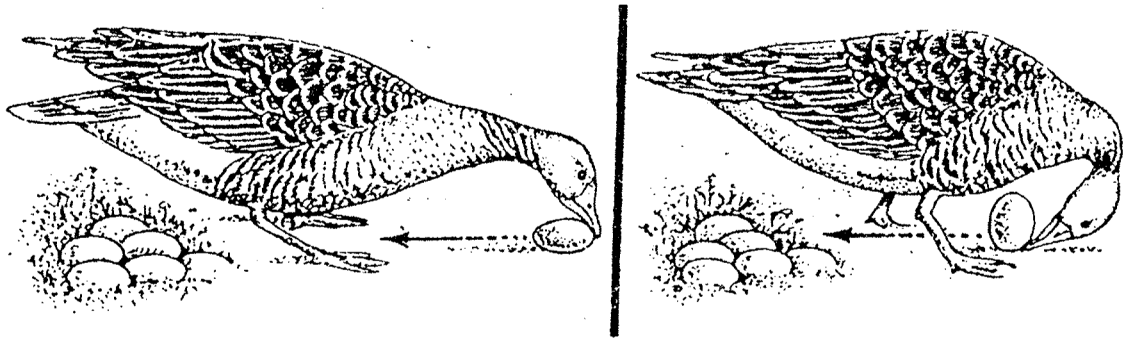
I. Behaviour of animal may be innate or learned.

Innate behaviour / Instinct / Stereotyped behaviour / Fixed action patterns :

An animal is born with a number of inherited responses. Such internally 'programmed' behaviour-are called inmate behaviour. These are performed without prior practice or experience. Innate behaviours are modified and perfected through learning in the light of experience. 'Instincts' are unlearned species specific motor patterns (Lorenz) e.g. tongue flicking of a new born snake. Two to three week old human infants smile when an adult face comes close to their own face. These stereotyped behaviours were called fixed action pattern by Lorenz.

A fixed action pattern FAP is a motor response that may be initiated by a very specific environmental stimulus, but can continue to completion without the influence of external stimuli e.g. Female greylag goose *(Anseranser)* will retrieve an egg that has rolled out of her nest by reaching beyond it and rolling it towards her with

the underside of her bill. If the egg is removed once the rolling behaviour has begun, the goose will continue the retrieval response until the 'imaginary' egg is back to the nest.



The FAPs are genetically determined, unlearned, species specific, stereotypic responses which once begun will continue to end. If an animal can perform a behaviour after being reared in isolation (deprivation experiment), the behaviour is a FAP. George Barlow (1968) has suggested the term 'model action pattern (MAP) instead of FAP.

Phases of innate behaviour :

Wallace Craig (1918) has made distinctions between different phases of innate behaviour.

1. ***Appetitive / searching / preliminary phase***

   Animals when born have a search image in mind or an appetite for the correct specific stimulus that would trigger appropriate behaviour. For example a new born will turn its head and open mouth in search of breast. When hungry an animal will ignore resting and mating places and will look for feeding place. This phase is goal oriented.

2. ***Consumatory phase :***

   When the goal is reached actual behaviour starts. The stereotyped behaviour that completes the sequence is called consumatory phase e.g. new born after comes in contact with the nipple starts breast sucking.

3. ***Quiescent phase or inactive phase :***

   Consumatory act brings about a sudden drop in the animals motivation to perform the behaviour. Behaviour stops and the animal (e.g. baby) passes into a resting phase.

II. Behaviour may also be classified as egocentric and altruistic.

Egocentris

A behaviour which is done for the welfare of oneself i.e. directed towards self is called egocentric behaviour. Sometimes this is also called selfish behaviour e.g. comfort movement, autogrooming etc.

Altruistic behaviour :

A behaviour which is done for the welfare of other individuals sometimes done at the cost of donor or performer is called altruistic behaviour. This is also called sacrificial or suicidal (though wrongly) behaviour e.g. parental care, behaviour of worker termite.

Learning :

Learning e. ables animals to adjust their innate behaviour patterns (FAP) to changing situations. Innate behaviour is modified through learning in the light of experience. By discovering what responses give best results animals modify their innate behaviour pattern. Thus learning is important in perfecting skill. For example, nest building is innate in birds but what materials are suitable for the purpose is learnt.

Phases of learning

According to Hunter (1964) learning has 4 phases :

1. The process of *acquiring* some knowledge or experience. This is the process of establishing association or *acquisition.*
2. The process of *remembering* the past experience.
3. The process of *failing* to remember past experience *(habituation).*
4. The process of *storage* of acquired experience or retention through *reinforcement.*

Types or learning :

According to Thorpe (1965) there are six types of learning

1. Habituation
2. Conditioned reflex
3. Trial and error learning
4. Latent learning
5. Insight learning
6. Imprinting

Some modern authors (e.g. Goodenough, McGuire and Wallace, 1993; Starr and Tuggart 1998) prefer to use classical *conditioning* instead of conditioned reflex and *operant conditioning in* place of **trial** and error learning. Starr and Taggart prefer *spatial* learning in lieu of latent learning. Often two kinds of trig[and error learning are recognised viz., *instrumental* and *operant* conditioning.

Habituation

This is the simplest form of learning common in all animals from Protozoa to Man. It involves loss of an old trait

(innate response) rather than the acquisition of a new response. In contrast *extinction is* the loss of an acquisition (learned response). Learning to ignore stimuli that have no accompanying consequence is habituation.

An animal repeatedly exposed to a given stimulus mot associated with any positive or negative consequence (reward or punishment) soon learns to ignore it. For example, A Hydra kept in water in a petridish will quickly retract its body and tentacles if the dish is tapped. But after a few taps its reactions begin to slow down and stops.

Adaptive value :

Habituation is important in the development of an animal (particularly young ones). Through it they are able to sort out relevant and irrelevant stimuli. It is of adaptive value because it conserves energy & time for other important activities.

Conditioning :

Classical conditioning :

Also called Pavlovian conditioning since Ivan Pavlov first experimentally demonstrated this. In this an animal learns to associate an automatic unconditioned response (UCR) normally elicited by an unconditional UC stimulus (UCS) which is in born, with a new stimulus that does not trigger the response (conditioned stimulus).

1. Sight of food ⟶ Salivation
   (UCS) (UCR)
2. Food + preceded by sound of bell or flesh of light ⟶ Salivation
   (UCS) + (CS) (UCR)
3. Sound of bell or flesh of light ⟶ Salivation
   (CS) (CR)

If the CS is presented many time without the UCS, the response to the new stimulus (CR) will be gradually lost. This is called *extinction.*

The adaptive value of classical conditioning is to prepare animals for important events.

Trial and error learning :

Toad explores the environment for food by flicking its tongue to any dark object. By repeated trial and error it learns to avoid bumble bee which stings but eats other insects.

Instrumental learning and operant conditioning :

Thomdike (1898) experimented with hungry cats which were required to press a latch or pull a string to open the door to come out to the cage to obtain food when it was presented. He called this trial and error and accidental success. *By* repeated trial and failure they were accidentally successful in opening the door. This is now called *instrumental learning.* It occurs in response to some obvious stimulus which could be seen and differs from *operant conditioning* where it is spontaneous action in the absence of any obvious stimulus. *B. E.* Skinner (1937) devised an apparatus, 'Skinner box', for the study of operant conditioning. A pigeon is placed inside the box which has a key (an illuminated button) and a mechanism for delivering grain to the pigeon when the key is pushed. The hungry pigeon pecks the wall of the box and accidentally presses the button and is rewarded with the food. By repeated trial and error it learns to press the key when it is hungry. Here the stimulus (food grains) remained unseen.

Imprinting :

Imprinting is a rapid form of associative learning that occurs only at a specific, brief and highly sensitive time in the early phase of life called' critical period' by Lorenz (1935). For example attachment of new born with the mother. A new born gosling considers the first object that appears over it after it hatches out as to be its mother. In nature this is normally a female goose. For incubated gosling this is usually a person. In case the person has not turned up in time it is the box or incubator to which the gosling develops attachment. Such unusual imprinting has been seen in many wild mammals in Africa. New born rhinoceros, antelopes, zebra have been seen chasing a horse back rider or an automobile. They might have seen these objects before their mother which might have fled away sensing approaching danger soon after giving birth. In human being often a shift of attachment from mother to forter mother or ayah is seen.

The earliest work on imprinting was done by Spalding between 1872 - 75 who published six papers before he died in 1887 at the age of 37. He observed that 2 or 3 days old chicks would follow any moving object and develop an attachment to it. Now this phenomenon is called "following response". Unfortunately his work was forgotten until rediscovered by J. B. S. Haldane in 1954.

Heninroth (1910) used the term imprinting for the first time but it was Lorenz who gave the concept of imprinting its final shape and highlighted the importance of critical period and suggested that unlike ordinary learning imprinting is irreversible. Lorenz recognised that imprinting was confined to a particular sensitive period of development which varies from one species to other. For example in chicks it is between 5 hrs. to 25 hrs. with peak at 15 hrs. (Hesse, 1959). In puppies sensitive period is upto 3 to 10 weeks.

A perfect score is the score obtained in laboratory test of 'following score'.

Although imprinting usually takes place in the early life occasionally it occurs in the adulthood also: For example mother goats lick their offspring just after birth and gets imprinted with its own kid which results in its rejection of alien kids.

Types of imprinting :

(a) Depending upon the signal it may be classified as :

1. Visual imprinting :

Imprinting takes place by looking at the object e.g. "following response" of birds.

2. Auditory imprinting :

Imprinting occurs by way of hearing. e.g. Responding to call of mother cow by calf.

3. Olfactory imprinting :

Imprinting occurs through smell. Common in mammals. New born recognise the mother by its smell.

4. Tactile imprenting :

Imprenting takes place by the touch. Common in primates. Baby stops crying when put on to the lap of mother.

(b) Filial and sexual imprinting :

Experimental evidences have further revealed that there may be more than one type of sensitive period. Both filial and sexual behaviour may be influenced by imprinting. Consequently imprinting may be classified as : (1)

Filial imprinting and (2) sexual imprinting. Filial imprinting is the development of attachment between offspring and mothers and sexual imprinting is the development of attachment with the opposite sex. Example of former is the following response and of the latter is the preference of an individual to the individual of opposite sex. Sexual imprinting occurs later than that of filial imprinting. According to Vidal (1976) in cockerels sensitive period for the sexual imprinting is between 31 to 45 days whereas the same for filial imprinting is upto 30 days. Unlike filial imprinting sexual imprinting occurs in both altricial and precocial species.

Significance of imprinting :

1. Imprinting is important in learning about the important primary objects such as parent, sex etc. particularly in species with precocial offspring. They have only a short time to learn about the primary objects.
2. According to Lorenz it is important in species recognition. Although later studies show that this is not essential for the same. It is important that the parent care for their own offspring and mating occurs with the members of same species.
3. Imprinting is important in recognition of dose kin so that mating with the siblings and interbreeding may be avoided (Bateson, 1979).
4. It is important in special behaviour such as homing of Salmon / Hilsa to their native streams for spawning.
5. It is important in behavioural isolation and speciation.

Review Questions

1. Distinguished between :

   (a) Stimulus & releaser, (b) Egocentric and Altruistic behaviour, (c) Motivational stimuli and terminating stimuli, (d) Vitalistic & mechanistic approaches, (e) Extinction & habituation, (f) Instrumental learning and operant conditioning, (h) Innate and learned behaviour, (i) Appetitive phase and searching phase.

2. Write notes on :

   (a) Hierarchy of drives (b) Fixed Action Pattern, (c) Ethogram (d) Classical conditioning, (e) Sign stimulus, (f) Critical period in Imprinting, (g) Following response.

3. Define imprinting. Distinguish between filial and sexual imprinting. Add a note on the significance of imprinting.

4. What is behaviour? Discuss innate nature of behaviour and add a note on the survival value of learning.

SUGGESTED READINGS

1. Animal Behaviour by L. C. Drickamer, S. H. Vessey and D. Meikle. 4th Ed. 1996. W. C. Brown Publishers, London.

2. Ecology & Behaviour by C. Starr and R. Taggart, 1998. Wadsworth Publishing Company, New York.

3. Animal Behaviour by D. Farland, 1987. Longman Scientific and Technical "Essesc".

4. Perspectives on Animal Behaviour by J. Goodenough, B. McGuire & R. Wallace 1993. John Willey & Sons, Inc. New York.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-A Unit-II

Module No. - 29

Egocentric behaviour:

A behaviour performed for the welfare of self is called egocentric behaviour, Maintenance behaviour, habitat selection, feeding, escape and defence behaviour etc. are some such behaviours.

Maintenance behaviour:

(I) Body surface maintenance:

These are performed to remove unwanted materials, for toilet purposes and for beautification. Some important forms are mentioned below:

(1) Body care

a) Bathing:

There are different forms of bathing.

i) Water bathing:

Performed to remove dirt and parasite, general cleaning and cooling of the body. This may be done by taking a dip in the water body (birds, mammals) or splashing water on the body (elephant) or simply may be rain bathing (parrot).

ii) Wallowing or mud bathing:

Performed by pigs and buffalo where body is black and almost naked. Mud has a cooling effect and also removes ectoparasites.

iii) Sand bathing:

It involves rolling on sand (Donkey, Zebu) or splashing sand over body (elephant). Horse role over grass. In kangaroo rats lipid secreted by sebeceous gland is removed by sand bathing.

iv) Fire bathing:

Crow, eagle, vultures etc. sit over hot chimneys to remove ectoparasite.

b) Powdering:

Base of powder down feather is broken by biting and powder which is antiseptic is splashed. Done by parrots during dry season.

c) Cleaning:

Insects clean eyes and body by antennae & legs. Honey bee clears antenna by dragging it through the antenna cleaner located in foreleg.

d) Moving tail:

By moving tails quadripeds remove flies from the body.

e) Grooming:

Birds and mammals spend considerable time in grooming activities. Autogrooming is the self grooming involving mails or teeth. Allogrooming is a social behaviour where one individual grooms other individual and as such is an altruistic activity. Some important forms of grooming are listed below —

	In birds		**In mammals**
i)	Bill wiping	i)	Licking (by tongue)
ii)	Feather sething	ii)	Pawing (moving a prelicked paw over the body)
iii)	Nibbling of feather	iii)	Licing
iv)	Preening	iv)	Biting fur
		v)	Combing

(II) Comfort behaviour:

These are performed to provide comfort to the body. More common in mammals than in birds. Some important types are listed below.

1. Rubbing:

Body or some part of it is rubbed against each other (autorubbing) or against the body of other individual (allorubbing) or against some hard object in the environment (e.g., deer).

2. Rutting:

Rubbing of antlers against some hard object during breeding period. Provides comfort and increases cutaneous circulation (e.g., deer).

3. Scratching:

By claws, nails or teeth (monkey).

4. ***Shaking:***

Shaking of the body not only provides comfort but also removes water from the body surface (horse).

5. ***Stretching of the body***

6. ***Arching the back***

7. ***Moving head up and body***

8. ***Nibbling:***

It is soft bitting without inflicting wound.

9. ***Yawning:***

Common in mammals, also seen in snakes & fish. Done in sleep or after sleep to remove fatigue of tissues and muscles, for removal of accumulated CO_2 in lungs and intake of fresh supply of O_2. Fish yawn to take large quantity of water for respiration.

10. ***Panting:***

Done by carnivores (dogs), has a cooling effect.

11. ***Sunbathing:***

Seen in mammals (seals, whales, walrus) birds (flamingo spread wings, pelicans stand on one leg), crocodiles and in frogs and sharks where body surface is exposed to sun. It helps in temperature regulation and Vitamin D synthesis.

12. ***Sleeping:***

It is an important maintenance behaviour which helps in repairing wear & tear in body and also removes fatigue and accumulated lactic acid during stretching following sleep. Sleep walking is an unusual behaviour.

(III) Anting:

It is a special kind of behaviour seen mostly in birds. Habit of rubbing their plumage with ants or tucking ants in the plumage is called 'anting'. 160 species of birds have been recorded to perform anting. Some mammals like cats and grey squirrels are also known to practice anting (rubbing fur or body on ant trail. At least 40 kinds of substitute materials (like burnt cigarette butts, lemon flesh, mealworm, bee etc.) have been recorded to be used in anting instead of ant. Earliest recording of anting dates back to J.J. Audubon (1831).

Types of anting:

(1) Active anting:

The action of applying ant or any substitute material to their plumage.

(2) Passive anting:

The action of sitting or wallowing among ant or ant trail.

(3) Vacuum anting:

The actions which birds some time make at the sight of ants or other substances such as smoke, but without actually using them.

Theories explaining anting:

Several theories have been put forth to explain practice of anting.

A. Chisholm (1973) has reviewed these theories

(1) Food carrying theory:

Anting is a form of food carrying in which ant is tucked into the plumage and carried to nest.

(2) Food cleansing theory:

Birds wipe acid from the ant before eating these.

(3) Skin cleansing or parasite repelling theory:

Birds apply acid onto the plumage to clean these of parasite as crushing releases formide acid from ants.

(4) Aromatic use theory:

The birds appreciate the odour of ant acid on their bodies and this is used like scent.

(5) Stimulant agent theory:

Acid acts as a stimulant or soothing agent.

Chisholm (1959) has dismissed the first three theories as substitute materials are also used for anting and as it is performed even in the absence of parasites. There is no proof for the fourth theory. He has experimentally showed that birds did not use ants which were deep frozen or dried by heat for anting when offered and as such he concluded that anting is performed for stimulation and as soothing phenomenon. According to Kelso and Nice (1963) anting is an instinctive action since it may be performed even in the absence of mite infestation.

HABITAT SELECTION AND SHELTER SEEKING:

All organisms from the lowly *Paramecium* to highly evolved man have one behavioural trait in common — shelter seeking. All seek optimum environmental conditions and avoid dangerous and injurious situations. Desert animals seek the shade of rock or comfort of burrow during hot day time. Shelter may be divided into two broad groups: (1) Normal shelter, which is inhabited during normal conditions and (2) emergency shelter, which is sought to avoid sudden danger.

Types of shelter or cover:

Based on the behaviour of animals Leopold (1933) recognised five different types of cover —

(1) Winter cover:

Provides protection from snow and cold. A place to remain out of sight and for hibernation. White tailed deer has a define preference for coniferous forest during winter.

(2) Escape cover:

Vegetation in which animal retreats to avoid predators. Spotted deer normally graze in open areas but retreat to woods when alarmed.

(3) Loafing cover:

It is a place where animal spend restful minutes during the day. Blockbuck rest under the shade of tree during noontime.

(4) Nesting cover:

Provides protection for both youngs and parents particularly during breeding season.

(5) Roosting cover:

It is a secured cover for sleeping and spending night.

Basis of shelter seeking:

Basic to shelter seeking behaviour is habitat selection. The underlying questions here are —

(1) Why one habitat is preferred over other?

(2) How the preferred habitat is selection?

(3) How does the choice change?

The whole question of habitat selection is very much associated with resource availability, niche size, niche overlap and competition. A natural habitat may not be occupied by a species in the persence of competition and it may be competitively excluded from that area.

Habitat selection may partly be a psychological process (Miller, 1942). The lack of singing perches and nesting sites may prevent birds from colonizing that area.

Recognition by some conspicuous feature may also play an important role. For example salmon recognize their home stream by their olfactory experience gained as fry (Hasler, 1966). Imprinting plays an important role.

There are experimental evidences which suggest that the habitat selection is innate and is independant of early experience.

Experimental evidences in favour of the innate nature of habitat selection:

Harris (1952) experimented with two subspecies of deer mice, *Peromyscus maniculatus bairdi*, restricted to open field habitats and *P.M. grcilis*, restricted to forests. By presenting artificial habitats, one of forest and other of grassland to the laboratory bred individuals, he found that the two sub-species selected the appropriate artificial habitat. Since the experimental individuals had no previous experience with either habitat and yet exhibited a preference for the type of habitat they normally occupy in the wild state, he concluded that the habitat selection was basically genetic.

The problem was probed in more deeply by Wecker (1963) who investigated the role of early experience in determining the habitat preference of *P.M. bairdi* in natural condition. Six groups of mice were tested. In separate experiments these were released in a 16000 square feet rectangular enclosure with its long axis lying halfway across a woods-field habitat were recorded. Findings of his experiments are summarized below —

	Hereditary background	**Early experience**	**Habitat preference**
1.	Grassland	Grassland	Grassland
2.	Laboratory	Grassland	Grassland
3.	Grassland	Laboratory	Grassland
4.	Grassland	Forest	Grassland
5.	Laboratory	Laboratory	None
6.	Laboratory	Forest	None

The results proved that the choice of the grassland habitat by *P.M. bairdi* is normally predetermined by

heredity i.e., it is innate in nature. Early field experience can reinforce this innate preference but it is not necessary for subsequent habitat selection. Early experience in other environment (forest and laboratory) was not sufficient to reverse the normal selection of grassland by mice. However he found that confinement of deer mice in the laboratory for 12 to 20 generations resulted in an apparent reduction of hereditary control over habitat selection.

In a similar experiment Partridge (1976) presented a choice between oak and pine branches to coal tits reared in aviary. The former selected pine while the later selected oak branches which corresponded to their natural habitat.

Plasticity in habitat selection and habital switches:

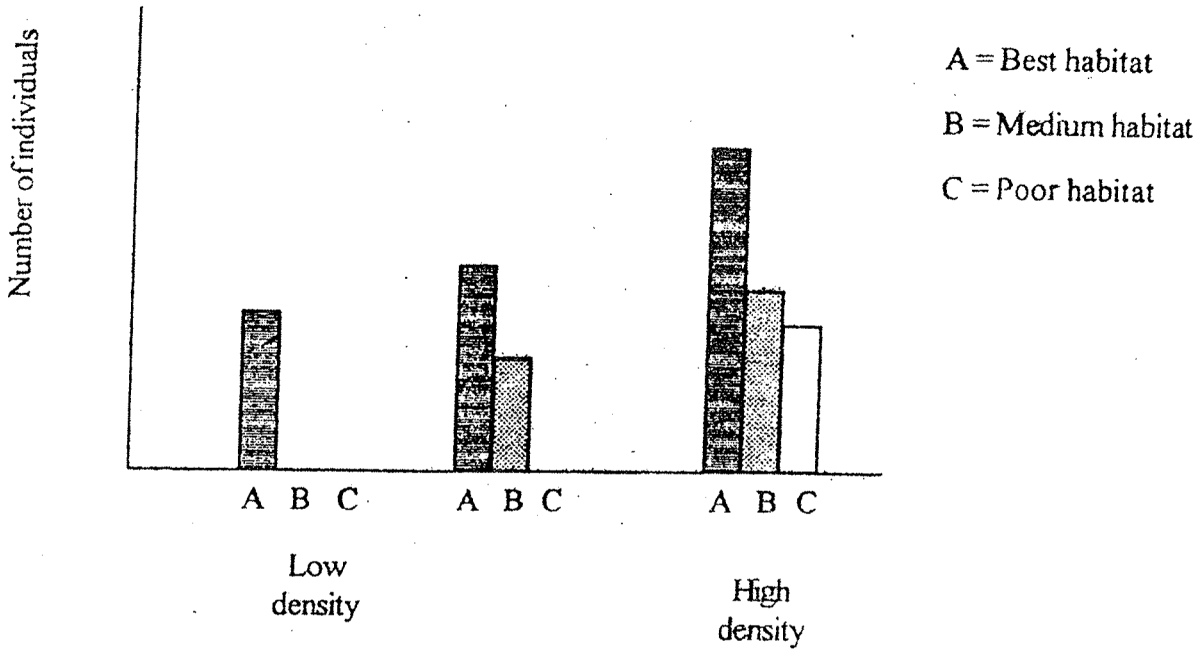
Habitat selection is not extremely rigid. Most species exhibit some plasticity which enable them to spread to unusual habitat and broaden their ecological niche. The most dramatic examples of habitat switches may be encountered in adaptation to urban artefacts. Rats have moved into sewers, pigeons onto window ledges, crows onto statues, bee eaters onto telephone wire, bedbug into beds & seats of cinema hall and borers into furnitures.

How does the choice change:

Change in habitat choice may be related to the availability and limiting effect of resources. There are two main theories explaining this.

(1) Theory of optimal foraging:

This is the classical theory. When resources are not limiting then the species feeds on the best type of food and chooses best type of habitat and ignores others. When resources are limiting then the species expands its niche to include other types expands its niche to include other types of food and chose alternate habitats. The situation is best explained through Fretwell's model (1972).



Fretwell's model

(2) Rosenzweig's theory:

This theory was put forth by Rosenzweig (1981) who opined that when resources are limiting species contract their niches as a result of interspecific competition. The dominant species will not change its habitat choice whereas the subordinate one will change it. Rosenzweig (1985), however, opined that no single theory can explain the change in habitat choice.

ESCAPE AND DEFENSE BEHAVIOUR:

Of all the behaviour pattern of animals the most generalised and the least species specific are those involving escape and defense. THe responses — escape movement, distress calls alarm reactions are similar in many different species. This behaviour lacks specificity because no selective advantage is gained by having differences in alarm reactions between species. However, against predators defence strategies are sometimes species specific. Some social herbivores clump together when threatened (e.g. zebra) while others scatter (e.g. Patas monkey). The defense may be specific to type of predator. For example chicken will head for a covered ground and crouch to

escape aerial predator but fly up to trees to escape ground predators. Each response is accompanied with different types of warning calls.

Alarm system and alarm reactions:

All animals except some top carnivores need some alarm system to detact and escape predators. Top carnivores do not need this for predators but like others they also require it to detect and escape unfavourable and dangerous situations such as forest fire.

Danger or enemy acts as a stimulus. Responses are alarm reactions of different kinds such as escape, distress calls etc.

Defence against predator:

Antipredator mechanism may be classified as primary and secondary.

Primary defenses operate whether a predator is present or absent and function to decrease the probability of an encounter with a potential predator. These include, cryptic colouration, warning colouration and Batesian mimicry.

Secondary defenses operate during an encounter with the predator and function to enhance an animal chances of surviving encounter. These include different types of individual and social strategies like distraction, creating confusion, mobbing, release of toxic substances etc.

Enemy recognition / detection of danger:

Enemy recognition and fear are innate in nature. Most species have "neophobia" and avoid unfamiliar or new situations. However, the ability to distinguish between what to fear and what not to fear is learned. Habituation plays an important role in this.

All the sense organs play important role in enemy recognition. Visual and auditory recognition are common in herbivores while olfactory recognition is vital is carnivores. Nice and Pelwyk (1941) and Hartley (1950) investigated the importance of visual recognition. Young birds when presented with models of predators (owl like outline, pattern & colouration) showed strong alarm reactions.

Reaction to enemies:

How an animal would react to an enemy depends upon a number of factors such as circumstances in which meeting takes place, behaviour of enemy, internal state of animal and its psychological state.

If the predator is well fed its reaction will be different to prey as compared to the starved state. Enemy will behave in different way depending upon the degree of intimidation. A prey will react in different ways depending upon the distance between itself and the enemy or predator. If the enemy or potential predator is within 'fight distance' it will trigger escape. If the approaching enemy is not detected until it is within the 'charge distance' it will evoke charge responces viz., threat or attack.

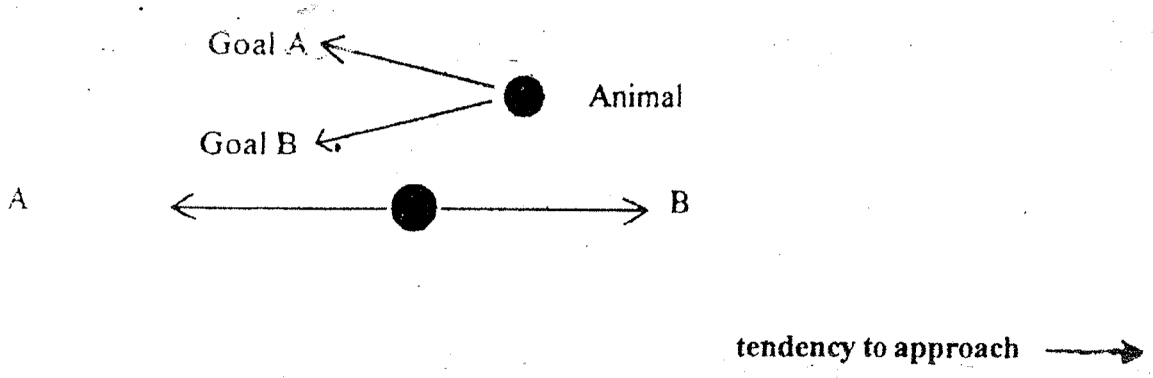
The reactions also depend upon the mood, motivation and level of reproductive hormone. Rise in the level of male hormone evokes more agonistic reactions. Psychological state is also important in this respect. If an animal is in desperate state it would show strong charge response. Similarly conflict in mind also plays vital role.

Conflict:

Conflict occurs when two tendencies complete for dominance in the control of behaviour. There may be three logical types of conflict.

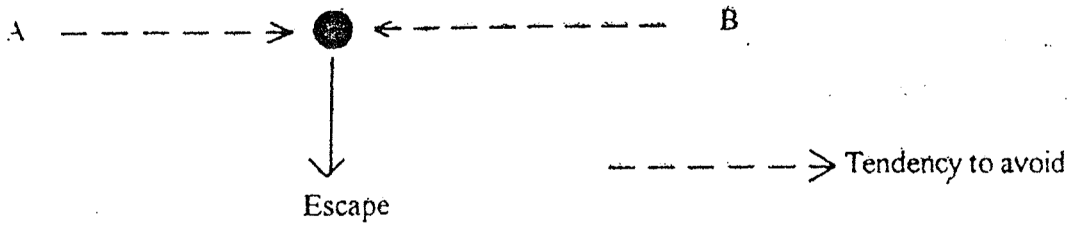
(1) Approach-approach conflict:

It occurs when two simultaneous tendencies direct towards different goals. This is unstable since ultimately the animal will have increased tendency to approach a goal.



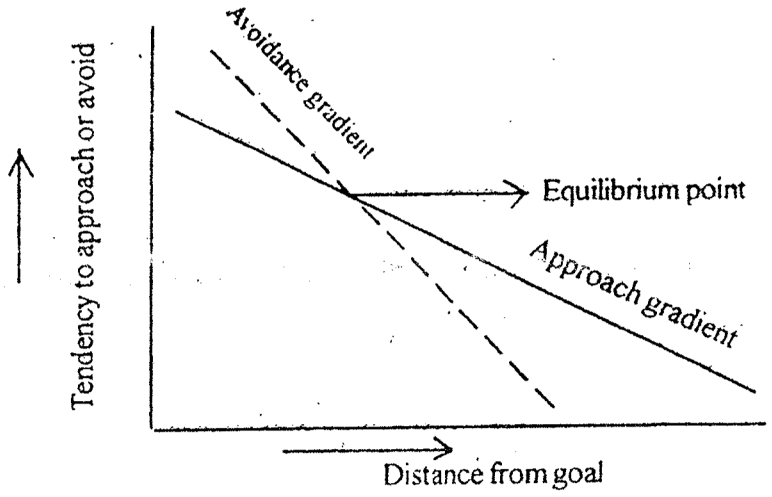
(2) Avoidance-avoidance conflict:

Occurs when two avoidance tendencies appear simultaneously. Since the tendency to avoid the objection increases with the proximity of the object the animal will ultimately escape at right angle to the line between the object.



(3) Approach-avoidance conflict:

Occurs when an animal has simultaneous tendencies to approach and avoid the same object/goal. As is illustrated in the goal gradient figure the tendency to avoid is stronger than the tendency to approach near the object and with the increase in distance the reverse becomes true. In between these is a point of equilibrium. This type of conflict is very common in animal behaviour.



Escape reactions / courses of action:

An animal may adopt some individual strategy or can adopt collectively some group or social strategy against animal.

(a) Individual strategies:

(1) **Sound production:**

Animal may produce alarm or distress call or may cry out warning.

(2) **Colour change:**

Some animals (e.g. *Sepia*) change colur according to background so that they may remain concealed.

(3) **Posture:**

Animals may stop movement (freez) or feign death (playing possum) at the sight of danger. Most predators usually attack moving prey. For example wolves will not attack prey that stands motionless (Mech 1970). Some animals may adopt a warning posture to thwart away the enemy. Intimidation display by enhancing body size is quite common.

(4) **Expression:**

Encounter with enemy may result into fear, warning or agonistic expression.

(5) **Deception:**

Larvae of several moth look like snakes. When threatened they inflate head and move it like 'hood' of snake to deceive enemy. Similarly eye spots on the hind wings of moth are made prominent at the sight of prey which gives it an appearance of the face of a large animal.

(6) **Distraction:**

Autonomy, the ability to break off a body part when attacked is a good antipredor defence mechanism. When a lizard is chased by a predator it waves its tail prominently. As soon as it is caught by the tail the lizard resorts to tail autonomy. Not only it breaks away from the attacker but the moving tail continues to distract the attracker.

(7) **Release of toxic substances:**

Some insects spray toxic substances when alarmed. For example *Brachinus* sp. spray a mixture of quinnones and hydrogen peroxide. such mechanism is seen in some amphibions & reptiles also.

(8) Speed:

Development of speed is perhaps the best defensive mechanism. Running and stotting (stiff legged jumping in deer and antelopes) are the best examples.

(b) Social strategies:

(1) Confusion effect:

Predators prefer to attack single prey. When attacked groups become either compact (zebra) or scatter (Patas monkey) and the predator is confused it is unable to isolate a prey.

(2) Distraction display:

Sometimes one individual makes itself prominent to distract the attention of the predator. In the mean time other individuals of the group escape. This is an altruistic behaviour.

(3) Alarm signal:

When predator approaches a group of prey one or more individuals give signal to alert other members of the group. Deer may raise the tail and birds may cry out distress call.

(4) Mobbing:

It is a mechanism of group defense in which prey harasses or attacks the predator forming a group. This involves visual and vocal displays, e.g. crow.

Sequence of actions:

When an animal encounters an enemy it adopts the following logical sequences. Initially it tries to hide. It seen it freezes or feigns death. Then it tries to escape or resorts to distraction display. If it caught unaware it cries out warning or threats the approaching enemy (by exposing teeth, claw or horn). When everything fails the final course of action is attack.

SOCIAL BEHAVIOUR

(I) Social dominance and hierarchy:

A tendency exists among many animals to form groups. Many of these animal groups have some form of organization based on intra-specific aggressiveness and intolerance and on the dominance of one individual over other. Two opposing forces are at work at the same time — mutual attraction verses social intolerance. There is a positive reaction leading to crowding and a negative reaction against crowding. Social interactions among individuals in a population involve agonistic or aggressive behaviour which forms the basis of social organization resulting into dominance. Each individual occupies a position in the group based on dominance and submissiveness. A social hierarchy is formed as a consequence.

Social dominance:

One animal is dominant over other if it controls the behaviour of the animal (Scott, 1966). In another sense, a prediction is being made about the outcome of future competitive interactions (Rowell, 1974). A dominant individual is likely to win in competitive interaction.

Dominance hierarchy is most often seen in arthropods and vertebrates which live in group but has been reported even in non-social sea anemones also (Brace et al., 1979). Among closely related species there may be clearcut hierarchy in one species (*Cercopithecus aethiops*) but not in other (*C. subaeus*).

Peck order:

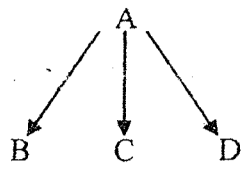
Schjelderup-Ebbe (1922) observed that in domestic chickens pecking following dominance i.e., birds peck others of lower dominance or rank. An 'alpha' individual pecks all other individuals. A 'beat' individual pecks all but alpha individual and so on to the 'omega' which is totally subordinate. This was called peck order which in its simplest form a straight line or linear peck order

$$\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \omega$$

Types of dominance hierarchy:

(1) **Simple types:**

All the subordinates are of equal status. Example: Mice



(2) **Linear hierarchy:**

As seen in domestic chicken.

(3) Triangular relationship:

First individual dominant over the second, the second over third and the third over the first. In such cases individual of lower rank can peck an individual of higher rank.

Peak order may be classified in another way as —

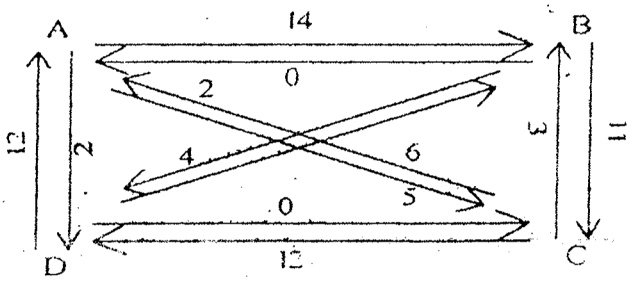
(a) Peck right hierarchy:

It is formed when all the aggression goes from dominant to subordinate

$$A \rightarrow B \rightarrow C \rightarrow D$$

(b) Peck dominance hierarchy:

It is formed when majority of aggressions go from dominant to subordinate. During encounter threats and pecks are dealt with both the individuals and individual with more pecks is regarded as dominant. Such relations result into a complex peck order.



→ indicates direction of attack and number indicate instances of win.

Factors associated with dominance and hierarchy:

The position of the individuals in the social hierarchy may be influenced by a number of factors such as sex, level of male hormone, strength, size, weight, maturity, previous fighting experience and social rank, injury, fatigue etc. Main factor is the agonistic ability of the individual.

In heterosexual groups there may be separate hierarchies for males and females with dominance of males

over females (e.g. Baboon). However, the most aggressive male may not be the most active in dominating the female (e.g. Red cross bill), in some species (e.g. Redstart) females dominate males while in others (e.g. Snow goos) there is no dominance based on sex.

In white tailed deer adult bucks dominated other deer and doe dominated the fawns. In this scheme larger individual dominated the smaller rather than the sex.

In *Macaca mulatta* there are separate hierarchies for males & females. By the end of first year infants in a group develop a linear peak order. The position is determined by the mother's rank. Infants are ranked just below the mothers.

	A	→	B	→	C	→	D
?	↓		↓		↓		↓
Infant	a	→	b	→	c	→	c

While in natal groups the male infant is ranked below mother but when in leaves the natal group and forms a ne group new hierarchy is formed.

Rise is social hierarchy:

Once social hierarchy is well established within a group, newcomers and subordinate individuals rise in rank with great difficulty. Removal (death) or injury of dominant individual causes a scramble for the alpha position. Sometimes several subordinant individuals associate together and by maintaining a mutual social rank may revolt against the top ranking individual.

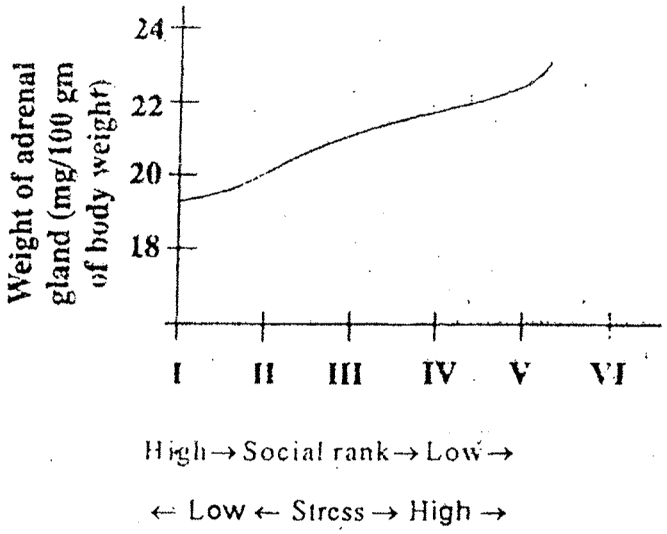
Rise in hierarchy often is related to the breeding and level of hormone. A male house finch which occupied the bottom rung of social ladder during winter but comes into the breeding condition first rises to top of the hierarchy in breeding season due to rise in the level of testosterone.

Aggression and social dominance:

Dominance is intimately associated with aggressive behaviour. Agonistic behaviours are influenced by the rise in the level of male hormone. Injecting testosterone in subordinate individual induces aggressiveness in it. However, this is not the only cause of aggressiveness. Shortage of food, space and mate may increase intraspecific competition and hostility between individuals. Dominant individual has the first choice of resources and subordinates are forced with no other alternative than to starve, leave the area or to revolt. According to Julian Huxley (1945) shortage of resource is tolerated upon a certain level by the individuals and further decrease is resisted. Individuals

are left with only three options, surrender and remain as subordinates, fight the dominant individual till death or leave the area and remain in periphery as 'floating population' waiting for the opportunity to challange the old or injured leader (e.g. Arctic ground squirrel).

Myers et al (1971) have experimentally shown that crowding and shortage of apace and resource result in social & physiological stress in rabbits which are reflected as increased aggressiveness, curtailment of reproduction in subordinates which show histopathological signs of stress such as increased size of adrenal gland and inflamation of liver.



In mice (*Mus musculas*) weight of a adrenal gland increases in individuals of lower rank. In mice kept in isolation (without stress) weight of adrenal gland was similar to that of 'I'.

Advantages of dominance:

Subordinate individuals suffer from stress, higher level of adrenal cortical hormores which elevate the level of blood sugar. They also suffer from the reduction in their defence mechanism and reproductive hormones, are melnourished and as such have high mortality. Compared to these dominant individuals are well fed and healthy, have low mortality and as such are better fitted from the point of view of natural selection.

Cooperative social groups:

Not all aggregations are based on social dominance. Some groups show no strong evidence of social rank among individuals and exhibit relatively strong group co-operation.

In harvester ants, perhaps through imprinting, individuals learn the odour of their own colony. Foraging ants returning to their own nest are recognised by their nestmate. But if an ant enters an alien nest it is threatened, seized and killed. Here the colony acts as a social unit. Members of the same social unit do not show any agonistic behaviour between themselves.

Prairie dogs live in aggregations called prairie dog town. Each town is subdivided into cohesive, co-operative units called 'coterie' containing 2-32 individuals. All coterie members use the same territory without any conflict but will not allow any trespassing and will defend the trritory as a social unit. There is hardly any division of labour and some evidence of the existance of a leader in coterie who is more aggressive.

Eusocial organization:

The most highly organised natural societies are seen in eusocial insects (ants, honey bee, termites) with well defined division of labour. Members of the colony are inseparably bound by behaviour and physiology and fail to survive if separated. Non reproductive castes (workers, soldiers) show altruism. All behaviours in these societies are innate.

TERRITORIALITY

Social organization which results i space being divided among and occupied exclusively by a social unit with a defended boundary is territoriality. This is found in birds, many fish and insects and in several mammals. The behaviour of maintaining a territory is called territoriality.

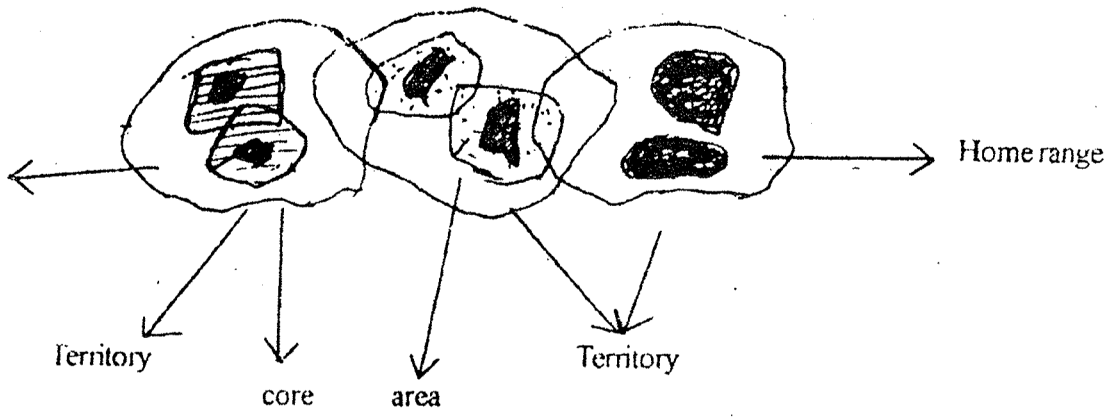
What is territory?

Territory is a defended, more or less fixed and exclusive area maintained by an individual or a social unit occupying it.

Home range and core area:

Territory should not be confused with home range and core areas. The home range is an area in which normally an animal lives and carries one normal activities. Within the home range there is often an area in hich most

activities are concentrated. This is called the core area. In some cases the core area may be the area immediately surrounding the nest site or feeding and drinking zone. Home ranges of individuals may overlap, such overlap is less likely in core areas and prevented by agonistic behaviour by the residents in case of territory.



Distinction between territory and home range:.

	Territory		**Home range**
(1)	Has more or less fixed boundary.	(1)	Do not have a fixed boundary.
(2)	Territories of individuals and social units do not overlap.	(2)	Home range of individuals or groups may overlap
(3)	Territory is always defended from the intruders.	(3)	May or may not be defended in part or in whole.
(4)	Always associated with agonistic behaviours : threat or attack	(4)	May not be associated with agenistic behavior
(5)	Advertisement behaviours (scent, song) are always present.	(5)	May not be associated with advertisement behavior.
(6)	Always there is social dominance and hierarchy.	(6)	Individuals tend to avoid contact by using different parts.
(7)	Often territorial boundary is marked by the dominant individual.	(7)	There is no marking of the boundary of home range.

Type of Territories:

Territories may be classified according to the activities that take place in them (Nice, 1941).

(1) General purpose / multipurpose territory:

All activities take place within territorial boundaries. It is well established during breeding seasons. After breeding season territorial defense breaks down. It is common among song birds and in some mammals like muskrat.

(2) Mating and nesting territory:

Used for the above mentioned purposes, feeding is done else where. Example nest of three spined stickleback and of hawk.

(3) Mating territory:

Only mating is done in territory. For example singing ground of woodcock and leks of prairie grouse. Leks are symbolic territories defended by male only for mating. Resources of the area are not defended.

(4) Nesting territory:

Ony the nesting site is defended for parental care. Swallows defended territory the size of which is determined by the distance the bird can strike from the nest with its beak.

(5) Feeding territory:

Only food resource is defended as in Humming birds.

(6) Roosting territory:

Only roosting (sleeping) site is defended. *Example:* Monkey.

Territoriality and social dominance:

It is difficult to draw a sharp line between the two which represent degrees of manifestation of the same basic behavioural pattern. In domestic fowls inhabiting large areas during breeding season territoriality is seen among males. When confined to small areas dominant males adjust from territorial behaviour to social dominance. When space is increased the dominant males again become territorial.

Territorial making:

Often territorial boundary is marked by some characteristic odour of the dominant resident.

For example *Antilope cervicapra* markes the boundary with the secretion of preorbital glands. Tiger does territorial marking by way of urination.

Cost and benefit of territoriality:

Benefits are —

1) Longer lasting food supply to residents since few individuals have access to it.
2) Increased access to mates since females have easy access in the territory of dominant males.
3) A greater opportunity to rear offsprings at a high quality site.

Costs include:

1) Considerable energy is needed to patrol and defend or for dominance display.
2) It also takes away time from other activities such as foraging.

Economically speaking territoriality is expected to occur only when benefits outweight cost.

Review questions

I) Distinguish between:

(1) Loafing cover and roosting cover
(2) Territory and home range.
(3) Flight distance and charge distance.
(4) Autogrooming and allogrooming.
(5) Primary and secondary defense.
(6) Peck right hierarchy and peck dominance hierarchy.
(7) Mating territory and nesting territory.
(8) Deception and distraction.

II) Write notes on:

(1) Co-operative social group.
(2) Peck order.
(3) Types of territory.

(4) Cost and benefits of dominance.

(5) Grooming.

(6) Comfort movement.

(7) Body care.

(8) Conflict.

(9) Types of cover.

(10) Territoriality.

III) **Discuss with the help of suitable examples the basis of habitat selection. Add a note on the role of early experience in this regard.**

IV) **What is anting? Mention different types of anting. Discuss various theories explaining anting.**

V) **Discuss escape and defense behaviour in animals with reference to enemy recognition and reaction to enemies.**

VI) **What is dominance? Discuss important factors which influence the position of individual in social hierarchy.**

VII) **Discuss individual and social strategies relating to escape.**

VIII) **What do you understand by habitat switch? What cause shift in habitat choice.**

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-A Unit-II

Module No. - 30

REPRODUCTIVE BEHAVIOUR

Types of reproduction :

On the basis of the type of reproduction animals may be classified as asexual, sexual and parthenogenetic. Lower organisms may reproduce asxually through fission or through budding. Sexually reproducing creatures may be **hermaphrodite,** with both male and female functions in the same individual, or may be **gonochoris** with separate male and female individuals. A hermaphrodite may be a **simultaneous** one with male and female organs developing simultaneously in the same individual (as in some snails and most worms) or may be **sequential** in which case first one sexual function appears, then the other. When male function is followed by the female function the individual is said to be **protandrous** (same mollusks and echinoderms). When female function is followed by the male function then the individual is **protogynous** (some fishes).

Gonochorism may involve different types of mating system. The ways in which males and females associate varies greatly between species. Some form life long pair bonds to a single partner while others release their gametes in ocean without worrying, whose sperm will unite with whose egg. A simple classification based on the formation of pair bonds is given below:

		Male pairs with	
		1 female	> 1 female
Female pairs with	1 male	Polygyny	Polyandry
	>1 male	Monogamy	Polygyandry

Mating with several mates but with out forming any pairing bond is called **promiscuity.**

Parthenogenesis is a modified form of sexual reproduction in which embryo develops from an unfertilized egg. When haploid males are formed from unfertilized eggs then it is called **arrhenotoky** (e.g. Gamasid mites). When haploid females are formed it is called **thelytoky (e.g.** oribatid mite).

Sexual reproduction is of greater significance from evolutionary point of view.

REPRODUCTIVE STRATEGIES :

Natural selection favours species/individuals that produce maximum number of successful young in their life time.

The size and number of offspring produced varies from one spenies to other. Some produce large number of small offspring (e.g. insects) while others produce small number of large offspring (e. g. elephant). Similarly some species produce large number of offsprings in a single effort while there are others which reproduce throughout their life span.

Reproductive effort and allocation of energy

The nature of reproduction or reproductive effort is determined by the manner in which parents invest energy in the production of young.

Parents have finite amount of energy for reproduction at any given time. A certain minimum amount of energy is needed for offsprings to be viable. As energy is distributed among increasing number of eggs/offsprings, the fitness of Individual offsprings decline.

Energy available to and assimilated by parents is utilized in growth and maintenance and in reproduction. Reproduction includes all the activities associated with the preparation for reproduction **(pre reproductive efforts** such as, development of secondary sexual characters, fighting, ritualized dancing, nesting etc.), **reproduction proper** (such as production of eggs and litters, development of foetus within uterus, mating etc.) and **post reproductive** efforts (such as incubation, brooding, nursing and rearing of youngs, all sorts of parental cares etc.).

How much of the energy will be allocated for growth and maintenance and how much for reproduction is the central point of reproductive strategy adopted by an organism. If an organism directs more energy to reproduction it will have less energy available for growth and maintenance. As a result the individual grows slowly and often fails to survive. Tinkle (1969) has demonstrated that among lizards, species with larger fecundity have poor individual survival because more energy is channeled into egg production over a short period resulting in heavy physiological stress. Besides females that have larger clutches or produce several clutches a year are less able to escape predation. Similarly Lawlor (1976) found that the reproductive females of *Armadillium vulgare*, an isopod, had a lower growth rate than the non-reproductive females. While a female *Desmognathus ochropheus* invests 48% of its energy in reproduction including the energy stored in eggs and energy cost of brooding (Fitzpatrick, 1973), *Lacerta vivipara* invests only 7-9% annual energy assimilation in reproduction (Avery, 1975).

Basic types of reproductive efforts :

There are two extreme types of reproductive efforts, semelparity and iteroparity.

Semelparity :

To produce large number of offsprings in a single effort. Breeding once in the life span. By expending all energy in one reproductive effort they sacrifice future prospect of reproduction. This is typical of many insects. A special type of semelparity is seen in some (e.g. Cicada) where reproduction is delayed for many years and a single

bout of reproduction leads to death of the adult. This type is reproduction is called 'big bang' reproduction (Gadgil and Bossert, 1970).

Iteroparity

To produce a few young at one time and repeat reproductive efforts through life by conserving energy (e.g. penguin).

Most organism, however, strike a compromise between the two extremes. Frogs produce a reasonably large number of eggs per effort and also repeat their reproductive efforts.

Energy allocation and adjustment of clutch size:

There are various ways in which energy may be allocated in reproductive effort. Energy may be invested in pre & post reproductive efforts such as nest building or in parental care. It may be stored in eggs for the use of the developing youngs or through adjustment of clutch size.

D. Lack (1954) proposed that in birds clutch size has evolved through natural selection to correspond to the average largest number of young that parents can feed. Later M. Cody (1966) modified this concept by employing the principle of allocation of energy and proposed the hypothesis that clutch size results from different allocation of energy to egg production, avoidance of predator and competitive ability. Thus in temparate regions environmental instabilities keep the population below carrying capacity (κ) and natural selection favours an increase in the reproductive potential (γ) and therefore, clutch size on the average is larger than that in tropical regions. In tropics losses due to environmental stability are little and the populations are maintained at carrying capacity. As such more energy is invested in avoiding predation and towards intra and inter specific competition. Consequently little amount of energy is available for allocation in eggs and as such clutch size is smaller. This hypothesis is supported by other evidences too. For example lizards also have larger clutch size and mammals have larger litter size in temperate regions as compared to tropical regions.

γ selection and κ-selection

From this discussion on reproductive strategies two points are evident – (1) Among populations living in a harsh, severe and unpredictable environment in which mortality is largely density independent, individuals allocate more energy in reproduction and less in growth, maintenance and in competition. (2) Among populations living in stable, predictable and equitable environment in which mortality is largely due to density dependent factors and competition is keen, individuals allocate more energy to non reproductive activities. The former is known as '**γ-strategy** because the population remains in the rising part of the logistic growth curve with more emphasis in reproduction and the latter is known as κ-strategy because the population is maintained near carrying capacity with more emphasis on the survival of individuals.

κ-strategies and **γ-strategies** are under different selection pressures viz. , γ-selection and κ-selection. γ-selection favours productivity and κ-selection favours efficient use of environment.

A comparison of some attributes between γ and κ-selection is given in the following table :

Characteristics	γ selection	κ-selection
1. Duration of life	Short, usually < lyr	Long
2. Size of organism	Small	Large
3. Reproductive rate	High	Low
4. Number of offsprings produced	Large	Small
5. Development rate	Rapid	Slow
6. Means of dispersal	Highly developed	Poorly developed
7. Ability to colonize new area	Good colonizer	Poor colonizer
8. Parental care	Minimum	Pronounced
9. Reproductive effort	Semelparity	Iteropority
10. Ability to make use of harsh environment	Good	Poor
1. Relation with succession	Characteristic of early stage	Characteristic of later stage

Limitations of the concept :

1. The concept is rather relative same species may behave as γ-strategist in the early stage of colonization and as κ-strategist in the later stage.
2. Many spades show attributes of both γ-strategist & κ-strategists e.g. Cicada, Penguin etc.
3. 'κ' can not be expressed as function of life history trait whereas 'γ' is a function of life history trait such as age, fecundity, survivorship etc. Therefore, only γ is sensitive to selection pressure, κ is not influenced by selection pressure.
4. Thus γ & κ are not equivalent terms which "..... can be reduced to units of common currency" (*Stearns, 1977*).

SEX RATIO

When sexes are separate one may define a sex ratio within a population. Sex ratio is the number of males relative to females and in most species it tends toward a 1:1 ratio or is close to unity. Many species, however, deviate from this expected ratio of 50 percent of each gender. Population of wood lemming *(Myopus schisticolor)* in Finland contain only 25 percent males. In fig wasp *(Idarnes sp.)* the ratio of males to female is 22:235. The sex ratio of a population not only affects the potential reproductive rate but may affect social interaction and behaviour in many species.

Types of sex ratio :

The ratio at the time of conception i.e. among zygotes is referred to as **primary sex ratio** which tends to be 1 : 1. The ratio at the time of birth **is secondary sex ratio** which also mostly tends to 1 : 1 or is slightly tilted towards males. The sex ratio at the time of mating is the **operational sex ratio** which may be different from the primary and secondary sex ratio owing to differences in the mortality between the sexes. In most vertebrates operational sex ratio is tilted towards females.

In a study in the Western Canadian National Park Flook (1970) found that the ratio among foetuses of 'Elk' was 113B& to 100@&. Between $1^1/_2$ year to $2^1/_2$ years of age ratio of males abruptly dropped and it continued to do so unit it reached 85:100 and in some cases it dropped to 37:100. The greatest decline in the number of males occurred between the ages 7 to 14 years. The decline in number of females was much less rapid. In Indian tiger the sex ratio at the time of birth is 1B& : 1@& but in adult it shifts to 1B& : 4@& in favour of females (Schaller, 1969). Among passerine birds the ratio tends to remain tilted towards males but in Great tailed grackle of Texas although 50% of nestlings are males but the operational sex ratio in adults is only 29% males.

Reasons for shift in sex ratio :

It is not easy to explain why the sex ratio shifts from an equal ratio at birth to an unequal ratio in life. Perhaps a partial answer lies in factors related to the genetic determination of sex and to physiology and behaviour of two sexes.

Genetic cause :

The XY combination of chromosomes produce males in mammals and females in birds and some insects. In some insects males are XO. Adults with XY or XO are supposed to be more susceptible to disease, physiological stress and again the XX individuals because the harmful effects of a single recessive gene in the heterozygous condition may be marked by the homozygous condition.

Physiological and behavioural cause

Differences in the physiology and behaviour of different sexes also alter sex ratio. During breeding season male elks battle with each other for dominance of harem and mating. As a result of these activities the males end the

breeding season in poor state of health or in injured state. Similarly among birds females spend considerable energy in defending territory, building nests, incubating eggs and brooding the youngs.

Thus adult males in mammals and adult females in birds apparently expend more energy, subjected to greater stress and are more vulnerable to predation and other dangers. The ultimate effect is increased mortality and shift in sex ratio.

Mother / females influence sex ratio :

In bees, ants and wasp females are diploid and males are haploid. Here the sex ratio is under direct control of the mother by manipulating fertilization of the eggs.

Trivers and Willard (1973) have suggested that in some species females vary the sex ratio of offsprings in relation to their own breeding condition. Experimental confirmation of this idea has been provided by Maclure (1981) on wood rat *(Neotoma floridana).* In this speices normal sex ratio is 1 : 1 but when Meclure restricted food drastically during the first three weeks of lactation period, male offspring were selectively starved by the mothers (by rejecting their attempts to nurse) and the sex ratio after 3 weeks was altered to 0.5B& : 1@&. This she died because production of additional male will only increase competition and would be a wasted investment.

Temperature and sex ratio :

In *Xenopsylla cheopis* males outnumber females on rat when mean temperature of day is between 21°C – 25°C but in cooler days trend is reversed.

In some species sex determination is temperature dependent and thus temperature influences the sex ratio. In red eared slider turtle, *Trachemys scripta,* eggs when incubated at 20°C - 28.6°C produce all males. Between 29.4°C and 35°C produce all females and produce both sexes between 28.6°C - 29.4°C (1). Crews, *1996).*

Dispersal and sex ratio

Males and females often differ in their liability to disperse. Among birds usually females are the main dispersers but among mammals it is the males which disperse further from their place of birth. The tendency to stay near the place of birth is called **philopatry** and dispersal farther from the place of birth is called **natal dispersal.** As a consequence of philopatry and natal dispersal the population may become male dominated or female dominated.

Number of species with predominant dispersal by different sex *(Greenwood, 1980).*

Taxa	Mainly by males	Mainly by females	No sex difference
Aves	3	21	6
Mammalia	45	5	15

Among insects it is the male that is usually more active disperser. For example female winter month is wingless and only males migrate. But there are exceptions too. In beet leaf hopper it is the females which disperse further.

Local mate competition and sex ratio :

In some hymenopterans sex ratio is controlled facultatively in response to **local mate competition**, a situation in which males compete with sibling for mating and the sex ratio is shifted in favour of females. Production of greater number of males would be a wasted effort because one male can fertilize several females. In fig wasp males have no wings and mate with neighbouring females, which usually are their sisters (sibling mating), the ratio is heavily tilted towards females (22B& : 235@&).

Local resource competition and sex ratio :

Competition among offsprings for resource may lead to a male biased sex ratio in those forms where females are philopatric (home loving). South African bush babies produce more sons than daughters since production of more daughters would mean greater competition among females, including mother, for resource which are philopatric.

Significance of 1:1 sex ratio

According to R. Fisher (1930) when males & females occur in equal frequency the fitness are balanced and there is no selective pressure to alter sex ratio. However, when a sex is rarer in frequency it enjoys greater fitness because they compete with fewer individuals of the same sex for mating. Selection pressure under such situation will favour production of the individuals of rarer sex to bring it back to 1:1 sex ratio. Otherwise the rarer sex will disappear. This theory however fails to explain the situation in parthenogenetic forms.

SEXUAL SELECTION :

Darwin (1871) in his book 'The Descent of Man' considered sexual selection as a form of natural selection. According to him sexual selection depends on the advantages which certain individuals have over others of the some sex within a species solely in respect of reproduction. He reasoned that females make a definite choice of sexual partner and males have acquired certain characters or courtship behaviour to gain an advantage over other males not for the struggle for existence but solely for the purpose of attracting females.

Types of sexual selection :

There are two ways in which an individual can gain advantage over individual: through intrasexual selection by way of intersexual selection.

Intrasexual selection (selection within sex)

It is the direct competition between the males. Intrasexual selection is primarily a matter of male rivalry. There is no instance of female rivalry in animals.

In red deer, *Cervus elephus* stags grow antlers each year in rutting season. They challenge each other for ownership of females which are devoid of antlers. The successful male takes possession of all the females which are herded into harem. Other males are denied access to female. Females have no choice or little choice in the selection of males.

The challenge starts with roaring, initially slowly and then loudly. The challenger usually retreats if the challenged one could roar harder. This is done to assess the fighting potentiality of each other. If the challenger is able to match the roar of the harem owner, then the two approach and walk parallel to each other, assessing each other more closely with respect to the body size. Most of the contest end here. This is sometimes followed by a ritualized fighting interlocked trial strength. The final phase of contest is fighting proper which may be quite dangerous often resulting into permanent injury or even death. Usually older and bigger stag which have larger antler wins in the competition.

Since stages do not have antlers during non breeding season and since females do not have antlers it is unlikely that antlers are important protective organs against predator. Moreover, growing antlers each year is physiologically disadvantageous. For their growth, each year, large quantity of nutrients like phosphorus and calcium salts are required which cause burden on stag's metabolism. Therefore, it seems growing antler has more intrasexual selection advantage then the ordinary natural selection pressure.

Fighting proper versus assessment of the competitive ability of the opponent:

Competition for mate thus involves both fighting as well as assessment of competitive ability of the opponent. Fighting has some obvious disadvantages. Firstly it is risky and may lead to injury, permanent disability or even death. Secondly it may distract the animal from other important activities and behavior. While the harem owner is busy in fighting another male may attempt to steal or court some of the unguarded females. In either of the cases

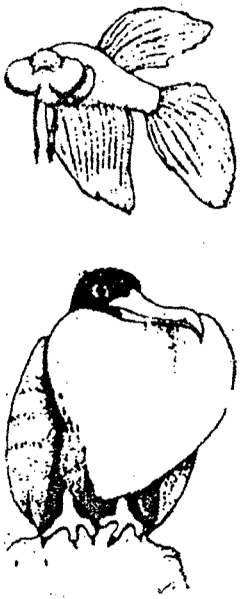
fighting is disadvantageous from the point of view of natural selection as well as sexual selection. Therefore, natural selection has led to the evolution of various modes of assessment of the fighting potentials of opponents. There is no point in engaging in a fight which is certain to be lost.

Modes of assessment of fighting potential :

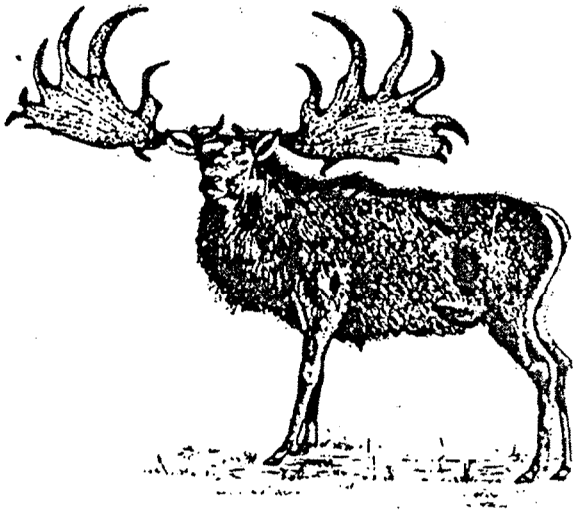
Various modes are discussed here under :

(1) Look for the direct indication :

Look for the direct indication of the fighting ability such as body size or weapon. The selective advantage is obvious. However, sometimes natural selection favours 'cheaters'. For example if an animal exaggregates its body size it may deter its rivals from fighting. For example male pigeon blows the neck region to look bigger in size. In simese fighting fish, *Betta splendens,* fins are exaggregated.



Body size is exaggerated in these courtship displays of the Siamese fighting fish (*Betta splendens*) and frigated bird (*Fregata magnificans*) (From *The Oxford Companion to Animal Behaviour,* 1981)



The extinct gaint deer *Megaloceros giganteus*

Exaggregation, sometimes, may be beyond the point of usefulness. The extinct giant deer, *Megaloceros giganteous,* which lived during ice age had antlers measuring more than 3 meters in span. Here intrasexual advantages overweighed the advantage in terms of natural selection and over size of antlers caused extinction. Not only the growing cost was huge, particularly when food was scarce in ice age, huge size of antler was a hindrance against running away from predators.

(2) Display of power :

Display of power like roaring in red deer, barking in dogs.

(3) Display of aggressiveness

Threatening postures are often adopted to drive away the rival, like raising trunk in elephant, showing of horns in deer or making canines prominent.

(4) Ritualized combat :

Ritualized trial of strength, like head on dashing and pushing by buffolos and sheeps, pushing matches of beetles, are also performed to assess fighting potentialities.

Role of females in male rivalry :

The female normally does not directly exercise a choice in male rivalry but her behaviour indirectly has that effect. In elephant seals, *Mirounga angustirostris,* when a male attempts to copulate a female she protests loudly drawing attention of other males who attempt to interfere. A male is successful in copulating only if he is dominant and can ward off the challenge. The female intensifies the competition among males by her protest and thus ensures a worthy male for her to mate.

Intersexual selection (selection between sexes)

Darwin assumed that only female selects her sexual partner and males compete for her attention. This is typically true for animal kingdom. Therefore, intersexual selection is also called **female choice.** But there are exceptions too.

(a) Female choosing male:

Darwin was convinced that females select mates with special features, those which are more attractive or have special behavioural ability like singing. Various views have been put forth subsequently to explain the phenomenon.

(1) Fisher's hypothesis (Runaway selection hypothesis)

Ronald Fisher (1930) provided the most popular explanation of intersexual selection. He pointed out that females which mate with attractive males will tend to have attractive sons provided the attractive characters are inherited. These sons, in turn, Will be successful in attracting more females and reproducing with them. Thus a female which selects a mate on the basis of sexual attractiveness *(epigamic qualifies)* will likely to produce more grandchildren as compared to a female which mates with a less attractive male. This is, therefore, advantageous to both females as well as well as the male mating with it from the point of view of reproductive selection.

What are epigamic characters?

Fisher suggested that initially females are attracted to male character, that have survival value. These characters then become exaggregated during evolution through intersexual selection. For example female birds

might show a preference for male birds with "well maintained plumage" since that would indicate not only a direct survival value in flying efficiently but also the fact that it could afford time and trouble to maintain his plumage in order. Evolution might subsequently result into formation of 'superplumage', as in peafowl or african paradise birds, which would be irrelevant for the survival value of male.

(2) Handicap hypothesis

Amos Zahavi (1975) suggested that females prefer males precisely because they carry a handicap (e.g. big horn, large tail & superplumage, big chelicera etc.). The reason is as they can carry a handicap they must be robust individuals and therefore must be a worthy male. Thus handicap is an advertisement for male quality.

The handicap hypothesis has been criticized by *T.R. Halliday (1978)* and *J. Maynard Smith (1976)* who believe that a handicap if inherited will ultimately effect survival of the individual adversely. *Halliday (1983)* on the contrary maintains that females choose males on the basis of two distinct classes of beneficial characters : material qualities and genetic qualities. In the former category falls such characters which would ensure survival of both females and the brood (like nest building ability, aggressiveness etc.). In the later category will fall epigamic features like attractiveness singing capability (crocking of frogs) etc. which are of little direct survival value.

(4) Parental investment hypothesis

Robert Trivers (1972) related sexual competition to parental investment in offspring. He used the term parental investment to indicate the effort in terms of time and resource in rearing offspring. He argued that where one sex invests considerably more than the other, members of the sex which invest less will compete among themselves for the opposite sex which will choose. In animals usually it is the female who pays the greater cost in the reproduction. She provides the developing embryo with the nutrition and cares for the young. Therefore it is important that she should mate with the correct male of the correct species and therefore she is the chooser.

(b) Male choosing female :

There are exceptions too where males of the species invest more than the females. In three spine stickleback, ***Gasterosteus aculeatus*** male establishes territory, defends it, build nest, dances, cares for eggs after fertilization and protect them from other males. Here male chooses the female which is ready to spawn and others are

Male and female African Paradise birds (*Vidua paradisea*). Outside the breeding season the male looks similar to the female (After Halliday, 1980)

driven off. Some authors (like *Ridley & Rechton, 1981)* believe that female stickleback also exercise a choice because they prefer to spawn with a male whose nest already contains eggs.

SEXUAL DIMORPHISM

Darwin (1871) was convinced that the differences between the sexes (sexual dimorphism) are the result of sexual selection. Males have developed characters and behaviour to attract, allure or excite females. Courage, aggressiveness, ornamentation, weapon, sound production and glands for emitting odour are some such male characters, Males are thus armed, ornamented and attractive as compared to females who are unarmed, unornamented and unattractive.

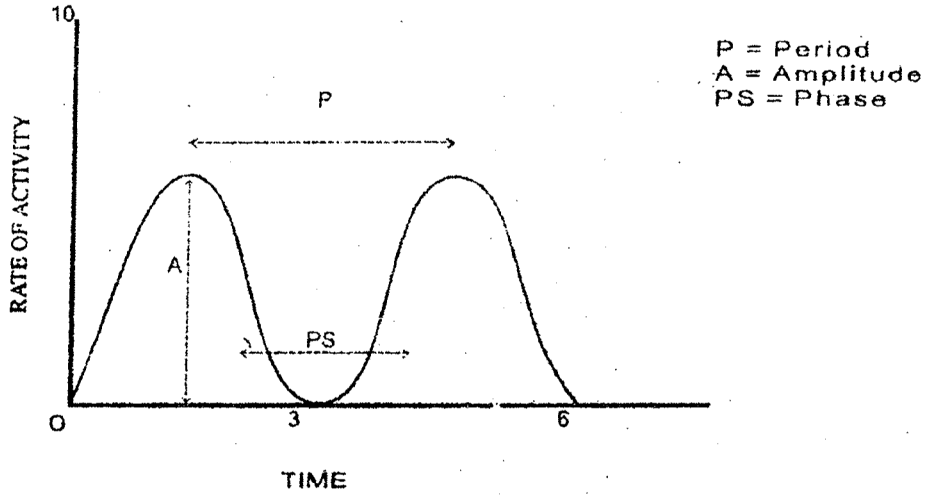
As long as males are competeting for the attention of females, the females are in a position to choose among alternative males and there is no selective pressure on them to become sexually attractive. A female can, therefore, afford to remain unattractive.

But this argument does not hold good for monogamous species which also have pronounced sexual dimorphism (some birds) and where opportunities of females to select male are severely limited. Besides, all sexual creatures are not dimorphic.

Perhaps sexual dimorphism is the result of natural selection and it is due to different parental roles played by each sex. For example female metatherians have special adaptations to retain new born in pouch and eutherians to feed babies with the milk. Males are incapable of doing so, therefore, their morphology is also different. Moreover bright colouration of males help in protection of females by distracting attention of the predator.

BIOLOGICAL RHYTHMS

Behavioural activities that occur at regular intervals are known as biological rhythms or biorhythms. Some well known examples are migration, courtship displays, resting behaviour of birds and periodic mating in oestrus mammals. Each biological rhythms is composed of repeating units called cycle. The length of time required to complete an entire cycle or the time interval is the rhythm's 'period'. The period may vary from several minutes (6-8min. feeding cycle in *Arenicola)* to about one year. The magnitude of the change in activity rate during a cycle is the 'amplitude'. Any specified recognizable part of a cycle is called a 'phase'.



Biological rhythms are the external manifestation of biological clock and are regulated by them.

Types of rhythms

1. Ultradian rhythms or epicycles

These are short duration cycle usually of a period of few minutes. *Arenicola marina* living in intertidal zones-have a feeding cycle of 6-8 minutes.

2. Tidal rhythms :

The cycle coincides with the tidal cycle. Usually the period from one low tide to another low tide is 12.4 hrs. Many species of crabs (e.g. fiddler crab, *Uca)* forage during the low tide in the intertidal zone and return to burrow during high tide.

3. Lunar rhythms and semilunar rhythms

Such rhythms are based on 29.4 day cycle (lunar) and 14.7 day cycle (semilunar) of the moon. The midge *Clunio marinus* of western Europe emerge from pupa, swarm in large number, mate and lay eggs on exposed sea weeds every 15 days at the ebb of spring tide. Laraee hatches in the marine environment. Ant lion *Myrmeleon obscurus* larvae show maximum activity at the time of full moon.

4. **Circadian rhythms :**

These cycle coincide with about 24 hrs. duration of the day. The term is derived from two Latin words circa = about & dies = day. Flying squirrel has a rhythm of slightly less than 24 hours where as hamster has a rhythm of slightly over 24 hours. Majority of the animals show ***daily activity cycle.*** Some exhibit peak activity during day time (diurnal), some are active during night (nocturnal) while others show peak activities during twilight, dawn and/ or dusk (crepuscular)

5. **Circannual rhythms:**

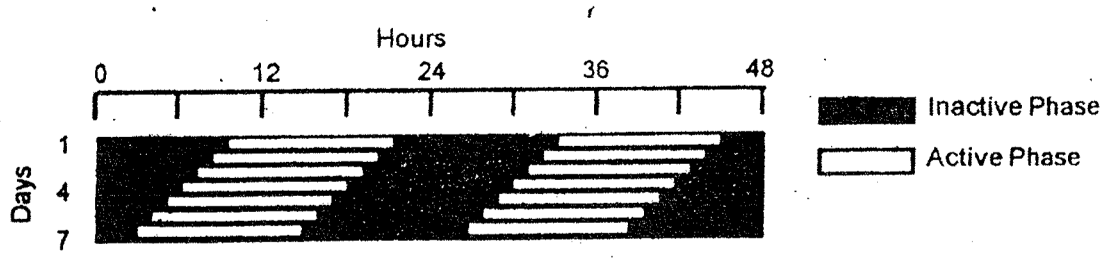
These rhythms are more or less of one year in duration. Hibernation, migration ofbirds, diapause in insects are examples of such rhythm.

BIOLOGYCAL CLOCK

Biological rhythms are regulated by biological clock. 'Biological clocks' are internal timing mechanisms that involve both self sustaining physiological pacemakers and environmental cyclic synchronizers (zeitgebers). Rhythms involving an internal clock or pace maker are ***endogenous rhythms*** (e.g. *Arenicola)* whereas those controlled by external factors like photoperiod or temperature are ***exogenous rhythms.*** Endogenous rhythms persist even in the absence of obvious environmental cues while exogenous rhythms fail to persist in absence of external cues.

Endogenous pace maker:

One of the basic characteristics of biologycal rhythm is the existence of an internal self - sustaining pace maker. If an animal is placed under constant conditions, for example constant darkness it will exhibit its natural periodicity independent of environmental cues. This is called ***free-running rhythm*** or free-running state. This is an indirect evidence of the existence of an endogenous pace maker. Since most of the animals show a circadian rhythm of slightly less than or more than 24 hours, a free run will result into a shift, delay or advancement of the peak activity or phase. ***Entrainment*** is the process by which such shift in daily rhythm is reset by environmental cues so that it is in phase with the 24 hours cycle. During entrainment, a circadian rhythm of more than 24 hours is ***phase advanced*** and a rhythm of less than 24 hours is ***phase delayed.*** Cues that provide information to animals of environmental variables are called *zeitgebers* (time givers). Zeitgebers are the entraining agents — those cyclic environmental cues that can entrain free-running endogenous pace maker.



Phase advance under constant darkness

Location of pace maker :

Experimental evidences suggest that pace maker is located in the optic lobes of cockroach and sea hares. In mammals the pacemaker system is located in the suprachiasmatic nuclei (SCN) and possibly a second one is situated in the ventromedial hypothalamus (VMH).

Advantages of biological clock

There are several reasons why it may be advantageous to have a biological clock rather than responding directly to environmental changes. There are —

1. Anticipation of the environmental changes with enough time to prepare for the behaviour.
2. Synchronization of the behaviour with some event that can not be sensed directly.
3. Continuous measurement of time so that time compensated orientation is possible. For example worker bee must be able to measure the time which has passed since it has discovered the nectar to communicate this information to its fellow bees.

Review questions :

1. Discuss the two basic types of reproductive strategies adopted by animals for allocation of energy in reproduction? Explain the relationship between energy allocation and adjustment of clutch size in birds.
2. Distinguish between intrasexual and intersexual selection. With the help of an example discuss the phenomenon of male rivalry in sexual selection. What are the various modes of the assessment of fighting potential of opponents? Do the females have any influence on the outcome of male rivalry?
3. With the help of suitable example discuss 'female choice' in sexual selection. Explain sexual dimorphism with reference to sexual selection.
4. Explain the phenomenon of 'female choice' and discuss the ideas of Ronald Fischer (1930), Amos Zahavi (1975) and Maynard Smith (1976) in this regard.

5. What is biological rhythm? Explain different types or Biological rhythm. Discuss biological clock with reference to endogenous pace maker.
6. Discuss various factors influencing shift in sex ratio.
7. Write notes on :
   (i) Tilting of secondary sex ratio in mammals in favour of females.
   (ii) Relationship between sex ratio and dispersal.
   (iii) Big bang reproduction.
   (iv) Types of mating system.
   (v) Female's influence on sex ratio.
   (vi) Entrainment & zeitgebers.
8. Distinguish between :
   (i) Arrhentotoky & Thelytoky.
   (ii) Semelparity and Iteroparity.
   (iii) γ-strategy and κ-strategy / γ-selection & κ-selection.
   (iv) Primary sex ratio and operational sex ratio.
   (v) Ritualized combat & fighting proper.
   (vi) Endogenous rhythm and Exogenous rhythm.
   (vii) Phase advanced and phase delayed entrainment.

Suggested reading :

A mentioned in module - 28.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Group-B Paper-III Unit-I
Module No. - 31

"IMMUNOLOGY"

INTRODUCTION :

The term immunology is derived from the Latin word "Immunitas". Historically immunity means protection from infectious diseases. The cells and molecules responsible for immunity constitute the immune system, and their collective and coordinated response to the introduction of foreign substances is the immune response. We know that many of the mechanisms of resistance to infections are also involved in the individual's response to noninfectious foreign substances. Furthermore, mechanisms that normally protect individuals from infections and eliminate foreign substances are themselves capable of causing tissue injury and disease in some situations. Therefore, a more inclusive and modern definition of immunity is a reaction to foreign substances, including microbes, as well as macromolecules such as proteins and polysaccharides, without implying a physiologic or pathologic consequence of such a reaction. Immunology is the study of immunity in this broader sense and of the cellular and molecular events that occur after an organism encounters microbes and other foreign macromolecules.

Historians often credit *Thucydides*, who lived in Athens during the fifth century B. C. as having first mentioned immunity to an infection that he called "Plague". The concept of immunity may have existed long before, as suggested by the ancient Chinese custom of making children inhale powders made from the crusts of skin lesions of patients recovering from smallpox. Immunology, in its modern form, is an experimental science, in which explanations of immunologic phenomenon are based on experimental observation and the conclusions drawn from them. The evolution of immunology as an experimental discipline has depended on our ability to manipulate the function of the immune system under controlled conditions. Historically, the first clear example of this, and one that remains among the most dramatic ever recorded, was *Edward Jenner's* successful vaccination against smallpox. Jenner,

an English physician noticed that milkmaids who had recovered from cowpox never contracted the more serious smallpox. Based on this observation, he injected the material from a cowpox Pustule into the arm of an 8-years old boy. When this boy was later intentionally inoculated with smallpox the disease did not develop.

Jenner's landmark treatise on Vaccination (Latin Vaccinus) was published in the year 1798. It led to the widespread inducing immunity to infectious diseases. An eloquent testament to the importance and progress of immunology was the announcement by the World Health Organization (WHO) in 1980 that smallpox was the first infectious disease that had been eradicated worldwide by a programme of vaccination.

The immune system works in three fundamentally different ways by humoral immunity, by cellular Immunity and by secretion of stimulatory proteins, called Lymphokines. The term "humor" refers to a fluid, and humoral immunity relies on molecules are proteins collectively called immunoglobulin (Ig). They constitute 20% of the proteins in the blood. A single immunoglobulin molecule is called an "antibody", but "antibody" is also used to mean many individually different molecules all directed against the same target molecule. Humoral immunity also involves complement, a set of proteins that are activated to kill bacteria both nonspecifically and in conjunction with anibody.

I. **Antigen :** It is a substance (usually protein) introduced into living organism that elicit antibody formation, or that induce a specific immune response. Antigens are large complex molecule, molecular weight more than 10KD, More appropriately these are called an 'Immunogen'.

II. **Antibody :**

a) When foreign substance (antigen) enter the blood stream they trigger a defence mechanism — the immune response which results in, synthesis of an extremely important groups of glycol-Protein which binds with the antigen. The protein structure is known as antibody.

b) Antibodies are globular proteins or glycoproteins in nature, present on B-cell membrane and also are secreted by plasma cells, serve as the effectors of humoral immunity for elimination of antigens.

III. **Antigenicity :** Antigenicity is the ability to bind with antibodies or with cells of immune system. This binding is highly specific the immune components are capable of recognizing various physicochemical aspects of the compound. The binding between antigen and immune components involves several weak forces operating over short distances.

IV. **Immunogenicity:** The ability to induced an immune response requires that a compound be foreign to immunized individual, possess a certain minimal molecular weight and a certain degree of chemical complexity.

B/T cells + antigen —* Effector B/T cells + memory B/T cells.

V. **Immunogen :** Substance that is capable of inducing an immune response (antibody response) is called immunogen, Sometimes it is used as a synoname of antigen but emphasis is on the ability to elicit a response rather than on reactivity with antibody. The most potent immunogen are reactivity with antibody. The most potent immunogen are proteins, polysaccharides but lipid, nuclic acid and synthetic polypeptides can also be immunogenic.

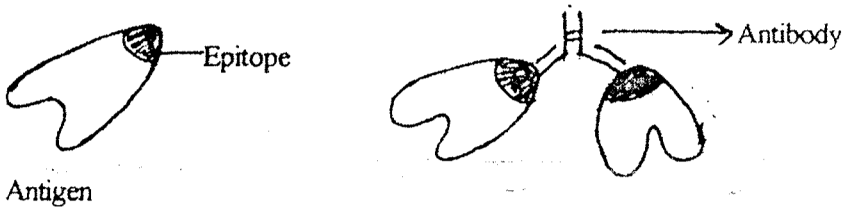
1. Nature of antigen :

i. Nonself foreign substance e.g.-proteins, mucoprotein, polysaccharides, virus, pollen grains, fungal spores and some glycolipids.

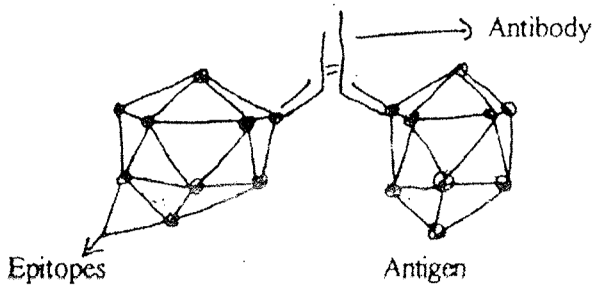
ii. When introduced it has an ability to produce specific antibody and induce an immune response.

iii. Antigens are large complex molecules with high molecular weight.

iv. An antigen may possess several epitopes (antigenic determinants). The number of epitopes on an antigen is termed as "valance". Thus antigen may become univalent, multivalent. Multivalent antigen generates greater immune response.



(a) Single epitope (Univalent)



(b) More than one epitope

v. Antigen can augment immune response when it binds with an adjuvant.

vi. Low molecular weight antigen binds with carrier molecule (high molecular weight protein) to become immunogenic. e.g. — hapten.

vii. Proteins and carbohydrates are usually more immunogenic than lipids and nuclic acid.

1. EPITOPE or Antigenic determinant: Sites either on or within the antigen with which antibodies & T-cell receptors reacts.

i. The epitopes of one antigen is usually different from those of another. Some antigen has repeated epitopes.

ii. Epitopes are molecular shape, recognized by antibodies and T-cell receptors of immune system.

iii. Each antibody receptor recognize one epitope rather than the whole antigen.

iv. The size of epitope recognize by a lymphocyte (B-cell) cannot be larger than the size of antibody binding site.

v. The epitopes are preferentially located on the membrane of cells that exist as transmembrane glycoprotein which are coded for —in man by a gene complex.called MHC located on chromosome-6.

vi. The antigen are called Human Leucocyte Antigen (HLA) because they are found on peripheral blood lymphocyte. Size and shape of antigen determinants and the chemical nature of the antibody determine the nature of interaction.

N. B.—*Most readily accessible epitope lends to induce a more powerful immune response called "Dominant epitope ".*

1.2 Structural aspects of biologic antigen :

An antigen can be defined as any substance that may be specifically bound by an antibody molecule or T-cell receptor. This differs from the original definition of antigen as a molecule that generates an antibody almost every kind of biological molecule, including simple sugars, lipids and hormones as well as macromolecules such as carbohydrate, phospholipids, nuclic acid and Protein can be serve as antigens. However, only macromolecules can initiate lymphocyte activation necessary for antibody response. Molecules that generate immune responses are

called "**Immunogen**". In order to generate antibody specific for small molecules, immunologist commonly attach such small molecules to macromolecules before immunization. In this system small molecules is called "hapten" and the macromolecule, usually a foreign protein (high molecular weight) is called a "carrier". The hapten-carrier complex, unlike free hapten can act as immunogen. In general macromolecules are much larger than the antigen binding region of an antibody. Therefore, the antibody binds to a specific portion of macromolecule called "antigen-determinant" or 'epitope'.

1.3 Allosteric effects :

Proteins commonly display different determinant that are spatially well separated and two seperate antibody molecules can bind to the same antigen without influencing each other; such determinant are said to be **non-overlapping**. In other cases, the first antibody bound to an antigen may sterically interfere with the binding of the second and the determinants of the antigen are said to be **overlapping**. In rarer cases, binding of the first antibody may cause a conformational change in the structure of antibody, influencing the binding of second antibody by means other than steric hindrance — such interaction are called Allosteric effects.

In the case of phospholipid or the complex carbohydrates the antigenic determinants are entirely a function of covalent structure of macromolecules.

However, in the case of nuclic acid and in the case of protein the non-covalent folding of the macromolecule may contribute to the formation of the determinant. In proteins, epitopes formed by adjacent amino acid residues in the covalent sequence are called "Linear determinants." It is estimated that, in a protein-antigen the size of the linear determinants may be accessible to the antibody in the native folded protein. If they appear on the surface or in a region of extended conformation More often, linear determinants may be inaccessible in native conformation and appear only when the protein in denatured. In contrast, conformational determinants are formed by amino acid residues from separated portion of linear amino acid sequence that a juxtaposed only upon folding. In theory denatured proteins could transiently give rise to conformational determinants. However, such determinants ate too short lived unless they are maintained by energetically favourable interaction such as those, found in native proteins.

Proteins may be subjected to covalent modification such as phosphorylation or specific proteolysis. These modifications by altering the covalent structure can produce new antigenic epitopes. Such epitopes designated as "new antigenic determinants.

1.4 Antigenic Determinant-Paratope :

The immune system does not recognize an infectious agent or foreign molecule as a whole. But reacts to structurally distinct areas. — the antigenic determinants. Thus exposer to a microorganism will generate an immune response to many different epitopes. The antiserum produce will contain different antibodies reactive with each determinant. This wilt ensure that the individual will be protected from the microorganism by producing a response to atleast a few of the possible determinant. If the host only reacted to the organisms as a whole then failure to react to this one site would have dire consequences that is it would not eliminate the pathogen.

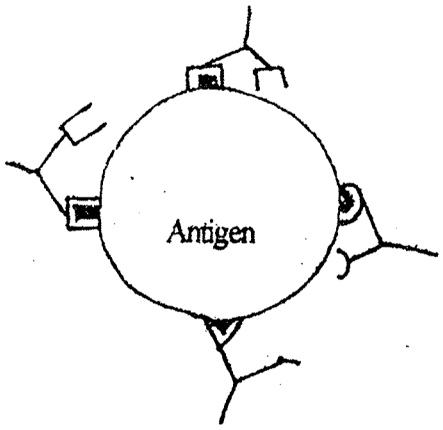


Fig. Str. of antigens-foreign material. antigen will contain a mosaic of str.-formed its chemical constituent that can induce an immune response. Ab with a complementary shape will be form that combine specifically with these epitopes on the antigen.

Response to antigen involves the specific interaction of components of the immune system, antibody and lymphocytes interact with specific epitopes on the antigens. The lymphocytes have receptors on their surface that functions as the recognition units — on **B-lymphocytes** surface bound Ig act as receptor and on **T-lymphocytes** the recognition unit is the T-cell recepor. The interaction between Ag-Ab is governed by the complementarity of the electron cloud surrounding the determinant. For over all configuration of the outer electrons, not the chemical nature of the constituent residues determines the shape of the epitope and its complementary paratope part of antibody or T-cell receptor that interacts with epitope.

There is no reason why the shape generated by the electron cloud surrounding a group of amino acid should not be indentical to that surrounding part of a lipid or polysaccharide. The better fit between the epitope and the paratope, the stronger the non-covalent bonds form and consequently the higher the affinity of the interaction.

Antigenic determinants have to be topographical that is composed of structures on the surface of molecule or cells, and can be constructed in two ways. They may be contained within a single segment or primary sequence or two or more antigenic determinants formed by a single sequence of amino acids — called sequential epitope. Those are assemble for residues are apart in the primary sequence but brought together on the surface by the foldling of the molecule into its native conformation i.. e. epitopes of a protein that are composed of amino acids that are close together in the three-dimentional structure of the protein but may not be near each other in the amino acid sequence — called **conformational epitopes or Non-sequentional epitope.**

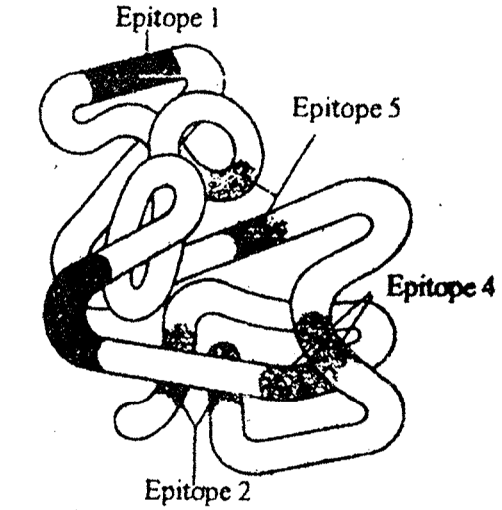


Fig. - 1 : Model of epitopes on lysozyme. The shaded areas are the specific epitopes. They are composed of chain segments that are either linear (epitopes 1 and 2) or conformational (epitopes 3-5). (Adapted from Klien) : *Immunology : The Science of Self—Nonself Discrimination.* (New York, Wiley, 1982, p-356.)

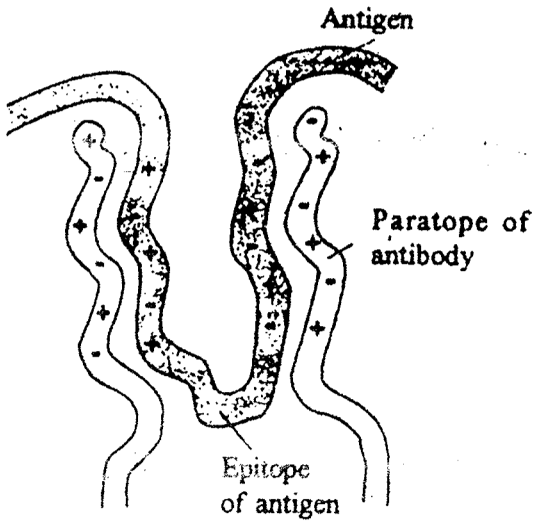


Fig. - 2 : The interaction of an epitope and its homologous paratope. The pattern of electrical charges influences the interaction between antigen and antibody. As the pattern becomes more complementary, the force of attraction becomes proportionately stronger, and the affinity of the two reactants increases.

(COURTESY : Hyde, Immunology)

2. REQUIREMENTS FOR IMMUNOGENICITY :

Immunogenicity is determined by four properties of the immunogen: its **Foreignness, molecular size, chemical composition & heterogenicity and susceptibility to antigen processing and presentation.**

- **Foreignness :**

In order elicit an immune response, a molecule must be recognized as nonself by the biological system. Much of the ability to recognize self molecules is thought to arise during lymphocyte development by expo of immature lymphocytes to self components. Any molecule that has not been exposed to immature lymphocytes during this critical period may be later recognized as nonself or foreign. When an antigen is introduced into an organism, the degree of its immunogenicity depends on the degree of its foreignness. Generally, the greater the phylogenic distance between two species, the greater the structural disparity between them.

e.g. - BSA is not immunogenic when it is injected to cow but strongly immunogenic to rabbit.

- **Molecular size :**

There is co-relation between the size of macromolecule and its immunogenicity. The best immunogen has a mol. wt. approaching 100 kd. A few substance, in exception, of mol. wt. less than 1 kd proven to be immunogenic.

Mol. wt. < 1 kd → Non-immunogenic (e.g.-Penicillin, progesterone)

Mol. wt. 1-6 kd → maybe immunogenic (insulin, ACTH)

Mol. wt. > 6 kd → immunogenic (albumin, tetanus toxin etc.)

- **Chemical composition & Heterogenicity :**

Synthetic homopolymer tends to lack of immunogenicity regardless of their size. Copolymers of sufficient size, if they contain two or more different amino acid, are immunogenic. The addition of aromatic amino acid, such tyrosine or phenylalanine, profoundly enhances the immunogenicity of these co-polymers.

e.g..-

glutamic $\bar{a}$ + lysine → require minimum mol. wt. 30-40 kd for immunogenicity.

glutamic $\bar{a}$ + lysine + tyrosine → require minimum mol. wt. 10-20 kd. for immunogenicity.

glutamic $\bar{a}$ + lysine + tyrosine + phenylalanine → require only 4 kd for immunogenicity.

All four levels of protein configuration — Primary, secondary, tertiary, quaternary — contribute contribute to structural complexity of protein and hence, affect its immunogenicity.

- **Susceptibility to Ag. processing and presentation :**

The development of both humoral and cellular immune response requires interaction of T-cell with antigen that has been processed and presented together with MHC molecule. To T_H, ag, must be presented with class II MHC molecule on the ag. presenting cells and to Tc cell ag. must be presented with class I MHC molecule. Macromolecules that can not be degraded and presented with MHC molecule are poor immunogenic. Degradative enzymes within ag presenting cells can degrade proteins containing L-amino acid not polymers of D-amino acids and thus, are poor immunogenic.

Large, insoluble macromolecule are more immunogenic than small soluble one because the larger molecules are more readily phagocytosed and processed.

TYPES OF IMMUNITY :

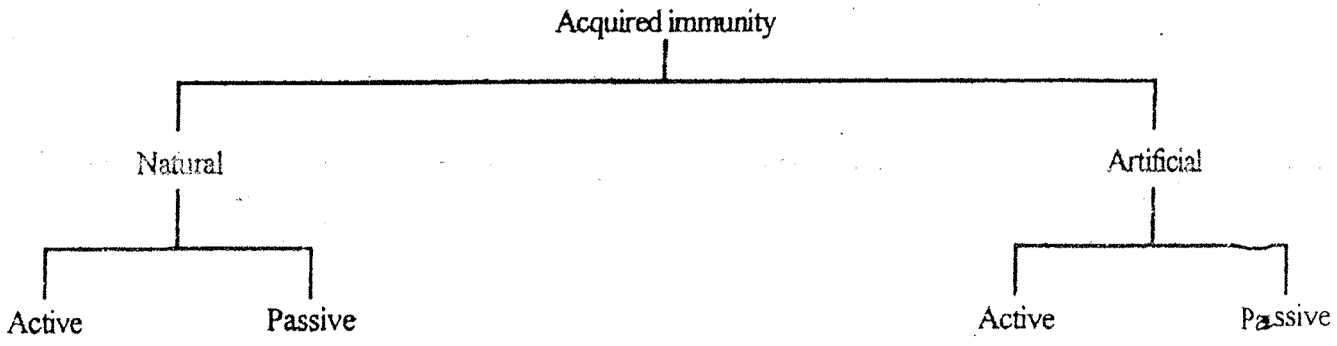
- **Acquired Immunity :**

In contrast to innate immunity, which is an attribute of every living organism, acquired immunity is most specialised form of immunity. It developer late in evolution and is formed only in vertebrates. The various elements that participate in innate immunity donor exhibit specificity against the foreign agents they encounter; In contract, acquired immunity always exhibits such specificity. As its name implies, acquired immunity is a consequence of an encounter with a foreign substance. The first encounter with a foreign substance that has penetrated the body triggers a chain of events that induces an immune response with specificity against that foreign substance (*E. Benjamini*).

Acquired immunity are two types —

(I) Naturally acquired immunity (NAI) and

(II) Artificially acquired immunity (AAI).



I. Natural Immunity (NAI):

It is naturally present by birth in an individual and is not acquired through previous contact with pathogen. This is actually inherited resistance to infection and is concerned with species, races or individuals. The power of resistance to infection varies with different individual of the same species and with age also.

a) ACTIVE : Naturally acquired active immunity occurs when an individuals immune system contacts an appropriate antigenic stimulus during normal daily activity.

The immune system responds by producing antibodies and sensitised lymphocytes that inactivate or destroy the antigen. The immune response produced may either be life long.

e.g. - Measles or Chicken pox.

b) PASSIVE : Naturally acquired passive immunity means transfer of antibodies from one host to another. For example - antibody of a Pregnant women, Pass across the placenta to her foetus. If the mother is immune to diseases like polio, the placental transfer renders the newborn immunity to polio.

Certain other antibodies (especially IgA) can pass through first secretion (colostrum, highly Proteineeious) of mammary gland.

Naturally acquired Passive immunity generally exist only for short period.

II. Artificially Acquired Immunity (AAI):

Acquired immunity is not present by birth with a individual. This type of immunity is acquired during the life time of an individual as a result of previous infection.

a) ACTIVE : Artificially acquired active immunity is an adaptive resistance that is built up when an individual is immunised by 'vaccine'. Previously prepared antigens are injected into the susceptible individual who produces antibodies and specialized lymphocytes. This process or technique is known as vaccination or immunization. A vaccine is nothing but a preparation of either killed microorganism or inactivated microbial toxin. These are injected or administered to induce immunity artificially.

This type (AAI) of immunity develops very slow but persist for several years.

In active immunization as the name implies, the immune system playes an active role, with poliferation of antigen-reactive 'T' and 'B' cells resulting in formation of memory cells. Active immunization with various types of vaccines has played an important role in the reduction of deaths from infectious diseases, especially among children.

b) PASSIVE : Artificially acquired passive immunity develops against an infective agent induce by the

administration of antibody against that agent. The antibody (Ab) is produced & derived from host and is introduced to another host.

We know, the first requirement for a compound to be immunogenic is foreignness. The more foreign the substance, the more immunogenic it is.

e.g. - A person bitten by snake might be injected with antibodies from a horse that is immune to snake venom.

This type of immunity lasts for short period usually a few weeks only. The antibody (Ab) molecules after entering into host constantly breakdown without forming new ones.

4. *DEFENSIVE BARRIERS :*

Innate immunity can be seen to comprise four types of defensive barriers, these are-

I. **Anatomical barrier**

II. **Physiological barrier**

III. **Phagocytic barrier**

IV. **Inflammatory barrier**

1. Anatomical barriers :

Anatomical barrier that tends to prevent the entry of microorganism, is an organisms first line of defence against infection. The skin and the surface of mucous membrane are under this category.

Skin : The skin consists of two distinct layers, — a thinner outer layer — the **epidermis** and a thicker inner laver — the **dermis**.

The epidermis contains several layers of tightly placed epithelial cells. The outer epidermal layer consists of dead cells and is filled with keratin. The dermis which is composed of connective tissue contains blood vessels, hair follicles, sebaceous glands and sweat glands.

The sabaceous gland is associated with hair follicles and produce an oily secretion called sebum, consists of **lactic acid** and **fatty acids** which maintain the pH of skin 3-5. This low pH inhibit the growth of microorganism. Not only does intact skin prevent the penetration of most pathogens, but its low pH also inhibit most bacterial growth.

The surface of mucosal membrane :

The conjunctival and alimentary, respiratory and urogenital tracts are lined by mucus membrane which consist of an outer epithelial layer and an underlying layer of connective tissue. Although most pathogens enter the body by binding to and penetrate mucous membrane, a number of nonspecific defense mechanism tend to prevent their entry. For example, saliva, tear, mucosal secretion act to wash away potential invader and also contain antipathogenic substances.

The viscous fluid, mucus, which is secreted by the epithelial cells of mucus membrane, entrap foreign microorganism. In lower respiratory tract & GI tract is covered by hair-like protrusion of epithelial cell membrane called cilia whose synchronous movement propels mucus-entrapped microorganism from this tract.

In addition nonpathogenic organisms tends to colonize epithelial cells of mucosal surface. These flora generally outcompete pathogens for attachment sites on the epithelial cell surface & for necessary nutrients.

II. Physiological Barriers :

The physiological barriers that contributes to innate immunity include temperature, pH and various soluble factors.

- # Many species are not susceptible to certain disease because their normal body temperature inhibits growth of pathogens.
- # Gastric acidity is an innate physiological barrier to injection because very few microorganism can survive in low pH of gastric contents.
- # Lysozyme, a hydrolytic enzyme found in mucus and in tears, is able to cleave peptidoglycan layer of bactrial cell wall.
- # Interferon comprises a group of proteins produced by virus-infection. It has ability to bind to nearby cells and induce a generalized antiviral state.
- # Complement is a group of serum protein persist in an inactive state when become active it can damage the membrane of pathogen, either destroying them or facilitating their clearance.

III. Phagocytic Barriers :

Phagocytosis is one type of endocytosis which is the general term of uptake by a cell of material from its environment. Phagocytosis is the ingestion of particular material, which may include whole pathogenic microorganism. In phagocytosis, a cell's plasma membrane expands around the particulate material to form a large vesicle, called phagosomes. Most phogocytosis is conducted by specialized cells, such as blood monocytes, neutrophils & macrophages etc.

IV. Inflammatory Barriers :

Tissue damage caused by a wound or by an invading pathogen induces a complex sequence of events collectively known as inflammatory response.

The Roman physician 'CELSUS' described four cardinal signs of inflammation as **rubor** (redness), **tumor** (swelling), **calor** (heat), **dolor** (pain). 'ALEN' added a fifth sign, **functio laesa** (loss of function).

The cardinal signs of inflammation reflects three major events of aninflammatory response :

Vasodialation — an increase in the diameter of vessels — of nearby capillaries occurs as the vessels carry blood away from the affected area constrict resulting in engorgement of capillary network. The engorged capillaries are responsible for tissue redness (erythema) and an increase in tissue temperature.

An increase in capillary permeability facilitates an influx of fluid and cells from the engorged capillary into the tissue. The fluid that accumulates (exudate) has a much heigher protein content than fluid normally released from the vasculature. Accumulation of exudate causes tissue swelling (edema).

Influx of phogocytes from the capillaries into the tissues is facilitated by the increased capillary permeability. The emigration of phagocytes is a complex series of events, — margination, diapedesis or extravasation and chemotaxis.

The events in the inflammatory response are initiated by a complex series of events involving a variety of chemical mediators. Among chemical mediator released in response to tissue damage are various serum proteins called **acute phase protein** whose concentration drammatically increase in tissue damaging infection.

Histamine is one of the principle mediator of inflammatory response which binds to receptors on nearby capillary and venules causing vasodialation and increased permiability. Another important inflammatory mediator is kinin, normally present in blood plasma in an inactive form. Tissue injury activates this peptide which cause vasodialation & increased permiability of capillaries. A particular kinin, called brady kinin also stimulates pain receptors in skin. Pain normally causes an individual to protect the injured area.

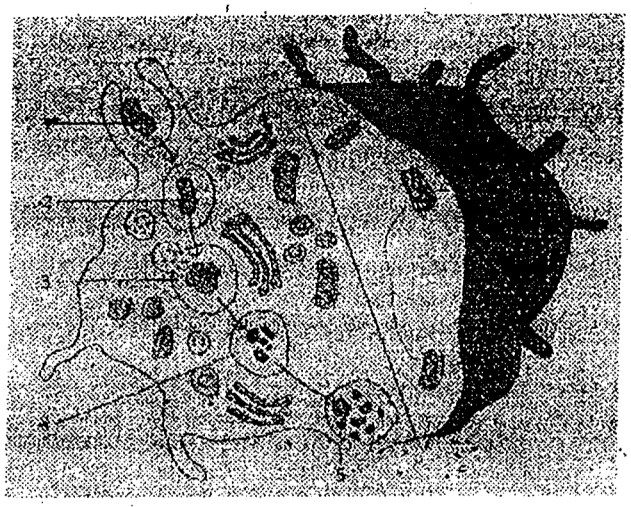


Fig. - 3 : Phagocytosis of bacteria. Schematic diagram of the steps in phagocytosis : (1) attachment of a bacterium to long membrane evaginations, called pseudopodia ; (2) ingestion of bacterium forming a phagosome, which moves toward a lysosome ; (3) fusion of the lysosome and phagosome, releasing lysosomal enzymes into the phagosome; (4) digestion ingested material ; and (5) release of digestion products from the cell.

(COURTESY : KUBY, Immunology)

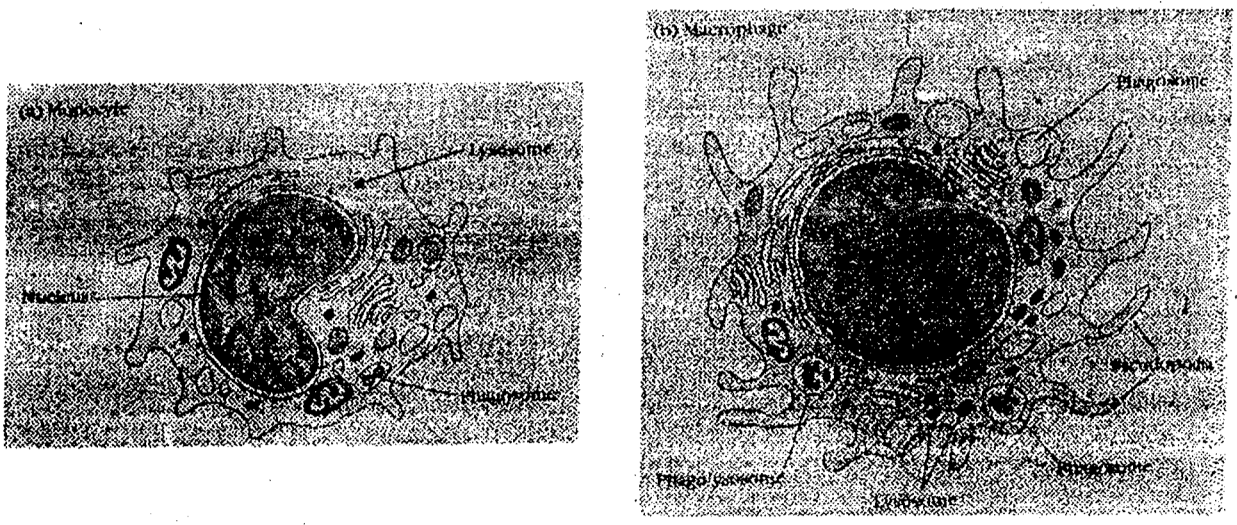


Fig. - 4 : Drawing showing typical morphology of a monocyte and macrophage. Macrophages are five- to tenfold larger than monocytes and contain more organelles, especially lysosomes.

(COURTESY : KUBY, Immunology)

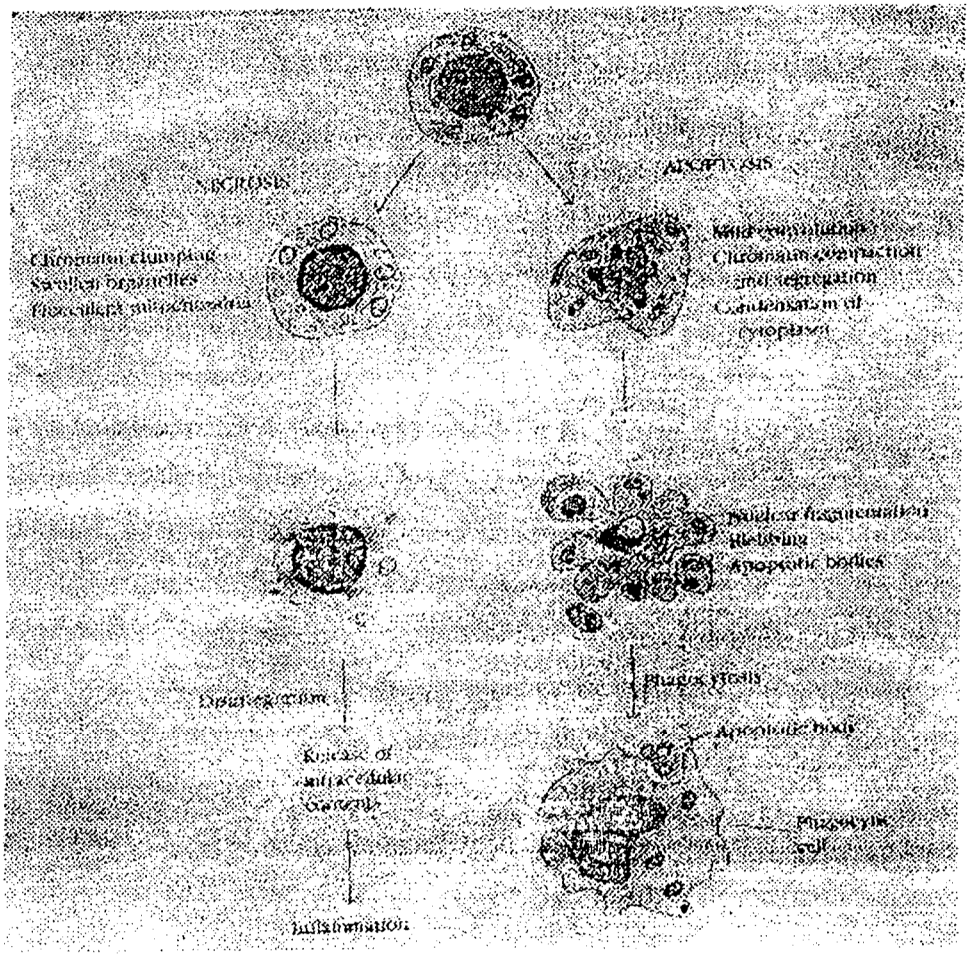


Fig. - 5 : Comparison of morphologic changes that occur in apoptosis and necrosis. Apoptosis, which is associated with the programmed cell death of hematopoietic cell, does not induce a localized inflammatory response. In contrast, necrosis, the process leading to death of injured cells, results in release of the intracellular contents, which induce a localized inflammatory response.

(COURTESY : KUBY, Immunology)

5. *BASIC STRUCTURE OF ANTIBODY MOLECULE :*

It has been known since the turn of the century that antibodies — the effector molecules of humoral immunity reside in the serum. The first evidene that antibodies were contained in particular serum protein fraction came from a classic experiment by **A. Tiselius** and **E.A. Kabat,** in 1939.

In the 1950's and 1960's experiments by **Rodney Porter** and by **Gerald Edelman** elucidated the basic structure of the immunoglobulin molecule and they got noble prize in the year 1972.

Antibody molecules have a common structure of four peptide chains (Fig.- 6). This structure consists of two identical Light (L) chains, polypeptides of about 25,000 mol. wt. and two identical heavy (H) chains, larger polypeptides of mol. wt. 50,000 or more. Each light chain is bound to a heavy chain by a disulfide bond, (s-s) a strong covalent linkage between the two polypeptides, and by a combination of such noncovalent interactions as salt linkages, hydrogen bonds, and hydrophobic bonds. Similar noncovalent interactions and disulfide bridges link the two identical heavy and light (H - L) chain combinations to each other to form the basic four-chain immunoglobulin structure, the dimer $(H - L)_2$. There are two region present in an immunoglobulin molecule, one is aminoterminal region (NH_3^+), in which immunogen can bind with antibody molecule and rest is carboxyterminal region (COO^-). The first 110 or so aminoacids of the amnio-terminal region (NH_3^+) of a light or heavy chain varies greatly among antibodies of different specificity. These segments of highly variable sequence are called 'V' region 'V_L' in light chain and 'V_H' in heavy chain. All of the differences in specificity displayed by different antibodies can be traced to difference in the aminoacid sequence of 'V' regions. In fact, most of the differences among antibodies fall within areas called **Complementary determining regions (CDRs),** and it is these CDRs, on both light and heavy chains, that constitute the antigen-binding site of the immunoglobulin molecules.

The regions of relatively constant sequence beyond the variable regions have been dubbed 'C' regions, 'C_L' on the light chain and 'C_H' on the heavy chain. Immunoglobulins are glycoproteins with few exceptions, un sites of attachment for carbohydrates are restricted to the constant region. We do not completely understand the role played by glycosylation of immunoglobulins, but some effects of carbohydrate attachment have emerged ; it affects the rate at which immunoglobulins are cleared from the serum by the liver, and it probably increase of immunoglobulins.

Different sites of Antibody :

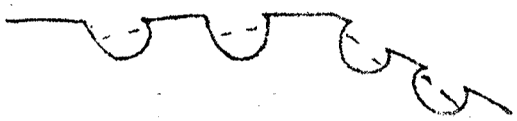
LIGHT CHAIN : Antibody molecule consist of two light chain (L), molecular weight of light chain is 20-25 KD and are composed of 200 amino acid. There are two distinct type of light chain, one is Kappa (K) light chain and other is Lambda (k) light chain. Any one individual of a species produce both type of light chain but the ratio

varies with the species. In mouse 95% K, 5%λ and in human 60% K and 40% λ. But in a particular Ig molecule the L-chain are always either both 'K' or both 'λ'. The cystein at the carboxyl end of both K and λ chain are extremely similar.

HEAVY CHAIN : Molecular weight of heavy chain is 50-55 KD and are composed of 400 amino acid. Five different heavy chains (H) have been identified from different Ig molecule. These are not only differ in antigenic reactivity and carbohydrate content but also in size and biological functions. Any individual of species makes all classes of H-chain in proportion characteristic of the sp. but in one antibody molecule both H-chain are identical (2α or 2μ).

Immunoglobulins	**Heavy chain (H)**
IgA	α
IgM	μ
IgG	γ
IgE	ε
IgD	δ

DOMAIN : The presence of interchain disulfide (S - S) bond approx. at regular interval of about 100 - 110 a.a. tends to the formation of loop in the peptide chains. This folded and globulin structure are known as domains. Each domain is designated by a letter that indicates whether it is an 'L' or 'H' chain and the number indicates its



position. The first domain on L and H chain is highly variable. The second and subsequent domains are more constant in amino acid sequence and are designated as CH_1, CH_2, CH_3 etc.

HINGE REGION: The **Hinge region** is the portion of the H chain between CH_1 and CH_2 domains rich in proline (Fig. - 6) residues. There is no homology between the hinge and other immunoglobulin heavy chain domains, and its sequence is unique for each immunoglobulin type and subclass. IgM and IgE donot posses a hinge but have one more C_H domain.

(a) In this region, interchain disulfide bonds form between the arms of the Fab fragments, preventing them from folding and therefore rendering this portion of the molecule highly **susceptible fragmentation by enzymatic attack.**

(b) The hinge region is highly flexible and allows for movement of the Fab arms in relation to each other. This mobility explains why native antibody molecules do not activate complement.

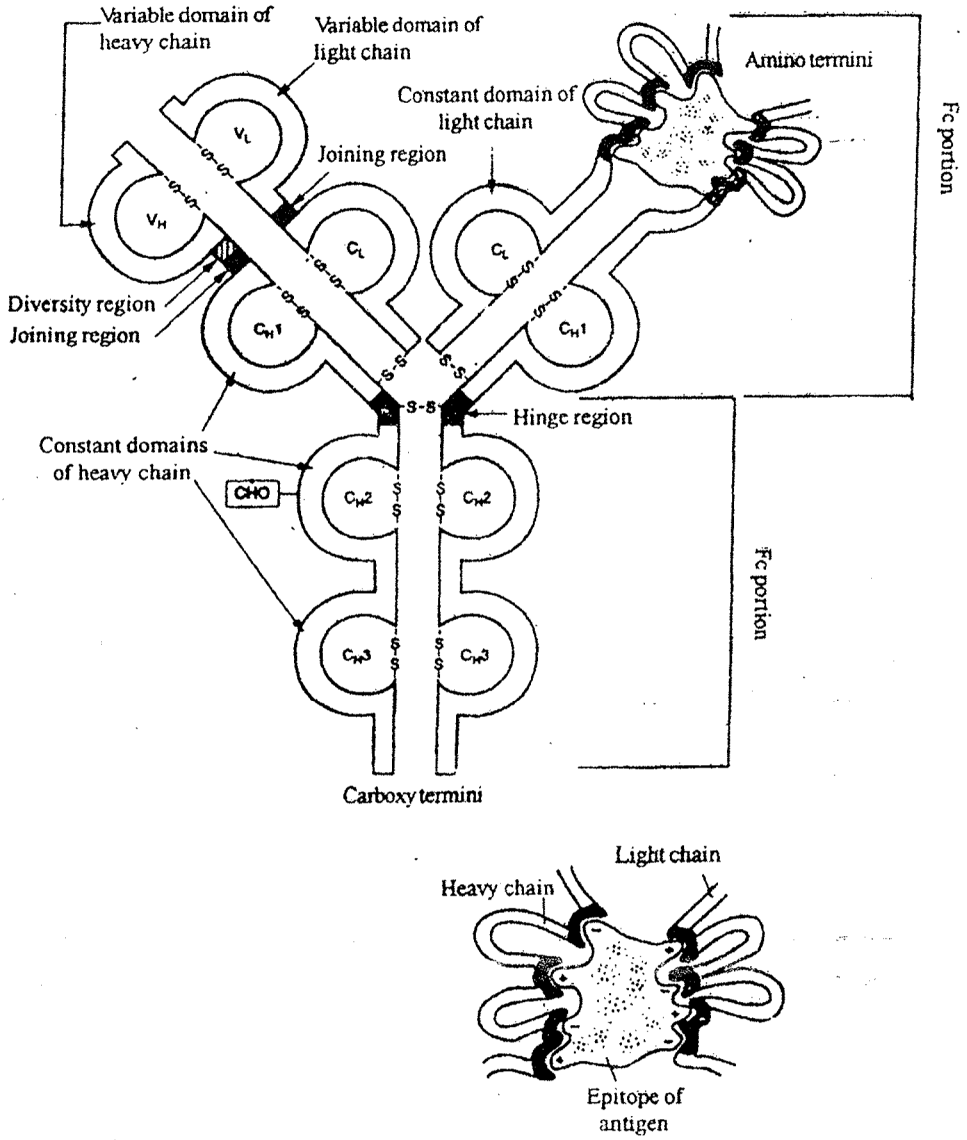


Fig. - 6 : Basic unit (monomer) of the IgG molecule, consisting of four polypeptide chains linked covalently by disulfide bond (S-S), with interchain disulfide linkages as well. The loops correspond to domains within each chain. V = variable domain ; C = constant domain ; L = light chain ; H = heavy chain ; CHO = carbohydrate side chain. The inset shows the hypervariable complementarity-determining regions as shaded. The framework regions of each variable domain in the paratopic region are not shaded. Similar hypervariable regions are found in the α and β chains of the T-cell receptor for antigen.

(COURTESY : KUBY, Immunology)

5.1 FUNCTION OF ANTIBODY :

a. OPSONIZATION : IgG is an opsonizing antibody (from the Greek **opsonin**, which means to prepare for eating). It reacts with epitopes on microorganisms via its Fab portions, but it is the Fc portion that confers the opsonizing property. Many phagocytic cells, including macrophages and polymorphonuclear phagocytes, bear receptors for the Fe portion of the IgG molecule. There cells adhere to the antibody — coated bacteria by virtue of their receptors for Fc. The net effect is a zipperlike closure of the surface membrane of the phagocytic cells around the organism, as receptors for Fc and the Fc region on the antibodies continue to combine, leading to the final engulfing and destruction of the microorganism. (Fig.- 3)

b. ACTIVATION OF COMPLEMENT : Activation of complement result in the release of several important biologically active molecules and leads to lysis if the antibody is bound to antigen on the surface of a cell. Some of the complement components are also opsonins ; they bind to the target antigen and thereby direct phagocytes which carry receptors specific for these opsonins, to focus their phagocytic activity on the target antigen. Other components from the activation of complement are chemotactic ; specifically, they attract phagocytic cells. All in all, the activation of complement by IgG has profound biologic effects on the host and on the target antigen, whether it is a live cell, a microorganism or a tumor cell.

c. ANTIBODY-DEPENDENT CELL-MEDIATED CYTOTOXICITY (ADCC) : The Ig molecule plays an important role in ADCC. In this form of cytotoxicity, the Fab portion binds with the target cell, whether it is a microorganism and the Fc portion binds with specific receptors for Fc that are found on certain large granular lymphocytic cells called natural killer (NK) cells. By this mechanism the IgG molecule focuses the killer cells on their target, and the killer cells destroy the target not by phagocytosis but with the various substances that they release.

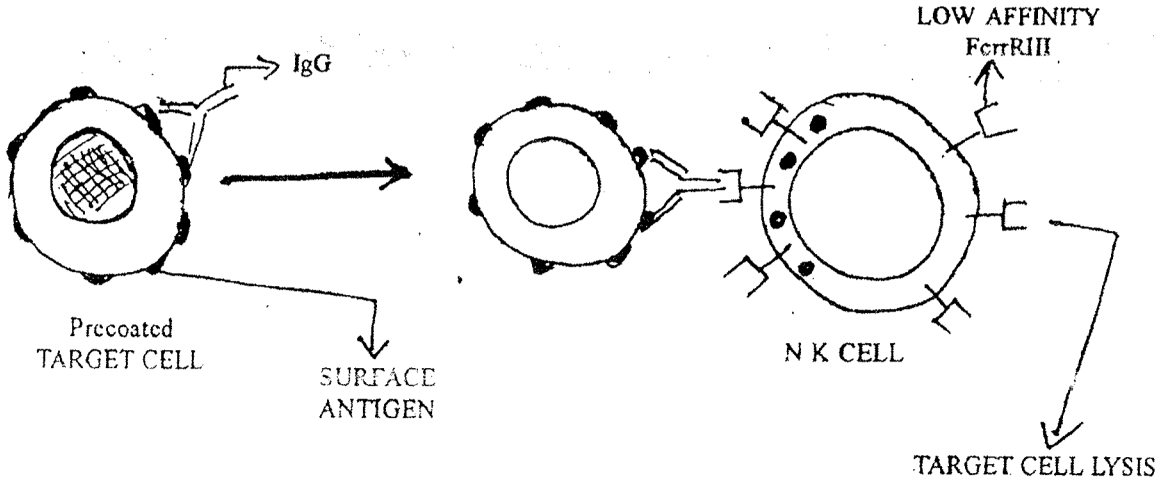


Fig. -7

d. **NEUTRALIZATION OF ANTIGEN BY SECRETED ANTIBODY :** The initial discovery on antibodies arose by analyzing the humoral factors that protect immunized hosts against microbial toxins. Many injurious agents, such as toxins, drugs, viruses, bacteria and other parasites, initiate cell injury by binding to specific cell surface receptors. Secreted antibodies can sterically hinder this interaction by binding to antigenic determinants on the agent (or, less commonly on the cell receptor), thereby neutralizing the toxic or infectious process. This action may mediated by antibodies of any isotype and experimentally can also be mediate by Fab or F (ab')2 fragments.

e. **HYPERSENSITIVE REACTION :** Fc portion of IgE binds with very high affinity to receptor found on mast cell and basophils. Antigen bind to Fab portion of IgE causing cross linkage to these calls. Induce degranulation and release of histamine, leukotrienes and other pharmacological active compound trigger the hypersensitive relations. (Fig.- 4-16 from Immunology by KUBY, page- 102)

f. **PASSAGE THROUGH PLACENTA :** IgG (except IgG2) isotype can pass through placenta, enabling the mother to transfer her immunity to the foetus.

Breast milk contain secretory IgA protect new born during 1st month of life when immune system of infants is not fully functional.

5.2 PROTEOLYTIC ACTIVITY OF ANTIBODY :

The hinge region in between CH1 and CH 2 domain which is highly proline rich is more susceptable to proteolytic attack.

a) When an antibody molecule is treated with enzyme **"Pepsin" (pH - 4-4.5)** it can cleave the molecule just below the interchain disulfide bond.

b) **"Papain"** enzyme cleaves the molecule just above the interchain disulfide bond i.e. in the aminoterminal region.

6. *IMMUNOGLOBULIN ISOTYPES :*

IgG → It is the major immunoglobulin in human serum constituting 80 % of the total immunoglobulin pool. IgG molecule is a monomer, consisting of 2 gamma (γ) heavy chain& 2 & or λ light chain IgG diffused more rapidly in the extra vascular body spaces than the other immunoglobulins. IgG is the only immunoglobulin molecule capable to cross the placenta and provide naturally occur passive immunity to new born.

IgG has four subclasses according to their decreasing average serum concentrations :

Serum concentration IgG_1 — 66 %

IgG_2 — 23 %

IgG_3 — 7%

IgG_4 — 4%

a) IgG_1, IgG_3 & IgG_4 readily crossed the placenta and play an important role in protectig the developing foetus.

b) IgG_3 is the most effective complement activator IgG_2 is less efficient and IgG_4 is not able to activate complement et al.

c) IgG_1 & IgG_3 bind with a high affinity to Fc receptor on phagocytic cells and thus mediates opsonization.

d) IgG_4 has an intermediate affinity and IgG_2, has an extremely low affinity.

IgG has four isotypic sus-classes-termed as IgG_1, IgG_2, IgG_3, IgG_4. The differences lies in the heavy chains (H chains) of the molecule. Certain, functional & biological differences exist between the class, for example IgG_3 does not combined with certain protein (staphylococcal A), IgG_2, does not cross the placenta realy and IgG_4 does not bind to complement in the classical pathway nor does it bind to monocytes.

IgA → IgA, constitutes only 10-15 % of the total immunoglobulin. It is the predominant immunoglobulin class in exteral secretion such as breast milk, saline, tears, and mucous of the bronchial & digestive tracts. This secretory IgA (SIgA) is always in the dimeric form and is composed of two basic 4 chain units (2 light chain & 2 α heavy chain), a 'J' chain and the secretory component.

The secretory component is a part of the molecule that transport the dimer produced by a submucosal plasma cell to surface. It is not only facilitates passage through out the epithelial cells but projects the secreted molecule from proteolytic digestion.

The 'J' chain polypeptide in IgA is identical to that found in pentameric IgM.

In man most of the serum IgA occurs as a monomer but in many other mammals it is mostly found as a dimer. The dimer is held together by a 'J' chain which is produced by the antibody-producing plasma cells.

IgA :

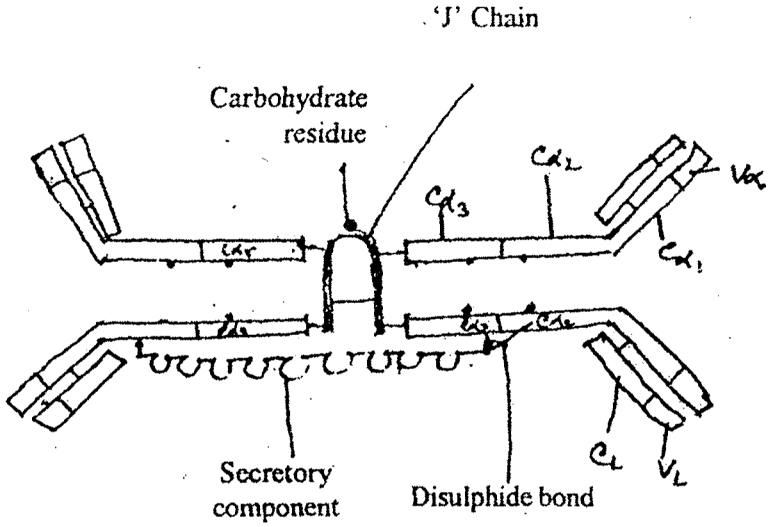


Fig. -8

A. Fig.- structure of secretory IgA.

Fig. I. In the dimeric form of IgA molecules are joint together by 'J'-chain. The secretory component held in place by noncovalent interaction and by a disulfide bond to the $C\alpha_2$ domain.

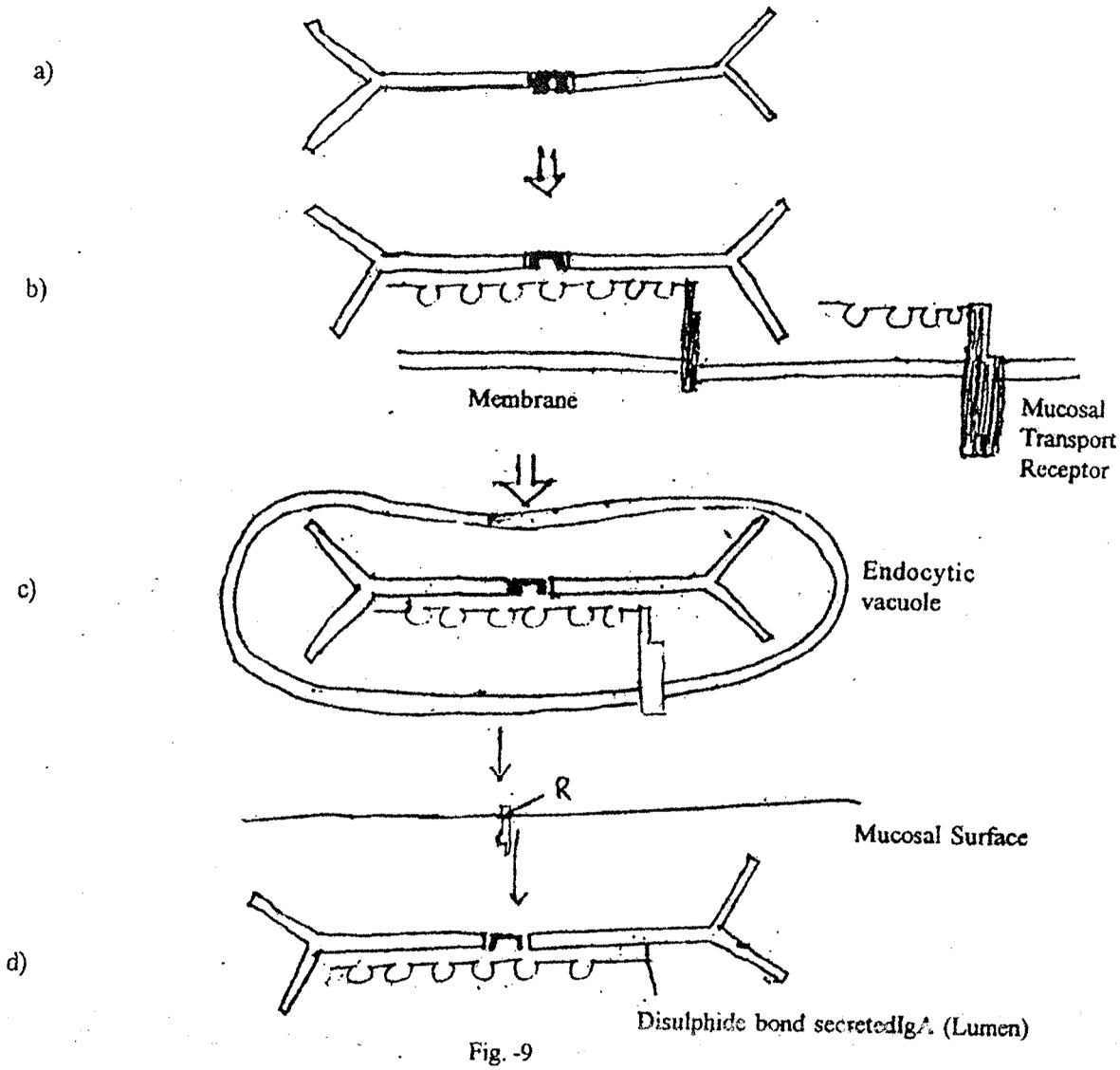


Fig. -9

Fig. II. - Transport of secretory IgA

Ref. : Immunology (7th Ed.) by Donald M. Weir & John Stewart.

IgM → IgM is the largest gest immunoglobulin molecule in size. This account for about 6 % ofthe total immunoglobulin pool, with an average serum concentration of 1.5 mg/ml. Monomeric IgM with the mol. wt. approximately 180,000 is expressed as membrane bond antibody from B cells. IgM is secreted by plasma cells as a pentamer in which 5 monomer units are held together by disulphide bonds that links their carboxy-terminal heavy chain domains (Cμ4) & (Cμ3 - Cμ3). The 5 monomer subunits are arranged with their Fc regions in the centre of

the pentamer and the 10 antigen binding sites on the periphery of the molecule. Each pentamer contains an additional Fc linked polypeptide called 'J' (joining) chain which is bonded to the carboxy-terminal cysteine residue of two of 10μ heavy chain.

IgM is the 1st immunoglobulin class produced in a primary respons to an antigen i.e., it is the fastest growing molecule. Because of its pentameric structure with 10 antigen binding sites, serum IgM has the higher valency than the other isotypes. Pentameric IgM is more efficient than other isotypes in binding such antigen because its contents high valency. For example, R.B.C.s are incubated with specific antibody, they clump together into large aggregates in a process called agglutination. To achieve this agglutination requires thousand times more IgG than IgM. IgM is also more efficient than IgG for activating complement. Complement activation requires 2 Fc region in close proximity and IgM molecules fulfills this requirement. Monomeric IgM has a hydrophobic sequence attached into the 'C' terminal region of the heavy chain to anchor the molecule into the cell membrane of the B lymphocytes. Since, IgM is large it does not cross the placenta and mainly confind to intravascular pole and the 1st antibody type be produced during an immune respons.

IgE → This immunoglobulin isotypes estimated about less than 1 % of the total immunoglubolin pool. It consist of 2 light chain K or λ & two heavy chains. Like the IgM molecule, IgE has an extra C_H domain and molecular wt. approximately 200,000 dalton. The extra domain of IgE heavy chain, by which it attach with usually high affinity to specific receptor on mast cell and basophils. When antigen reappears, it combines and cross links the IgE molecule on the mast cell granules and the ensuing (occur) symptom of anaphylaxis. The presence of serum components responsible for allergic reactions was first demonstrated by K. PRAUSNITZ & H. KUSTNER in 1921. The major biological role of IgE apparently involved in protection against invasion by parasites, such as helminths. A protection achieved by activation of the same acute inflammatory respons seen in a more pathologic form in the anaphylatic response.

NOTE : ANAPHYLAXIS.

Immediate hypersensitivity response to antigenic change, mediated by IgE & mast cells. It is a life-threatening allergic reaction, caused by release of pharmacologically active agents.

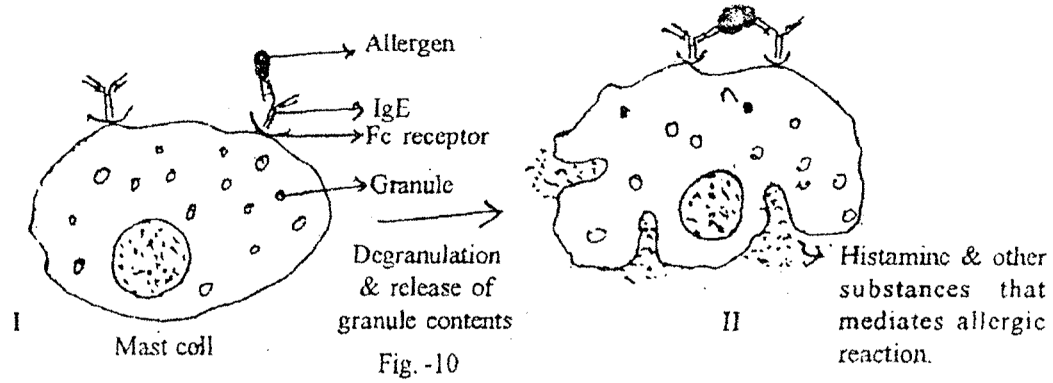


Fig. -10

IgD → This immunoglobulin isotype present in sodium in trace amount (30 µg / ml). It composed of the basic unit with heavy chain delta. Because of an extended hinge region this protein is very susceptable to proteolytic attack. The major function of IgD has not yet been determined. But together with IgM it found abundantly on the surface of B-Lymphocyte, perhaps it helps to differentiate B-Lymphocyte to plasma cells. IgD is a monomer and show similarity to IgG, it never produced any memory cells.

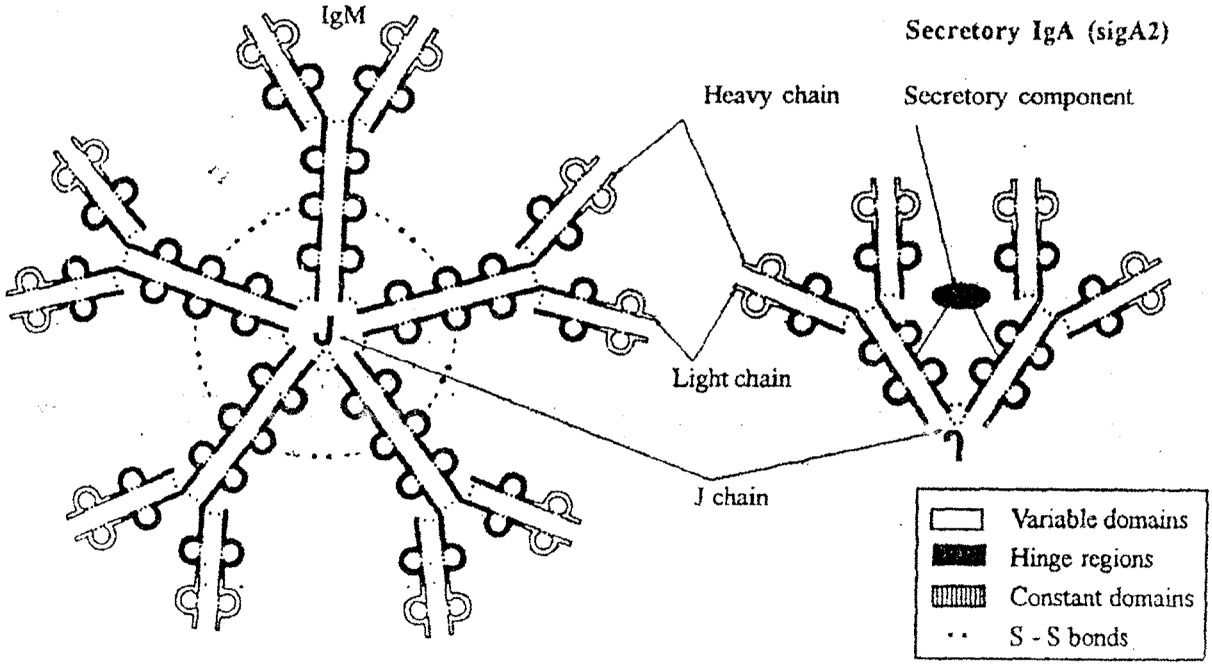


Fig. -11 : Structural models of IgM and secretory IgA (sigA). IgM has a pentameric structure linked by the J chain at the Fc fragment. The sigA molecule has a dimeric structure plus joining (J) chain plus secretory component. Shown is the dominant IgA2 subclass.

(COURTESY : KUBY, Immunology)

Table:

Antibody	H-chain	H-chain Domains (no)	Serum conc. (mg/ml)	Molecular weight (KD)
IgA_1	α_1	4	3	150
IgA_2	α_2	4	0.5	150
IgD	δ	4	0.04	180
IgE	ε	5	0.0003	190
IgM	μ	5	1.5	900
IgG_1	γ^1	4	9	150
IgG_2	γ^2	4	3	150

REVIEW QUESTIONS

Q.1 What do you mean by "Epitope"? Describe the nature of "Epitope".

Q.2 What do you mean by Immunogenicity? What are the basic requirements for Immunogenicity?

Q.3 Describe briefly different types of Immunity.

Q.4 Describe briefly about the defensive barriers in innate immunity

Q.5 Add a note on structure of Antibody (Ab) and mention its biological functions.

Q.6 What will be results if the immunoglobulin molecule is treated with enzyme pepsin, papain and plasmin.

Q.7 Discuss the structure and biological function of IgG, IgM and IgE.

Q.8 Write Notes on :

a) Allosteric effect.

b) Hypervariable region.

c) Vaccinization.

d) Hinge region.

e) Opsonization.

f) Secretory Component.

g) Antibody dependent cell mediated cytotoxicity (ADCC).

h) IgA.

i) Immunoglobulin domain.

j) T Chain.

SUGGESTED READINGS :

1. ABUL K. ABBAS, ANDREW K. LICHTMAN, Cellular and Molecular Immunology (3rd. Edn.)
2. DONALD M, WEIR, JOHN STEWART, Immunology (7th Edn.) ELI BENJAMINI, Immunology A short course (4th Edn.)
4. IVAN ROITT, Essential Immunology (Blackwell Science)
5. KUBY, Immunology, (W. H. FREEMAN AND COMPANY)
6. P. M. LYDYARD, A. WHELAN, Instant notes in IMMUNOLOGY.
7. RICHARD M: HYDE, Immunology, B.1. Waverly Pvt. Ltd., New Delhi.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I
Paper-III Group-B Unit-I
Module No. - 32

IMMUNOLOGY

1. A. ANTIGEN-ANTIBODY INTERACTION :

The antigen-antibody interaction is a bimolecular association similar to an enzyme-substrate interaction but with the important distinction that it does not lead to an irreversible chemical alteration in either the antibody or antigen and therefore is reversible. The interaction between an antibody and an antigen involves various non-covalent interactions between the antigenic determinant or epitope of the antigen and the variable-region (V_H/V_L) domain of the antibody molecule, particularly the hypervariable regions or complementary determining regions (CDRs). The exquisite specificity of antigen-antibody interactions has led to the development of a variety of immunogenic assays. These assays can be used to detect the presence of either antibody or antigen and have played vital roles in diagnosing diseases, monitoring the level of the humoral immune response, and identifying molecules of biological or medical interest. These assays differ in their speed and their sensitivity; some are strictly qualitative, and others are quantitative.

B. FORCES INVOLVED IN ANTIGEN — ANTIBODY INTERACTION :

It should be stressed immediately that the forces which hold antigen and antibody together are, in essence, no different from the so-called "nonspecific" interaction which occur between any two unrelated proteins (or other macromolecules).

These intermolecular forces may be classified under four headings.

1.1 Electrostatic :

These are due to the attraction between oppositely charged ionic groups on the two protein side-chains as, for example an ionized amino group (NH_2^+) on a lysine of one protein and an ionized carboxyl group (—COO-) of glutamate on the other. The force (F) of attraction is inversely proportional to the square of the distance (d) between the charges, i.e., $F \propto 1/K_D d^2$ where K_D is the dielectric constant. Thus as the charges come closer

together, the attractive force increases considerably if we halve the distance apart, we quadruple the attraction. Furthermore, since the dielectric constant of water is extremely high, the exclusion of water molecule through the contiguity of the interacting residues to which we have already drawn attention, would greatly increase F.

1.2 Hydrogen bonding :

The formation of reversible hydrogen bridges between hydrophilic groups such as OH, NH_2 and COOH depends very much upon the close approach of the two molecules carrying these groups. Although H-bonds are relatively weak, because they are essentially electrostatic in nature, exclusion of water between the reacting side chain would greatly enhance the binding energy through the gross reduction in dielectric constant.

1.3 Hydrophobic :

Non-polar hydrophobic groups such as the side-chains of valine, leucine and phenylalanine tend to associate in an aqueous environment. The driving force for this hydrophobic bonding derives from the fact that water in contact with hydrophobic molecules, with which it cannot H-bond, will associate with other water molecule but the number of configurations which allow H-bonds to form will not be as great as that occurring when they are surrounded completely by other water molecules i.e., the entropy is lower. The greater the area of contact between water and hydrophobic surface, the lower the entropy and the higher the energy state. Thus if hydrophobic groups on two proteins come together so as to exclude water molecules between them, the net surface in contact with water is reduced and the proteins take up a lower energy state than when they are separated. It has been estimated that hydrophobic forces may contribute up to 50% of the total strength of the antigen-antibody bond, although it must be appreciated that the relative contributions of these different types of molecular interaction will vary significantly from one complex to another.

1.4 Van der Waals :

These are the forces between molecules which depends upon interaction between the external 'electron clouds'. The deviation of gaseous molecules of nitrogen and hydrogen from ideal behaviour according to the kinetic theory is attributable to the Van der Waals attractions between them. The nature of this interaction is difficult to describe in non-mathematical terms but it has been linked to a temporary perturbation of electrons in one molecule effectively forming a dipole which induces a dipolar perturbation in the other molecule, the two dipole then having a force of attraction between them; as the displaced electrons swing back through the equilibrium position and beyond the dipoles oscillate. The force of attraction is inversely proportional to the 7th power of the distance i.e., $F \propto 1/d^7$ and as a result this rises very rapidly as the interacting molecules come closer together.

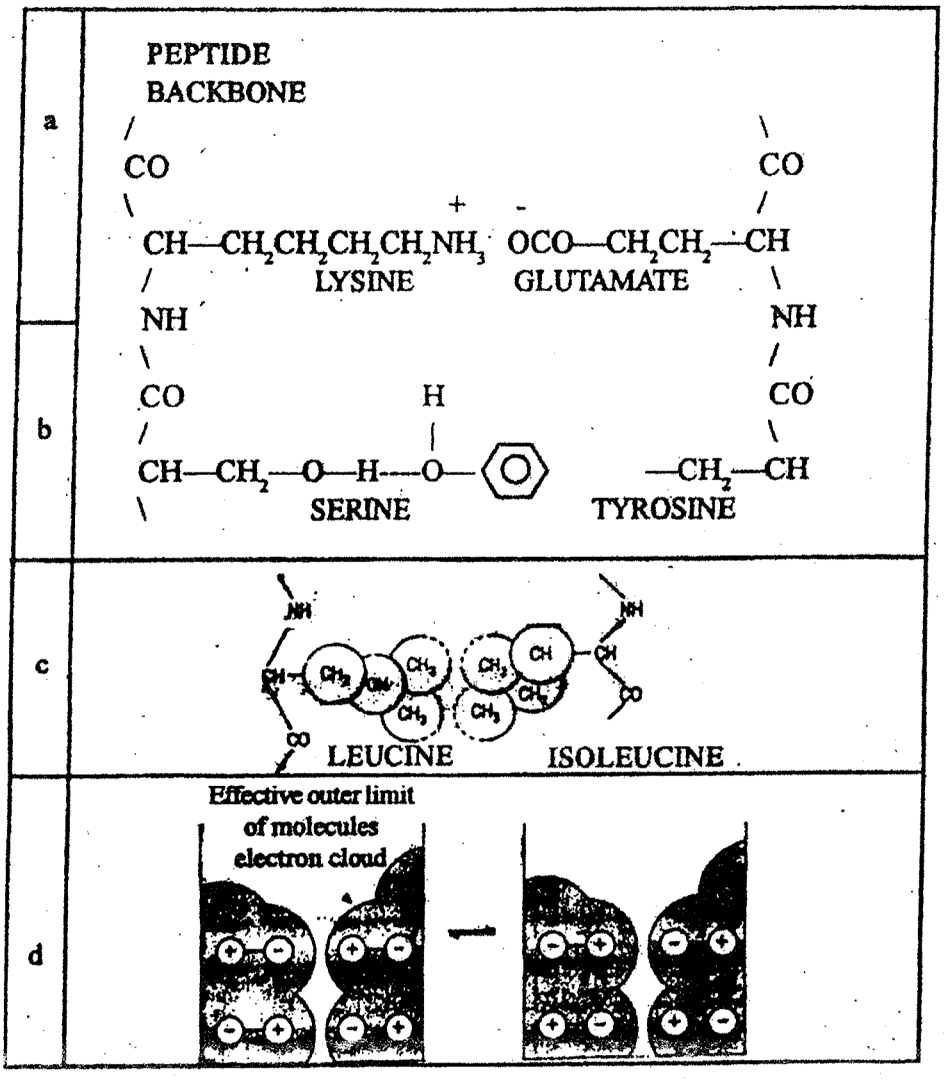


Fig. - 1 : Protein-protein interactions. (a) Coulombic attraction between oppositely changed ionic groupings. (b) Hydrogen bonding between two proteins. (c) Hydrophobic bonding, the region in which the water milecules are in contact with the hydrophobic groups (—) is considerably reduced when the hydrophobic groups on two proteins are in contact with each other (——) and the lower energy of this system makes this a more probable state than separation of the hydrophobic groups. (d) Van der Waals force : the interaction between the electrons in the external orbitals of two different macromolecules may be envisaged for simplicity) as the attraction between induced oscillating dipoles in the two-electron clouds.

Visualizing concepts

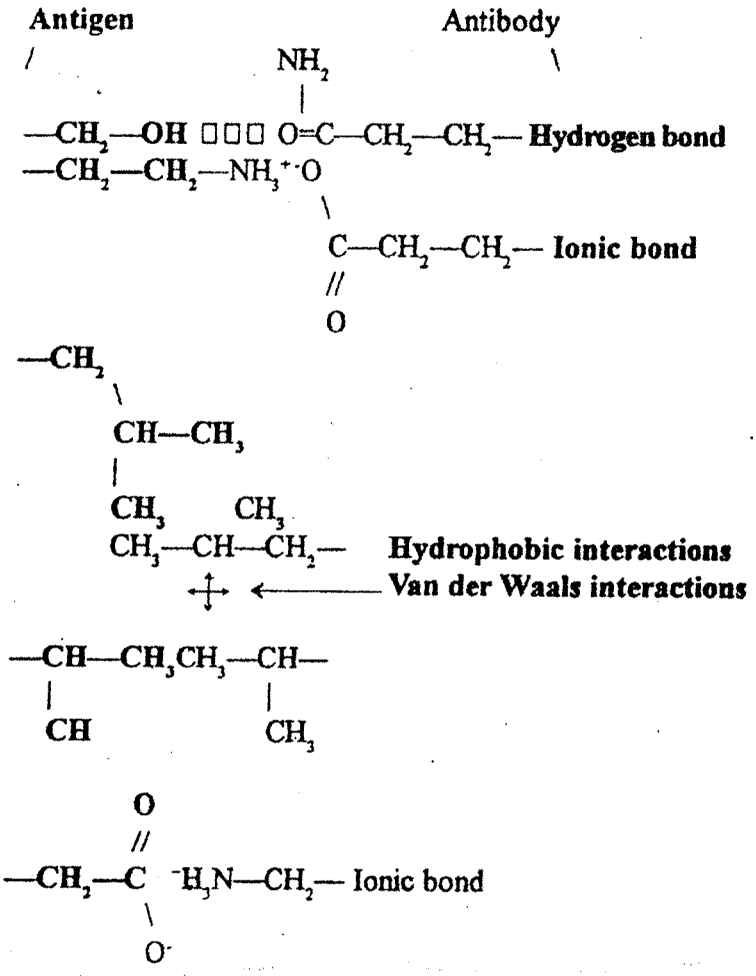


Fig. - 2 : The interaction between an antibody and an antigen (shown here in bold type) depends on four types of noncovalent forces: (1) hydrogen bonds in which a hydrogen atom is shared between two electronegative atoms, (2) ionic bonds between oppositely charged residues, (3) hydrophobic interactions in which water forces hydrophobic groups together to maximize hydrogen bonding of water molecules, and (4) Van der Waals interactions between the outer electron clouds of two atoms. In an aqueous environment noncovalent interactions are extremely weak and depend upon close structural complementarity between antibody and antigen.

(2) Antibody Affinity :

The strength of the total noncovalent interactions between a single antigen - binding site on an antibody and a single epitope is the affinity of the antibody for that epitope. Low- affinity antibodies bind antigen weakly and tend to dissociate readily, whereas high affinity antibodies bind antigen more tightly and remain bound longer. The association between a binding site on an antibody (Ab) with a monovalent antigen (Ag) can be described by the equation –

$$Ag + Ab \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} Ab - Ag$$

Where k_1 is the forward (association) rate constant and k_{-1} is the reverse (dissociation) rate constant. The ratio of k_1 / k_{-1} is the association constant K, a measure of affinity. It can be calculated from the ratio of the concentration of bound Ag – Ab complex to the concentration of unbound antigen and antibody, as follows

$$K = \frac{k_1}{k_{-1}} = \frac{Ag - Ab}{[Ab][Ag]}$$

The value of 'K' varies from different Ag Ab complexes and depends upon both k_1, which is expressed in liters / mole / second (L / mol / S), and k_{-1}, which is expressed in 1/second. For small haptens, the forward rate constant can be extremely high; in some cases k_1 values can be as high as 4×10^8 L / mol / S , approaching the theoretical upper limit of diffusion – limited reactions (10^9 L / mol / S). For larger proteins (Ags) however, k_1 is smaller, with values in the range of 10^5 L/mol/S. The rate at which bound antigen leaves an antibody's binding site (i.e. the dissociation rate constant k_{-1}) plays a major role in detecting the affinity of antibody's for an antigen.

The association constant K can be determined by **equilibrium dialysis.** In this procedure a dialysis chamber containing two equal compartments separated by a semipermeable membrane is used. Antibody is placed in one compartment, and a radioactively labeled ligand that is small enough to pass through the semipermeable membrane is placed in the other compartment. (Fig. – 3a) Suitable ligand include haptens as well as oligosaccharides and oligopeptides composing the epitope of complex polysaccharide or protein antigens. In the absence of antibody, ligand added to compartment B will equilibrate on both sides of the membrane. In the presence of antibody, however, part of the labeled ligand will be bound to the antibody at equilibrium, and the unbound ligand will be equally distributed in both compartments. Thus the total concentration of ligand will be greater in the compartment containing antibody. The difference in the ligand concentration in the two compartments represents the concentration of ligand bound to the antibody (i.e. the conc. of Ag – Ab complex). The higher the affinity of the antibody, the more ligand that is bound.

Since the total concentration of antibody in the equilibrium dialysis chamber is known, the equilibrium equation can be rewritten as :

$$K = \frac{[Ab - Ag]}{[Ab][Ag]} = \frac{r}{(n-r)(c)}$$

Where r = the ratio of the concentration of bound ligand of total antibody concentration, c = concentration of free ligand and n= number of binding sites per antibody molecule. This expression can be rearranged to give the **"scatchard equation"**

$$\frac{r}{c} = K_n - K_r$$

Values for 'r' and 'c' can be obtained by repeating the equilibrium dialysis with the same concentration of antibody but with different concentrations of ligand If 'K' is a constant, that is, if all the antibodies within the dialysis chamber have the same affinity for the ligand, then a Scatchard plot of r/c versus r will yield a straight line with a slope of -K (Fig – 4a). As the concentration of unbound ligand c increase r/c approaches o, and r approaches n, the **valency**.

For most antibody preparations, K is not a constant because antibodies (unless they are monoclonal) are heterogeneous and have a range of affinities. a **scatchard plot** of heterogeneous antibody yields a curved line whose slope is constantly changing, reflecting the antibody heterogeneity (Fig – 4b). With this type of **Scatchard plot,** it is possible to determine the average affinity constant K_0 by determining the value of K when half of the antigen – binding sites are filled:

$$K_0 = \frac{1}{(2-1)c} = \frac{1}{c}$$

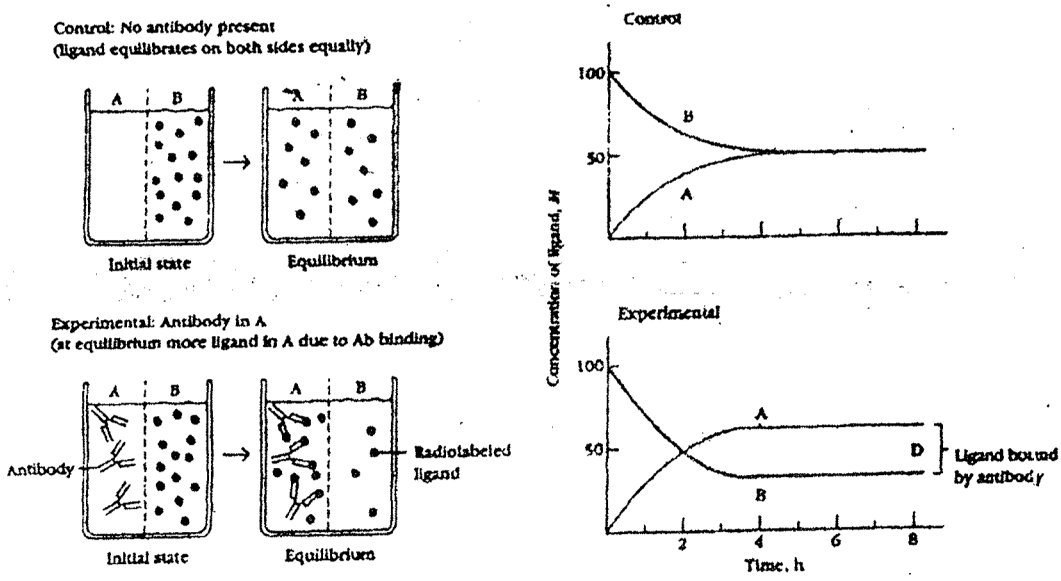


Fig. - 3 : Determination of antibody affinity by equilibrium dialysis. (a) The dialysis chamber contains two compartments (A and B) separated by a semipermeable membrane. Antibody is added to one compartment and a radiolabeled ligand to another. At equilibrium the concentration of radioactivity in both compartments is measured. (b) Plot of concentration of ligand in each compartment with time. At equilibrium the difference in the concentration of radioactive ligand in the two compartments represents the amount of ligand bound to antibody.

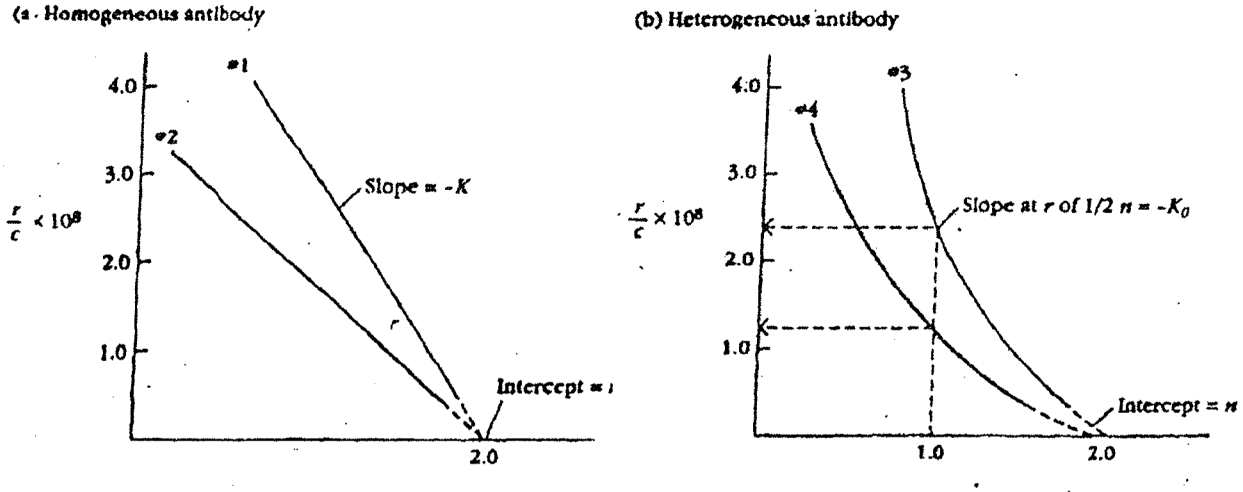


Fig. - 4 : Scatchard plots are based on repeated equilibrium dialyses with a constant concentration of antibody and varying concentration of ligand. In these plots, r = moles bound ligand / mole antibody and c = free ligand. From a Scatchard plot, both the equilibrium constant (K) and the number of binding sites per antibody molecule (n), or its valency, can be obtained. (a) If all antibodies have the same affinity, then a Scatchard plot yields a straight line with a slope of –K. The Y intercept is the valence of the antibody, which is 2 for IgG. In this graph antibody #1 has a higher affinity than abtibody # 2. (b) If the antibodies have a range of affinities, a Scatchard plot yields a curved line, whose slope is constantly changing. The average affinity constant K_0 can be calculated by determining the value of K when one half of the binding sites are occupied (i.e.when r = 1). In this graph antisera # 3 has a higher affinity ($K_0 = 2.4 \times 10^9$) than antisera # 4 ($K_0 = 1.25 \times 10^9$).

3. ANTIBODY AVIDITY :

The affinity at one binding site does not always reflect the true strength of the antibody antigen interaction. When complex antigen containing multiple, repeating antigenic determinants are mixed with antibodies containing multiple binding sites, the interaction of antibody with antigen at one site will increase the probability of reaction at a second site. The strength of such multiple interaction between a multivalent antibody and antigen is called the **avidity.**

The avidity of an antibody is a better measure of its binding capacity within biological systems (e.g. the reaction of an antibody with antigenic determinants on a virus or bacterial cell) than is the affinity of its individual binding sites. High avidity can compensate for low affinity. For example, secreted pentameric IgM often has lower affinity than IgG, but the high avidity of, resulting from its multivalence, enables it to bind antigen effectively.

4. PRECIPITATION REACTIONS :

The interaction between an antibody and a soluble antigen in aqueaus solution forms a lattice that eventually develops into a visible precipitate. Antibodies that thus aggregate soluable antigens are called **precipitins.** Although

formation of the soluble Ag —Ab complex occurs within minutes, formation of the visible precipitate occurs more slowly and often takes a day or two to reach completion. The precipitate develops as neighboring antibody molecules within the lattic form ionic bonds with each other, causing the lattice to lose its charge and thus become insoluble.

Formation of an Ag —Ab lattice depends on the valency of both the antibody and antigen.

- The antibody must be bivalent; a precipitate will not form with monovalent Fab fragments.
- The antigen must either be bivalent or polyvalent; that is, it must have at least two copies ofthe same epitope, or have different epitopes that react with different antibodies present in polyclonal antisera.

4.1 Precipitation Reactions in Liquids :

A quantitative precipitation reaction can be performed by placing a constant amount of antibody in a series of tubes and adding increasing amounts of antigen to the tubes. After the Precipitate forms, each tube is centrifuged to pellet the precipitate, the supernatant is poured off, and the amount of precipitate is measured. Plotting the amount of Precipitate against increasing antigen concentrations yields a preciptin curve. As Fig — 5 shows, excess of either antibody or antigen interfere with maximal precipitation, which occurs in the so called **"equivalence zone".** When the ratio of antibody to antigen is optimal. As a large multimolecular lattice is formed at equivalence, the complex increases in size and Preciptates out of solution. In the region of antibody excess, unreacted antibody is found in the supernatant along with small soluble complex consisting of multiple molecules of antibody bound to a single molecule of antigen. In the region of antigen excess, unreacted antigen can be detected and small complexes are again observed, this time consisting of one or two molecules of antigen bound to a single molecule of antibody. Although the quantitative precipitation reaction in Seldom used experimentally today, the principles of antigen excess, antibody excess and equivalence apply to many Ag —Ab reactions. The precipitation reaction can also be used as a rapid test for the presence of antibody or antigen. **The interfacial (ring) precipitin test** is performed by adding antiserum to a small tube and layering antigen on top. If the antiserum contains antibodies specific for the test antigen, then the antibody and antigen diffuse toward each other and form a visible band of precipitation at the interface within a few minutes (Fig. 6).

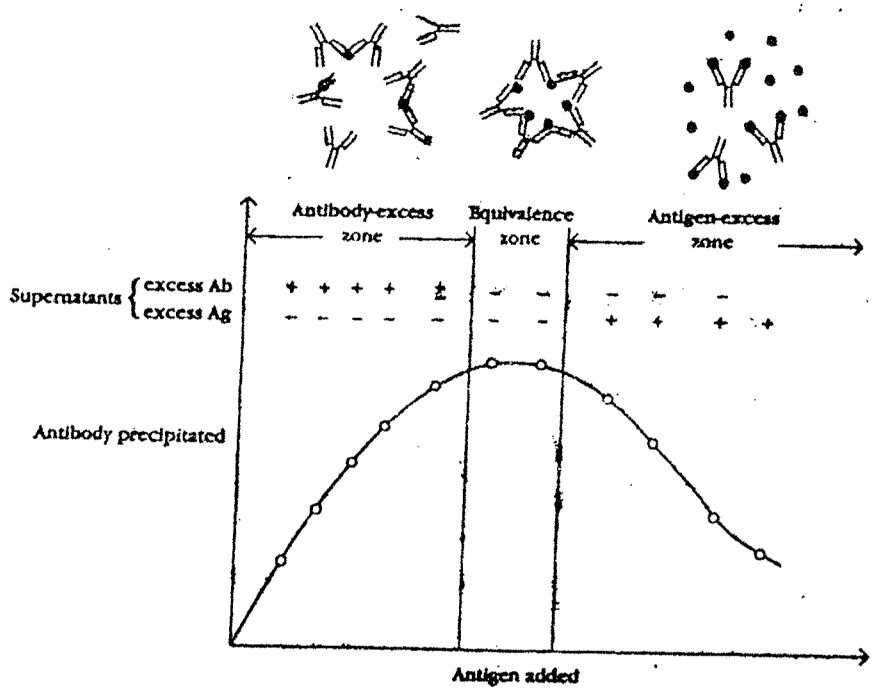


Fig. - 5 : A precipitation curve for a system of one antigen and its antibodies. This plot of the amount of antibody precipitated versus increasing antigen concentrations (at constant total antibody) reveals three zones : a zone of antibody excess in which precipitation is inhibited and excess antibody can be detected in the supernatant; an equivalence zone of maximal precipitation in which antibody and antigen form large insoluble complexes (shaded in blue) and neither antibody nor antigen can be detected in the supernatant; and a zone of antigen excess in which precipitation is inhibited and excess antigen can be detected in the supernatant.

Fig. - 6 : The interfacial (ring) precipitin test is a rapid, qualitative method for determining the presence of antibody or antigen. The antiserum is placed in the bottom of a tube, and then the antigen solution is carefully layered on top. Formation of a visible line of precipitation in the tube at the extreme right indicates a positive reaction. The other four tubes are various controls (e.g. normal serum with antigen and antiserum with buffer). [From J. S. Garvey et al., 1977, *Methods in Immunology*, 3d ed., W. A. Benjamin Inc., Advanced Book Program.]

4.2 Precipitation Reactions in Gells :

Immune precipitates can form not only in solution but also in an agar matrix. When antigen and antibody diffuse toward one another in agar or when antibody is incorporated into the agar and antigen diffuses into the antibody- containing matrix, a visible line of precipitation will forms. As in a precipitation reaction in fluid, visible precipitation occurs in the region of equivalence, whereas no visible precipitates forms in regions of antibody or antigen excess. These **immunodiffusion reactions** can be used to determine relative concentrations of antibodies or antigens, to compare antigens, or to determine the relative purity of an antigen preparation. Two frequently used immunodiffusion techniques are **radial immunodiffusion** (Mancini method) and **double immunodiffusion** (Ouchterlony method); both are carried out in a semisolid medium like agar (Fig. - 7).

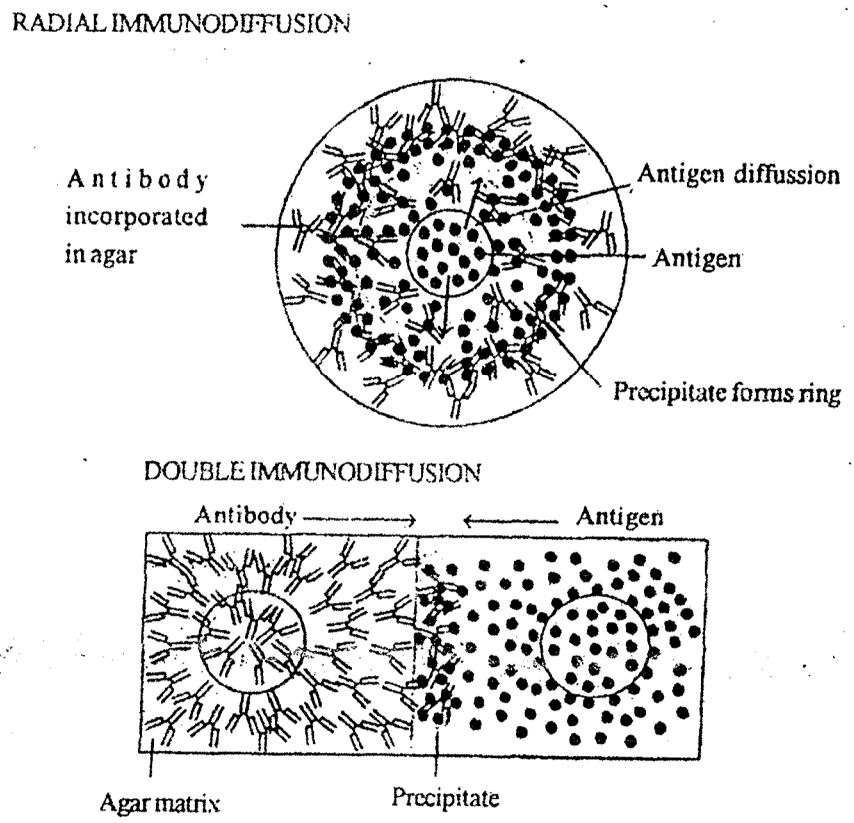


Fig. - 7 : Diagrammatic representation of radial (Mancini) and double immunodiffusion (Ouchterlony) in a gel. In both cases, large insoluble complexes form in the agar in the zone of equivalence, which are visible as a line of precipitation (blue region). Only the antigen diffuses in radial immunodiffusion, whereas both the antibody and antigen diffuse in double immunodiffusion.

*(N.B.– For Radial and Double immunodiffusion see **Immunology by Eli Benjamini**, chapter-5).*

5. Agglutination Reactions :

The interaction between antibody and a particulate antigen results in visible clumping called **agglutination.** Antibodies that produce such reactions are called **agglutinins.** Agglutination reactions are similar in principle to precipitation reactions. Just as antibody excess inhibits precipitation reactions, an excess of antibody inhibits agglutination reactions; this inhibition is called the **PROZONE EFFECT.**

Several mechanism can cause the prozone effect. First, high levels of antibody increases the likelihood that a single antibody molecule will find to two or more epitopes on a single particulate antigens. The prozone effect can also occur at high concentrations of antibodies that bind to the antigen but do not induce agglutination; these antibodies, called **incomplete antibodies,** are often of the IgG class. At high concentrations of IgG, in complete antibodies may occupy all of the antigenic sites, thus blocking access by IgM, which is a good agglutinin. The lack of agglutinating activity of an incomplete antibody may be due to restricted flexibility in the hinge region, making it difficult for the antibody to assume the required angle for optimal cross-linking of epitopes on two or more particulate antigens. Alternatively, the density of epitope distribution or the location of some epitopes in deep pockets of a particulate antigen may make it difficult for antibodies, specific for there epitopes, to agglutinate certain particulate antigens.

5.1 Hemagglutination :

Agglutination reactions are routinely performed to type red blood cells (RBCs). In typing for the ABO antigens, RBCs are mixed on a slide with antisera to the A and B blesd group antigens. If the antigen is present on the cells, they agglutinate forming a visible clump on the slide. Determination of which antigens are present on donor and recipient RBCs is the basis for matching blood types for transfusions.

At neutral PH, red blood cells are surrounded by a negative in cloud that makes the cells repel one another; this repulsive force is called the **zeta potential.** Because of its side and pentameric nature, IgM can overcome the zeta potential and cross-link red blood cells, leading to agglutination. The smaller size and bivalency of IgG makes it less able to over come the zeta potential. For this reason, IgM is more effective than IgG in agglutination, (i.e in agglutinating red blood cells).

5.2 Bacterial agglutination :

The bacterial infection often elicits the production of serum antibodies specific for surface antigens on the bacterial cells. The presence of such antibodies can be detected by bacterial agglutination reactions. Serum from a patient thought to be infected with a given bacterium is serially diluted with in a series of tubes to which the bacteria is added. The last tube showing visible agglutination will reflect the serum antibody **titer** of the patient. The agglutinin titer is defined as the reciprocal of the last serum dilution that elicits a positive agglutination reaction.

The agglutinin titer of an antiserum can be used to diagnose a bacterial infection. Patients with typhoid, fever, for example, show a significant rise in the agglutination titer to **salmonella typhi.** Agglutination reactions also provide a way to type bacteria. For instance, different species of the bacterium salmonella can be distinguished by agglutination reactions with a pance of typing antisera.

5.3 Passive Agglutination :

The sensitivity and simplicity of agglutination reaction can be extended to soluble antegens by the technique of passive **hemagglutination.** In this technique, a soluble antigen is mixed with red blood cells that have been treated with tannic acid or chromium chloride, both of which promote adsorption of the antigen to the surface of the cells. Serum containing antibody is serially diluted into microliter plate wells, and the antigen-coated red blood cells are added to each well; agglutination is assessed by the size of the characteristic spread pattern of agglutinated red blood cells on the bottom of the well.

Passive hemagglutination is far more sensitive than precipitation reactions and can detect antibody concentrations as low as 0.001 μg/ml. The sensitivities of precipitation and hemagglutination can be compared by testing an antiserum to lien ovalbumin in a tube precipitation reaction and in passive hemagglutination with ovalbumin coated red blood cells. Delution of the antiserum by 1:5 results in loss of precipitation ability, whereas the antiserum still ill functions in passive agglutination also can be performed with antigen-coated particles of latex or the mineral colloid bentonite.

5.4 Agglutination Inhibition

A modification ofthe agglutination reaction, called agglutination inhibition, provides a highly sensitive assay to detect small quantities of an antigen.

Agglutination inhibition assays are widely used in clinical laboratories to determine if an individual has been exposed to certain types of viruses that came agglutination of real blood cells. If an individuals serum contain specific antiviral antibodies, then the antibodies will bind to the virus and interfere with hemagglutination by the virus. This technique is commonly used in premarital testing to detrmine the immune status of women to rubella virus. The reciprocal of the least serum dilution to show inhibition of rubella hemagglutination is the titer of the serum. A titer greater than 10 (1 : 10 dilution) indicates than a woman is immune to rubella, where as a titer less than 10 is indicative of a lack of immunity and the need for immunization with the rubella vaccine.

6. HAPTENS :

The pioneering work of ***Karl Landsteiner in*** the 1920s and 1930s created a simple, chemically defined system for studying the binding of an individual antibody to a unique epitope on a complex protein antigen. Landsteiner employed various haptens, small organic molecules that are antigenic but not immunogenic. Chemical coupling of a hapten to a large protein, called a carrier, yields an immunogenic hapten carrier conjugate. Animals immunized with such a conjugate produce antibodies specific for

a) the hapten determinant

b) Unaltered epitope on the carrier protein and

c) new epitopes formed by combined parts of both the hapten and carrier.

By itself, a hapten can not function as an immunogenic epitope. But when multiple molecules of a single hapten are coupled to a carrier protein, the hapten becomes accessible to the immune system and often functions as the immunodominant determinant.

TABLE :

HAPTEN	CARRIER PROTEIN
(1) Dinitrophenol	Bovine gamma globulin
(2) Trinitrophenyl	B S A
(3) 5 – Nitrophenyl acetic acid	Keyhole Limpet hemocyanin
(4) Azophenyl arsonate	Ovalbumin
(5) Phenyllactoside	Human gamma globulin.

Source: *Immunology, Kuby*

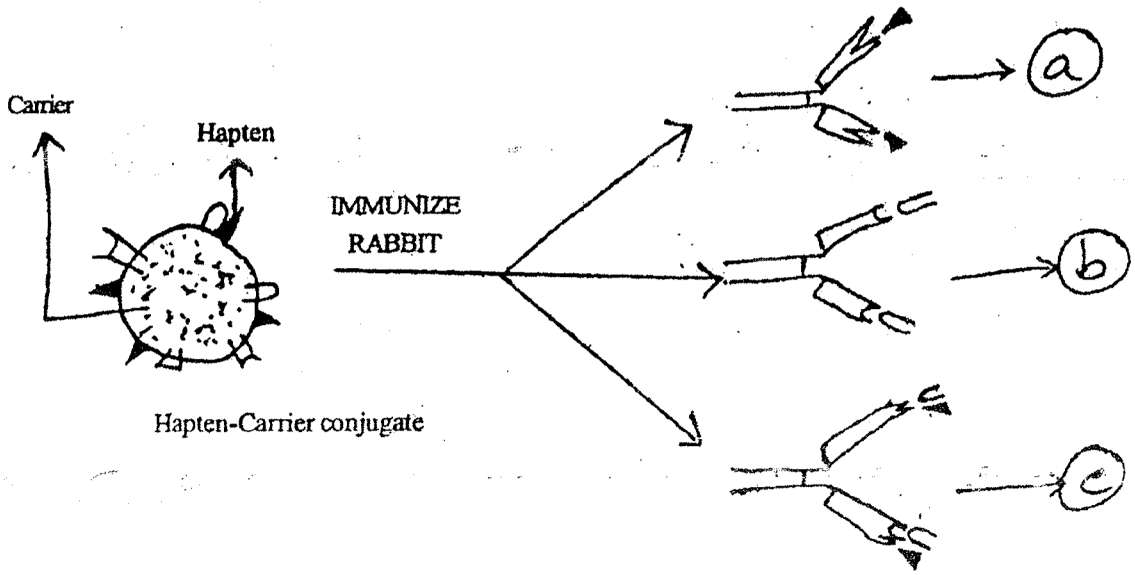


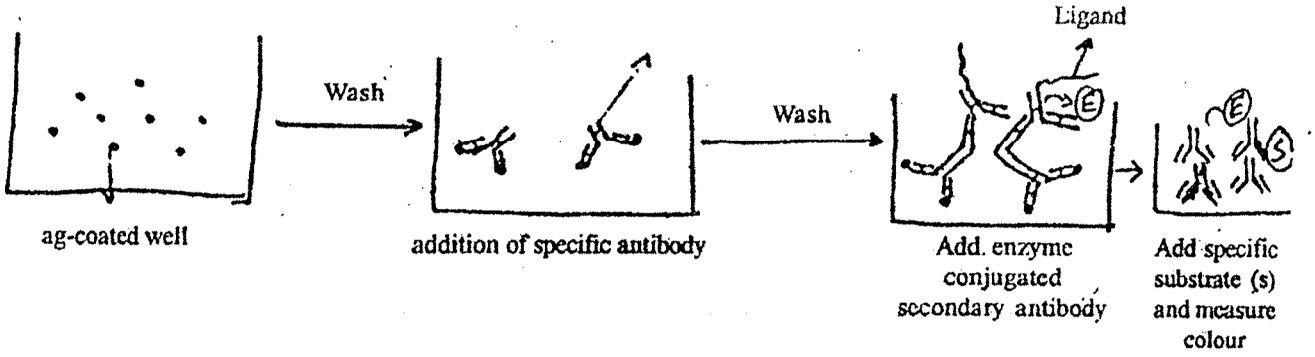
Fig. - 8 : a) Antibodies to hapten, b) Antibodies to carrier, c) Antibodies to conjugate of hapten and carrier.

7. ELISA :

ELISA (Enzyme Linked Immuno Sorbent Assay) is similar in principle to RIA but depends on an enzyme rather than radioactive level. An enzyme conjugated to an antibody reacts with a colourless substrate to generate coloured reaction product. A number of enzyme including alkaline phosphatase horse radish peroxidase (or FITC conjugated) when mixed with suitable substrate, each of these generate a coloured reaction products.

7.1 Indirect ELISA :

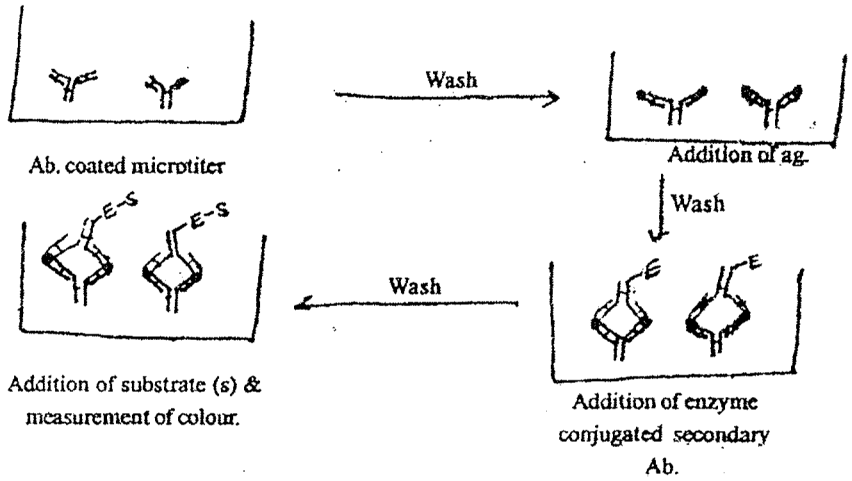
It is used to detect or quantitate antibody. Serum or some other sample containing primary antibody (Ab1) is added to an antigen - coated microtriter well and allow to react with the bound antigen. After free Ab1 is washed away (generally using phosphate Buffer saline or PBS : pH 8.4). The presence of antibody bound to antigen is detected by adding enzyme conjugated secondary antiisotype antibody (Ab_2) which bind to Ab_1. After free Ab_2 is washed away, a substrate for enzyme is added. Here Ab_2 treated Ab_1 / ag complex as a foreign material. The colored reaction product that forms, is measured by specialized spectrophotometric plate reader.



Presence of serum antibody against HIV is detected by indirect ELISA. Recombinant envelope and core protein of HIV antigen reacts with serum antibody in microtites plate.

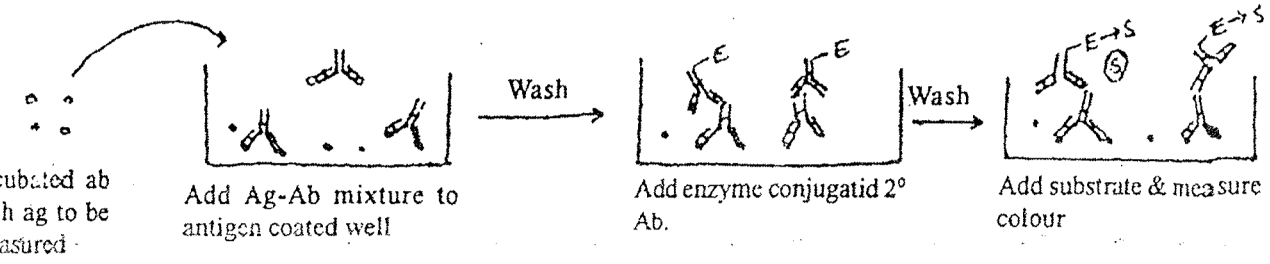
7.2 Sandwich ELISA or Direct ELISA :

This technique allows to detect or quantitation of antigen. In this case the antibody is immobilized on a microtiter well. A sample containing antigen is added and allowed to react with bound antibody. After washing (using proper using buffer), a second enzyme linked antibody specific for a different epitope on the antigen is added and allowed to react with the bound antigen. After the free secondary antigen is removed by washing, substrate is added and coloured reaction product is measured.



7.3 Competitive ELISA :

Antigen can also be quantified by another variation called competitive ELISA. In this technique antibody is first incubated in a solution with a sample containing antigen. The antigen - antibody mixture is then added to an antigen coated microtiter well. The more antigen present in the sample, the less free antibody will be available to bind antigen coated well. Addition of an enzyme linked Ab_2 specific for the isotype of Ab_1 can be used to quantify the amount of primary antibody bound to the well as indirect ELISA. In the competitive as say, higher the concentration of antigen in the original sample, the lower the absorbance.



In the competitive ELISA, the concentration of antigen is inversely proportional to the colour produced.

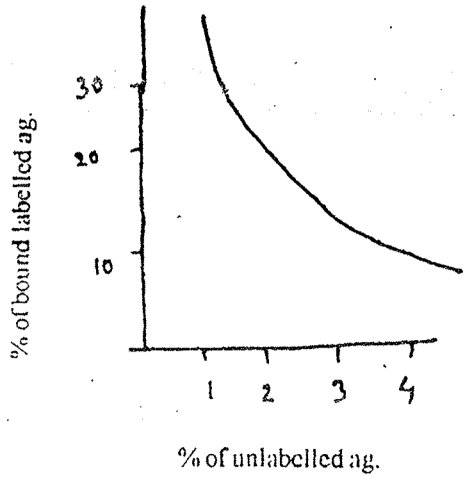
8. Radio Immuno Assay (RIA):

RIA is a highly sensitive technique that can detect antigen or antibody at concentration less than 0.001 μg/ml. The technique was first developed by *S. A, Berson and Rosalyn Yalow (1960).* This method is widely used in biochemistry and also in clinical field for diagonistic purpose. RIA is more sensitive than ELISA. The quantitative analysis of hormones, serum proteins, steroids and drugs can be done with the help of this method.

The principle of RIA involves competitive binding of radiolabelled antigen with unlabelled antigen to a high affinity of antibody. The antigen is generally labelled with r-emitting isotopes such as I^{125}, the labelled ag. is mixed with antibody at a concentration that just saturates the ag. binding sites of the antibody molecule and then increasing the amount of unlabelled ag. of unknown concentration are added. The antibody does not distinguish labelled ag. from unlabelled one. So, the two kinds of ag. (labelled and unlabelled) complete for available binding sites on the ab. molecule. With increasing concentration of unlabelled ag. more labelled ag. will be displaced from the binding sites. By measuring the amount of labelled ag. free in solution, it is possible to determine the concentration of unlabelled ag.

Several methods have been developed for seperating the bound ag. from the free one.

e.g.- If ag - ab complex contain rabbit IgG molecule, then goat antirabbit IgG can precipitate the complex. The amount of free labelled ag. remaining can be quantitate in y-counter. A standard curve is then plotted of the percentage of bound labelled ag. vs. known concentration of unlabelled ag. unlabelled ag. can be determined from this curve.



Sometimes the mixture of this ag. - ab. complex are passed through sephadex column, when the bound ab remain attached to the sephadex beads and unbound ag. can be washed out. Then the labelled ag. are eluted from the column.

Isotypes		Halflife
H^3	—	12.36 yrs.
C^{14}	—	5760 yrs.
P^{32}	—	14-20 days.
S^{35}	—	87.20 days.
I^{125}	—	60 days.
I^{131}	—	8.05 days.

9. SOUTHERN BLOTTING

The technique of Southern blotting, named after its originator *Edward Southern,* can identify specific restriction fragments in a complex mixture of restriction fragments. The DNA to be analyzed, such as the total DNA of an organism, is digested to completion with a restriction enzyme. For an organism with a complex genome, this digestion may generate millions of specific restriction fragments. The complex mixture of fragments is subjected to gel electrophoresis to separate the fragment according to size. However, many different fragments are of exactly the same length, and these do not separate from each other.

Even though all the fragments are not resolved by gel electrophoresis, an individual fragment that is complementary to a specific DNA clone can be detected. The restriction fragments present in the gel are denatured with alkali and transferred onto a nitrocellulose filter or nylon membrane by blotting. This procedure preserves the distribution of the fragments in the gel, creating a replica of the gel on the filter, much like the replica filter produced from plaques of a λ library. The filter then is incubated under hybridization conditions. With a specific radiolabeled DNA probe usually generated from a cloned restriction fragment. The DNA restriction fragment that is complementary to the probe hybridizes, and its location on the filter can be revealed by autoradiography.

Southern blotting permits a comparison between the restriction map of DNA isolated directly from an organism and the restriction map of cloned DNA. It is also used to map restriction sites in genomic DNA next to the sequence of a cloned DNA fragment.

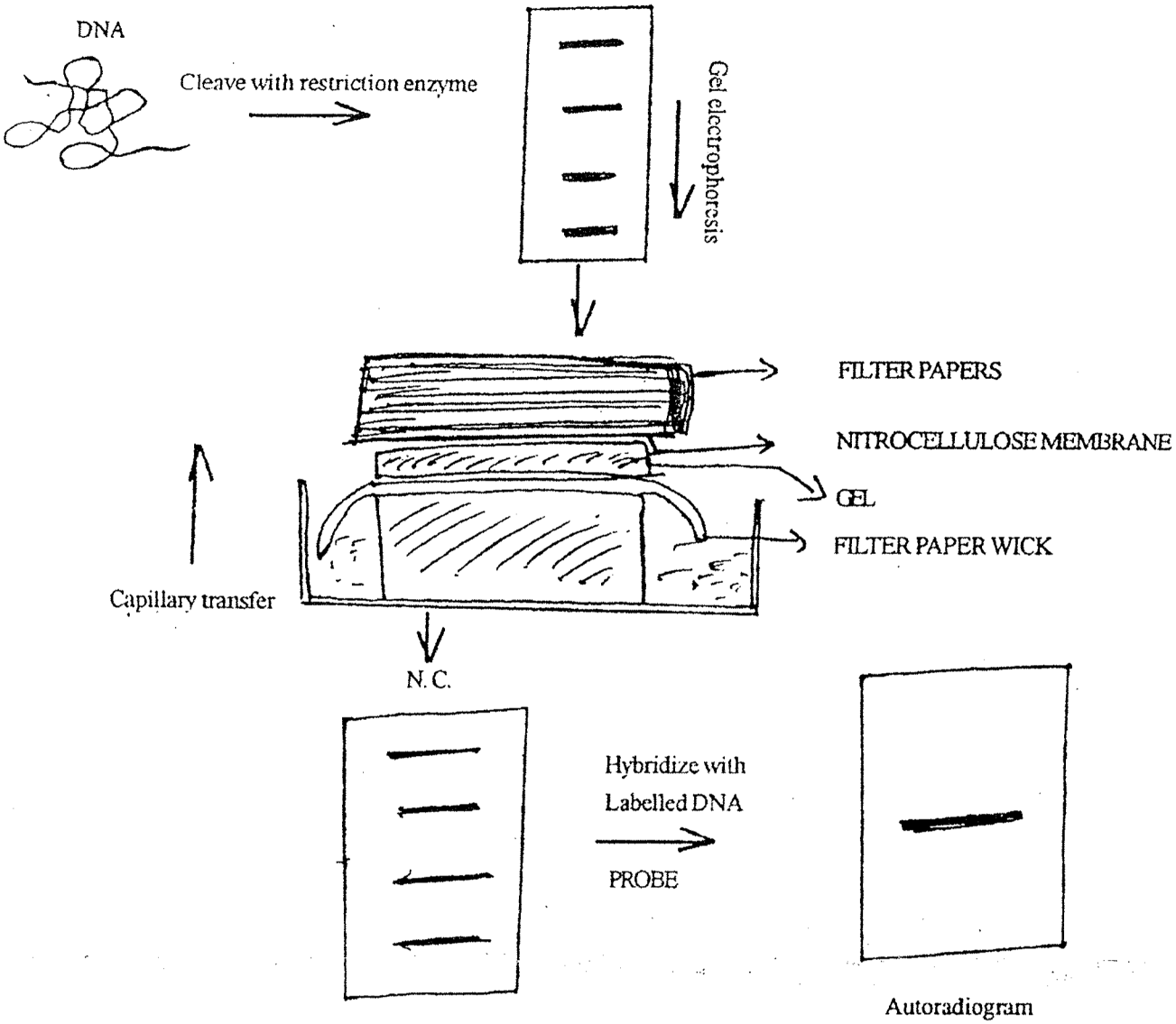


Fig.9 :Steps of Southern blotting.
(See E. M. Southern, 1975 J. Mol. Biol. 98 : 508)

10. WESTERN BLOTTING:

One of the most powerful methods for detecting a particular protein in a complex mixture combines the superior resolving power of gel electrophoresis, the specificity of antibodies, and the sensitivity of enzyme assays, called Western blotting or Immunoblotting. This three - step procedure is commonly used to separate proteins and then identify a specific protein of interest.

The first step in an immunoblot assay is to separate a protein mixture on an SDS - polyacrylamide gel. Then, a paper - thin membrane made from nitrocellulose, which tenaciously binds most proteins, is applied to the face of the gel. When an electric field is applied, protein are driven out of the gel and transferred to the membrane. This process is called blotting because the membrane picks up the proteins as if it were a blot impression of the gel. In the second step, the membrane is soaked in a solution of an antibody (Ab_1) specific for the protein of interest. Only the band containing this protein binds the antibody, forming a layer of antibody molecule. In the final step, the membrane is "developed" with an enzyme - linked antibody (Ab_2) to identify the band containing the protein of interest.

"Western blots" are used widely in research and clinical laboratories for the detection and characterization of antigens. A particularly useful example is the confirmatory diagnosis of HIV infection by the application of a patient's serum to the nitrocellulose sheets on which HIV antigens are bound.

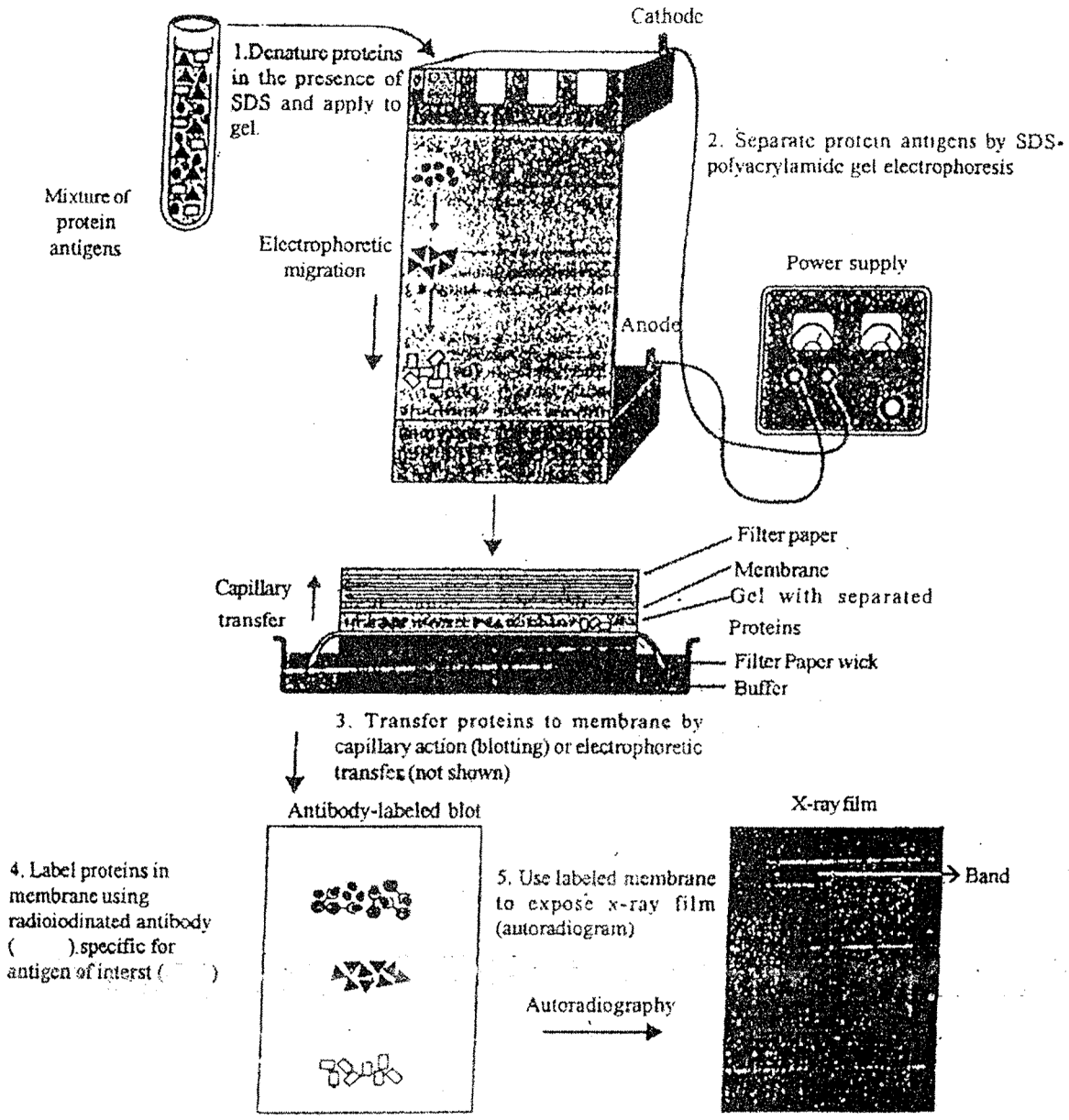


Fig. 10: Characterization of antigens by Western blotting. Protein antigens, separated by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophore is and transferred to a membrane, can be labeled with radioactive or (not shown) enzyme-coupled antibodies. Analysis of an antigen by Western blotting provides information similar to that obtained from immunoprecipitation followed by polyacrylamide gel electrophoresis. Some antibodies work only in one technique or the other.

(Courtesy : ABBAS, Cellular and Molecular Immunology)

(N. B. : For Fig. 1-7, Courtesy : KUBY, Immunology)

MODEL QUESTIONS :

1. Discuss briefly about the forces involved in antigen - antibody interaction.
2. What do you mean by "equivalence zone".? Add a note on precipitation reaction in gells.
3. a) What is prozone effect

   b) What do you mean by incomplete antibody.
4. a) What is zeta potential ?

   b) Discuss briefly about hemagglutination and bacterial agglutination.
5. Write short notes on
   a) Antibody affinity and avidity
   b) Double immuno diffusion
   c) Haptens
   d) ELISA
   e) RIA
   f) Western blotting
   g) Southern blotting

SUGGESTED READINGS :

1. ABUL K. ABBAS, ANDREW K. LICHTMAN, Cellular and Molecular Immunology 3rd. Edn.)
2. DONALD M. WEIR, JOHN STEWART, Immunology (7th Edn.)
3. DAVID BALTIMORE, MOLECULAR CELL BIOLOGY.
4. ELI BENJAMIN, IMMUNOLOGY – A short course (4th Edn.)
5. E. P. DIAMANDIS, T. K. CHRISTO POULOS, IMMUNOASSAY. (ACADEMIC PRESS)
6. IVAN ROITT, Essential Immunology (Black well Science)
7. KUBY, Immunology, (W. H. FREEMAN AND COMPANY)
8. KEITH WILSON AND WALKER, PRACTICAL BIOCHEMISTRY (Principles and techniques) [C. L. P Edn.]
9. P. M. LYDYARD, A. WHELAN, Instant notes in IMMUNOLOGY
10. RICHARD M. HYDE, Immunology (B. I. Waverly Pvt. Ltd,, New Delhi).
11. ROBERT M. COLLEMAN, Fundamental Immunology
12. V. S. MALIK, Antibody technique. (Academic Press).

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-B Unit-I

Module No. - 33

INTRODUCTION

The immune system protects us from attack by microbes and foreign antigen. There are of two major types of lymphocytes, B and T-lymphocytes. These two types of lymphocytes involved in the adaptive response. They have specific but different antigen receptors and other surface molecules necessary for interaction with other cells. B-cells mature into plasma cells which produce and secrete antibodies.

Humoral immune response is mediated by antibodies, which are produced by B-lymphocytes. The humoral immune system has the capacity to respond to different types of antigens by producing different classes of antibodies. T-lymphocytes helps the B-lymphocytes for the formation of antibodies. The complement system is one of the major effector functions of humoral immunity and is also an important effector mechanism of immunity of innate immunity. The name of the complement system is derived from experiments performed by Jules Bordet shortly after the discovery of Antibodies.

Phagocytes are specialized 'eating' cells of which there are two main types, neutrophils and macrophages. Monocytes are present in the blood stream and settle in the tissues as macrophages. Activated macrophages contain many lysosomes, and activation leads to increased phagocytosis and intracellular killing of ingested bacteria. In addition, macrophages play an important role in the 'processing' of antigen and its presentation to T-lymphocytes. Macrophages, B-lymphocytes and dendritic cells are also termed as antigen-presenting cells (APCs).

Cell mediated Immunity (CMI) is the mechanism of defence against intracellular microbes that are either ingested by and survive in phagocytes or infect non-phagocytic cells. This form of immunity is mediated by T-lymphocytes. There are two types of cell-mediated immune reaction. In one type, T-cells provide specificity and phagocytes eliminate the antigen. This reaction is initiated

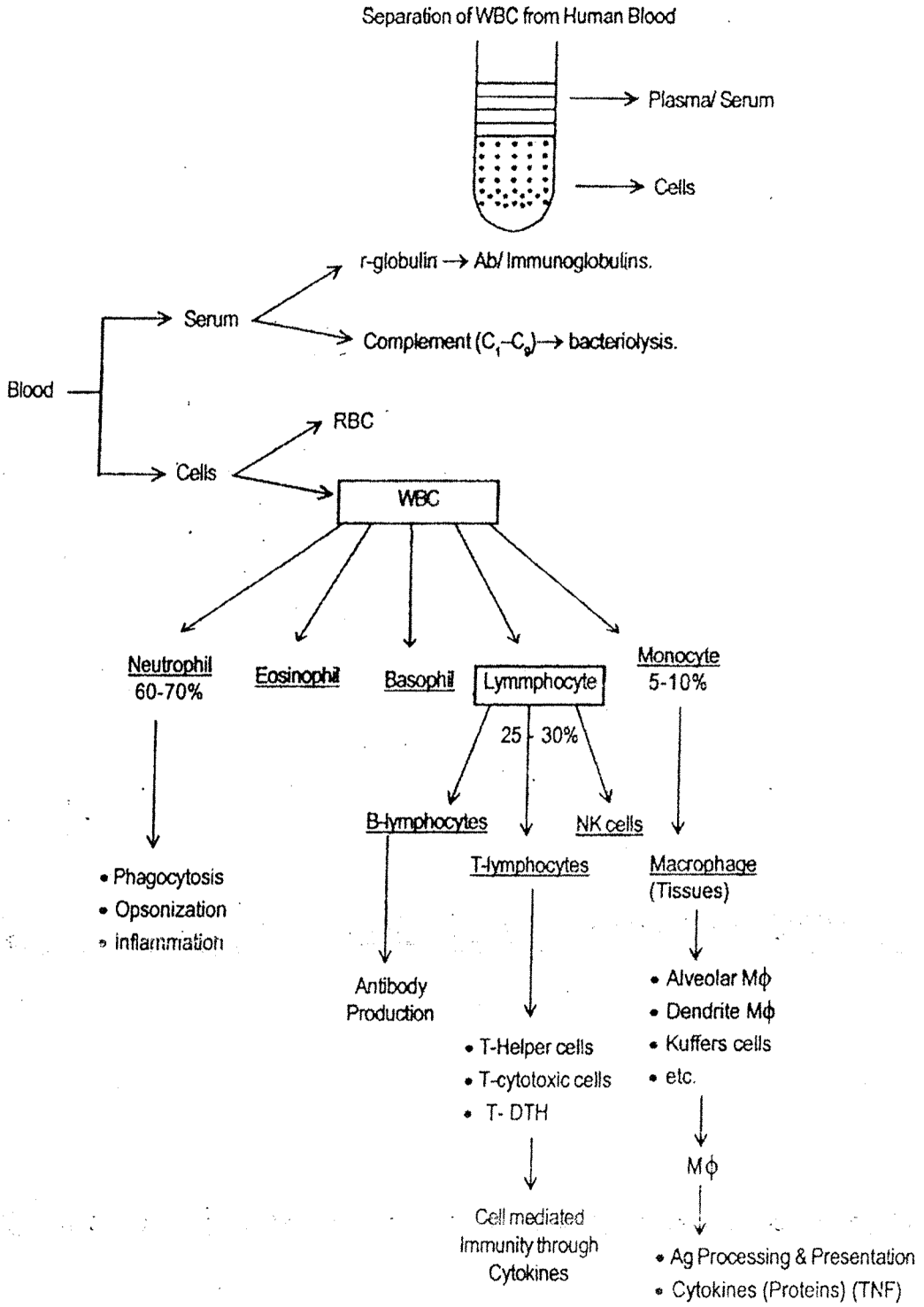


Fig. 1 : Separation Immunoglobulin and different types of wBC.

when effector T-cells of both the CD_4 and CD_8 subsets recognize peptide antigens of microbes ingested by macrophages. In the second-type of T-cell mediated immunity, CD_8+ cytolytic T-lymphocytes (CTLs) recognize peptide antigens of microbes, such as viruses, that infect and survive or replicate in the cytoplasm of any nuleated host cells. As a result of antigen recognition, the CTLs secrete granule proteins and express surface molecules that kill the infected cells, thus eliminating the reservoir of infection.

B-Lymphocytes :

Antibodies are produced by B-lymphocytes and plasma cells. B-cells are produced in the bone marrow and like T-cells migrate to the secondary lymphoid organs and tissues where they respond to foreign antigens. There are two kinds of B-cells.

B_1-cells : These cells arise early in ontogeny, express mainly IgM antibodies encoded by germ line antibody genes, mature independently of the bone marrow, generally recognize multimeric sugar/lipid antigens of microbes and are T-cell independant.

B_2-cells (Conventional B cells) : These cells are the B-cells primarily responsible for the development of humoral (antibody) mediated immunity. They are produced in the bone marrow and with the help of T-cells produce IgG, IgA and IgE antibodies.

Though antibodies are produced in the bone marrow but antibodies perform their effector functions at sites distant from their production. The reason for this distant effect is that antibodies enter the blood or mucosal secretions and are able to circulate to sites where antigens are located.

Antibodies inhibit, or neutralize, the infectivity of microbes and microbial toxins and exhibit following functions -

(i) Opsonization and phagocytosis.

(ii) Antibody dependant cell mediated cytotoxicity (ADCC).

(iii) Activation of complement (lysis of microbes).

B-cell receptors :

Antibodies are the antigen receptor on these cells and are associated with several other molecules which form the B-cell receptor complex. When activated by antigen and in most cases, with T-cell help, B-cells proliferate and mature into memory cells or plasma cells. Memory cells

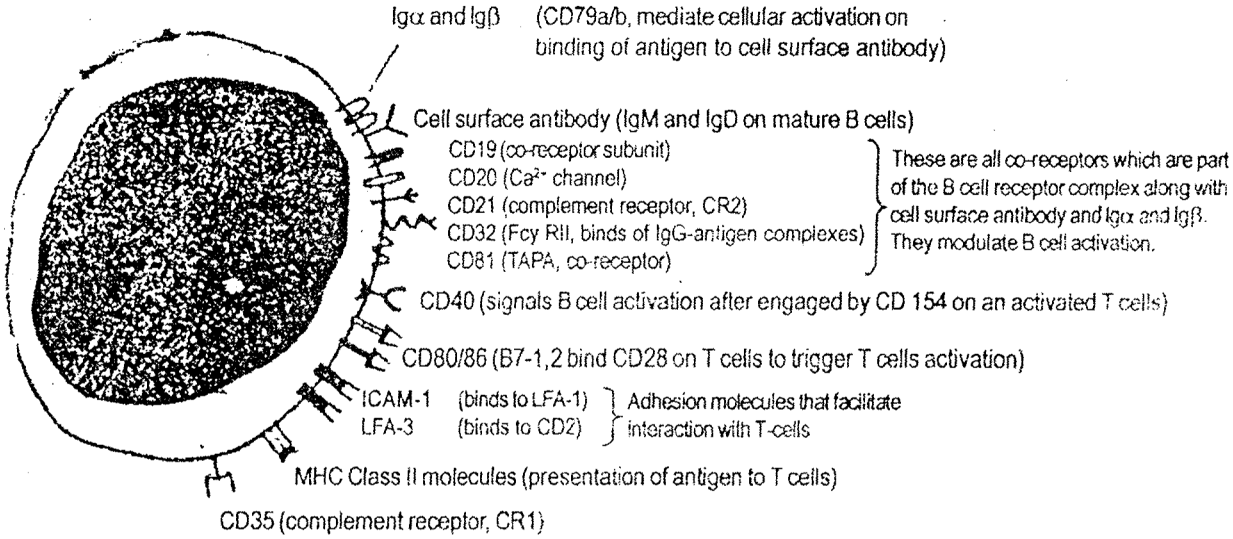


Fig. 2 : B cell surface molecules.

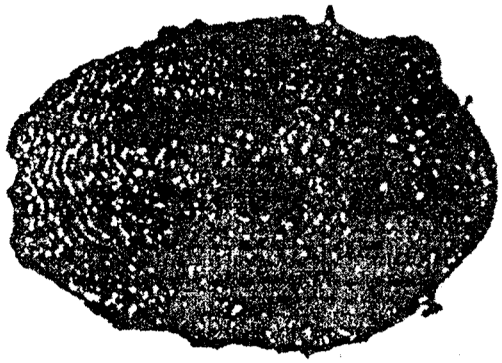


Fig. 3 : Ultrastructure of a plasma cell. Note the extensive rough endoplasmic reticulum for antibody production.

only produce antibody for expression on their cell surface and remain able to respond to antigen if it is reintroduced.

In contrast, plasma cells do not have cell surface antibody receptors. Rather these cells, which are derived from activated B-cells, function as factories producing and secreting large amounts of antibody of the same specificity as the antigen receptor on the stimulated parent B-cell. The morphology of a plasma cell (Fig. 2) shows the high degree of adaptation of this cell type to high rate

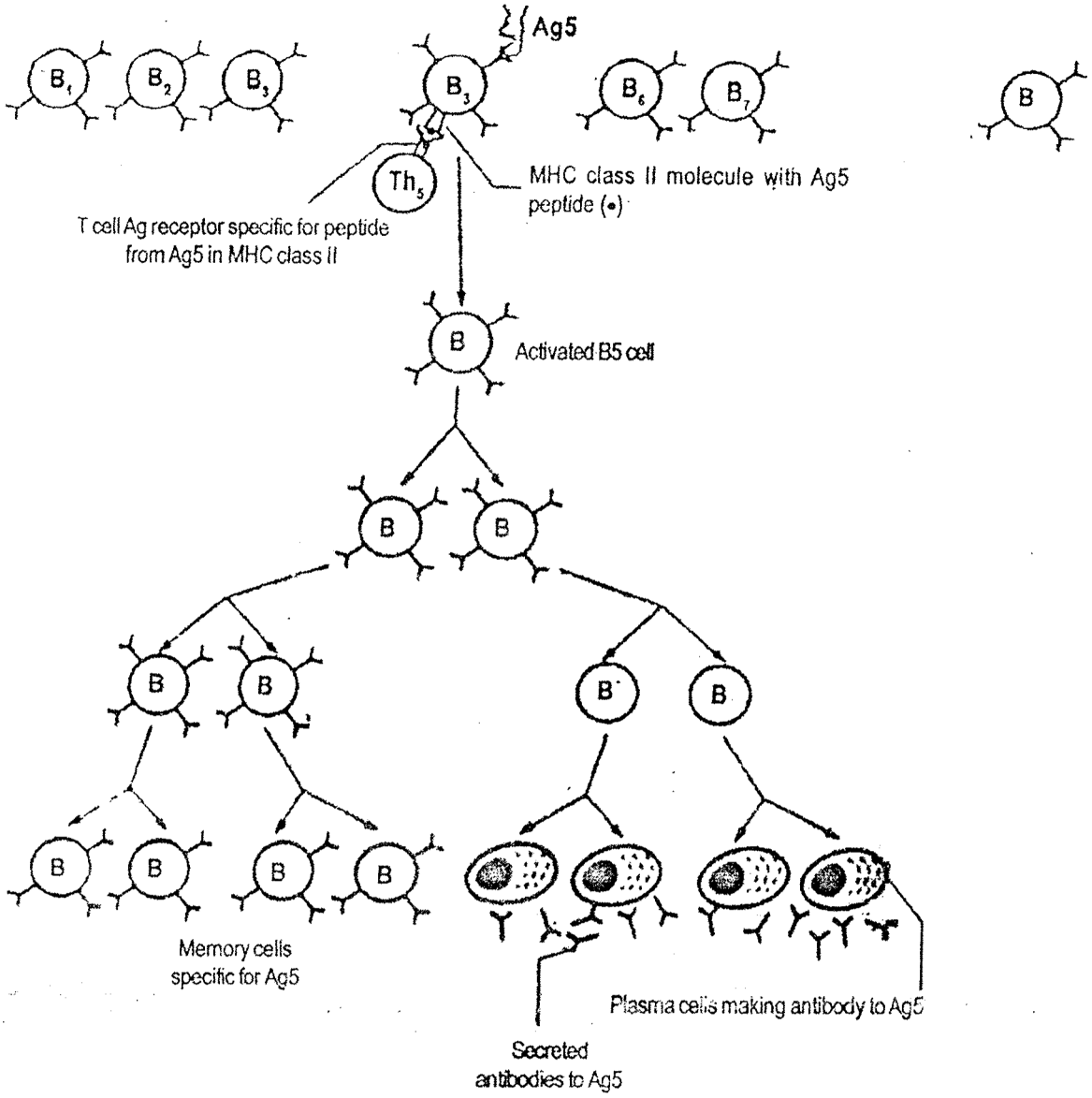


Fig. 4: Clonal selection, memory cells and plasma cells.

-glycoprotein (antibody) synthesis. This includes extensive endoplasmic reticulum, mitochondria, and golgi apparatus. It should be noted that a plasma cell only produces antibodies of one specificity and class/subclass.

When antigen is introduced into an individual, B-cells with receptors for that antigen bind

and internalize it into an endosomal compartment, and process and present it on MHC class II molecules to helper T-cells. These B-cells are triggered to proliferate, giving rising to clones of large numbers of daughter cells. Some of the cells of these expanding clones serve as memory cells, others differentiate and become plasma cells which make and secrete - large quantities of specific antibody. For example, on introduction of antigen (Ag5) into a person (Fig . 4), more than 10^6 B cells have the opportunity to interact with it. Only a very few B cells (e.g. B5) have receptors specific for this antigen. It binds Ag5, internalizes, and processes and presents it on MHC class II molecules on the surface of this B cell. T helper cells with specific receptors for a peptide from Ag5 in MHC class II bind to this complex and stimulate this B cell to clonally expand and differentiate into memory B cells and plasma cells which produce soluble antibody to Ag5. In addition, direct T cell interaction with the B cell induces class switching, which depending on the type of helper cell (Th1 vs Th2) and the cytokines it secretes, will result in production of antibody of the IgG, IgA or IgE classes.

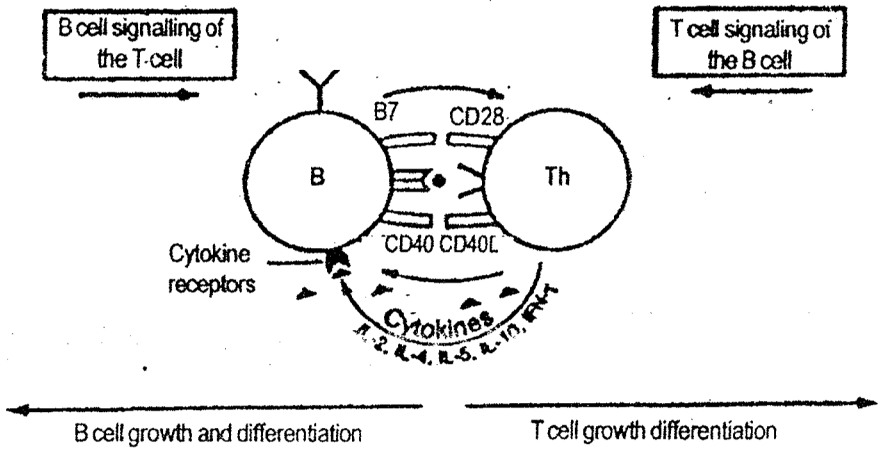


Fig. 5 : Reciprocal activation of T and B cells.

Mechanism of B-cell Activation :

More specifically, Th cells recognize antigenic peptides on the surface of antigen-specific B cells because the B cells are able to capture antigen specifically via membrane (m) IgM and mIgD. This feature of B cells, to capture, process and present specific antigen, makes them unique amongst

the antigen-presenting cells which normally take up antigen via scavenger and other receptors. The antigen is then endocytosed, degraded via the exogenous processing pathway and peptides are associated with class II MHC molecules.

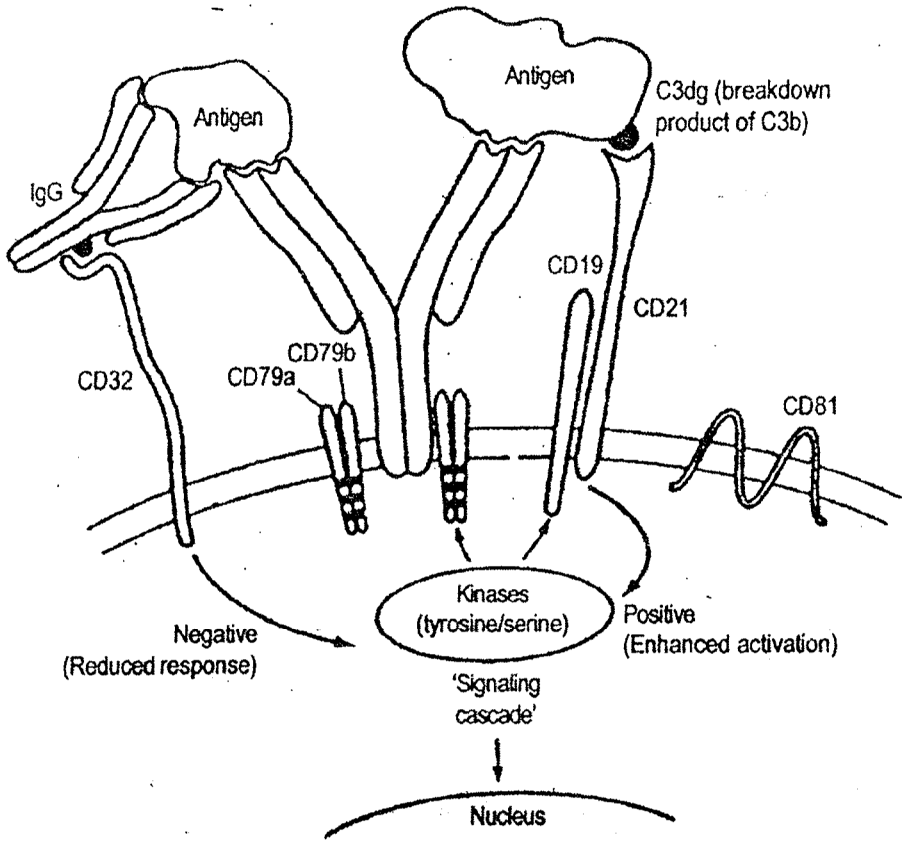


Fig. 6 : Activation of B cells via the BCR and co-receptor complex. Attachment of antigen results in activation of several kinases and CD79a/b are phosphorylated on their ITAMs. Binding of CD21 to antigen bound to complement (C3dg) regulates signaling via CD19 in a positive way, whilst interaction of CD32 with IgG antibody bound to antigen and the antigen receptor, provides a negative signal.

The cells whose TCR are specific for that peptide-MHC interactions and through engagement of adhesion molecules.

Once triggered via the TCR, the Th cells express CD40 ligand (CD40L), the ligand for the B cell surface molecule CD40. This Th cell now triggers the activation of the B cell via the CD40 surface receptor. As a result, the activated B cell reciprocally co-stimulates the Th cell via CD28. At this time, both the T cell and B cell are stimulated. T cells then produce cytokines including IL-2 (autocrine growth factor for the Th cells) and IL-4 and IL-5 (growth and differentiation factors for

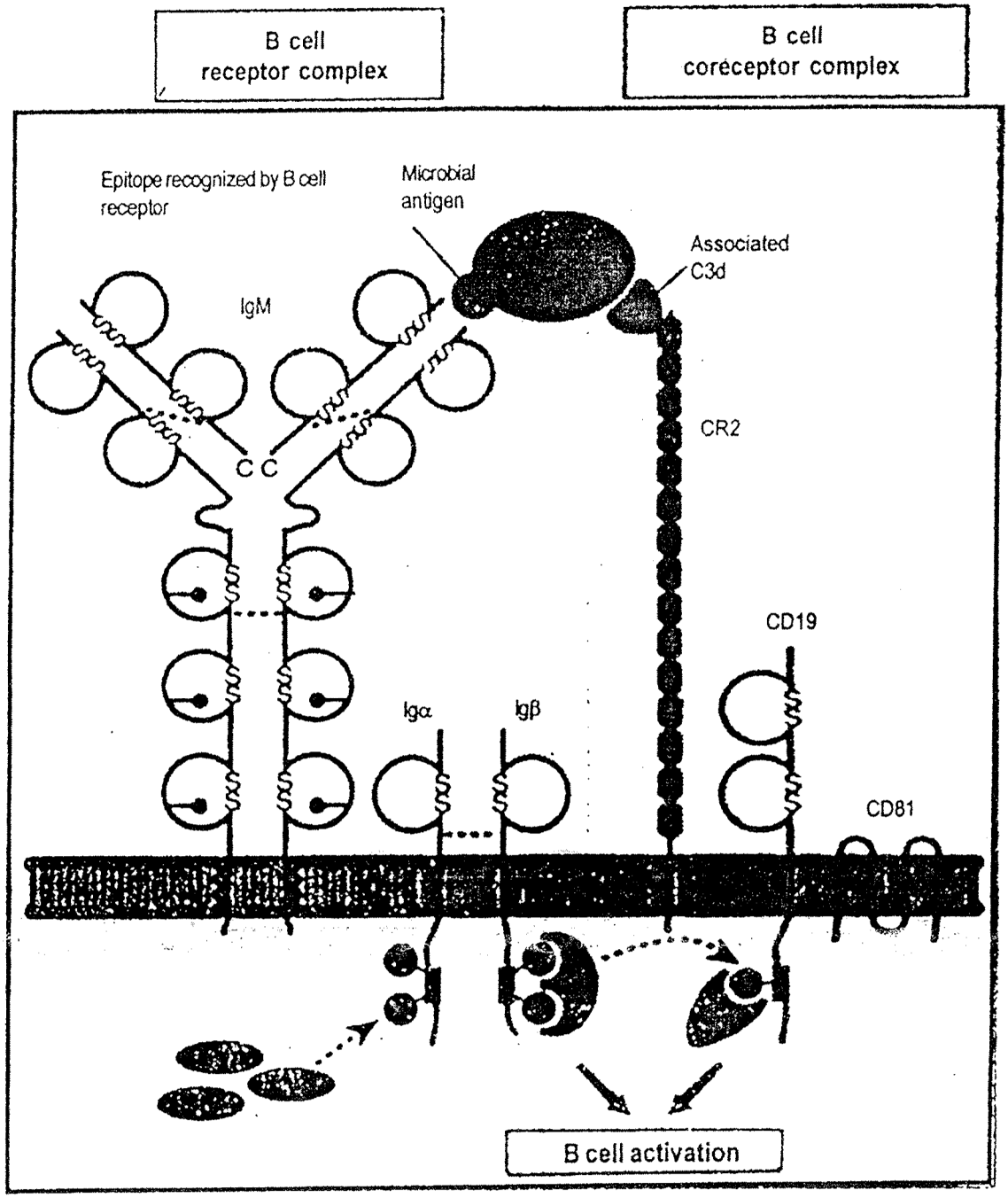


Figure 7 : B-cell-receptor and conception complex.

B cells express a complex of the CR2 complement receptor, CD19, and CD81. Microbial antigens that have bound complement fragment C3d can simultaneously engage both the CR2 molecule and the membrane Ig on the surface of a B cell. This leads to the initiation of signaling cascades from both the B cell receptor complex and the CD19 molecule. The response to C3d-antigen complexes is greatly enhanced compared with the response to antigen stimulation alone, Ig, immunoglobulin.

the activated B cells). As a result, both the T cells and B cells clonally expand and differentiate. Ligation of B cell CD40 by CD40L on T cells is also important in that it rescues B cells with mutated antibody receptors from death in germinal centers.

The transmembrane surface immunoglobulin antigen receptors on B cells, like the TcR on T cells have short intracytoplasmic tails unable to transduce signals themselves. Therefore, on engagement of the B cell antigen receptor, other molecules associated with these receptors mediate signaling. In particular, CD79a and b (Igα and Igβ) of the B cell receptor complex contain ITAMs which are phosphorylated during the early stages of activation and initiate the B cell signaling cascade. Other members of the B cell receptor complex modulate the initial signals mediated through antigen binding and enhance the strength of cell-cell interaction. For example, CD21 (CR-2) a complement receptor which binds C3d, may provide an additional positive signal in B cell activation as a result of binding to C3d associated with antigen. As with T cell activation, these processes in B cells are mediated through phosphatases and kinases. The importance of one kinase, btk, is indicated by the observation that mutation of the gene encoding it results in the absence of B cells (Bruton's agammaglobulinemia).

B cells like T cells require two signals for their activation. Binding of soluble antigen to the antibody receptor alone **(signal-1)** results in apoptosis. This is seen experimentally using antibodies to the sIgM on B cells. Proliferation is induced in the presence of another second signal such as that provided by IL-2. The physiological second signal **(Signal-2)** for B cell activation is provided by interaction of Th cell CD40L with CD40 on the B cell surface. This is also an absolute requirement for class switching. Other cytokines produced by Th cells induce appropriate signals important to differentiation of the B cells into plasma or memory cells. After initial B cell activation and following class switching to IgG, B cells are susceptible to regulation by concomitant binding of FcγRII (CD32) and the B cell antigen receptor. In particular, further activation of these cells can be inhibited by their binding of specific antigen attached to IgG. As a result, CD32 transmits a negative signal to the B cell and prevent its activation ('negative feedback').

TI antigens, which do not induce IgG responses (since their CD40 molecules are not ligated), receive their second signal via the 'mitogenic' component of T-II antigens and probably via antibodies on B cells by T-I antigens with repeating antigenic units allows clustering of co-receptors which is also likely to enhance signaling in activation of T-D B cells.

HUMORAL IMMUNE RESPONSE

Contact with antigen does not invariable lead to formation of antibody. A number of conditions must be met before antibody is produced. Among the important parameters to be consider when an attempt is made to elicit an immune response are the amount of antigen given, the time of injections, and the route of injections.

Kinetics of the Immune Response :

The first exposure of an individual to a particular antigen is referred to as the priming event and the measurable response that ensues is called the primary response. Primary response may be divided into several phases.

Primary response :

Latent or lag phase :

After initial injection of antigen a significant amount of time elapses before antibody is detectable in the serum. The length of this period is generally 1-2 weeks, depending on the species immunized, the antigen, and the type of adjuvant used. The length of the latent period is also heavily dependent on the sensivity of the assay used to measure the antibody.

The latent period includes the time taken for B-cells to make centact with the antigen, to proliferate, to differentiate and then to secrete antibody in sufficient quantity that can be detected in the serum.

The less sensitive the assay used for detection of antibody, the more antibody will be required for detection and the longer the apparent latent period will be.

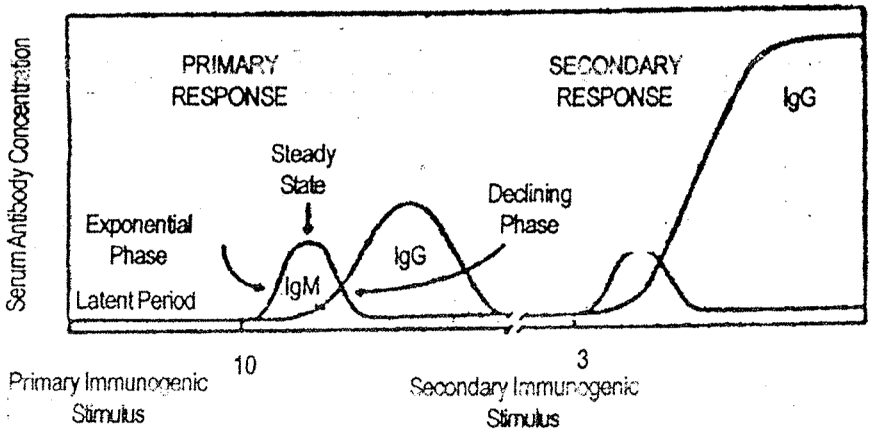


Fig. 8 : The kinetics of an Immune Response.

Exponential production phase :

During this phase, the concentration of antibody in the serum increases exponentially.

Declining Phase :

After an interval during which production and degradation of antibody balance each other (the steady state) the immune response begins to shut down, and the concentration of antibody in serum declines rapidly.

In this primary response, the first class of antibody detected is generally IgM, which in some instances may be the only class of immunoglobulin that is made. If production of IgG antibody ensues, its appearance is generally accompanied by a rapid cessation of production of IgM.

Secondary Response :

(i) Although production of antibody after a primary contact with antigen may cease entirely with in a few weeks, the immunized individual is left with a cellular memory (i.e. leftover memory cells) of this contact. This memory becomes apparent when a response is induced by a second injection of antigen.

(ii) After the second injection, the lag phase is considerably shorten and antibody may appear in less than half the time required for the primary response. The production of antibody is much greater, and higher concentrations of antibody are detectable in the serum. The production of antibody may also continue for a longer period, with persistant levels remaining in serum for months, or even years, later.

(iii) There is a marked change in the type and quality of antibody produced in the secondary response. There is a shift in class response, with IgG antibodies appearing at higher concentrations and with greater persistence, than IgM. In addition, a maturation of the response occurs, such that the average affinity (binding constant) of this antibodies for the antigen increases as the secondary response develops. The driving force for this increase in affinity may be a selection process during which B-cells compete with free antibody to capture a decreasing amount of antigen.

(iv) Thus, only those B-cells clones with high-affinity Ig-receptors on their surfaces will bind enough antigen to ensure that the B-cells are triggered to differentiate into plasma cells. The plasma cells, which arise from preferentially selected B-cells, synthesize the antibody with high affinity for antigen.

(v) The capacity to make a secondary or anamuestic (memory) response may persist for a long time (months in mice, years in human) and it provides an obvious selective advantage for an individual that survives the first contact with an invading pathogen.

(vi) Establishment of this memory for generating specific responses is, of course, the purpose of public health immunisation programs.

THE COMPLEMENT SYSTEM

1. INTRODUCTION:

The complement system is a protective system. Complement consists of a group of serum proteins, they act in concert in any orderly sequence. These are not immunoglobulins and their concentrations do not increase after immunization. They exhibit a variety of biological properties. One of which is the ability to lyse cells or microorganisms.

2. DISCOVERY:

Research on complement began in the 1890, when Jules Bordet at the Institute of Pasteur in Paris showed that sheep antiserum to the bacterium *vibrio cholerae* caused lysis of bacteria and that heating the antiserum at 56°C for 30 min. abolished the activity. Addition of normal fresh serum restore its lytic activity. He concluded that antibodies plus a heat component is necessary for the lysis. Bordet termed this heat labile component as "Alexine". Later Paul Ehrich designated this compound as "complement".

3. THE COMPONENTS OF COMPLEMENTS:

The complement component molecules (C) include C_1 (C_{1q}, C_{1r}, C_{1s}), C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9 as well as a set of molecules which are primarily associated with the alternative pathway, including factor B and Factor D. The peptide fragments formed by activation of a component are denoted by small letters. In most cases, the smaller fragment resulting from cleavage of a component is designated "a" and the larger fragment "b" (e.g.-C3a, C3b), note that C2 is an exception. C2a is the larger cleavage fragment. The larger fragments bind to the target near the site of activation and the smaller fragments diffuse from the site and can initiate localized inflammatory response by binding to specific receptors.

3.1 SITE OF SYNTHESIS:

The proteins and glycoproteins that compose the complement system are synthesized mainly by liver hepatocytes, although significant amounts of them are also produced by blood monocytes,

tissue macrophages and epithelial cells of the gastrointestinal and genitourinary tracts. These components constitute 5% (by weight) of the serum globulin fraction and most circulate in the serum in functionally inactive forms. Many components are proenzymes which are inactive untill proteolytic cleavage. The complement fragments interact with one another to form functional complexes. Those complexes that have enzymatic activity are designated by a bar over the number or symbol (e.g.- $\overline{C4b2a}, \overline{C3bBb}$).

3.2. COMPLEMENT ACTIVATION:

Complement can be activated directly by certain molecules associated with microbes (alternative and lectin pathway) or by antibodies bound to a microbe or any other antigen (classical pathway). The early steps in complement activation, culminating in formation of C_{5b}, can occur by the classical pathway, the alternative pathway or the lectin pathway. The final steps that lead to a membrane attack are the same in all pathways.

4. THE CLASSICAL PATHWAY:

Complement activation by the classical pathway commonly begins with the formation of soluble antigen-antibody complexes (immune complexes) or with the binding of antibody to antigen on a suitable target, such as a bacterial cell. IgM and certain subclasses of IgG (human IgG_1, IgG_2 & IgG_3) can activate the classical complement pathway, as can certain non-immunologic activators. The initial stage of activation involves C_1, C_2, C_3 and C_4, which are present in plasma in functionally inactive forms. Because the components were named inorder by their discovery and before their functional roles had been determined, the numbers in their names do not always reflect the order in which they out.

4.1. STRUCTURE OF C1-MACROMOLECULAR COMPLEX:

The complexing of antibody with antigen induces conformational changes in the F_c portion of the IgM molecule that expose a binding site for the C1 component of the complement system. C1 in serum is a macromolecular complex consisting of C_{1q} and two molecules each of C_{1r} and C_{1s} held together in a complex ($C_1qr_2s_2$) stabilized by Ca^{++} ions. The C_{1q} molecule is composed of 18

polypeptide chains that associated to form sixcollagen-like triple helical arms, the tips of which bind to exposed C_{1q} binding sites in the CH_2 domain of the antibody molecule. The $C_1qr_2s_2$ complex can exist in two configurations. When it is free and not bound to C_{1q}, it assumes an 'S'-shaped form; on binding to C_{1q}, $C_1qr_2s_2$ assumes a shape similar to figure '8'. Each C_{1r} and C_{1s} monomer contains a catalytic domain and an interaction domain, the later facilities interaction with C_{1q} or with each other.

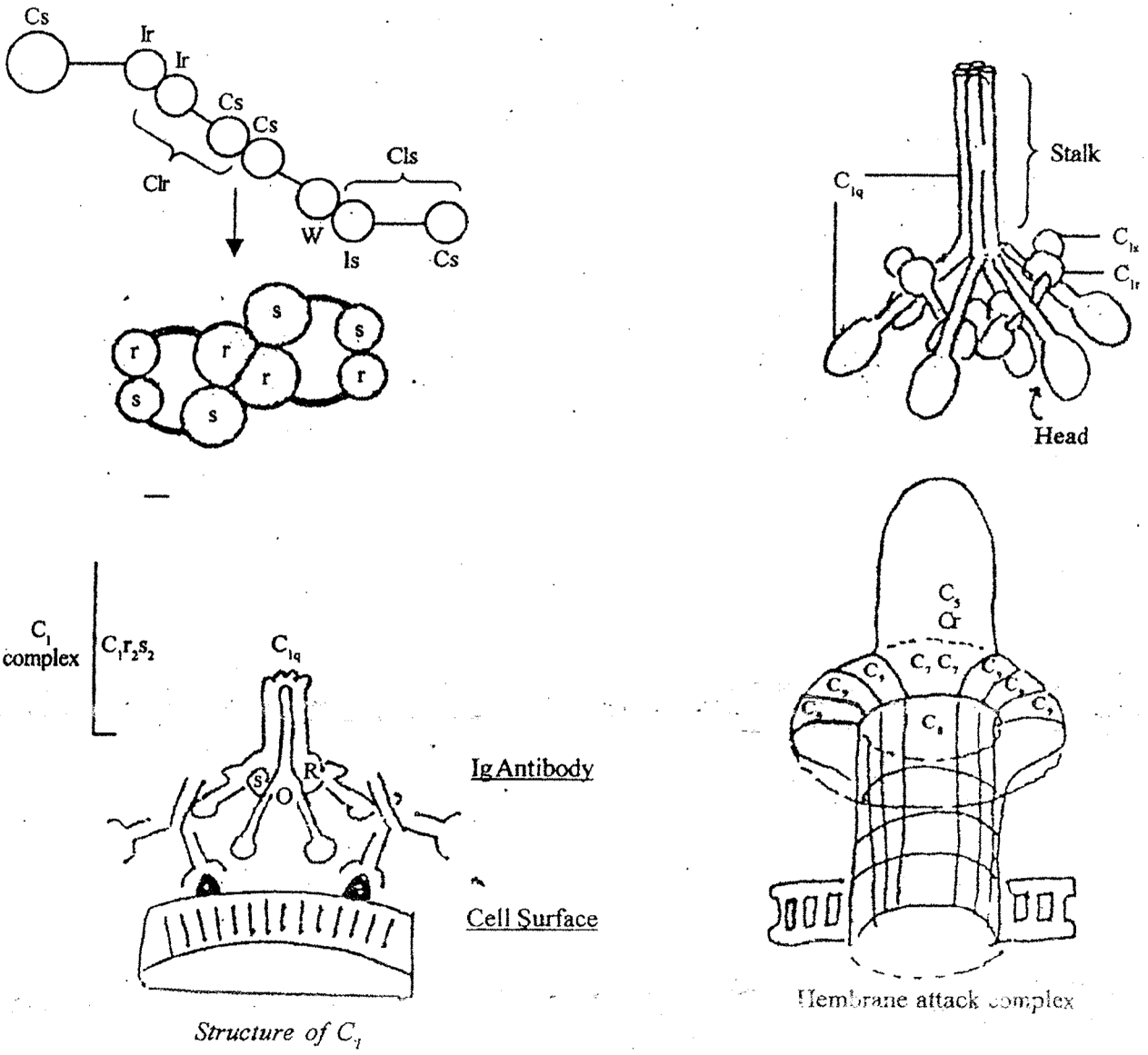


Fig. 9: Structure of C_1-Macromolecular complex

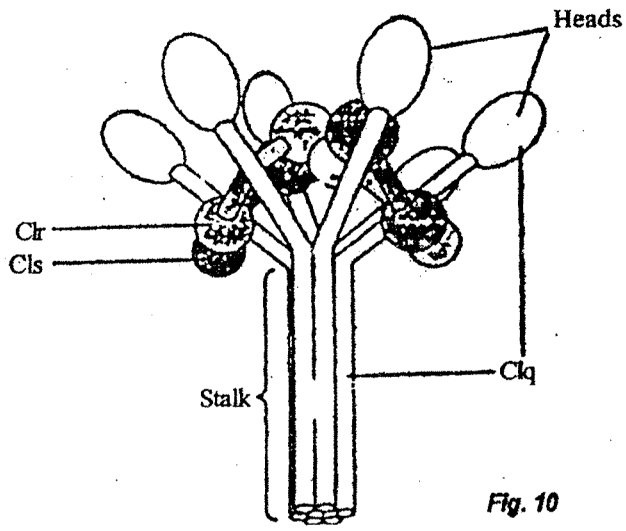


Fig. 10

STAPLE AND PLANAR FORM OF IgM :

Each C_1 molecules must bind by its C_{1q} globular heads, to atleast two F_c sites for a stable C_1– antibody interaction to occur. When pentameric IgM is bound to antigen on a target surface it assumes the so called 'staple' configuration, in which at least three binding sites for C_{1q} are exposed. Circulating IgM, however, exists as a planar configuration in which the C_{1q}– binding sites are not exposed. For this reason, circulating IgM cannot activate the complement cascade by itself. An IgG molecule on the otherhand, contains only a single C_{1q}– binding site in the CH_2 domain of the Fe, so that firm C_{1q} binding is achieved only when the two IgG molecules are within 30-40 nm of each other on a target surface or in a complex, providing two attachment sites for C_{1q}.

This difference in the structure of IgM and IgG accounts for the observation that a single molecule of IgM bound to a red blood cell is enough to activate the classical complement pathway and lyse the red blood cells, whereas some 1000 molecule of IgG are required if two molecules, randomly distributed, are to end up close enough to each other to initiate C_{1q} binding.

MECHANISM OF ACTIVATION :

The intermediates in the classical activation pathway are depicted schematically in the Figure. Binding of C_{1q} to F_c binding sites induces a conformational change in C_{1r} that converts C_{1r} to an

active serine protease enzyme, $\overline{C_{1r}}$, which then cleaves C_{1s} to a similar active enzyme, $\overline{C_{1s}}$. $\overline{C_{1s}}$ has two substrates, C_4 and C_2. The C_4 component is a glycoprotein containing three polypeptide chains α, β and γ. C_4 is activated when $\overline{C_{1s}}$ hydrolyzes a small fragment (C_{4a}) from the amino terminus of the α-chain, exposing a binding site on the larger fragment (C_{4b}). The C4b fragment attaches to the target surface in the vicinity of C_1, and the C_2 proenzyme then attaches to the exposed binding side on C4b, when the C_2 is then cleared by the neighbouring $\overline{C_{1s}}$; the smaller fragment (C2b) diffuses away. The resulting $\overline{C4b2a}$ complex is called C_3 convertage, referring to its role in converting the C_3 proenzyme into an enzymatically active form. The smaller fragment from C_4 cleavage, C4a, is an anaphylatoxin or mediator of inflammation, which does not participate directly in the complement cascade; the anaphylatoxins, which include the smaller fragments of C_4, C_3 and C_5.

ACTION OF C_3 CONVERTAGE AND C_5 CONVERTAGE :

The native C_3 component consists of two polypeptide α β chains, and hydrolysis of a short fragment (C_3a) from this C_3b. A single C_3 convertase molecule can generate over 200 molecule of C_3b, resulting in tremendous amplification at this step of the sequence. Some of the C_3b bind to $\overline{C4b2a}$ to form a *trimolecular complex,* $\overline{C4b2a3b}$, called C_5 convertase. The C_3b component of this complex binds C_5 and alters its conformation, so that the C4b2a component can cleare C_5 into C_5a, which diffuses away, and C_5b, which attaches to C_6 and initiate formation of the *membrane attack complex* (MAC). Some of the C_3b generated by C_3 convertase activity does not associate with $\overline{C4b2a}$; instead it diffuses away and then coats immune complexes and particulate antigens, functioning as an opsonins, C_3b may also bind directly to cell membrane.

THE ALTERNATIVE PATHWAY :

The alternative pathway is an antibody - independent means to generate bound C_5b, the same product as the classical pathway generates in a sequence for initiation. Because no anti body is required, the alternative pathway is a component of the immune system. This major pathway of complement activation involves four serum proteins: C3, factor B, factor D and properdin. Unlike

the classical pathway, the alternative pathway is initiated without requiring antibody, in most cases by various cell-surface constituents that are foreign to the host. For example, both gram-negative bacteria have cell wall constituents that can activate the alternative pathway. The intermediates in the alternative pathway for generating C_{5b} are depicted schematically in the Fig. 11.

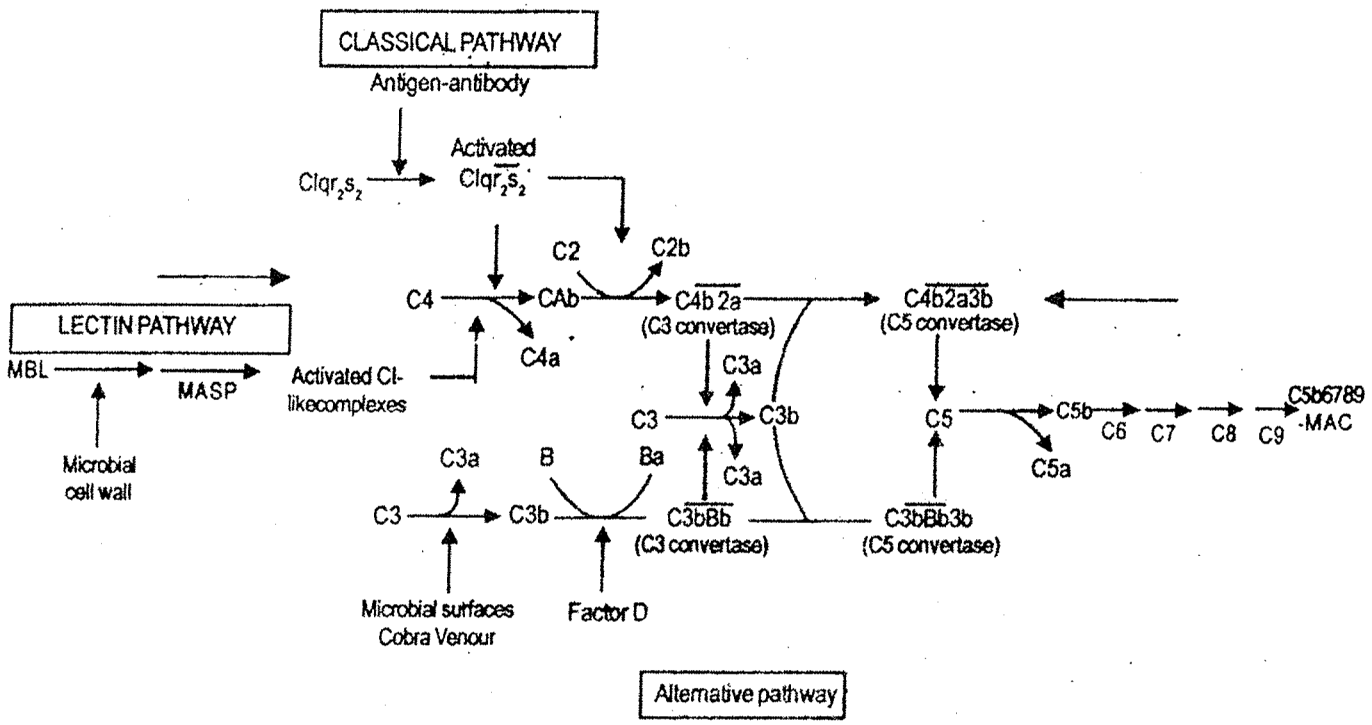


Fig. 11 : *Overview of the complement activation pathways. The classical pathway is initiated by binding of C1 to antigen-antibody complexes. The alternative pathway is initiated by binding of C3b to activating surfaces such as microbial cell walls. The lectin pathway is initiated by binding of the serum protein MBL to the surface of a pathogen. The alternative pathway utilizes some components (or factors) that are unique, whereas the classical and the lactin pathways have all components in common after C1. All three pathways generate C3 and C5 convertases and bound C5b, which is converted into a membrane-attack complex (MAC) by a common sequence of terminal reactions. Hydrolysis of C3 is the major amplification step in all pathways, generating large amounts of C3b, which forms part of C5 convertase. C3b also can difuse away from the activating surface and bind to immune complexes or foreign cell surfaces, where it functions as an opsonin.*

FACTOR B AND FACTOR D :

In the classical pathway C_3 is cleaved to C_{3a} and C_{3b} rapidly by the enzymatic activity of the C3 convertase. In the alternative pathway, serum C_3, which contains an unstable thioester bond, is subject to slow spontaneous hydrolysis to yield C_{3a} and C_{3b}. The C_{3b} component can bind to foreign surface

antigens (such as these on bacterial cells or viral particles) or over to the host's own cells. The membranes of most mammalian cells have high levels of sialic acid, which contributes to the rapid inactivation of bound C_{3b} molecules on host cells, consequently this binding rarely leads to further reactions on the host cell membrane. Because many foreign antigenic surfaces (eg, -bacterial cell walls and certain viral envelopes) have only low levels of sialic acid, C_{3b} bound to these surfaces remain active for a longer time. the C_{3b} present on the surface of the foreign cells can bind another serum protein called factor B by way of a Mg^{++} dependent bond. Binding to C_{3b} exposes a site on factor B that serves as the substrate for an enzymatically active serum protein called factor D. Factor D cleaves the C_{3b} - bound factor B, releasing a small fragment (Ba), which diffuses away, and generates C_{3b} Bb, the C_{3b}Bb complex has C_3 convertase activity and thus is analogous to the C_{4b2a} complex in the classical pathway. The C_3 convertase activity of C_{3b} Bb has a halfilife of only 5 mins. Unless the serum protein properdin binds to it, stabilizing it and intending the half life of this convertase activity to 30 minutes.

The C_{3b} Bb generated in the alternative pathway can activate unhydrolysed C_3 to generate more C_{3b} autocatalytically. As a result the initial steps are repeated and amplified, so that more than 2×10^6 molecules of C_{3b} can be deposited on an antigenic surface in less than 5 minutes. The C_3 convertase activity of C_{3b}Bb generates C_{3b}Bb complex, which exhibits C_5 convertase activity, analogous to the $\overline{C4b2a3b}$ complex in the classical pathway. The non-enzymatic C_3b component binds C_5, and the Bb component subsequently hydrolyzes the bound C_5 to generate C_5a and C_5b; the later binds to the antigenic surface.

THE LECTIN PATHWAY :

Recently, an additional means by which the complement cascade can be activated is called the lectin pathway. Lectins are protein that bind to the carbohydrate. The lectin pathway, like the alternative pathway, does not depend on antibody for its activation. However, the mechanism is more like that of the classical pathway, because after initiation, it proceeds, through the action of C_4 and C_2, to produce a C_3 convertase.

The pathway is activated by the binding of mannose-binding lectin (MBL) to mannose residues on glycoproteins or carbohydrates on the surface of microorganisms. MBL is acute phase protein

produced in inflammatory response. Its function in the complement pathway is similar to that of c_1q, which it resembles in structure. After MBL binds to the surface of a cell or pathogen, MBL-associated serine protease, or MASP, binds to it. The active complex formed by this association causes cleavage and activation of C_4. MASP has structural simiarilty to C_{1r} and C_{1s} and mimics their activities. Details of the lectin pathway is not fully known, but the means of activating the C_2–C_4 components to form C_5 convertase without need for specific antibody binding certainly represents an important innate defence mechanism comparable to the alternative pathway.

FORMATION OF MEMBRANE-ATTACK COMPLEX (MAC):

The terminal sequence of complement activation involves C_{5b}, C_6, C_7, C_8 and C_9, which interact sequentially to form a macromolecular structure called the membrane attack complex (MAC). This complex displaces the membrane phospholipids, forming a large transmembrane channel that disrupts the membrane of the target cell and enable ions and small molecules to diffuse through it freely.

In the classical, lectin and alternative pathways, a C_5 convertase cleaves C_5, which contains two protein chains (α and β). After binding of C_5 to the nonenzymatic C_{3b} component of the convertase, the amino terminus of the α-chain is cleaved. This generates the small C_{5a} fragment, which diffuses, away and the large C_{5b} fragment, which binds to the surface of the target cell and provides a binding site for the subsequent components of the membrane attack complex. The C_{5b} component is extremely labile and is inactivated within 2 minutes unless C_6 binds to it and stabilizes its activity.

All the complement reactions take place on the hydrophilic surface of membranes or on immune complexes in the field phase. As $C_{5b}6$ binds to C_7, the resulting complex undergoes a hydrophilic-amphiphilic structural transition that exposes hydrophobic regions, which serve as binding sites for membrane phospholipids. If the reaction occurs on a target cell membrane, the hydrophobic binding sites enable the $C_{5b}67$ complex to insert into the phospholipid bilayer. If, however, the reaction occurs on an immune complex or other noncellular activating surface, then the hydrophobic binding sites can not anchor the complex and it is released. Released $C_{5b}67$ complexes can insert into the membrane of nearby cells and mediate "innocent-bystander" lysis, although this rarely occurs. In a number of diseases in which immune complexes are produced, tissue damage may result from such innocent bystander lysis.

Binding of C_8 to membrane-bound $C_{5b}67$ induces a conformational change in C_8, so that it too undergoes a hydrophilic-amphiphilic structural transition, exposing a hydrophobic region, interacts with the plasma membrane. The $C_{5b}678$ complex creates a small pore 10 $\overset{0}{A}$ in diameter; formation of this pore can lead to lysis of red blood cells but not of nucleated cells. The final step in formation of the MAC is the binding and polymerization of C_9 a perforin like molecule, to the $C_{5b}678$ complex.

As many as 10-17 molecules of C_9 can be bound and polymerized by a single $C_{5b}678$ complex. During polymerization, the C_9 molecules undergo a hydrophilic amphophilic transition, so that they also can insert into membrane. The completed MAC, which has a tubular form and functional pore size of 70-100 $\overset{0}{A}$, consists of a $C_{5b}678$ complex surrounded by a poly C_9 complex. Since ions and small molecules can diffuse freely through the central channel of the MAC, the cell cannot maintain its osmotic stability and is killed by an influx of water and loss of electrolytes.

BIOLOGICAL CONSEQUENCES OF COMPLEMENT ACTIVATION :

Complement serves as an important mediator of the humoral response by amplifying the response and converting it into an effective defense mechanism to destroy invading microorganisms and virus, while other complement components or split products participate in the inflammatory response, opsonization of antigen, viral neutralization, clearance of immune complexes. Many of the biological activities of the complement system depend on the binding of complement fragments to complement receptors (e.g., CR1, CR2, CR3 and CR4), which are expressed by various cells.

CELL LYSIS :

The membrane-attack complex formed by complement activation is capable of lysing a broad spectrum of microorganisms, viruses, erythrocytes and nucleated cells. Because the alternative and lectin pathways of activation of generally occur without an initial antigen-antibody interaction, these pathways serve as important innate systems of nonspecific defense against infectious microorganisms. The requirement for an initial antigen-antibody reaction in the classical pathway supplements these non-specific innate defenses with a more specific defense mechanism. Nevertheless, antibody and complement do play a role in host defense against viruses and are often crucial in containing viral spread during acute infection and in protection against reinfection.

INFLAMMATORY RESPONSE :

The smaller fragments resulting from complement cleavage, C_{3a}, C_{4a} and C_{5a}, called anaphylatoxins, bind to receptors on most cells and blood basophils and induce degranulation with release of histamine and other pharmacologically active mediators. The anaphylatoxins induce smooth-muscle contraction and increase vascular permiability. The activities of these highly reactive anaphylatoxins are regulated by a serum protease called carboxypeptidase N, which cleaves an Arg residue from the C-terminus of the molecules, resulting in what are termed the des-Arg forms. The des-Arg forms of C_{3a} and C_{4a} are completely inactive while that of C_{5a} retains about 10% of its chemotactic activity and 1% of its ability to cause smooth muscle contraction.

Each of C_{3a}, C_{4a} and C_{5b67} can induce monocytes and neutrophils to adhere to vascular endothelial cells, extravasate through the endothelial lining of the capillary, and migrate toward the site of complement activation in the tissues.

OPSONIZATION OF ANTIGEN :

C_{3b} is the major opsonin of the complement system, although C_{4b} and C_{3b} also have opsonizing activity. The amplification that occurs with C_3 activation results in a coating of C_{3b} on immune complexes and particulate antigens. Phagocytic cells, as well as some other cells, express complement receptors (CR_1, CR_2 and CR_4) that bind C_{3b}, C_{4b} or C_{3b}. When antigen has been coated with C_{3b} during complement activation, the coated antigen binds to cells bearing CR_1. If the cell is a phagocyte, phagocytosis will be enhanced.

VIRAL NEUTRALIZATION :

The complement system plays in important role in host defense by neutralizing viral infectivity. Some viruses can activate the alternative, lectin or even the classical pathway in the absence of antibody. The complement system mediates viral neutralization by a number of mechanisms. Some degree of neutralization is achieved through the formation of layer viral aggregates, simply because these aggregates reduce the net number of infectious virus particles. Although antibody does play a role in the formation of viral aggregates, in vitro studies show that the C_{3b} component facilitates aggregate formation in the presence of as little as two molecules of antibody per virion.

The binding of antibody and or complement to the surface of a viral particle creates a thick

protein coating that can be visualized by electron microscopy. This coating neutralizes viral infectivity by blocking attachment to susceptible host cells. The deposits of antibody and complement on viral particles also facilitate binding of the viral particles also facilitate binding of the viral to cell processing Fc or type 1 complement receptors (CRI). Finally, complement is effective in lysing most, if not all, enveloped viruses, resulting in fragmentation of the envelope and disintegration of the nucleocapsid.

SOLUBILIZATION OF IMMUNE COMPLEX :

The role of the complement system in clearing immune complexes can be seen in patients with the autoimmune desease systemic lupus erythromatosus (SLE). These individuals produce large quantities of immune complexes and suffer tissue damage as a result of complement-mediated lysis and the induction of type II or type III hypersensitivity. Although complement plays a significant role in the development of tissue damage in SLE, the paradoxical finding is that deficiencies in C_1, C_2, C_4 and CRI predispose an individual to SLE; indeed 90%, of the individuals who completely lack C_4 develop SLE. The complement deficiencies are thought to interfere with effective solubilization and clearance of immune complexes, the result is the persistence of the complexes and subsequent tissue damage by the very system whose deficiency was to blame.

The coating of soluble immune complexes with C_3b is thought to facilitate their binding to CRI on erythrocytes. Although red blood cells express lower levels of CRI (~5 x 10^2/cell) than granulocytes do (~5 x 10^4/cell), there are about 10^3 RBC for every WBC; Therefore erythrocytes account for about 90% of the CRI in the blood. For this reason erythrocytes play an important role in binding C3b coated immunecomplexes and varrying these complexes to the liver and spleen. In these organs, immune complexes are strriped from the red blood cells and are phagocytosed, thereby preventing their deposition in tissues.

SHORT SUMMARY :

The complement system includes serum and membrane proteins that interact in a highly regulated manner to produce biologically active protein products. The system includes three convergent proteolytic pathways, compased of several different zymogens. The classical pathway is initiated by antigen-antibody complexes, the alternative pathway is usually activated directly by the surface of infection organisms.

PHAGOCYTES

Mobile phagocytes:

Phagocytes are specialized 'eating' cells (phago - I eat, *Latin*) of which there are two main types, neutrophils and macrophages. Neutrophils, often called polymorphonuclear cells (PMNs) because of the multilobed nature of their nuclei (*Fig.*), are mobile phagocytes that comprise the majority of blood leukocytes (about 8×10^6 ml^{-1} of blood). They have a very short half-life (6 hours) and die in the blood stream by apoptosis (programmed cell death). They are granular leukocytes which stain with neutral dyes and have a different function from those granulocytes which stain with eosin (eosinophils), or basic dyes (basophils). Their granules contain peroxidase, alkaline and acid phosphatases, and defensins (small antibiotic peptides) which are involved in microbial killing. They have receptors for chemoattractive factors released from microbes, e.g. muramyl dipeptide (MDP), and for complement components activated by microbes. Their main function is to patrol the body via the blood stream in search of invading macrobes. As such they are pivotal cells in acute inflammation. Like the majority of cells involved in the immune system, these phagocytes are produced in the bone marrow (Table-1).

Table 1. Cells of the mononuclear phagocyte system

Cells	Location
Monocytes	Blood stream
Kupffer cells	Liver
Mesengial cells	Kidney
Alveolar macrophages	Lungs
Microglial cells	Brain
Sinus macrophages	Spleen, lymph nodes
Serosal macrophages	Peritoneal cavity

The mononuclear phagocyte system:

The mononuclear phagocyte system (previously called the reticuloendothelial system), is the name given to the widely distributed tissue bound phagocytic system, whose major function is to

dispose of microbes and dead body cells through the process of phagocytosis. Monocytes are blood borne precursors of the major tissue phagocytes, macrophages (Fig.). Different organs/tissues each have their versions of monocyte-derived phagocytic cells (Table 1).

The phagocytic process:

This is a multistep process which ultimately concludes with the demise of a microbe. The stages in this process are listed in Table 2. Phagocytosis is the major mechanism by which microbes are removed from the body and is especially important for defence against extracellular microbes.

Opsonization:

This is the process of making a microbe easier to phagocytose. A number of molecules called 'opsonins' (to 'make more tasty' - Greek) do this by coating the microbes. They aid attachment of the microbe to the phagocyte and also trigger activation of phagocytosis. Opsonins include the complement component C3b and antibody itself, the latter acting as a bridge between the innate and adaptive immune systems. Phagocytes use their surface receptors which bind to C3b, or which bind to the Fc region of IgG antibody (Fc receptors, FcR) to attach to C3b or IgG coating the microbes, respectively.

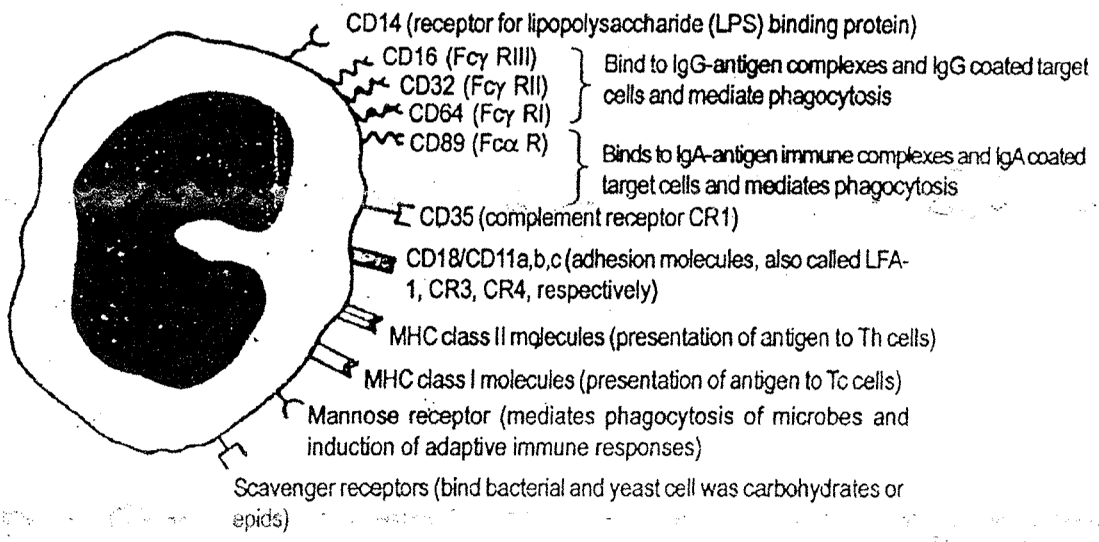


Fig 12 : Surface molecules on monocytes.

ANTIGEN PROCESSING AND PRESENTATION

Cells That Function in Antigen Presentation:

Since all cells expressing either class I or class II MHC molecules can present peptides to T cells, strictly speaking they all could be designated as antigen presenting cells. However, by convention, cells that display peptide associated with class I MHC molecules to $CD8^+T_C$ cells are referred to as *target cells*; cells that display peptides associated with class II MHC molecules to $CD4^+T_H$ cells are called **antigen-presenting cells (APCs)**. This convention is followed throughout this text.

A variety of cells can function as antigen-presenting cells. The distinguishing feature of these cells is their ability to express class II MHC molecules and to deliver a constimulatory signal. Three cell types are classified as *Professional* antigen-presenting cells: dendritic cells, macrophages, and B lymphocytes. These cells differ in their mechanisms of antigen uptake, in whether they constitutively express class II MHC molecules, and in their co-stimulatory activity:

- Dendritic cells are the most effective of the antigen-presenting cells. Because these cells constitutively express a high level of class II MHC molecules and costimulatory activity, they can activate naive T_H cells.
- Macrophages must be activated by phagocytoses of microorganisms before they express class II MHC molecules or the co-stimulatory B7 membrane molecule.
- B cells constitutively express class II MHC molecules but must be activated before they express the co-stimulatory B7 molecule.

Several other cell types, classified as *nonprofessional* antigen-presenting cells, can be induced to express class II MHC molecules or a co-stimulatory signal. Many of these cells function in antigen presentation only for short periods of time during a sustained inflammatory response.

Because nearly all nucleated cells express class I MHC molecules, virtually and nucleated cell is able to function as a target cell presenting endogenous antigens to T_C cells. Most often, target cells are cells that have been infected by a virus or some other intracellular microorganism. However, target cells can also be altered self-cells such as cancer cells, aging body cells, or allogeneic cells from a graft.

Intracellular (endogenous) and extracellular (exogenous) antigens present different challenges to the immune system. Extracellular antigens are eliminated by secreted antibody, whereas intracellular antigens are most effectively eliminated by cytotoxic T lymphocytes (CTLs). To mediate these responses, the immune system uses two different antigen-presenting pathways: (a) endogenous antigens are processed in the *cytosolic pathway* and presented on the membrane with class I MHC molecules; (b) exogenous antigens are processed in the *endocytic pathway* and presented on the membrane with class II MHC molecules (Fig. below).

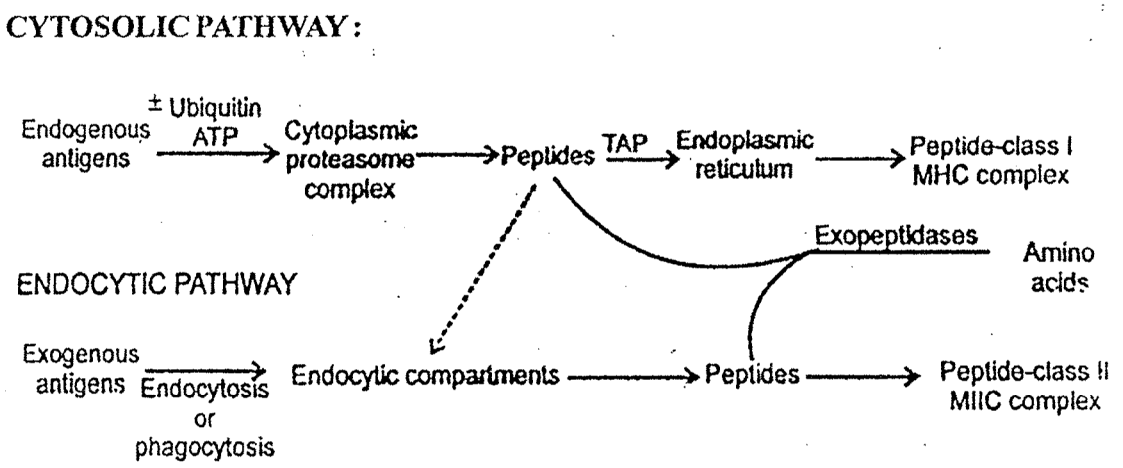


Fig 13 : Overview of cytosolic and endocytic pathways for processing antigen. The proteasome complex contains enzymes that cleave peptide bonds, converting proteins into peptides. The antigenic peptides bonds, coverting proteins into peptides. The antigenic peptides from proteasome cleavage and those from endocytic compartments associated with class I or class II MHC molecules, and the peptide-MHC complexes then are transported to the cell membrane. TAP (transporter to antigenic peptides) transport the peptides to the endoplasmic reticulum. It should be pointed out that the ultimate fate of most peptides in the cell is neither of these pathways, but rather to be degraded completely into amino acids.

Endogenous Antigens:

The Cytosolic Pathway

Endogenous antigens, such as those produced by a virus replicating within a cell, are degraded within the cytoplasm into peptides that can associated with class I MHC molecules. Soluble protein antigens that are experimentally delivered into the cytoplasm of a cell also have been shown to be degraded into peptides that can be presented to T_c cells by class I MHC molecules. (One way to deliver soluble protein antigens into the cytoplasm is to load them inside membrane liposomes; as

these fuse with the plasma membrane, the antigen is released into the cytoplasm). The pathway by which endogenous antigens are degraded for presentation with class I MHC molecules utilizes pathways involved in the normal turnover of intracellular proteins.

Peptide Generation by Proteasomes:

In eukaryotic cells, protein levels are carefully regulated. Even protein is subject to continuous turnover and is degraded at a rate that is generally expressed in terms of its half-life. Some proteins (e.g., transcription factors, cyclins, and key metabolic enzymes) have very short half-lives, denatured, misfolded, or otherwise abnormal proteins also are degraded rapidly. Intracellular proteins are degraded into short peptides by a cytosolic proteolytic system present in all cells. Those proteins targeted for proteolysis often have a small protein, called *ubiquitin*, attached to them (Fig.). Ubiquitin-protein conjugates can be degraded by a multifunctional protease complex called a proteasome. Each proteasome is a large (265), cylindrical particle consisting of four rings of protein subunits with a central channel of 10-50 A. A proteasome can cleave peptide bonds between 2 or 3 different amino acid combinations in an ATP-dependent process (Fig.). Degradation of ubiquitin-protein complexes is thought to occur within the central hollow of the proteasome, thereby preventing lysis of other proteins within th cytoplasm.

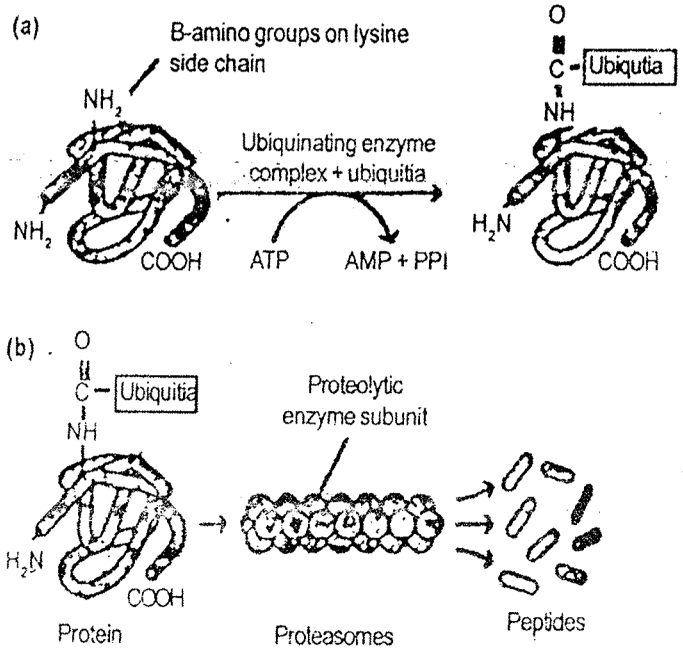


Fig 14 : Cytosolic proteolytic system for degradation of intracellular proteins. (a) Proteins to be degraded often are covalently linked to a small protein called ubiquitin, is this reaction, which requires ATP, a ubiquinating enzyme complex links several ubiquitin molecules to a lysin eamino group near the amino terminus of the protein. (b) Degradation of ubiquitin-protein complexes occurs within the central channel of proteasomes, generating a variety of peptides. Proteasomes are large cylindrical particles whose subunits catalyze cleavage of peptide bonds.

Experimental evidence indicates that the immune system utilizes this general pathway of protein degradation to produce small peptides for presentation with class I MHC molecules. The immune system appears to modify the proteasome by the addition of two subunits to the proteasome: LMP2 and LMP7. These subunits are encoded within the MHC gene cluster and are induced by increased levels of IFN-γ. An additional modifier that is added to the proteasome is also induced by IFN-γ but is not MHC encoded. This modifier was originally designated MECL-I but has also been called LMP10. The peptidase activities of proteasomes containing LMP2, LMP7 and LMP10 preferentially generate peptides that bind to MHC class I molecules. Such proteasomes, for example, show increased hydrolysis of peptide bonds that follow basic and/or hydrophobic residues. Peptides that bind to class I MHC molecules terminate almost exclusively with hydrophobic or basic residues.

Peptide Transport from the Cytosol to the Rough Endoplasmic Reticulum (RER):

Insight into the role that peptide transport, the delivery of peptides to the MHC molecule, plays in the cytosolic processing pathway came from studies of cell lines with defects in peptide presentation by class I MHC molecules. One such mutant cell line, called RMA-S, expresses about 5% of the normal levels of class I MHC molecules on its membrane. Although RMA-S cells synthesize normal levels of class I α chains and β_2-microglobulin, both molecules remain intracellular instead of appearing on the membrane. A clue to the mutation in the RMA-S cell line was the discovery by A. Townsend and his colleagues showed that "feeding" these cells peptides restored their level of membrane-associated class I MHC molecules to normal. These investigators suggested that peptides might be required to stabilize the interaction between the class I α chains and β_2-microglobulin. This led them to speculate that the RMA-S cell line might have a defect in peptide transport. The ability to restore expression of class I MHC molecules on the membrane by feeding the cells predigested peptides would be compatible with this type of defect.

Subsequent experiments with the RMA-S cell line showed that the defect in these cells is in the protein that transports peptides from the cytoplasm to the RER where class I molecules are synthesized. When RMA-S cells were transfected with a functional gene encoding the transporter

protein, the cells began to express class I molecules on the membrane. The transporter protein, designated TAP (for transporters associated with antigen processing) is a membrane-spanning heterodimer consisting of two proteins: TAP1 and TAP2 (Fig.). In addition to their multiple transmembrane segments, the TAP1 and TAP2 proteins each have one hydrophobic domain, which is thought to project through the membrane into the lumen of the RER, and one ATP-binding domain, which projects into the cytosol. Both TAP1 and TAP2 belong to the family of ATP-binding cassette proteins found in the membranes of many cells, including bacteria; these proteins mediate ATP-dependent transport of amino acids, sugars, ions, and peptides.

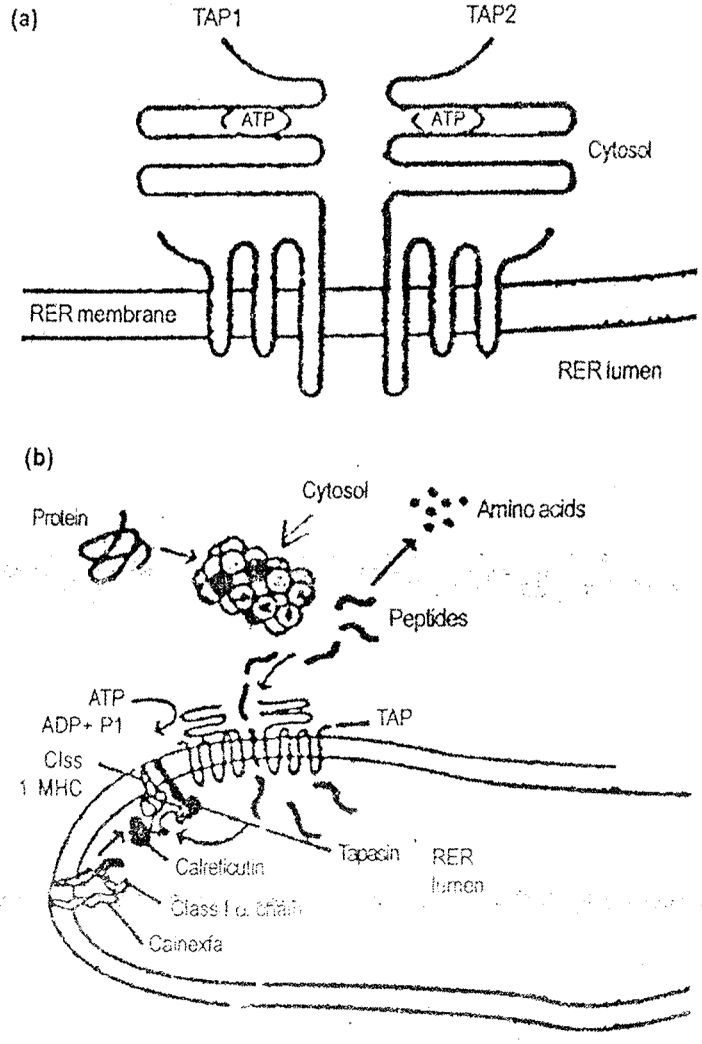


Fig 15 : Generation of antigenic peptide-class I MHC complexes in the cytosolic pathway. (a) Schematic diagram of TAP (transporters associated with antigen processing), a heterodimer anchored in the membrane of the rough endoplasmic reticulum (RER). The two chains are encoded by TAP1 and TAP2. The cytosolic domain in each TAP subunit contains an ATP-binding site, and peptide transport depends on the hydrolysis of ATP. (b) In the cytosol, association of LMP2, LMP7 and LMP10 (black spheres) with a proteasome changes its catalytic specificity to favour production of peptides that bind to class I MHC molecules. Within the RER membrane, newly synthesized class I α chain associates with calnexin until β_2-microglobulin binds to the α chain. The class I α chain-β_2-microglobulin heterodimer then binds to calreticulin and the TAP-associated protein tapasin. When a peptide delivered by TAP is bound to the class I molecule, folding of MHC class I is complete and it is released from the RER and transported through the Golgi to the surface of the cell.

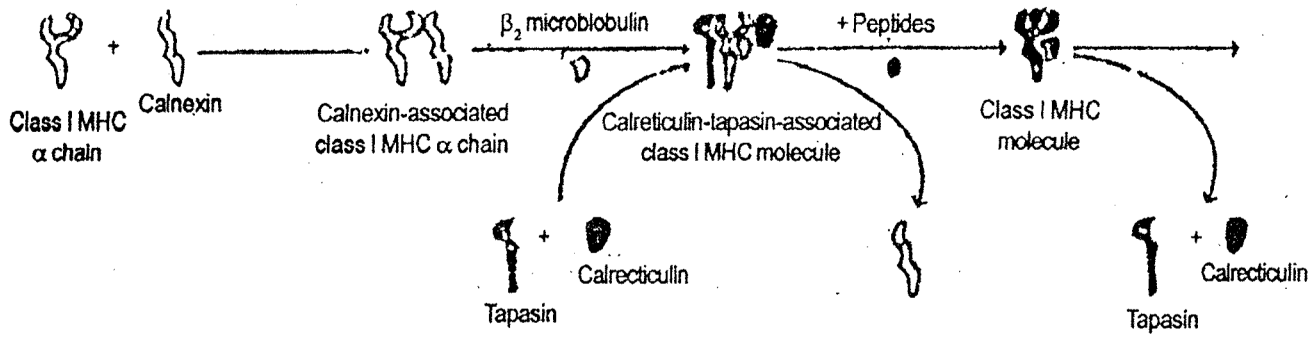


Fig 16 : Assembly and stabilization of class I MHC molecules. Newly formed class I α chain associate with calnexin, a molecular chaperones in the RER membrane. Subsequent binding to β_2-microglobulin release calnexin and allows binding to the chaperonin calreticulin and to tapasin, which is associated with the peptide transporter TAP. This association promotes binding of an antigenic peptide, which stabilizes the class I molecule-peptide complex, allowing its release from the RER.

Peptides generated in the cytosol by the proteasome are translocated by TAP into the RER by a process that requires the hydrolysis ATP (Fig.). TAP has the highest affinity for peptides containing 8-13 amino acids, which is the optimal peptide length for class I MHC binding. In addition, TAP appears to favor peptides with hydrophobic or basic carboxyl-terminal amino acids, the preferred anchor residues for class I MHC molecules. Thus, TAP is uniquely designed to transport peptides that will interact with class I MHC molecules.

The TAP1 and TAP2 genes map within the class II MHC region adjacent to the LMP2 and LMP7 genes (see Fig.). Both the transporter genes and these LMP genes are polymorphic, that is, different allelic forms of these genes exist within the population. Allelic differences in LMP-mediated proteolytic cleavage of protein antigens or in the transport of different peptides from the cytosol into the RER may contribute to the observed variation among individuals in their response to different endogenous antigens.

Assembly of Peptides with Class I MHC Molecules:

Like other proteins, the α chain and β_2-microglobulin components of the class I MHC molecule are synthesized on polysomes along the rough endoplasmic reticulum (RER). Assembly of these

components into a stable class I MHC molecular complex that can exit the RER requires the presence of a peptide in the binding groove of the class I molecule. The assembly process leading to a stable class I MHC molecule takes place in several steps and involves the participation of *molecular chaperones*, which facilitate the folding of polypeptides. The molecular chaperone first involved in class I MHC assembly is calnexin, a resident membrane protein of the endoplasmic reticulum. Calnexin associates with the free class I α chain and promotes its folding. When β_2-microglobulin binds to the α chain, calnexin is released and the class I molecule associates with the chaperonin *calreticulin* and with *tapasin*. Tapasin (TAP-associated protein) brings the TAP transporter into proximity with the class I molecule and allows it to acquire an antigenic peptide (Fig.). The physical association of the α chain-β_2-microglobulin heterodimer with the TAP protein (see Fig.) promotes peptide capture by the class I molecule before the peptides are exposed to the luminal environment of the RER. Peptides not bound by class I molecules are rapidly degraded. As a consequence of peptide binding, the class I molecule displays increased stability and can dissociate from calreticulin and tapasin, exit from the RER, and proceed to the cell surface via the Golgi.

Overview of the Class I Endogenous Pathway:

The current model of the pathway by which endogenous antigens are processed and presented is shown on the left side of Figure . As mentioned earlier, endogenous antigens are thought to be degraded within the cytoplasm by the proteasome complex. Peptides are then transported across the membrane of the RER by an ATP-binding transporter called TAP. The calnexin-associated class I MHC α chain binds to β_2-microglobulin, dissociates from calnexin, and binds to calreticulin and to tapasin, which is associated with the transmembrane TAP protein (ERp57 may be involved also at this step). This class I MHC complex then captures an antigenic peptide, which allows dissociation of the MHC peptide complex from the chaperones. Finally, the class I MHC molecule-peptide is transported from the RER through the Golgi complex to the plasma membrane.

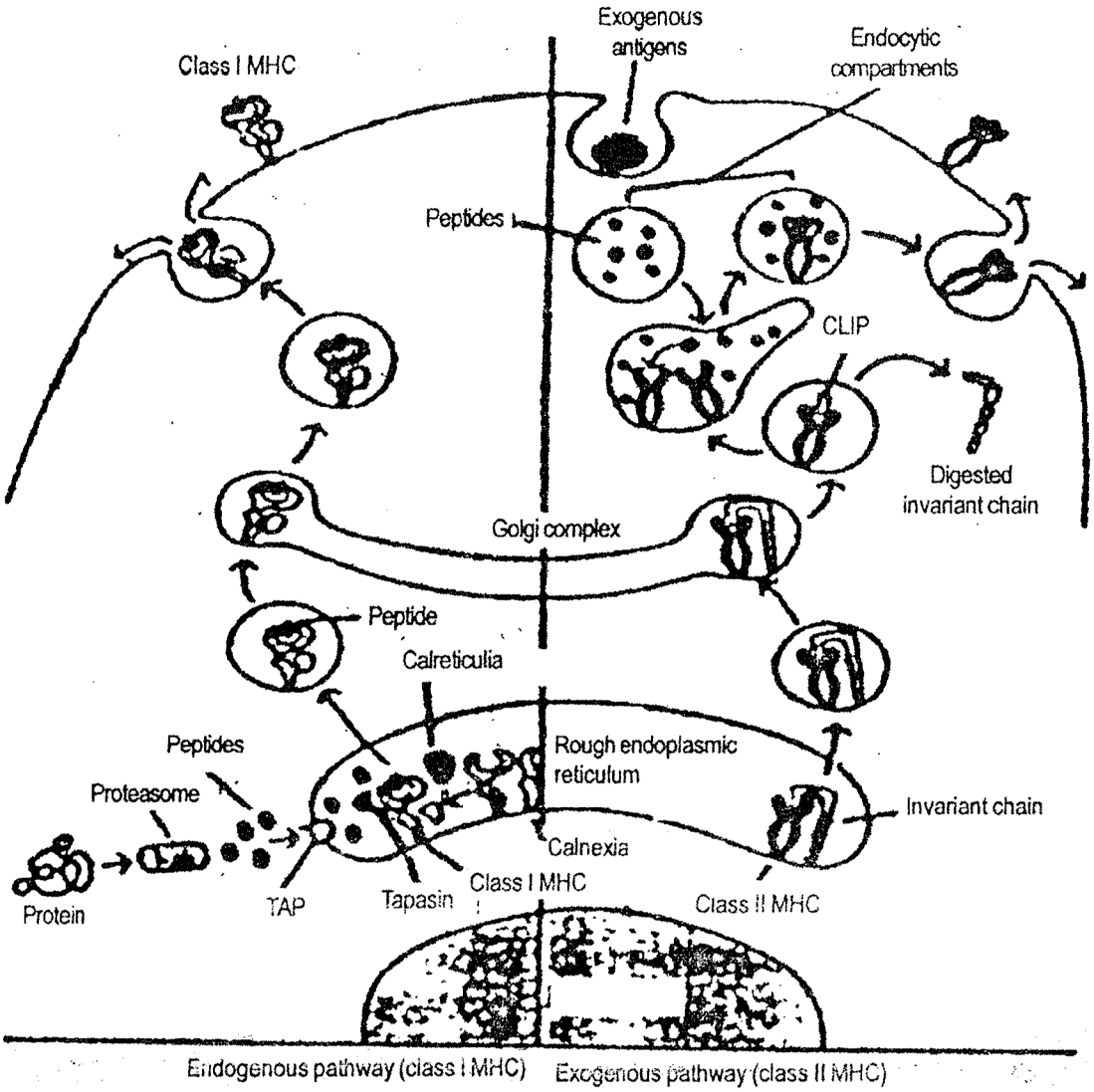


Fig. 17 : Model of separate antigen-presenting pathways for endogenous (green) and exogenous (red) antigens. The mode of antigen entry into cells and the site of antigen processing appear to determine whether antigenic peptides associate with class I MHC molecules in the rough endoplasmic reticulum or with class II molecules in endocytic compartments.

Exogenous Antigens

The Endocytic Pathway:

Antigen-presenting cells can internalize antigen by phagocytosis, endocytosis, or both Macrophages internalize antigen by both processes, whereas most other APCs are not phagocytic or

are poorly phagocytic and therefore internalize exogenous antigen only by endocytosis or pinocytosis). B cells, for example, internalize antigen very effectively by receptor-mediated endocytosis using antigen specific membrane antibody as the receptor.

Peptide Generation in Endocytic Vesicles:

Once an antigen is internalized, it is degraded into peptides within compartments of the endocytic processing pathway. As the experiment shown in Figure demonstrated, internalized antigen takes 1-3 h to transverse the endocytic pathway and appear at the cell surface in the form of peptide-class II MHC complexes. The endocytic pathway appears to involve three increasingly acidic compartments: early endosomes (pH 6.0 - 6.5); late endosomes or endolysosomes (pH 5.0 - 6.0); and lysosomes (pH 4.5-5.0). Internalized antigen moves from early to late endosomes and finally to lysosomes, encountering hydrolytic enzymes and a lower pH in each compartment (Fig. 18). Lysosomes, for example, contain a unique collection of more than 40 acid-dependent hydrolases, including proteases, nucleases, glycosidases, lipases, phospholipases, and phosphatases. Within the compartments of the endocytic pathway, antigen is degraded into oligopeptides of about 13-18 residues, which bind to class II MHC molecules. Because the hydrolytic enzymes are optimally active under acidic conditions (low pH), antigen processing can be inhibited by chemical agents that increse the pH of the compartments (e.g., chloroquine) as well as by protease inhibitors (e.g., leupeptin).

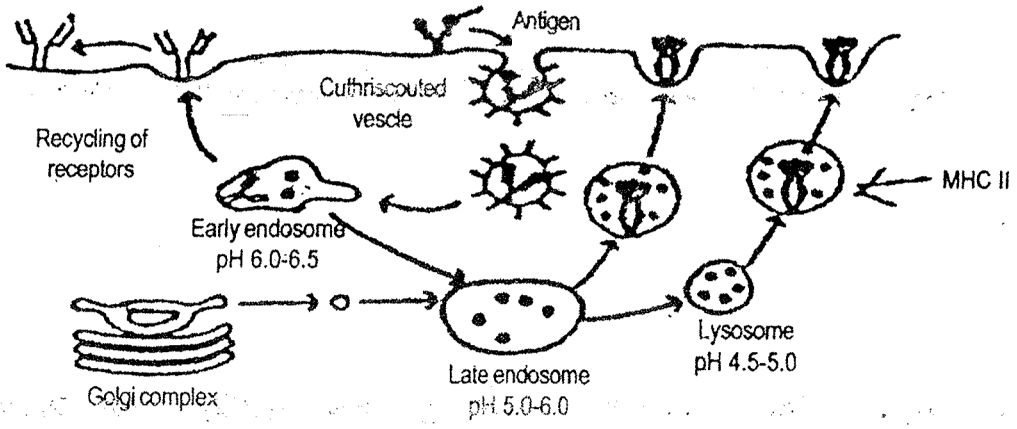


Fig 18 : Generation of antigenic peptides in the endocytic processing pathway. Internalized exogenous antigen moves through several acidic compartments, in which it is degraded into peptides that ultimately associate with class II MHC molecules transported in vesicles from the Golgi complex. The cell shown here is a B cell, which internalizes antigen by receptor-mediated endocytosis, with the membrane-bound antibody functioning as an antigen-specific receptor.

The mechanism by which internalized antigen moves from one endocytic compartment to the next has not been conclusively demonstrated. Some have suggested that early endosomes from the periphery move inward to become late endosomes and finally lysosomes. Others have suggested that small transport vesicles carry antigens from one compartment to the next, in a manner similar to the way in which small transport vesicles carry proteins from one compartment of the Golgi complex to the next. Eventually the endocytic compartments, or portions of them, return to the cell periphery, where they fuse with the plasma membrane. In this way, the surface receptors may be recycled.

Transport of Class II MHC Molecules to Endocytic Vesicles:

Since antigen-presenting cells express both class I and class II MHC molecules, some mechanism must exist to prevent class II MHC molecules from binding to the same set of antigenic peptides as the class I molecules. When a class II MHC molecule is synthesized within the RER, it associates with another protein called the **invariant** (II) chain. Three pairs of the class II αβ chains associated with a preassembled in variant chain trimer. This trimeric protein interacts with the peptide-binding cleft of the class II molecule, preventing any endogenously derived peptides from binding to the cleft while the class II molecule is within the RER (see right side of previous Figure-17). In addition to its role in preventing peptide binding to class II MHC molecules, the invariant chain also appears to be involved in the folding of the class II α and β chains, their exit from the RER, and the subsequent routing of class II molecules to the endocytic processing pathway from the trans-Golgi network.

The role of the invariant chain in the routing of class II molecules has been demonstrated in transfection experiments with cells that lack the genes encoding class II MHC molecules and the invariant chain. Immunofluorescent labeling of such cells transfected only with class II MHC genes revealed class II molecules localized within the Golgi complex. However, in cells transfected with both the class II MHC genes and invariant-chain gene, the class II molecules were localized in the cytoplasmic vesicular structures of the endocytic pathway. The invariant chain contains sorting signals in its cytoplasmic tail that directs the transport of the class II MHC complex from the trans-Golgi network to the endocytic compartments.

Assembly of Peptides with Class II MHC Molecules:

Recent experiments indicate that most class II MHC invariant chain complexes are formed, through the Golgi complex and trans-Golgi network to early endosomes. Here, within the endocytic pathway, these complexes move from early endosomes to late endosomes, and finally to lysosomes. As the proteolytic activity increases in each successive compartment, the invariant chain is gradually degraded. However, a short fragment of the invariant chain termed CLIP (*for class II-associated invariant chain peptide*) remains bound to the class II molecule after the invariant chain has been cleaved within the endosomal compartment. CLIP physically occupies the peptide-binding groove of the class II MHC molecule, presumably preventing any premature binding of antigenic peptide (see Fig.).

A nonclassical class II MHC molecule, called HLA-DM plays a role in the removal of CLIP and in the subsequent loading of class II molecules with antigenic peptides. HLA-DM is required to catalyze the exchange of CLIP with antigenic peptides (Figure). The reaction between HLA-DM and the class II CLIP complex facilitating exchange of CLIP for another peptide is inhibited in the presence of HLA-DO, which binds to HLA-DM and blocks the exchange reaction.

MHC class II genes encoding HLA-DM have been identified in the mouse and rabbit, indicating that HLA-DM is widely conserved among mammalian species. Like other class II MHC molecules, HLA-DM is a heterodimer of α and β chains. However, unlike other class II molecules, HLA-DM is not polymorphic and is not expressed at the cell membrane but is found predominantly within the endosomal compartment. The DMα and DMβ genes are located near the TAP and LMP genes in the MHC complex of humans. HLA-DO, like HLA-DM, is a nonclassical and nonpolymorphic class II molecule that is also found in the MHC of other species. DO differs from DM in that it is expressed only by B cells and the thymus.

Earlier descriptions of the class II processing pathway suggested that CLIP did not bind to HLA class II molecules in a stable fashion, and that this instability fostered its replacement by another peptide. Recent structural evidence makes this mechanism unlikely. An HLA-DR3 molecule associated with CLIP was isolated from a cell line that did not express HLA-DM and was therefore defective in antigen processing. This complex of HLA-DR3-CLIP was crystalized and its three-dimensional structure determined. Comparison of the class II CLIP complex with other DR molecules bound to

antigenic peptides shows that CLIP binds to class II in the same manner as do antigenic peptides that form a stable complex.

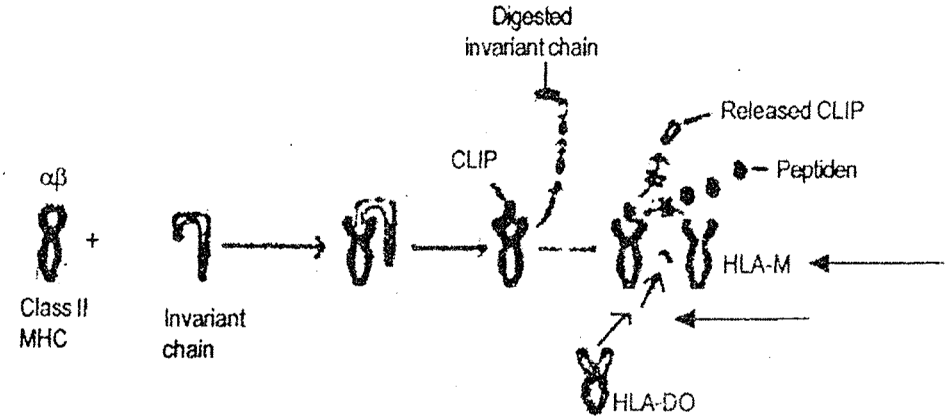


Fig 19 : Assembly of class II MHC molecules. Within the rough endoplasmic reticulum, a newly synthesized class II MHC molecule binds an invariant chain. The bound invariant chain prevents premature binding of peptides to the class II molecule and helps to direct the complex to endocytic compartment containing peptides derived from exogenous antigens. Digestion of the invariant chain leaves CLIP, a small fragment remaining in the binding groove of the class II MHC molecule. HLA-DM, a nonclassical MHC class II molecule expressed within endosomal compartments, mediates exchange of antigenic peptides for CLIP. The nonclassical class II molecule HLA-DO may act as a negative regulator of class II antigen processing by binding to HLA-DM and inhibiting its role in the dissociation of CLIP from class II molecules.

As with class I MHC molecules, peptide binding is required to maintain the structure and stability of class II MHC molecules. Once a peptide has bound, the peptide-class II complex is transported to the plasma membrane, where the neutral pH appears to enable the complex to assume a compact, stable form. Peptide is bound so strongly in this compact form that it is difficult to replace a class II-bound peptide on the membrane with another peptide at physiologic conditions.

Overview of the Class II Exogenous pathway:

The current model of the pathway for processing and presenting exogenous antigens is summarized in the right side of Figure. As described above, newly synthesized class II α and β chains associate with an invariant chain within the RER. This association precludes binding of class II MHC molecules with peptides derived from endogenous antigens or other intracellular proteins. Class II molecules, together with the associated invariant chains, are thought to be routed from the RER through the Golgi complex to the endocytic pathway. Once inside an endocytic compartment, the invariant chain is degraded by proteolytic enzymes, leaving the CLIP fragment bound to the peptide-

binding cleft of the class II molecules. Within the acidic endocytic compartment, CLIP associated with the class II molecule is replaced with an antigenic peptide, a process catalyzed by HLA-DM. The resulting class II MHC-peptide complexes move to the plasma membrane.

Presentation of Nonpeptide Bacterial Antigens:

Although it is well known that nonprotein antigens are recognized by the immune system, we have so far limited the discussion to peptide antigens and their presentation by classical class I and II MHC molecules. At present little is known about the mechanisms used for the processing and presentation of antigens that are not peptides. Some recent progress has been made in this area. It has been shown that antigenic lipid and glycolipid molecules derived from pathogenic bacteria such as *Mycobacterium tuberculosis* are presented by members of the CD1 family of nonclassical class I molecules.

The CD1 family of molecules associates with β_2-microglobulin and has general structural similarity to class I MHC molecules CDI gene products are found on antigen-presenting cells. There are five genes encoding human CDI molecules (CDIa-e), which are not located within the MHC but on a different chromosome. These genes have been identified in all mammalian species studied. Sequence identity of CDI with classical class I is considerably lower than that among the class I molecules. For example, the $\alpha 3$ domain of CDI molecules is closer in sequence to β_2-microglobulin than to the $\alpha 3$ of MHC class I. CDI molecules are recognized by T cells in the absence of foreign antigens, and self restriction can be demonstrated in these reactions.

Examination of antigens presented by CDI molecules revealed them to be lipid components (mycolic acid) of the *M. tuberculosis* cell wall. Further studies of CDI presentation indicated that a glycolipid (lipoarabinomannan) from *Mycobacterium leprae* could also be presented by these molecules. The pathway by which the CDI molecules process antigen differs from that of the classical class I pathway in that TAP-deficient cells are able to present CDI-associated antigens. The data concerning CDI antigen presentation point out the existence of a third pathway for the processing of antigens, a pathway with distinct intracellular steps that donot involve the molecules found to facilitate class I antigen processing. Exactly how the CD pathway complements better understood class I and class II pathways and if and where it intersects with them remain in question.

CELL MEDIATED IMMUNITY

T-lymphocytes :

T-cell precursors migrate from the bone marrow to the thymus, where some mature into antigen specific T-cells. These cells then migrate to the secondary lymphod organs and tissues of the body. Some T-cells reside, at least temporarily, in T-cell dependent areas of tissues. T-cells can be identified using monoclonal antibodies specific for characteristic molecules such as T-cell receptor (TcR) or CD_3 (Fig .-20). These cells function to control intracellular microbes and to provide help for B-cell (antibody) responses. Two different kinds of T-cells are involved in these functions.

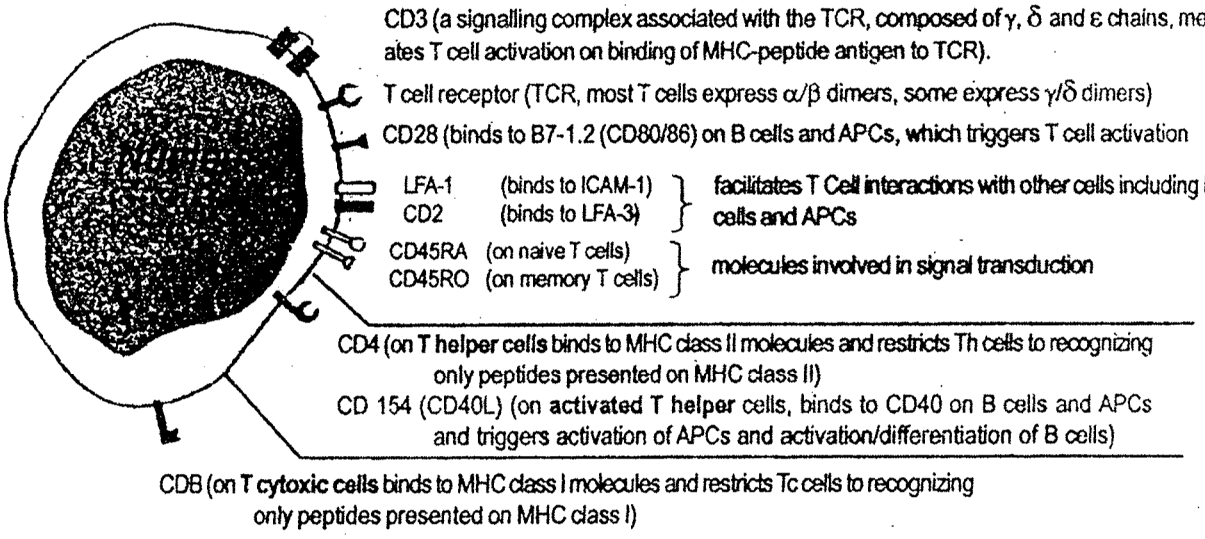


Fig 20 : T cell surface molecules.

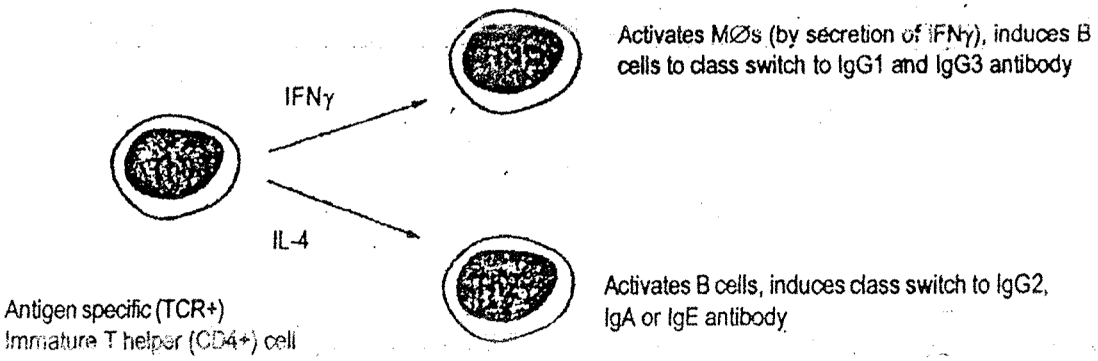


Fig 21 : Two types of T helper cells. In the presence of Th1 inducing cytokines (e.g. IFNγ TNFα) Tho cells differentiate to become Th1 cells whereas in the presence of Th2 inducing cytokines (e.g. IL-4, IL-13, IL-5, IL-101 Th0 cells differentiate into Th2 cells)

T-helper (Th) cells provide help for B-cells through direct cell surface signalling and by producing cytokines which are critical to B-cell growth and differentiation. Th cells, in addition to TcR and CD_3, also express cell surface CD_4 molecules which can bind to MHC class II molecules (an interaction required for these cells to be activated by antigen). Th cells are further subdivided into Th1 and Th2 cells depending on their ability to help in the development of different immune responses which in turn related to this cytokines profiles (Fig.-21). The percentages of these cells in the peripheral blood are - T-helper cells (55%), T-cytotokic (25%), B-cells (10%) and NK-cells (10%).

T-cytotoxic (Tc) cells mediate killing of infected cells, primarily those infected with virus. These cells express, in addition to TcR and CD_3, a surface molecule, CD_8, that binds to MHC class I and is important for these cells to interact effectively with virally infected cells.

Mechanism of CMI :

CMI function to enhance the microbicidal actions of phagocytes and to thus eliminate microbes. The T-cells that are responsible for macrophage activation are differentiated CD_4^+ helper T-cells of

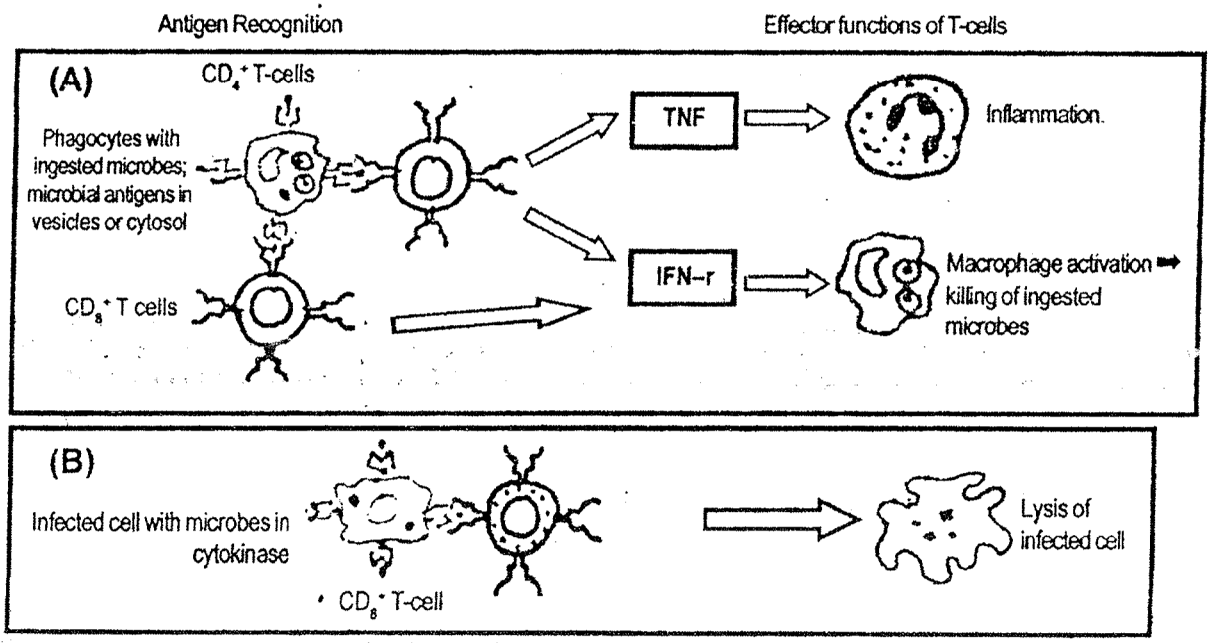


Fig 22 : Types of T-cell mediated immune reactions.

(A) CD_4^+ TH1 cells and CD_4^+ T cells recognise class II MHC-associated or class I mHC-associated peptide antigens of phagocytosed microbes respectively and produce cytokines that stimulate inflammation and activate the phagocytes to kill the microbes.

(B) CD_8^+ cytolytic T lymphocytes (CTLs) recognize class-MHC-associated peptide antigens of microbes replicating in the cytoplasm of infected cells and kill the cells.

the TH_1 subset, as well as CD_8^+ T-lymphocytes, their shared characteristic being the ability to secrete the macrophage-activating cytokine interferon-γ (IFN-γ) (Fig. 22) T-cell also produce tumor necrosis factor (TNF) and lymphotoxin, which promote leukocyte recruitment and inflammation.

The process of T-cell mediated macrophage activation may be divided into a number of steps.

(a) induction, in which naive T-lymphocytes recognize antigens in peripheral lymph nodes and are primed or stimulated, to proliferate and differentiate into effector cells.

(b) the migration of effector T-cell and other leukocytes to the site of antigen challenger and

(c) macrophage activation, the key event of CMI, which is responsible for elimination of the microbe and repair of infected tissue.

Some of the antigen-activated CD_4^+lymphocytes differentiate into TH_1 cells, which are the specific effector cells of CMI, and this process is stimulated by IL-12 produced by APCs (Fig. 23). IL-12 is produced by macrophages and dendritic cells in response to three types of stimuli- microbial products, CD40 engagement and the cytokine IFNγ. Bacterial lipopolysaccharide (LPS) is a powerful stimulus for IL-12 production by macrophages during innate immunity. It is likely that other components of microbes, and perhaps even the process of phagocytosis, stimulate IL-12 production. Upon antigen stimulation, T-lymphocytes rapidly express the protein called CD40 ligand (CD40L), which binds to CD40 on macrophages and dendritic cells and stimulate IL-12 secretion. IFN-γ produced by NK cells during innate immune reactions and by T cells during adaptive immune responses also stimulates IL-12 production by macrphages. IL-12 acts on antigen stimulated CD_4^+ T cells and induce the differentiation of there lymphocytes to the TH_1 subset.

The effector function of activated macrophages include the following :

(a) Activated macrophages kill phagocytosed and extracellular microbes, mainly by producing microbicidal reactive oxygen species (ROS), nitric oxide (NO) and lysosomal enzymes.

(b) Activated macrophages stimulate acute inflammation through the secretion of cytokines (TNF, IL-1, IL-12 etc.). The collective action of these macrophage-derived cytokines and lipid mediators is to produce a local inflammation that is rich in neutrophil, which phagocytose and destroy infectious organisms.

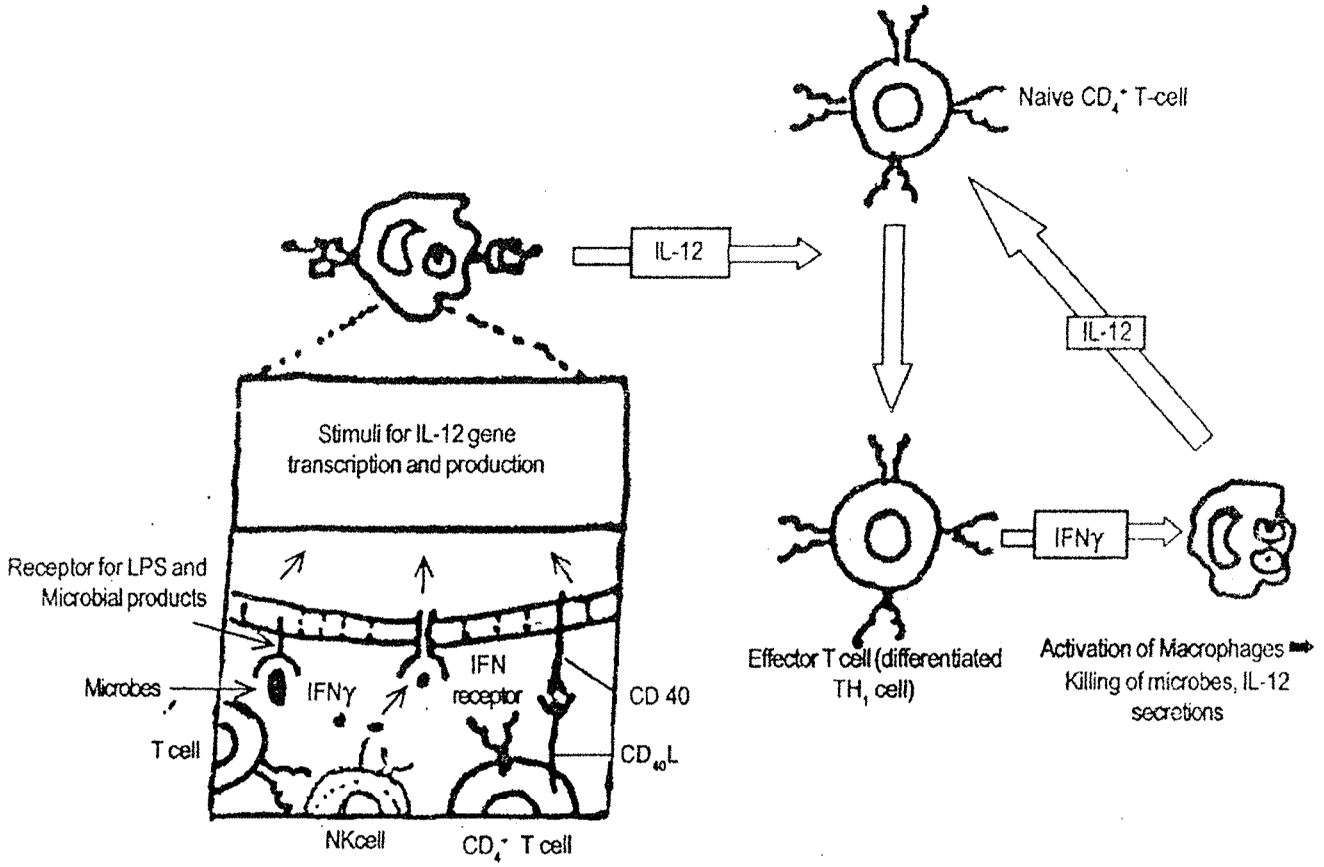


Fig 23: T-cell mediated Macrophage activation.

(c) The activated macrophage acts as an endogenous surgeon to canterize the wound and thereby eliminate antigen and resolve the inflammatory reaction.

(d) In addition to their effector functions, activated macrophages become more efficient APCs because of high levels of molecules involved in antigen - processing such as components of the proteosome and cathepsins, increased surface expression of class II MHC molecules and costimulators (B_7) and production of cytokines IL-12 that stimulate T-lymphocyte proliferation and differentiation. These macrophage responses function to enhance T-cell activation, thus serving as amplification mechanisms for CMI.

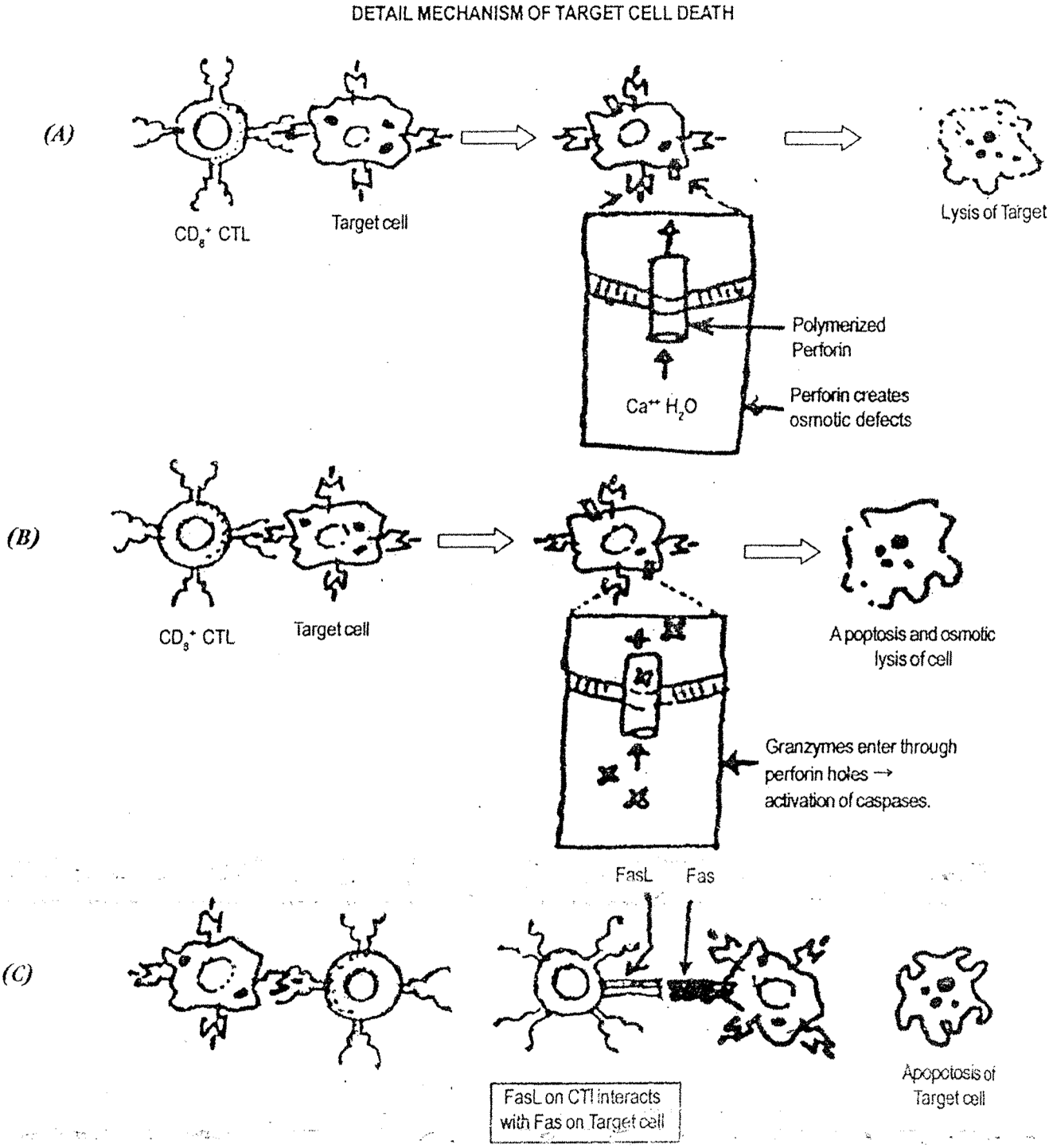


Fig. 24 : Mechanism of cytolytic T-lymphocyte (CTL) mediated Lysis of Target cell.

A. *Perforin is exocytosed in CTL granules and opolymerizes in the target cell plasma membrane to form pores that allow the entry of water and ions and result in target cell death.*
B. *Granzymes are exocytosed in CTL granules enters target cells through perforin pores and induce target cell apoptosis.*
C. *Fasligand (FasL) is expressed on activated CTLs, engages Fas on the surface of target cells, and induce apoptosis.*

Cytolytic T-lymphocytes :

The principal mechanism of CTL-mediated cytolysis is the delivery of cytotoxic granule proteins to the target cell recognized by the CTL. The two granule proteins that are most important for cytolysis are perforin and granzymes. Perforin is a pore-forming protein that is present as a monomer in the granules of CTLS (and NK cells). The target cell dies partly because the influx of water induces osmotic welling and partly because the high concentration of calcium ions that enters the cell may trigger apoptosis.

Granzymes (granule enzymes, serine proteases) are also exocytosed from CTLs and they enter target cells mainly through perforin-created "holes". Once inside the target cells, granzyme B proteolytically cleaves and thereby activates cellular emymes called caspases, which in turn cleave several substrates and induce target cell apoptosis. CTL use a second mechanism of cytolysis that is mediated by interactions of membrane molecules on the CTLs and target cells. Upon activation, CTLs express a membrane protein, called Fas ligand (FasL), that binds to its target protein Fas, which is expressed on many cell types. This interaction also results in activation of caspases and apoptosis of target cells. By activating an apoptotic death pathway in target cells, CTLs can initiate the destruction of microbial DNA, as well as the target cell genome, thereby eliminating potentially infectious DNA (Fig.-24).

T CELL ACTIVATION

Accessory molecules :

Pathogens or antigens infecting peripheral sites are typically trapped in the lymph nodes directly downstream of the site of infection. Blood borne pathogens are trapped in the spleen. These secondary lymphoid organs contain APC (dendritic cells and macrophages) that efficiently trap antigen for processing and presentation. Naive T-cells recirculate through these sites looking for appropriate processed antigen.

Initial recognition of processed antigen by T-cells is via the T-cell antigen receptor. To strengthen the cellular associations between the antigen-specifec T-cells and the APC and accessory

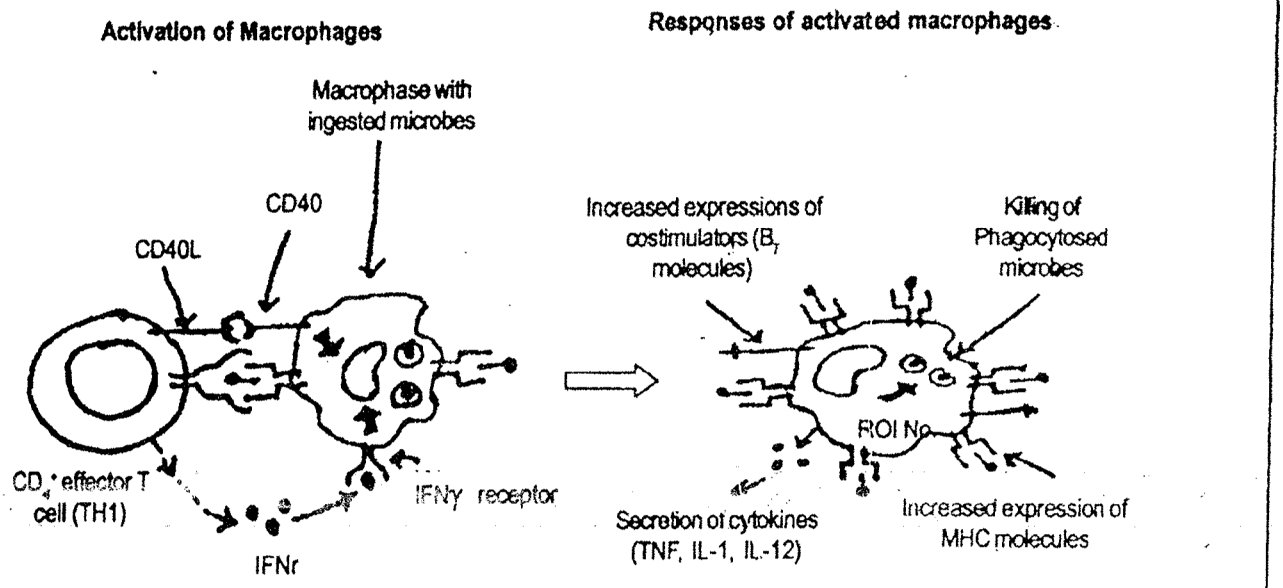


Fig 25 : Activation and effector functions of macrophages in CMI. Macrophages are activated by CD40-CD40L interactions and by IFNr and perform several functions that kill microbes, stimulate inflammation and enhance the antigen-presenting capacity of the cells.

molecules provide additional linkages between the APC and the T-cell (Fig.). CD4 binds to the constant region domain of class II MHC molecules thereby strengthening the association of the TCR with peptide class II MHC molecules. Likewise. CD8 binds to class I MHC molecules to strengthen the

association of the TCR with class I MHC molecules to strengthen the association of the TCR with class I MHC molecules. In addition to engagement of these ligand-receptor pairs (Fig.), additional adhesion molecules, integrins, become engaged. These include the intercellular adhesion molecules (ICAMs) and lymphocyte function-associated antigens (LFA). Some of these accessory molecules are also important in regulating early activation events through signaling, e.g. CD4 and CD8, and are sometimes termed co-receptors, as they are part of the T-cell receptor complex.

Co-stimulatory molecules : two signals are required for Th cell activation :

Ligation of the TCR on its own does not stimulate T-cell clonal expansion or lymphokine production. The full activation of antigen-specific T-cells requires two signals. Signal one is provided by the engagement of the T-cell antigen receptor and signal two is provided by engagement of a co-stimulatory molecule. The best characterized co-stimulatory molecule is B7, which is on many APC and binds to CD28 on the T-cell. Signals emanating from this TCR and CD28 synergize to induce T-cell lymphokine production and T-cell proliferation. If the T-cell receives signal 1 (TCR binding) and not signal 2 (co-stimulation) the T-cell is turned off (Fig.).

Superantigens and Th cell activation :

Some protein products of bacteria and viruses produce proteins known as super-antigens that bind to lateral surfaces of the MHC class II molecules (not in the peptide binding groove) and the V region of the β subunit of the TCR. These antigens are not processed into peptides as conventional antigens, but are able to bind to specific family of TCR. In a sense they 'glu' T-cells to APC (Fig.) and

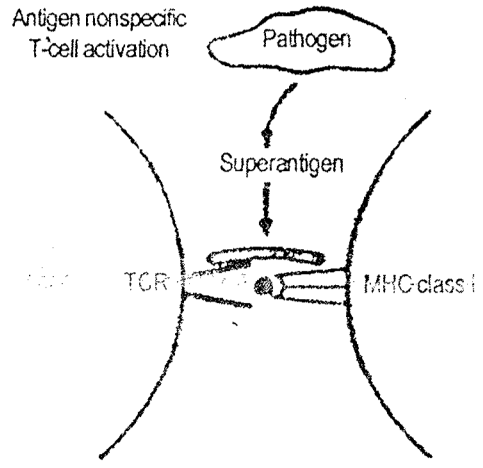


Fig 26 : Superantigen activation of T cells by bridging TCR and MHC class II

cause stimulation of the T-cell. However, these T-cells are not specific for the pathogen, since all members of a particular family of TCR are activated. The consequence of binding to a large percentage of the T-cells is the massive production of cytokines leading, in some cases, to lymphokine-induced vascular leakage and shock. Among the bacterial superantigens are the staphylococcal enterotoxins (SE) that cause common food poisoning and the toxic shocks syndrome toxin (TSST).

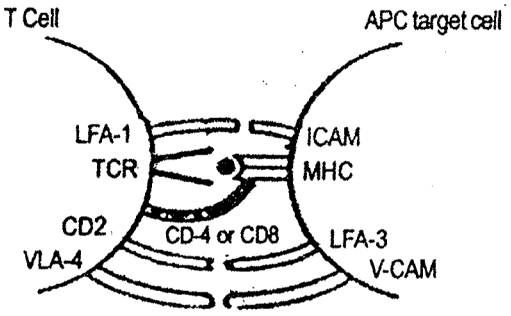


Fig. 27 : Pairs of molecules which strengthen the association of T-cells with antigen presenting and target cells.

Activation of cytotoxic T cells :

$CD8^+$ cytotoxic T-cells also need to be activated to produce functional CTLs containing granules and perforin. This requires attachment of their TcR to MHC class I-peptide complexes on APCs (Signal 1). In addition, a second signal is provided, the nature of which is unclear but can only be supplied by APCs that have been conditioned by Th cells (Fig.). As with Th activation, cytokines

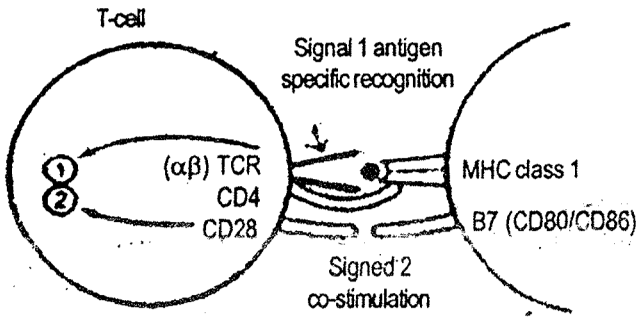


Fig 28 : The role of co-stimulation in T cell activation.

produced by Th cells and APCs are important for Tc activation. Of note, although other Th conditioned APCs may be able to provide the necessary signals for Tc activation, dendritic cells are the only cells which have a significant cross-over of processed antigen into exogenous and endogenous pathways. Moreover, they are, in general, the most efficient of the antigen presenting cells.

Biochemical Events leading to T lymphocyte activation :

In order for a T-cells to fulfill its function it has to convey a signal from the cell surface via the cytosol into the nucleus to give rise to specific gene transcription. This signal transduction is brought about by several molecules at the cell surface (Fig.). The key players in this amplification process are enzymes which phosphorytate (kinases) and dephosphorylate (phosphatases) particular

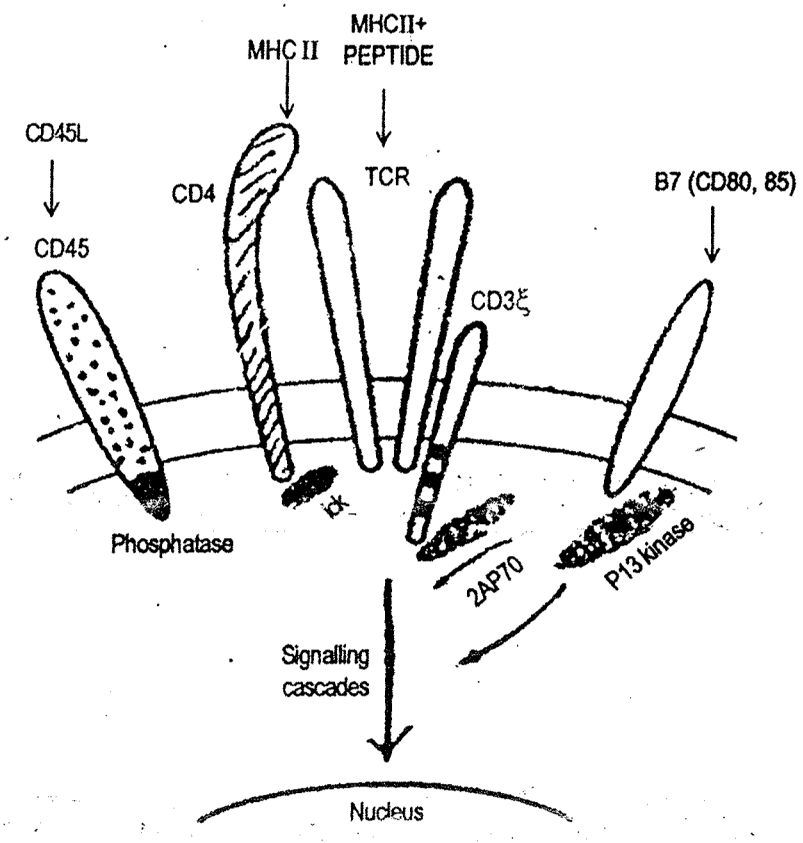


Fig 29 : CD 45 and B7 (CD80/86) are important cell surface molecules involved in T cell activation. Ligation of surface receptors leads to activation of CD45 (a phosphatase) and several kinases (e.g. lck, ZAP 70, P13 kinase). Activity of these enzymes results in 'signaling cascades' leading ultimately to transcription of specific genes. Co-stimulation via CD28 (is necessary for cell activation and proliferation.

amino acids thus activating them sequentially leading eventually to activation of specific transcription factors in the nucleus and production of functional proteins. Initiation of the process in through activation of a CD4 associated kinase (lymphocyte kinase-lck) by CD45 (a phosphatase) one of its functions being to phosphorylate tyrosines of the immunostimulatory tyrosine-based activation motifs (ITAMS) on the zeta chain of CD3 (Fig.). This, in turn, activates another protein tyrosine kinase ZAP 70 which activates an enzyme responsible for activating the phosphatidyl inositol pathway and further downstream events leading to the activation of transcription factors regulating production of IL-2 and other important proteins.

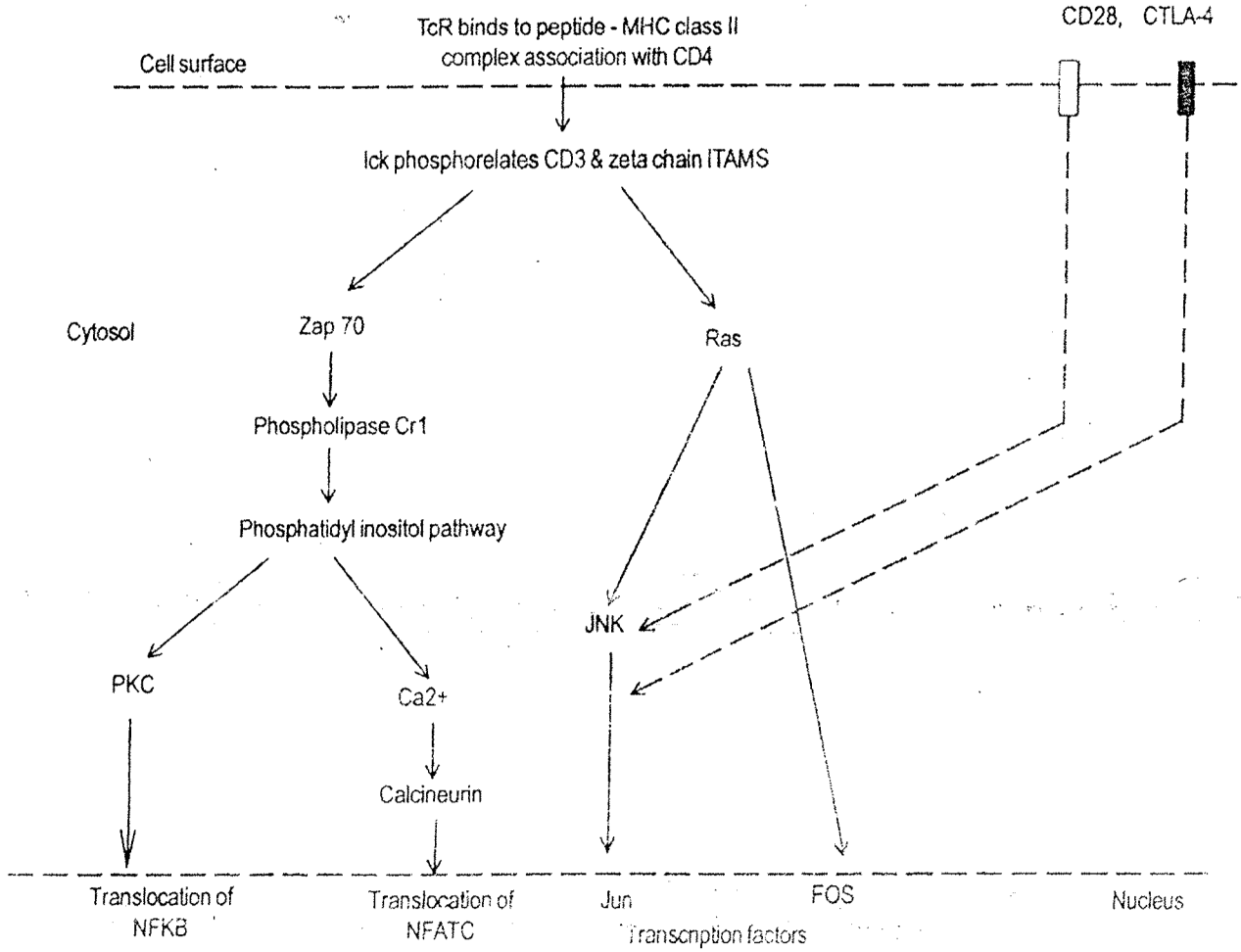


Fig 30: Signaling pathways leading to Th cell activation. Ligation of CD28 on the T cell (Signal 2) modulates the JNK biochemical pathway in a positive way and ligation of CTLA-4 is thought to negatively influence the activation pathway thus resulting in damping down of the Th cell response. The important outcome of these processes is the production of receptors e.g. Il2 R and cytokines by the cell, PKC, protein kinase C, JNK, June Kinase).

SUMMARY

Lymphocytes provide both the specificity and memory which are characteristic of the adaptive immune response. They are matured in the primary lymphoid organs (thymus, bone marrow) and function in the secondary lymphoid organs (spleen, lymph nodes etc.) where they recognize and respond to foreign antigens. There are three types of lymphocytes – NK cells, T-cells and B-cells, although only T and B cells have true antigen specificity and memory.

B-cell receptor Complex (BCR) consists of the antigen receptor, Ig, in association with two other polypeptides, Igα and Igβ. Igα and Igβ are signalling molecules for the BCR and are also required for assembly and expression of Ig. B-cell co-receptors, including CD_{21}, CD_{32}, CD_{19} and CD_{81}, associate with BCR complex especially when both the BCR and one or more of the co-receptors are linked through an antigen-complement/antibody complex.

B-cell activation is initiated by the clustering of antigen receptors (membrane IgM and IgD or naive B-cells) by the binding of multivalent antigen. Membrane Ig-associated signalling molecules Igα and Igβ transduce signals on antigen binding to the Ig, and these signals lead to activation of transcription factors and expression of various genes.

Helper T cell – dependent B-cell responses to protein antigen are initiated by antigen presentation by B-cells and activation of helper T cells at the edge of lymphoid follicles in peripheral lymphoid organs. Activated helper T-cells, express CD40L, which engages CD40 on the B-cells and the T-cells secrete cytokines that bind to cytokine receptors on B-cells. The combination of CD40 and cytokine signals stimulates B-cell proliferation and differentiation into antibody secreting cells.

In primary immune response, naive B-cells are stimulated by antigen, became activated, and differentiate into antibody secreting cells that are specific to the antigen. Long-lived memory B-cells are generated during the primary response. A secondary immune response is elicited when the

same antigen stimulates these memory B-cells, leading to more rapid proliferation and differentiation and production of greater quantities of specific antibody than are produced in the primary response.

The complement system consists of serum and membrane proteins that interact in a highly regulated manner to produce biologically active protein products. There are three pathways of complement activation. The two major pathways of complement activation are the alternative pathway, which is activated on microbial surfaces in the absence of antibody, and the classical pathway, which is activated by antigen – antibody complexes. The third pathway is known as lectin pathway. The two pathways (alternate & classical) generate enzymes that cleave the C_3 protein, and products of C_3 become covalently attached to microbial surface on antibodies, so subsequent steps of complement activation are limited to these sites. All pathways converge after C_3 proteolysis and association of several complement proteins, culminating in MAC formation. It is an important mediators of the humorul immune response and responsible for the lysis of the invading organisms.

The most abundant mobile phagocyte is the **neutrophil** (polymorphonuclear cells). This granular leukocytes comprises the majority of white blood cells. It patrols the blood stream in search of invading microbes. Another phagocyte monocytes are present in the blood stream and settle in the tissues as macrophages. Macrophage like cells in the liver called Kuffer cells and in the brain microglial cells. Phagocytosis is the major mechanism by which microbes are removed from the body by the above mentioned phagocytes.

Macrophages, B-lymphocytes and **Dendritic** cells are also called professional antigen-presenting cells (APC), capture extracellular protein antigens, internalize and process them and display class II-associated peptides to T-helper ($CD4^+$) cells. All nucleated cells can present class I associated proteins, derived from cytosolic proteins such as viral and tumor antigens, to cytotoxic T-cells ($CD8^+$).

For **Class I-associated antigen presentation**, cytosolic proteins are proteolytically degraded in the proteosome, generating peptides with features that enable them to find to class I molecules.

Peptides are delivered from the cytoplasm to the ER by an ATP-dependent transporter called TAP. Newly synthesized Class I MHC-β_2-microglobulin dimers in the ER are attached to the TAP complex and receive peptides transported into the ER. Class I MHC molecules with bound peptides are stable and move out of the ER, through the Golgi complex, to the cell surface.

For **class II-associated antigen presentation**, extracellular proteins are internalized into endosomes, where these proteins are proteolytically cleaved by enzymes that function at acidic pH. Newly synthesized class II MHC molecules associated with invariant chain are transported from the ER to the endosomal vesicles. Here the invariant chain is proteolytically cleaved, and a small peptide remnant of the invariant chain called CLIP, is removed from the peptide-binding cleft of the MHC molecule by the DM molecules. The peptides that are generated from extracellular proteins then bind to the available cleft of the class II MHC molecule, and the trimeric complex (class II MHC α and β chains and peptides) moves to and is displayed on the surfaces of the cell.

The **T-cell receptor (TCR)** for antigen is only found on the T-cell membrane and is composed of two polypeptide chains, α and β. Each of these glycoproteins is made up of constant and variable regions, like those of IgG and together the α and β chain variable regions compare the antigen binding site. Some T-cells, whose function is not clear, express a TCR consisting of γ and δ chains.

The **T-cell receptor complex** of the antigen receptor, the $\alpha\beta$ or $\gamma\delta$ dimer, plus CD_3, a signalling complex composed of γ, δ and ε-chains. CD_4 on T-cells binds to the non-polymorphic region of MHC class II on APCs restricting Th cells to recognizing only peptides presented on MHC class II molecules. CD_8^+ on cytolytic T-cells binds the non-polymorphic region of MHC class I, restricting killing to cells presenting peptide in MHC class I.

Cell mediated Immunity (CMI) is the adaptive immune response against intracellular microbes. It is mediated by T-lymphocytes and can be transferred from immunized to naive individuals by T-cells and not by antibodies.

There are two main forms of cell-mediated immune reactions. In one, which is exemplified by DTH reactions, CD_4^+ TH 1 cells, as well as CD_8^+ CTLs kill any nucleated cell that contains foreign antigens (microbial or tumor antigens) in the cytosol. Cell-mediated immune reactions mediated by CD_4^+ T-cells consist of several steps: T-cell recognition of cell-associated antigens in peripheral lympoid organs, clonal expansion of the T-cells and their differentiation into TH_1 cells, migration of the effector T cells to the site of infection or antigen challenge, leukocyte recruitment with resultant inflammation, and activation of macrophages to eliminate the microbe or antigen. T-cell proliferation and differentiation are dependent on costimulators and cytokines (mainly IL-12) produced by APCs, migration of T-cells to sites of infection is mediated by adhesion molecules binding to their ligands on activated-endothelium, and macrophage activation is mediated by IFN-r and CD40L-CD40 interactions. If the infection is not fully resolved, activated macrophages cause tissue injury and fibrosis.

CD_8^+ CTLs kill cells that express peptides derived from cytosolic antigens (e.g. – viral antigens) that are presented in association with Class I MHC molecules. CTLs differentiate from Pre-CTLs (naïve-CD_8^+ T cells) in response to antigen and costimulators provided by professional APCs or cytokines secreted by helper T-cells. Differentiated CTLs kill target cells by there principal mechanisms : granule exocytosis, which release perforin, a membrane pore-forming protein that polymerizes in the target cell membrane and cause osmotic lysis of target cells;entry of serine proteases (granzymes) into the target cell through the perforin channel, and subsequent granzyme-triggard apoptotic death of the target cell; and expression of FasL, which engage Fas on the target cell membrane, also triggering apoptosis of the target cell (See Fig.-24). Helper T-cells of the TH-2 subset function to control TH-1 mediated macrophage activation. TH-2 cells also induce inflammatory reactions consisting of eosinophils, mast cells and neutrophils.

QUESTION : Immunology

1. Name the different receptors present on the B-lymphocytes. Mention their functions.
2. Name the different receptors present on the T-lymphocytes. Mention their functions.
3. How complement is activated by the alternative pathway after microbial infection and able to lyse the microbes?
4. What is the kinetics of immune response?
5. Write the primary and secondary immune response with suitable curve.
6. (a) What is complement?
   (b) How complements are activated in the body?
   (c) What are the biological activities of the components of complement?
7. (a) Classify T-cells.
   (b) Describe their role in (i) recognition of antigen (ii) killing of infected cells and (iii) secretion of cytokins.
8. (i) Describe the mechanism of B-cell activation.
   (ii) Discuss the biochemical-events leading to T-cell activation?
9. (a) What is antigen processing and presentation?
   (b) Describe cytosolic and endocytic pathway.
10. (a) What are the functions of phagocytes?
    (b) How macrophages are activated and produce effector T-cell?
11. **Short notes :**

    Opsonization, ADCC, B-cell receptor, T-cells receptor, Antibody diversity, Plasma cell, NK-cell, Apoptisis, Cytotoxic T-cell, TH_1 & TH_2 cells, IFNγ, IL-12, phagocytes.

Reference :

1. Kuby Immunology (5th Edition)-R.A. Goldsby, T.J. Kindt and B.A. Osborne.
2. Cellular and Molecular Immunology.-A.K. Abbas, A.H. Lichtman and J.S. Pober.
3. Immunology a short course - Benjamini & Leskowitz.
4. Instant notes in Immunology - P.M.Lydyand, A. Whelar & M.W. Fanger.

----- 0 -----

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-B Unit-II

Module No. - 34

Unit I : Introduction to Biostatistics

Structure :

1.0. Objectives.
1.1. Introduction and definitions.
1.2. Population and samples.
1.3. Variables.
1.4. Frequency distributions.
1.5. Graphical representations of frequency distributions.
1.6. Summary.
1.7. Glossary.
1.8. Questions.

1.0. OBJECTIVES

After studying this unit, it should be possible for you to:

* define statistics,
* define biostatistics and state its aims,
* understand and explain the terms population, sample and simple random sampling,
* understand and explain parameter and statistic,
* arrange the data in frequency distributions,
* present the data in graphical and diagrammatic forms, and
* identify the class of variable to which the property being studied belongs.

1.1. INTRODUCTION AND DEFINITIONS

In this unit, you will learn the definitions of statistics and biostatistics and their aims. You will also come to understand what is meant by population and sample. You will realise why samples have to be drawn from a population for any scientific study, what is meant by population parameter and sample statistic, and how sample statistic may be used as an estimate of the population parameter. You will learn about different classes of variables studied in scientific investigations. You will learn also to classify and tabulate the data, collected in a study, into frequency distributions, and to present the data in graphical and diagrammatic forms for comparative studies and easy comprehension.

1.1.1. *Statistics : the science*

Perhaps all of you know that statistics is a science under the discipline of mathematics. Statistics applies the fundamental principles and laws of mathematics to the data collected from natural observations, surveys and planned experiments. This helps in presenting the data in meaningful forms, quantifying and analyzing the data, interpreting the data, and making inferences as well as predictions. Thus, statistics may be defined as a science serving as a tool in collecting, analyzing and interpreting the data collected in any investigation.

1.1.2. *Biostatistics*

Now that you know what statistics means, you should be in a position to understand what is biostatistics and how to define it. When statistical principles, laws and methods are applied to biological data for analyzing, exploring and interpreting problems of living organisms, such applications of statistics in the study of biology come under the discipline of biostatistics. Its methodology has been proved to be essential for any scientific work in any branch of life sciences.

1.1.3. *Aims of biostatistics*

Biostatistics aims at :

* drawing a sample of a small number of animals from a vast population with well-defined methods to make the sample representative of the population,

* testing how far the experimental method to be used gives consistent and precise results,
* subjecting the sample to the experimental method in a well-designed way to eliminate experimental errors,
* presenting the data, consisting of experimental observations, in meaningful forms by tabulating and graphically expressing them,
* finding out how far the experimental findings in the sample can be generalized for the entire population by working out the probability of obtaining the results accidentally due to the use of only a small sample instead of the entire population,
* finding out whether or not two or more properties bear any relation to each other in a population of animals, and
* predicting the value of a property in an animal from the known value of some other property in the same animal.

1.2. POPULATION AND SAMPLES

1.2.1. *Population*

In statistics, a population or universe consists of all such living organisms, events, cases or inanimate objects of a specified type as possess the property or characteristic under investigation, at a given time. For example, the population of Dalmatian dogs includes all such dogs living on the earth at the time; the population of White Leghorn hens laying cholesterol-free eggs consists of all such hens living in the entire world at the time. You may easily realize that some populations are so large and spread out as to make it almost impossible to measure its size by counting all its members. Such populations are called *infinite populations;* e.g., the population of all tharparkar-breed cattle in the world, or that of all blue whales in the marine water. But some populations, which you may call *finite populations,* are quite small and located within a narrow area so that their sizes may be precisely known by counting all their members; e.g., the population of tigers in a particular animal reserve park, or the population of horses with the mounted police forces in India. In case of both types of population, however, whenever you think or talk about a population, you must take into your consideration each and every member of that population, without a single exception.

1.2.2. *Samples*

It should be evident to you that it is neither feasible nor advisable to subject an entire population, whether finite or infinite, to a scientific investigation or experiment; apart from the immense resources, vast manpower and large finance needed for any such venture, there is always a possibility of inadvertant and accidental exclusion of some members of the population from the study, thus leading to errors which can be neither conceived nor provided for, because the investigator is not aware of the omission. Hence, it is preferable to choose a small number of the members of that population for subjecting them to the study. A *sample* consists of a small number of animals or cases, drawn from the population by statistically formulated methods of *random sampling* and used in an experiment, survey or investigation as a representative of the population. In *simple random sampling*, the choice of individuals for the sample is not left to the subjective selection by the investigator or anyone else; instead, the individuals are chosen from the entire population, using the laws of probability, and thereby fulfilling the following criteria:

* Each member of the population has the same probability of getting chosen for the sample.
* The choice of any member of the population does not influence the choice or rejection of any other, nor does it influence the choice of others.
* Different types of individuals get included in the sample in almost the same proportions in which they appear in the population.

Such a random sample may be subjected to an experiment, survey or study as a representative of the relevant population.

1.2.3. *Statistics or sample statistics*

Think of a sample of fifty cockroaches, in which you have measured the bodylengths. So, you have 50 bodylength *scores* (values) as the data from a sample whose sample size (n) was 50. You may of course expect and find that most of the scores would vary from each other. For subsequent work, it is more convenient and helpful if you work out some summary numerical values like the mean or the median from these numerous scores of the sample and use any such single summary value, instead of the entire set of scores in the data.

A *statistic* or *sample statistic* is a summary numerical value like mean or median of the scores of a sample. It is precisely obtained for a sample because it is worked out directly from the actual (raw) scores of the sample. Now, imagine that you have drawn several samples from the same population of cockroaches. From sample to sample, the individual animals are different and so, their scores are different. Thus, any summary value or statistic, say the mean, is expected to vary from sample to sample - the sample statistics (plural of statistic), therefore, mostly differ in different samples and lie scattered around some central value in case of samples from the same population. These sample statistics are classified into three main classes according to their properties and uses:

(i) Some of them like mean, standard deviation and correlation coefficient describe the characteristics of a sample, such as a central point of the scores, an estimate of spread of the scattered scores and a relationship between two or more properties in the sample (*descriptive statistics*).

(ii) Some like standard error and variance find use in comparing more than one sample, and in making inferences from the differences between samples (*sampling statistics*).

(iii) Some statistics such as *b* coefficients are used in predicting the likely score of a property in an individual from the known score of someother property in the latter (*predictive statistics*).

1.2.4. *Parameter*

If you wish to work out the summary values like mean and standard deviation for an entire population, you cannot directly compute them from raw scores because almost never do you use the entire population in your study. However, you can work out an estimate of a summary value for the population, using the sample statistics computed from the raw scores of a sample from that population. Such estimated summary values for the population are called *parameters*. Thus, the mean ($\bar{X}$) of the scores of a sample is a statistic (sample mean) while the population mean (μ) estimated from $\bar{X}$ is a parametric mean. Because the population does not change like the samples, the parameters are considered constant for the respective populations.

1.3. VARIABLES

You are aware that in most experiments, investigations or surveys, properties, characters or phenomena which are studied show changes or variations - they may vary from animal to animal, organ to organ, and cell to cell even at the same moment (*spatial variations*), or from time to time in the same animal, the same blood vessel and the same cell or tissue (*temporal variations*). All such properties or phenomena are called *variables.* Thus, height, weight, blood volume, wing length, fur color, femur length, blood sugar, urinary urea – all are variables. Application of different statistical methods depends heavily on the particular class of variables to which belongs the property you are investigating.

1.3.1. *Qualitative or nominal variables*

Individuals of a population may be distinguished from each other from their qualitative differences in these variables. So, they may be identified as either equal to or different from each other about these variables. But these variables cannot be measured and expressed in numerical scores. Even, one animal cannot be considered higher or lower than another with respect to these variables. Such qualitative or numerical variables include sex, blood group, eye color and hair color.

1.3.2. *Ordinal variables*

Individuals of a population may be distinguished from each other from their qualitative differences in these variables which, however, cannot be quantitatively measured and expressed. Although similar to nominal variables in these respects, the ordinal variables differ from the latter because individuals of a population can be identified as higher or lower than each other with respect to an ordinal variable. So, individuals can be given ranks in either ascending or descending orders with respect to these variables. Examples: alertness, leadership quality, ferocity and prowess.

1.3.3. *Continuous measurement variables*

Variables such as length, volume, mass, temperature, pressure, area and time can be measured quantitatively. Individual scores of these variables are not confined to whole units only; instead, they may go upto any small fraction of the whole unit depending on the fineness and accuracy of the measuring test or scale. Thus, some or other individuals may be expected to have scores corresponding to even infinitely small fractional units in the scale for any such variable - such a scale, therefore, has no real gap where no individual score may exist. You may realize that the individual scores of such a variable not only show which is higher and which is lower, but also give us quantitatively by how much

each score exceeds or falls short of another score. Biological examples of such continuous measurement variables include femur length, wing length, body temperature, arterial or venous O_2 tension, blood or urine volume, blood pressure, pulmonary alveolar surface area, vital capacity, lean body mass, blood sugar, blood or urinary pH, and blood circulation time.

1.3.4. *Discontinuous measurement variables*

Biological variables such as RBC count, WBC count, platelet count, sperm count, heart or pulse rate, respiratory rate, number of appendages or digits of animal and litter size belong to this class. No doubt, these variables may be measured quantitatively, but their scores cannot be reasonably expected to go to fractional units and must always be in whole numbers of units. So, there cannot be any individual score going upto a fraction of the whole unit, and there would be real gaps in the scale of such scores between the whole units, where no individual score can occur. Nevertheless, individual scores of such variables enable you not only to know which of them is higher or lower than another, but also to find out by how much one is higher or lower than another.

1.4. FREQUENCY DISTRIBUTIONS

The data, particularly of continuous and discontinuous measurement variables from any experiment, survey or test consists or numerical scores, as numerous as the sample size (n). For meaningful presentation, analysis and interpretation, it is often convenient to arrange such raw scores into a number of classes in order to present the frequencies of scores as belonging to the different classes so formed. This is known as a frequency distribution of a measurement variable.

1.4.1. *Continuous frequency distribution*

Scores of a *continuous variable* in a sample are often tabulated in such a frequency distribution. The entire *range* of scores from the lowest to the highest one is divided for this purpose into a number of *class intervals*, each of the same length, i.e., each covering an equal range of the scores which may amount to 3, 5, 7 or 10 scores usually. Against each class interval, the *frequency* or the number of occurrences of all scores belonging to that interval are entered to complete the table. Such grouping of scores into different class intervals changes the data consisting of ungrouped and individual raw scores into *grouped data*. This is particularly helpful in revealing inherent characteristics and features of the data, and in subsequent statistical computations and analysis, particularly where the sample size is large. You should bear in mind that once the data have been changed to a frequency distribution, individual raw scores will not be used in subsequent computation;

instead, all the cases in any class interval are to be assumed to have the score of the *midpoint* (X_c) of that interval, which should be computed for each interval. Moreover, in case of a continuous variable, there should be no gap between the limits of successive class intervals. So, from the scores limits, i.e., the lowest and the highest scores of an interval, true limits or class boundaries (X_l and X_u for the lower and upper true limits respectively) must also be computed.

Let us assume that you are working out a frequency distribution of wing length scores of 30 cockroaches, as given in Example 1.4.1 below. The range of scores is first worked out from the lowest and the highest scores of the data. Depending on the range and the sample size, the entire range is next divided into a number of class intervals and they are shown in both score limits and true limits. For any class interval,

$$X_l = \frac{1}{2} \text{ [(its lower score limit) + (upper score limit of the next lower interval)]}$$

$$X_U = \frac{1}{2} \text{ [(its upper score limit) + (lower score limit of the next higher interval)]}$$

$$X_C = X_l + \frac{1}{2}(X_U - X_l), \text{ or}$$

$$X_C = \text{lowest score} + \frac{1}{2} \text{ [(highest score) – (lowest score)].}$$

The interval length (i) of any interval is obtained as follows:

$i = X_U - X_l$.

***Example* 1.4.1.** Tabulate the following wing length scores of a sample of cockroaches into a frequency distribution.

Wing lengths (mm) : 34, 37, 36, 25, 27, 45, 29, 20, 33, 27, 24, 39, 42, 46, 37, 31, 41, 34, 33, 28, 42, 24, 33, 30, 37, 22, 40, 35, 32, 38.

Solution :

Highest score : 46.

Lowest score : 20.

Range : (46 – 20) + 1 = 27 scores.

Proposed interval length (i) : 4.

No. of intervals : $\frac{\text{Range}}{i} = \frac{27}{4} \approx 7.$

Class intervals in score limits are entered in column 1 of Table 1.1. True limits and midpoint of each interval are worked out and entered in columns 2 and 3, respectively. For example, for the interval with the score limits of 28-31,

$$X_1 = \frac{1}{2}(28+27) = 27.5;$$

$$X_U = \frac{1}{2}(31+32) = 31.5;$$

$$X_C = 27.5 + \frac{1}{2}(31.5 - 27.5) = 29.5, \text{ or}$$

$$X_C = 28 + \frac{1}{2}(31 - 28) = 29.5.$$

Each score is then entered as a tally against its class interval in column 4. After all the scores have been thus entered, the total number of tallies for each interval is entered as the frequencey (f) of that interval in Column 5.

Sample size (n) = $\Sigma f = 30$.

Table 1.1. Tabulation of a continuous frequency distribution of wing length scores (mm) of a sample of cockroaches.

Class intervals (score limits)	Class intervals (true limits)	X_C	Tallies	f
20 - 23	19.5 - 23.5	21.5	||	2
24 - 27	23.5 - 27.5	25.5	||||	5
28 - 31	27.5 - 31.5	29.5	||||	4
32 - 35	31.5 - 35.5	33.5	|||| ||	7
36 - 39	35.5 - 39.5	37.5	|||| |	6
40 - 43	39.5 - 43.5	41.5	||||	4
44 - 47	43.5 - 47.5	45.5	|\	2
Σ				30 (n)

1.4.2. *Discontinuous frequency distribution*

This is the frequency distribution of the scores of a discontinuous measurement variable such as cell count, heart rate, respiratory rate, litter size and number of digits. The procedure for working it out is almost the same as that for a continuous frequency distribution, except that no true limit or class boundary is worked out for any class interval or used in the computations for the discontinuous frequency distribution. This is because the variable here is a discontinuous measurement variable with true gaps in its scale of scores due to the occurrence of scores in whole units only. For the tabulation of such a frequency distribution, you may see Example 1.4.2. given below.

***Example* 1.4.2.**

Tabulate the following heart rate scores (heart beat / minute) of a sample of 25 chimpanzees into a frequency distribution.

68, 72, 67, 74, 75, 82, 83, 73, 67, 68, 84, 76, 65, 63, 61, 77, 73,

83, 80, 79, 74, 77, 70, 75, 71.

***Solution* :**

Highest score	:	84.
Lowest score	:	61.
Range	:	(84 - 61) + 1 = 24 scores.
Proposed interval length (i)	:	5.
No. of intervals	:	$\frac{\text{Range}}{i} = \frac{24}{5} \approx 5.$

Class intervals in score limits are entered in Table 1.2. Midpoint (X_c) of each interval is worked out and entered in the table. For example, for the class interval of 71-75,

$$X_c = 71 + \frac{1}{2}(75 - 71) = 73.$$

Each score is entered as a tally against its class interval in the table. The total number of tallies for each interval is entered as the frequency (f) of that interval.

Sample size (n) = $\Sigma f = 25$.

Table 1.2. Tabulation of a discontinuous frequency distribution of heart rates.

Class intervals	X_c	Tallies	f
61 - 65	63	|||	3
66 - 70	68	|||||	5
71 - 75	73	||||| |||	8
76 - 80	78	|||||	5
81 - 85	83	||||	4
Σ			25 (n)

1.4.3. *Qualitative frequency distribution*

For statistical treatment of data of a qualitative or nominal variable, a qualitative frequency distribution may be framed by tabulating the frequencies or numbers of occurrences of individual cases in a sample against the classes of the nominal variable. An example is given below. You may notice that such variables as sex, hair color, blood group, eye color, race, phenotype and fur color cannot be measured quantitatively to yield scores. But the numbers of cases occurring in different classes of its frequency distribution form *enumeration data* in numerical units and can be subjected to statistical investigations. Also note that the classes of such a qualitative variable cannot be given score limits, nor can you work out any midpoint of any of those classes.

***Example* 1.4.3.**

In a sample of 492 *Drosophila* flies, grey-bodied red-eyed, black-bodied red-eyed, grey-bodied scarlet-eyed and black-bodied scarlet-eyed flies were found to number 269, 92, 88 and 43, respectively. Present the data as a qualitative frequency distribution of the four phenotypes.

Solution :

In Table 1.3, the phenotypes are shown as the classes of the relevant qualitative variable and the respective frequencies of cases are entered against them.

Table 1.3. A qualitative frequency distribution of *Drosophila* phenotypes.

Phenotypes	Frequencies
Grey-bodied red-eyed	269
Black-bodied red-eyed	92
Grey-bodied scarlet-eyed	88
Black-bodied scarlet-eyed	43
Σ	492 (*n*)

1.5. GRAPHICAL PRESENTATION OF FREQUENCY DISTRIBUTIONS

Presentation of data in the form of tabulated frequency distributions does not always reveal the features and characteristics of the data clearly and is not always convenient for comparison betweeen two sets of data. A better way of presentation often consists of some graphical and pictorial forms into which the frequency distribution may be transformed. Such graphic representations include frequency polygons, bar diagrams, histograms, and cumulative frequency or cumulative percentage curves.

1.5.1. *Frequency polygon*

It is a graphical area diagram of a quantitative frequency distribution of a continuous measurement variable. For drawing it, the scores of the continuous variable are scaled along the abscissa or X axis on a graph paper, and the midpoints (X_c) of all the class intervals in the data as well as two additional X_c scores for the intervals immediately below and above the intervals, shown in the data, are marked on the X axis. You may easily see that each of these two additional intervals bears a frequency of zero number of cases – they are included for bringing the sides of the polygon to the base or zero point of the Y axis scale. The ordinate or Y axis is scaled in frequencies (*f*). The frequency of each interval, including the zero frequency of each of the two additional intervals, is plotted against the midpoint (X_c) of that interval. The plotted points are joined by straight lines to complete the frequency polygon which encloses an area representing the sample size (*n*).

Such frequency polygons, drawn for more than one sample, may be conveniently superimposed to compare the frequency distributions of a variable in different samples. However,

jagged contour of a frequency polygon does not clearly indicate the proportionate frequencies in different class intervals. The jaggedness can be moderated by modifying the frequency polygon into a *smoothed frequency polygon*. For this, the smoothed frequenoy (f_s) of each interval is computed as a mean of the observed frequencies (f_o) of the relevant interval and the two intervals immediately preceding and following it. For any interval, therefore,

$$f_s = \frac{1}{3}\ [(f_o \text{ of that interval}) + (f_o \text{ of next lower interval}) + (f_o \text{ of next higher interval})].$$

Each computed f_s is then plotted against the X_c of the relevant interval and the plotted points are joined by lines to complete the smoothed polygon.

***Example* 1.5.1.**

Draw a frequency polygon with the frequency distribution computed in Example 1.4.1. Also compute smoothed frequencies and draw a smoothed frequency polygon for the same distribution.

Solution :

A frequency polygon is drawn, plotting the original frequencies (f_o) against the respective midpoints (X_c) of class intervals of Example 1.4.1. (Table 1.4. and Fig. 1.1). The plotted points are joined to give the original frequency polygon. Note that two additional class intervals, respectively below and above the original intervals and each showing zero frequency, have been included in Table 1.4.

Next, the smoothed frequency (f_s) is computed for each interval and entered in Table 1.4. For example, for the interval 20 - 23,

$$f_s = \frac{1}{3}\ [(f_o \text{ of the interval } 20 - 23) + (f_o \text{ of the interval } 16 - 19) + (f_o \text{ of the interval } 24 - 27)]$$

$$= \frac{1}{3}\ (2 + 0 + 5) = 2.3.$$

The computed f_s scores are plotted against the respective X_c scores in Fig. 1.1. and the plotted points are joined to give the smoothed polygon.

Table 1.4. Original and smoothed frequency distributions of the data of Example 1.4.1.

Class intervals (score limits)	Class intervals (true limits)	X_C	f_o	f_s
16 - 19	15.5 - 19.5	17.5	0	0.7
20 - 23	19.5 - 23.5	21.5	2	2.3
24 - 27	23.5 - 27.5	25.5	5	3.7
28 - 31	27.5 - 31.5	29.5	4	5.3
32 - 35	31.5 - 35.5	33.5	7	5.7
36 - 39	35.5 - 39.5	37.5	6	5.7
40 - 43	39.5 - 43.5	41.5	4	4.0
44 - 47	43.5 - 47.5	45.5	2	2.0
48 - 51	47.5 - 51.5	49.5	0	0.7

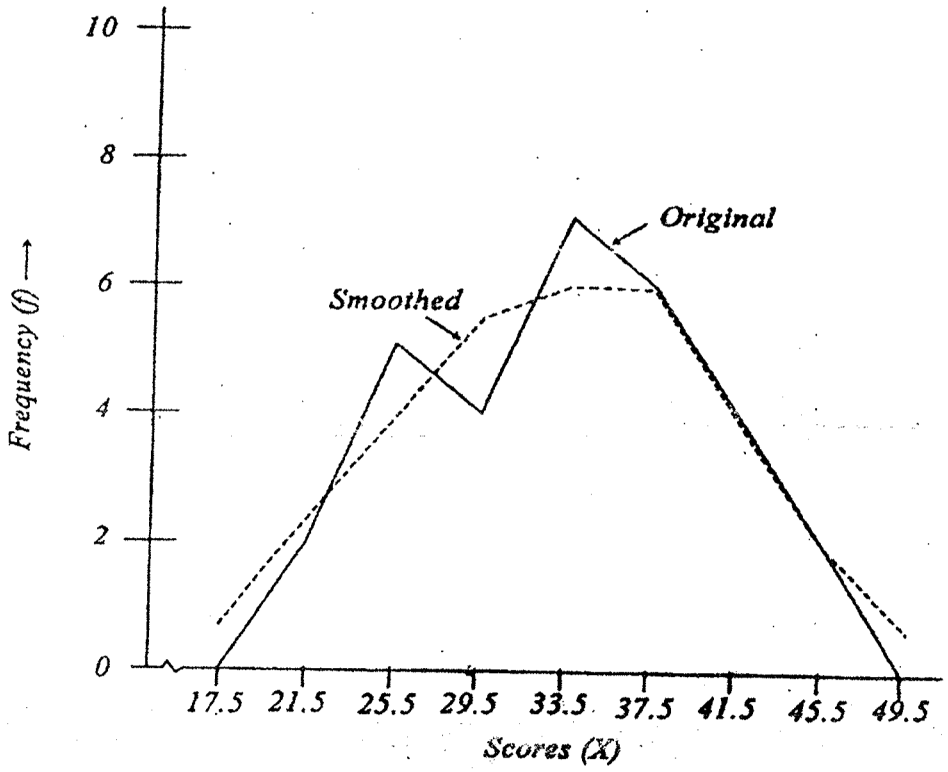


Fig. 1.1. Original and smoothed frequency polygons of wing length distribution of a sample of cockroaches (Table 1.4).

1.5.2. *Histogram*

This is an area diagram of a quantitative frequency distribution of a continuous variable. For drawing a histogram, the scores of the variable are scaled along the abscissa or X axis, marking only the true limits (X_l and X_U) of the class intervals on that axis. The frequencies (f) of the scores are scaled along the ordinate or Y axis. At each of the true limits of a class interval, an ordinate is raised till the two ordinates at its true limits reach the level of f of that interval, read from the Y axis scale. A horizontal line is then drawn at the level of the corresponding f score to join the two ordinates, thus completing a bar or column with its base and height corresponding to the interval length and the frequency of the relevant interval respectively. In this way, bars or columns are drawn for all intervals with no gap between the bars, because the continuous variable has no true gap in its scale. If all the intervals have equal lengths (i), the bases of all the bars will be equal; so the area of each bar will be proportional to only the height or frequency of that interval. This continuous set of bars is called a histogram. Unlike the frequency polygon, the histogram gives a distinct idea about the proportionate distribution of frequencies in the class intervals; but it is less convenient for comparing more than one distribution, because two histograms cannot be superimposed on each other.

Fig. 1.2. shows a histogram of the distribution of the wing length data of Tahle 1.4.

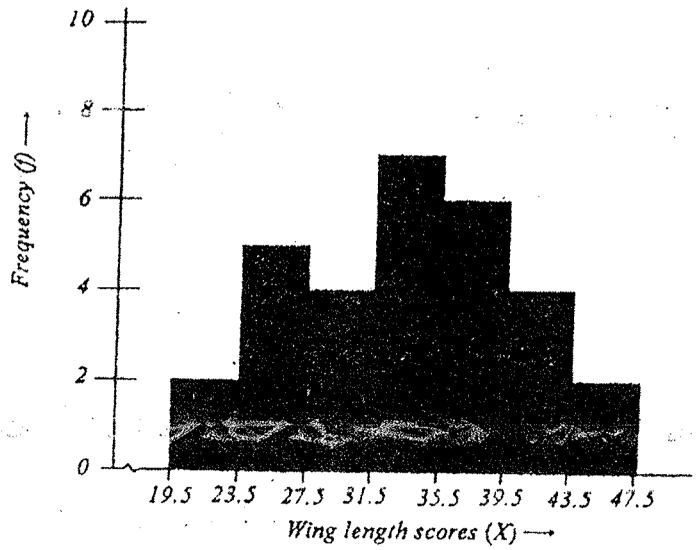


Fig. 1.2. Histogram of the frequency distribution of Example 1.4.1. and Table 1.4.

1.5.3. *Bar diagram*

This is used for presenting and comparing discontinuous and qualitative frequency distributions of respectively discontinuous and nominal variables, with true gaps between their classes. *Simple bar diagram* shows the classwise distribution of frequencies in a single sample. For this, as many parallel bars are drawn as the number of classes of the variable, with an equal but arbitrary width of their bases, with lengths determined by the frequencies of the respective classes, and with small gaps between consecutive bars to indicate true gaps between the classes. The frequencies (*f*) are scaled parallel to the long axis of the bars. The areas of the bars are proportional only to their lengths determined by their respective frequencies in the class intervals and so, gives a relative idea about classwise distribution of frequencies. But a number of samples cannot be compared by superimposition of their bar diagrams. Instead, a *multiple bar diagram* is used for such purpose. In it, the frequency distribution of each sample is presented by a simple bar diagram, and as many such sets of bars are used as the number of samples, separating each set from the neighbouring sets by wider gaps (Fig. 1.3).

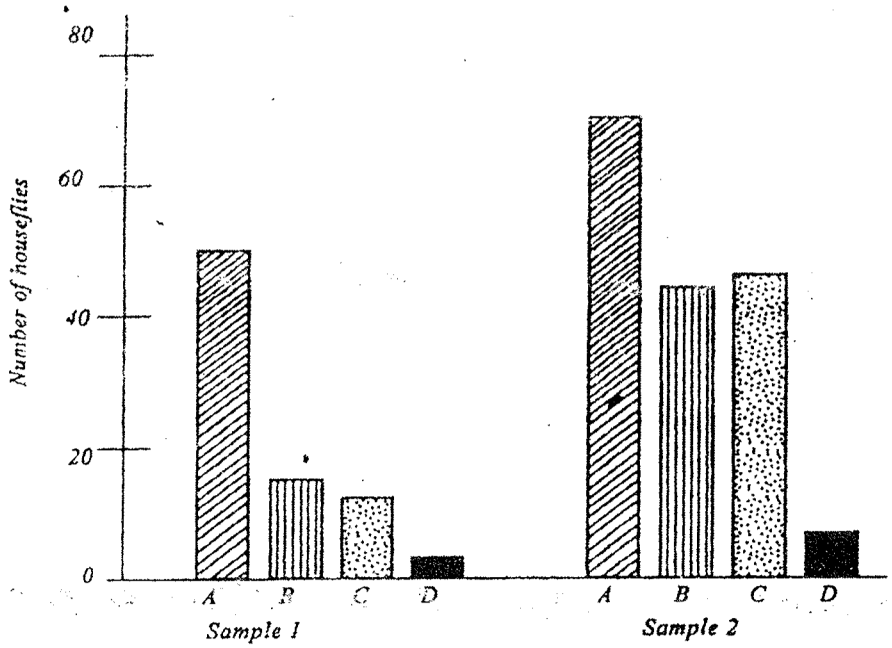


Fig. 1.3. A multiple bar diagram showing the distribution of phenotypes of houseflies in two samples. A: grey body - red eye. B:black body - red eye. C: grey body - scarlet eye. D: black body - scarlet eye.

1.5.4. *Cumulative frequencies and percentages and their ogives*

Cumulative frequency (*cf*) of any class interval of a frequency distribution is the total frequency upto the upper limit of that interval, commencing from the lowest class interval. Where $f_1, f_2, \ldots\ldots$ are the frequencies of class intervals starting from the lowest, cf_k of the kth interval is given by:

$$cf_k = f_1 + f_2 + \ldots\ldots + f_k.$$

Cumulative percentage (*cP*) of any class interval is its *cf* expressed as a percentage of the sample size (n). Thus, for the kth interval,

$$cP_k = \frac{cf_k}{n} \times 100.$$

Scaling the *cf* scores along the Y axis and the upper true limits of intervals (X_U) along the X axis, a *cumulative frequency ogive* is drawn by plotting the *cf* of each interval against its X_U. A *cumulative percentage ogive* is drawn similarly by scaling *cP* scores, instead of *cf* scores, along the Y axis. In both cases, the X_U of the interval just lower than the lowest interval should be included in the X scale so that the ogive may start from the zero value of the *cf* or *cP* scale on the Y axis. *cP* ogives find use in the graphic determination of median, percentiles and percentile ranks; *cf* may also be used in computing percentile ransks.

***Example* 1.5.2.**

Use the frequency distribution of wing length scores (mm) of a sample of cockroaches, given in Example 1.4.1, to compute their *cf* and *cP* and draw the cumulative percentage ogive.

Solution :

The wing length frequency distribution is entered in Table 1.5, including an additional class interval below the lowest.

Table 1.5. Computation of *cf* and *cP* scores of wing length distribution.

Class intervals	True upper limits	*f*	*cf*	*cP*
16-19 (15.5-19.5)	19.5	0	0	0
20-23 (19.5-23.5)	23.5	2	2	6.7
24-27 (23.5-27.5)	27.5	5	7	23.3

Table 1.5. (Continued)

Class intervals	True upper limits	f	cf	cP
28-31 (27.5-31.5)	31.5	4	11	36.7
32-35 (31.5 - 35.5)	35.5	7	18	60.0
36-39 (35.5-39.5)	39.5	6	24	80.0
40-43 (39.5-43.5)	43.5	4	28	93.3
44-47 (43.5-47.5)	47.5	2	30(n)	100.0

The cf and cP scores of all the class intervals are computed and entered in Table 1.5. For example, for the class inerval 28-31,

$$cf = 0+2+5+4 = 11; \quad cp = \frac{11}{30} \times 100 = 36.7.$$

A cP ogive is then drawn scaling the cP scores along the ordinate and the X_u scores along the abscissa, and plotting each cP score against the corresponding X_u score (Fig. 1.4.).

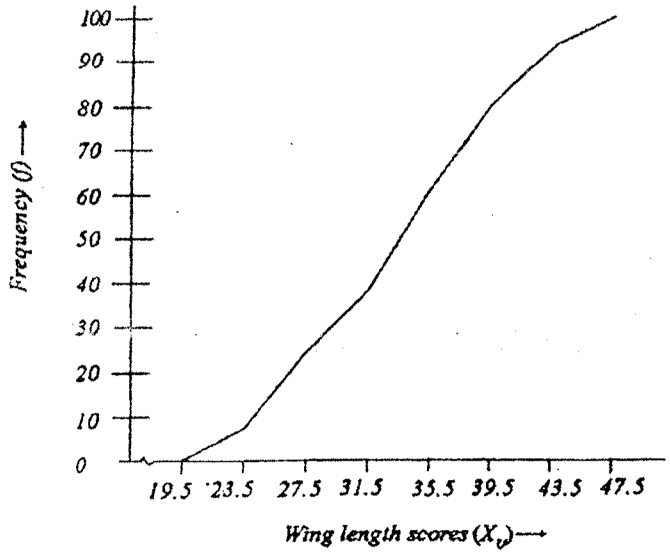


Fig. 1.4. The cumulative percentage ogive of the wing length distribution of Table 1.5.

1.6. SUMMARY :

In this unit, statistics and biostatistics have been defined. You have learnt about the concept of random samples and why a sample should be used instead of the entire population for any

scientific investigation. Variables have been defined and classified into qualitative variables, ordinal variables, continuous variables and discontinuous variables with examples and principal characteristics. You have been told about the grouping and tabulation of data into continuous, discontinuous and qualitative frequency distributions. Presentation of data in graphical and diagrammatic forms such as frequency polygons, histograms and bar diagrams have been described with their relative merits and demerits. You have read about computations of cumulative frequencies and cumulative percentages, their presentation in graphical forms called ogives and their uses.

Statistic and parameter have been defined. A broad classification of statistics has been made into descriptive, sampling and predictive statistics with examples.

In the next unit, you will learn more about some descriptive and sampling statistics, their computations and uses. This will lead you to the application of statistical method in the analysis and interpretation of scientific observations.

1.7. GLOSSARY

Bar diagram : A set of parallel columns or bars separated from each other by short distances, used for graphical presentation of the frequency distribution in the classes of a discontinuous or qualitative variable.

Biostatistics : The science of application of statistical principles and methods to biological sciences.

Cumulative frequency (cf) : The sum of frequencies of class intervals from the lowest interval and upto the upper limit of the class interval under consideration.

Cumulative percentage (cP) : The cumulative frequency of a class interval expressed as a percentage of the total frequency of the entire sample.

Continuous variables : A variable whose scores are not limited to whole numbers of units only, and can be measured and expressed even in infinitely small fractions of units.

Descriptive statistics : A sample statistic as the mean and standard deviation, which describes the properties of a particular sample.

Discontinuous variables : A variable whose scores are limited only to whole numbers of the unit and cannot be measured or expressed in fractional units.

Frequency (*f*) : The total number of occurrences of a score or case in a particular class interval or in the entire sample.

Frequency distribution : A tabulated presentation of the distribution of scores in a sample, showing their frequencies in the different class intervals.

Frequency polygon : The graphical presentation of a continuous frequency distribution in the form of an area diagram obtained by plotting the frequencies of the class intervals against their respective midpoints.

Histogram : A set of parallel bars or columns with no intervening distances between them, used for graphical presentation of the frequency distribution of a continuous variable in different class intervals of a sample.

Ogive : A curve for the graphical representation of the classwise distribution of either cumulative frequencies or cumulative percenrages in a sample.

Ordinal variable : A variable which cannot be measured quantitatively and expressed in numerical units although the individuals of a sample can be ranked with respect to the variable.

Parameter : Summary values such as mean, median and standard deviation of all the scores of a variable in the population.

Population : All individuals, cases or events, having specified characteristics or properties, and existing in the entire universe at a given time.

Predictive statistic : A sample statistic such as the regression coefficient, used in predicting the probable score of a variable in an individual.

Qualitative variable : A variable which can neither be measured quantitatively, nor be used in ranking individuals with respect to it.

Quantitative variable : A variable which can be measured quantitatively.

Sample : A limited number of individuals or cases, chosen from a population and used in any scientific investigation as the representative of that population.

Sampling statistic : A statistic such as a standard error which compares the properties of more than one sample and interpretes their differences.

Statistic : Any single summary value such as mean and standard deviation, computed from the scores of a sample.

Statistics : (*a*) Plural of statistic (see above). (*b*) The science of application of mathematical principles and methods for the planning and execution of any survey or investigation and analysis and interpretation of its observations.

Variable : Any property, characteristic or event which changes from time to time and/or from site to site, either in quantity or in quality.

1.8. QUESTIONS

1. (a) Describe how you would tabulate the raw scores of a continuous variable into a frequency distribution.

   (b) How would you plot a frequency polygon from a frequency distribution? Mention the advantages and disadvantages of using the polygon in presenting the data.

2. (a) How would you compute cumulative frequencies and cumulative percentages, starting from raw scores?

   (b) Describe how you would present cumulative frequencies and percentages graphically.

   (c) Mention the principal applications of cumulative frequencies and percentages.

3. (a) Explain the terms sample and population. Mention why a sample is used instead of the population in scientific studies. What are the basic criteria of simple random sampling?

   (b) Write briefly about statistic and parameter. Give a brief classification of sample statistics.

4. (a) Classify variables, giving the characteristics and examples of each class.

   (b) Describe the construction of a histogram, a simple bar diagram and a multiple bar diagram. Mention when you would draw each of them to represent frequency distributions.

5. (a) Tabulate the following interorbital width scores (in mm) of a sample of pigeons in a frequency distribution and draw its frequency polygon.

11.4, 11.9, 12.1, 11.4, 12.1, 10.9, 11.4, 11.2, 12.5, 12.1, 11.5, 10.5, 13.5, 12.2, 12.0, 12.6, 11.4, 12.2, 11.3, 12.7, 11.9, 11.1, 12.3, 11.9, 11.9, 11.5, 11.0, 10.7, 11.6, 10.8, 11.5, 13.2, 11.8, 12.2, 11.9, 12.5, 11.0, 12.4, 11.1, 11.0.

(b) Draw a histogram presenting the frequency distribution of the scores given above.

6. (a) Group the following femur length scores (mm x 10^{-2}) of a sample of aphids into a frequency distribution showing its class intervals with the respective true limits and midpoints.

39, 38, 36, 42, 34, 32, 43, 45, 38, 33, 39, 36, 40, 37, 42, 44, 36, 35, 40, 41, 33, 43, 31, 39, 35.

(b) Work out the cumulative frequency and the cumulative percentage of each class interval upto its true upper limit.

(c) Draw the cumulative frequency and percentage ogives of the given data.

7. Draw a multiple bar diagram to present the following frequency distributions of phenotypes in two samples of *Drosophila*.

Phenotypes :	Black-body red-eye	Grey-body scarlet-eye	Black-body scarlet-eye	Grey-body red-eye
Frequencies :				
(i) Sample 1 :	60	62	30	208
(ii) Sample 2 :	64	56	20	182

- - - - -

ZOOLOGY

Module No. - 34

Unit II : MEASURES OF CENTRAL TENDENCY, DISPERSION AND SAMPLING ERROR

Structure :

2.0. Objectives.
2.1. Introduction.
2.2. Measures of central tendency.
2.3. Measures of dispersion.
2.4. Sampling errors and their measures.
2.5. Summary.
2.6. Glossary.
2.7. Questions.

2.0. OBJECTIVES

It is intended that after going through this unit, you should be able to do the following:

* understand and explain the measures of central tendency, percentiles and measures of dispersion,
* describe the principal properties of mean, median, percentiles and mode as the statistics of location,
* compute the mean, median and mode for different types of data,
* understand and describe the principal properties of variance, standard deviation, quartile deviation and coefficient of variation as measures of dispersion,
* compute *SD, Q, MS* and *CV* from sample scores,
* understand what and why are the sampling errors of statistics, and
* define, describe and compute standard errors as estimates of sampling errors of different statistics.

2.1. INTRODUCTION

In the preceding unit, you have learnt that statistics (singular: statistic) are some numerical scores computed as summary values from the raw scores of a sample for use in describing the properties of the sample, comparing them with those of other samples and applying them for further analysis and interpretation of the data. You have also been acquainted with the broad classification of the statistics into *descriptive statistics* for describing the properties of a sample, *sampling statistics* for comparing more than one sample and making inferences from the data, and *predictive statistics* for predicting the score of one variable from the observed score of another in the same individual.

In this unit, you will learn further about some of the descriptive and sampling statistics, their computations and uses.

2.1.1. Classes of descriptive statistics

We shall deal first with *descriptive statistics*. They are further divided into three subclasses, viz., statistics of location, statistics of dispersion and statistics of correlation. *Statistics of location* show where specific points of the frequency distribution of a sample are situated on the scale of its scores. They include mean, median, mode, percentiles and quartiles. Some of these, e.g., mean, median and mode, give the positions of some central scores of the distribution on the score scale and are collectively known as *central values* or *measures of central tendency. Statistics of dispersion* include range, variance, standard deviation, quartile deviation, and coefficient of variation. They are different measures of the dispersion of scores around specific central values. *Statistics of correlation* measure the magnitude and algebraic sign of the relationship between two or more variables in the individuals of a sample; you will learn about the correlation statistics in a subsequent module.

2.2. MEASURES OF CENTRAL TENDENCY

Statistics such as mean, median and mode are called the central values or measures of central tendency because they are some centrally located scores of the frequency distribution, around which the individual scores of the sample tend to gather. They are included in the statistics

of location as they indicate the location of the frequency distribution on the scale of scores. In fact, they are different kinds of average scores computed from the observed raw scores of the sample. So, they are useful in computing the statistics of dispersion as the measures of scattering of scores around them, and as the representative central scores of the sample while comparing different samples. Moreover, the position of any individual score in the frequency distribution of the sample may be easily understood from its deviation from any measure of central tendency. For example, if the mean fasting blood sugar of a sample of goats is 60 mg/ 100 ml and two individual goats of the sample have fasting blood sugar values of 49 and 90 mg respectively, evidently the former has a somewhat lower blood sugar and the latter has much higher blood sugar than the average value for the sample. Again, if the mean femur lengths of two samples of cockroaches amount to 18 and 27 mm respectively, it is evident that the individual animals of the first sample have, on average, shorter femurs than those of the second one.

2.2.1. *Mean*

Mean ($\overline{X}$) is the most commonly computed and used statistic of location. It can be worked out directly by dividing the sum of all the individual scores (X) of a sample by the total number (n) of individuals or scores in the sample. It is thus defined as the arithmetic average of the scores of a sample. It may also be defined as the score each individual would have possessed if the total score (ΣX) of the sample would have been distributed equally among all its individuals.

If the total number of cases or scores in the sample, i.e., the sample size (n) is small, the scores (X) need not be grouped into a frequency distribution with regular class intervals and may be used directly for computing their mean. For such *ungrouped data* not divided into class intervals,

$$\overline{X} = \frac{\Sigma X}{n}$$

If the sample is not so large as to require the grouping of its scores into a frequency distribution or if the scores are too widely separated from each other to be grouped into classes, but all or most of them occur more than once each, then each individual score (X_i) is multiplied by the number of its occurrence, i.e., its frequency (f_i), to give $f_i X_i$. The sum ($\Sigma f_i X_i$) of all such products is divided

by the sample size (n) to give the mean $(\bar{X})$. [See Example 2.2.1].

$$\bar{X} = \frac{\Sigma f_i X_i}{n} = \frac{f_1 X_1 + f_2 X_2 + \ldots\ldots + f_k X_k}{n}$$

where $X_1, X_2, \ldots\ldots X_k$ are the respective separate scores.

The scores of a large sample are grouped into a frequency distribution, working out its class intervals, each ranging over a number of scores, and their class marks or midpoints (X_c). You have learnt in the previous unit that in such *grouped data*, the midpoint X_c of each class interval is taken as the score of each case whose frequency (f) is recorded in that interval. Hence, the mean of such a frequency distribution, whether continuous or discontinuous, is worked out by dividing the sum of the products of the respective frequencies (f) and midpoints (X_c) of all the class intervals with the sample size (n). [See Example 2.2.2.].

$$\bar{X} = \frac{\Sigma f X_c}{n}.$$

If there are k number of samples with $\bar{X}_1, \bar{X}_2, \ldots\ldots \bar{X}_k$ as their respective means and n_1, n_2, ...n_k as their respective sample sizes, you may compute the weighted mean $(\bar{X}_w)$ of all the samples by giving the weights of sample sizes to the respective means $(\bar{X}_i)$.

$$\bar{X}_w = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \ldots\ldots + n_k \bar{X}_k}{n_1 + n_2 + \ldots\ldots + n_k}$$

or, $$\bar{X}_w = \frac{\Sigma n_i \bar{X}_i}{\Sigma n_i}.$$

Some important properties of the mean are mentioned below.

(i) Because the mean is computed by dividing the sum of all scores of a sample with the sample size, the product of the mean and the sample size gives the sum of all the scores: $\Sigma X = n\bar{X}$.

(ii) The algebraic sum of the differences between the individual scores and the mean would always amount to zero : $\Sigma(X - \bar{X}) = 0$.

(iii) The algebraic sum of the squared differences between the individual scores and the mean, $\Sigma(X - \bar{X})^2$, is always a positive integer.

(iv) The algebraic sum of such differences, each raised to an odd number of power such as 3, 5 and 7, e.g., $\Sigma(X-\bar{X})^3$, amounts to zero in a symmetric frequency distribution, but is either a positive or a negative integer according as the positive (or high-value) end or the negative (or low-value) end of an asymmetric distribution has more extreme scores than the other end.

(v) The mean of a sample having a perfectly symmetrical frequency distribution coincides with its median and mode, and is a very dependable measure of central tendency. But it is higher or lower than both median and mode according as the frequency distribution has more extreme scores in its high-value or low-value tail respectively. Because even a single extreme score in one of the tails would deflect the mean towards that tail, mean is not dependable as a measure of central tendency for such asymmetric distributions.

(vi) The mean has the same unit as the scores from which it is computed.

The parametric or population mean (μ) is usually not directly worked out because seldom the entire population is subjected to a study. However, estimates of the parametric mean may be made from the sample mean $(\bar{X})$.

***Example* 2.2.1.**

Compute the mean of the following interorbital width scores (mm) of a sample of pigeons using their respective frequencies.

11.4, 12.1, 10.9, 11.4, 12.5, 11.9, 12.7, 11.9, 10.9, 12.1, 12.7, 11.9, 12.5, 12.1, 11.4, 11.9, 12.8, 12.5, 10.8, 11.9.

***Solution* :**

The scores (X_i), the tallies for their respective occurrences and the total number of tallies giving the frequency (f_i) of each score are entered in the columns 1, 2 and 3, respectively in Table 2.1. The products f_iX_i of the respective frequencies and scores are worked out and entered in column 4. The totals of the columns 3 and 4 give the sample size (n) and Σf_iX_i.

Table 2.1. Table for computing the mean from the frequencies of scores of ungrouped data.

Scores (X_i)	Tallies	Frequencies (f)	f_iX_i
10.8	/	1	10.8
10.9	//	2	21.8
11.4	///	3	34.2
11.9	/////	5	59.5
12.1	///	3	36.3
12.5	///	3	37.5
12.7	//	2	25.4
12.8	/	1	12.8
Σ		20 (n)	238.3 (Σf_iX_i)

$$\overline{X} = \frac{\Sigma f_i X_i}{n} = \frac{238.3}{20} = 11.915 \cong 11.9\text{mm}.$$

***Example* 2.2.2.**

Compute the mean of the winglength scores (mm) of a sample of cockroaches, given in Example 1.4.1., using the frequency distribution tabulated in Table 1.1.

Solution :

The class intervals, midpoints (X_c) and frequencies (f) given in Table 1.1. are entered in the first three columns of Table 2.2.

The frequency (f) of each class interval is multiplied with the midpoint (X_c) of that interval and the product (fX_c) is entered in the fourth column of Table 2.2. The totals of the third and fourth columns give the sample size (n) and the sum of all the fX_c scores.

Table 2.2. Table for computing the mean of winglength data grouped into a frequency distribution.

Class intervals	X_c	f	fX_c
20-23	21.5	2	43.0
24-27	25.5	5	127.5
28-31	29.5	4	118.0
32-35	33.5	7	234.5
36-39	37.5	6	225.0
40-43	41.5	4	166.0
44-47	45.5	2	91.0
Σ	-	30 (n)	1005.0 (ΣfX_c)

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{1005.0}{30} = 33.5 \text{ mm}$$

2.2.2. *Mode*

Mode (M_o) of a variable is that score which occurs the highest number of times in the sample, i.e., has the highest frequency. You may easily see that a sample does not possess a mode if its scores do not differ at all in their frequency of occurrence; e.g., where the scores are like 33, 33, 35, 35, 38, 38, 42, 42, 43 and 43 – each score occurring twice. On the contrary, a sample with one, two or many scores, occurring most frequently, has as many modes and its frequency distribution is accordingly called *unimodal, bimodal* or *multimodal*. It should also be evident that the mode has the same unit as that of the raw scores of a sample.

Unlike the mean and the median, the mode is unaffected by the presence of more extreme

scores at one end of the frequency distribution. But while the mode coincides with the median and the mean in a symmetric frequency distribution, the median and the mean are either higher or lower than the mode in an asymmetric frequency distribution according as its high-value or low-value end has more extreme scores. This has led to the use of mode in working out the degree and direction of asymmetry of a frequency distribution. Otherwise, mode has little application in zoological experiments.

Mode is computed fairly easily. For *ungrouped data,* mode is straightway the most frequent score in the sample. For *grouped data* of a frequency distribution with *equal interval lengths* (i), mode is worked out using the lower limit (X_1) and the frequency (f_m) of the modal class having the highest frequency, the frequencies (f_{m-1} and f_{m+1} respectively) of the immediate premodal and postmodal classes, and the class interval length (i). [See Example 2.2.3.]

$$M_o = X_1 + \frac{f_m - f_{m-1}}{2f_m - (f_{m-1} + f_{m+1})} \times i.$$

But if the interval length is unequal, the mode of a frequency distribution has to be computed approximately from its mean ($\bar{X}$) and median (Mdn).

$$M_o = 3Mdn - 2\bar{X}.$$

Example **2.2.3.**

Work out the mode of the winglength scores (mm) of a sample of cockroaches, presented in the frequency distribution of Table 1.1.

Solution :

The class intervals with their score limits as well as true limits, and their respective frequencies are entered in the first three columns of Table 2.3.

Table 2.3. Table for computing the mode of winglength data grouped into a frequency distribution of equal interval length.

Class intervals (score limits)	Class intervals (true limits)	Frequencies (f)	Remarks
20 - 23	19.5 - 23.5	2	
24 - 27	23.5 - 27.5	5	
28 - 31	27.5 - 31.5	4	Premodal class
32 - 35	31.5 - 35.5	7	Modal class
36 - 39	35.5 - 39.5	6	Postmodal class
40 - 43	39.5 - 43.5	4	
44 - 47	43.5 - 47.5	2	

The interval length (i) is obtained from the difference between the lower limits of any two successive intervals. For example, those of the intervals 28-31 and 32-35 amount respectively to 28 and 32;

$\therefore i = 32 - 28 = 4.$

The modal, premodal and postmodal classes are identified and their respective frequencies are noted.

$f_m = 7; \qquad f_{m-1} = 4; \qquad f_{m+1} = 6.$

The lower true limit (X_1) of the modal class is noted : $X_1 = 31.5.$

$$M_o = X_1 + \frac{f_m - f_{m-1}}{2f_m - (f_{m-1} + f_{m+1})} \times i$$

$$= 31.5 + \frac{7-4}{2 \times 7 - (4+6)} \times 4$$

$$= 34.5 \text{ mm.}$$

2.2.3. *Median - a percentile and quartile*

Median (*Mdn*) is that score of a sample which has equal numbers of scores on its two sides. It thus divides the frequency distribution into two equal halves and has 50%, two quarters or half the scores of the sample below it. While it coincides with both mean and mode in a symmetrical distribution, the median lies between these other two measures of central tendency because it is deflected - but much less than the mean - towards that end of the frequency distribution which has more extreme scores. Because of the last property, median is a better representative of a central value than the mean in asymmetric frequency distributions, and indicates a more precise midpoint of such distributions. Moreover, being less deflected than the mean by an asymmetry of the distribution, the median can be used to measure the degree and direction of the asymmetry. Like the mean and the mode, the median also bears the unit of the corresponding raw scores. Besides being a measure of central tendency, the median also belongs to two other types of statistics of location, viz., percentiles and quartiles.

Percentiles and quartiles are also very useful statistics of location, giving the locations of particular points of a frequency distribution on the scale of scores of the relevant variable. A *percentile* is that score below which lies a given percentage of all the scores of a continuous frequency distribution. For example, the 50th percentile (P_{50}), the 25th percentile (P_{25}) and the 75th percentile (P_{75}) are those scores of a sample, which have below them 50%, 25% and 75% of the total number of scores respectively. P_{50} is the same as the median and is thus a measure of central tendency; but the other percentiles, though statistics of location, are not measures of central tendency. A *quartile* is that score which has below it a given number of quarters (or one-fourth fractions) of all the scores of a continuous frequency distribution. For example, the 1st, 2nd, 3rd and 4th quartiles (Q_1, Q_2, Q_3 and Q_4, respectively) are those scores of a sample which have below them one-fourth, two-fourths, three-fourths and all the four quarters of the total scores, respectively. You may see that Q_1, Q_2, Q_3 and Q_4 are identical with P_{25}, P_{50}, P_{75} and P_{100}, respectively. Moreover, Q_2, identical with median, is a measure of central tendency; but the other quartiles, though statistics of location, are not measures of central tendency. Changes in asymmetry and peakedness of a distribution bring about changes in its percentiles and quartiles.

You have just seen that the median and the quartiles are basically different percentiles. So, all of them can be computed just like percentiles for the grouped data of a *continuous frequency distribution.* [See example 2.2.4.] Where P_p is the required percentile, n is the sample size, p is the proportion of total cases and pn the number of cases or scores below P_p, X_1 is the true lower limit and f_p the frequency of scores of the interval in which P_p falls, cf_1 is the cumulative frequency from the lowest interval upto X_1, and i is the length of each interval,

$$P_p = X_1 + i \times \frac{pn - cf_1}{f_p}.$$

Thus, for some specific percentiles or quartiles,

$$Mdn \text{ or } P_{50} \text{ or } Q_2 = X_1 + i \times \frac{0.50n - cf_1}{f_{50}};$$

$$P_{25} \text{ or } Q_1 = X_1 + i \times \frac{0.25n - cf_1}{f_{25}};$$

$$P_{75} \text{ or } Q_3 = X_1 + i \times \frac{0.75n - cf_1}{f_{75}};$$

$$P_{10} = X_1 + i \times \frac{0.10n - cf_1}{f_{10}};$$

$$P_{100} \text{ or } Q_4 = X_1 + i \times \frac{n - cf_1}{f_{100}}.$$

To compute the *Mdn* or P_{50} from *ungrouped data*, the location of the *Mdn* is first worked out from the sample size (n) :

$$Mdn = \frac{n+1}{2} \text{th score.}$$

By counting serially now from either the lowest or the highest score, the $(n+1)/2$th score has to be reached. If n is an odd number, the $(n+1)/2$th score coincides with a score occurring in the

data and that score is taken as the median. If n is an even number, the $(n+1)/2$th score falls between two scores, their average being the median. But if the $(n+1)/2$th score falls within a set of identical scores, *Mdn* has to be worked out by interpolation (see Example 2.2.5.).

***Example* 2.2.4.**

Work out the median, the 25th percentile and the third quartile of the frequency distribution of winglength scores (mm) of a sample of cockroaches, presented in Table 2.3.

Solution :

The class intervals with their score limits, true limits and frequencies are entered in the first three columns of Table 2.4. The cumulative frequency upto the true upper limit of each interval is worked out and entered in the fourth column of Table 2.4. For example, for the class interval 24-27,

$$cf = f_1 + f_2 = 2 + 5 = 7.$$

Table 2.4. Table for computing percentiles of winglength data grouped into a frequency distribution.

Class intervals (score limits)	Class intervals (true limits)	Frequencies (f)	Cumulative frequencies (cf)
20 - 23	19.5 - 23.5	2	2
24 - 27	23.5 - 27.5	5	7
28 - 31	27.5 - 31.5	4	11
32 - 35	31.5 - 35.5	7	18
36 - 39	35.5 - 39.5	6	24
40 - 43	39.5 - 43.5	4	28
44 - 47	43.5 - 47.5	2	30 (n)
Σ		30 (n)	

The length (i) of the class intervals is worked out. For example, using the lower score limits of the intervals, 28 - 31 and 24 - 27, $i = 28 - 24 = 4$.

(1) *Mdn* or P_{50} :

$p = 0.50$. $n = 30$. $\therefore pn = 0.50 \times 30 = 15$.

Counting off the pn number of scores, viz., 15, starting from the lowest interval, we reach the interval 32 - 35 in which the *Mdn* lies. The frequency (f_{50}) of this interval is found to be 7.

$X_1 = 31.5$. $cf_1 = 11$. $f_{50} = 7$.

$$\therefore P_{50} = X_1 + i \times \frac{pn - cf_1}{f_{50}} = 31.5 + 4 \times \frac{0.50 \times 30 - 11}{7}$$

$= 33.8$ mm.

(2) P_{25} or Q_1 :

$p = 0.25$. $n = 30$. $\therefore pn = 0.25 \times 30 = 7.5$.

Interval for P_{25} : $28 - 31$. $i = 4$.

$X_1 = 27.5$. $cf_1 = 7$. $f_{25} = 4$.

$$\therefore P_{25} = X_1 + i \times \frac{pn - cf_1}{f_{25}} = 27.5 + 4 \times \frac{0.25 \times 30 - 7}{4}$$

$= 28.0$ mm.

(3) Q_3 or P_{75} :

$p = 0.75$. $n = 30$. $\therefore pn = 0.75 \times 30 = 22.5$.

Interval for P_{75} : $36 - 39$. $i = 4$.

$X_1 = 35.5$. $cf_1 = 18$. $f_{75} = 6$.

$$\therefore P_{75} = X_1 + i \times \frac{pn - cf_1}{f_{75}} = 35.5 + 4 \times \frac{0.75 \times 30 - 18}{6}$$

$= 38.5$ mm.

***Exmaple* 2.2.5.**

Find the median of the following femur length scores ($m \times 10^{-2}$) of a sample of aphids.

32, 32, 33, 34,36, 36, 36, 38, 40, 40, 42.

Solution :

$n = 11$. $Mdn = \frac{n+1}{2}$ th score $= \frac{11+1}{2}$ th or 6th score.

Counting the scores serially from the lowest one, we reach the 6th score which is 36 - the second one of a set of three identical scores.

The score 36 is assumed to extend over one unit from 35.5 to 36.5, which is occupied by three such successive identical scores. So, each of these three scores is taken to extend over (1.0÷3) or 0.33 units in this interval with its true lower limit (X_1) at 35.5. In reaching the 6th score which is the median, one of these three identical scores have been counted off.

$$\therefore Mdn = X_1 + 0.33 \times 1 = 35.5 + 0.33 \times 1$$

$$= 35.83 \times 10^{-2} \text{ mm.}$$

***Example* 2.2.6.**

Compute the median of following winglength scores (mm) of a sample of cockroaches.

25, 27, 27, 29, 32, 34, 35, 36, 37, 38, 38, 39, 40, 40, 41, 42.

Solution :

Sample size (n) = 16.

$Mdn = \frac{n+1}{2}$ th score $= \frac{16+1}{2}$ th score = 8.5th score.

As the 8.5th score falls between the 8th score (36) and the 9th score (37) starting from the lowest score 25 in the sample,

$$Mdn = \frac{36+37}{2} = 36.5 \text{ mm.}$$

2.3. MEASURES OF DISPERSION

Dispersion is the scattering of raw scores of a sample around a measure of central tendency such as mean and median. *Measures of dispersion* consist of sample statistics such as range, standard deviation, quartile deviation, coefficient of variation and variance. Some of these statistics of dispersion are computed directly from the raw scores of a sample and have the same unit as that of the corresponding raw scores, or such a unit raised to a power. They are included in the class of *absolute measures of dispersion* and include the range, variance, standard deviation and quartile deviation. There are some other statistics of dispersion such as coefficient of variation, which are computed from some absolute measures of dispersion, expressing them as either percentages or proportions of some measures of central tendency. These do not have the units of their raw scores and are called *relative measures of dispersion.* Because they do not have the units of raw scores and are expressed as percentages of the measures of central tendency, they can be applied to compare the dispersions of two sets of scores, given in different units, as also the dispersions of two sets of scores with widely differing magnitudes of their central values.

2.3.1. *Range*

Range is that span along the scale of scores of a measurement variable which is occupied by its scores in a sample. Thus, it extends from the lowest to the highest score in the sample, is given by the difference between those two extreme scores, and is in the same unit as that of the corresponding raw scores. For example, the lowest and the highest winglength scores of a sample of cockroaches (Example 1.4.1) being 20 mm and 46 mm respectively, the range of scores in this sample is as follows.

Range = (highest score) – (lowest score)

= 46 mm – 20 mm = 26 mm.

But except the two extreme scores in a sample, range supplies no information about the intermediate scores. Moreover, it is affected by even a single extreme score. Besides, it gives a span instead of a particular score. Thus, it finds little use in further statistical tests.

2.3.2. *Variance*

Variance or *mean square* (s^2 or *MS*) is a very important absolute measure of dispersion with wide statistical applications. It is the arithmetic average of the squared differences between raw scores of a sample and its mean, and is given in squared units of raw scores. The population variance (σ^2) may be estimated from the sample variance (s^2). However, the squared unit of variance restricts its use partly in ordinary everyday routine work.

For *ungrouped data* of a large sample,

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n} = \frac{\Sigma x^2}{n}$$

where x is the deviation of any raw score from the mean and is called the *error term* of that score, and the sum of squared deviations, viz., $\Sigma(X - \bar{X})^2$ or Σx^2, which is the numerator of the above expression, is called *the sum of squares.*

Alternatively, working directly with raw scores,

$$s^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2;$$

$$\text{or, } s^2 = \frac{n\Sigma X^2 - (\Sigma X)^2}{n^2}.$$

However, for small samples with sample sizes (n) of 30 or less, in particular, computation of

variance with n in the denominator creates an undesirable *downward bias* for variance, because the smaller the sample, the lower is the number of extreme scores in it and consequently, the lower are the deviations $(X-\bar{X})$ of those scores from the mean. Hence, to compensate for this downward bias of variance, the latter is computed using the *degree of freedom* $(n-1)$ instead of the sample size n as the denominator.

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} = \frac{\Sigma x^2}{n-1}$$

$$\text{or, } s^2 = \frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}$$

While the square root of variance is the *standard deviation* which is consequently known as the root mean square also, the ratio of variance and mean is the *coefficient of dispersion* (*CD*) : $CD = s^2/\bar{X}$. Both these are also measures of dispersion; they are in the units of the corresponding raw scores.

An important application of variance is in the test called the *analysis of variances* (anova) for finding whether or not the means of a number of samples differ significantly from each other.

See Examples 2.3.1 – 2.3.3 for computation of variance.

2.3.3. *Standard deviation*

This is the most frequently used absolute measure of dispersion. It may be defined as the positive square root of variance, i.e., of the mean squared deviation. So, unlike the latter which is in squared unit, standard deviation (*SD* or *s*) is expressed in the same unit as the corresponding raw scores. This makes it more convenient than variance for everytime statistical work. Parametric standard deviation (σ) is seldom directly computed; rather, its estimates may be made from the sample *SD*. Some important properties of the latter are as follows.

Both *SD* and variance take into account the dispersions of all the scores from the central value (mean), giving weight to each score according to its distance from the latter. A sample would

ıave an *SD* of 0 if all its scores are identical. The *SD* suffers a change if even a single score is ıltered. It follows from this that the *SD* varies from sample to sample drawn from the same population; ıowever, its sample-wise variations are of a lower magnitude than those of other measures of lispersion. Wider and narrower dispersions of scores are indicated respectively by higher and ower magnitudes of standard deviations. The *SD* shows a tendency of being lowered with the lecline in sample size because of a higher probability of missing the rare cases with more extreme ıcores during sampling. You will find below that this may necessitate a modification of the ːomputation formula for *SD* of small samples.

The standard deviation (s) may be computed from the *ungrouped scores* of a large sample (n>30) using either (a) their mean or (b) only the raw scores (X).

(a) $s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$;

(b) $s = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n^2}}$.

But when the sample is small, the relatively rare cases with more extreme scores may fail to ɔe picked up from the population for the sample. This makes the dispersion of scores narrower in he sample and correspondingly lowers the sum of squares, viz., $\Sigma(X-\bar{X})^2$, in the numerator of the ːomputational formula for both variance and *SD*. This results in a tendency of downward bias for ɔoth of the latter. To compensate for this, the denominator of the formula is changed from the ıample size (n) to the degree of freedom (n-1) in case of small samples (n < 30). The sample *SD*, ːomputed using (n-1) in the denominator as follows, is called the *unbiased SD*. The latter is often ːomputed now even for large samples because their smaller sizes than the population cause some lownward bias, howsoever small, even in their *SD*.

(a) $s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}}$;

(b) $s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{(n-1)}}$.

For *grouped data* of a frequency distribution, the sum of squares is worked out as the sum of the products of the class frequencies (f) and the respective squared deviations of the classmarks or midpoints (X_c) from the mean $(\bar{X})$, because all the cases in a class interval are taken to have the score of its X_c. Thus,

(a) for large samples : $s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n}}$;

(b) for unbiased *SD* of both small and large samples : $s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n-1}}$.

***Example* 2.3.1.**

Compute the variance and unbiased *SD* of the following femur length scores (mm × 10^{-2}) of a sample of aphids.

32, 38, 40, 36, 34, 41, 40, 37, 34, 31, 33.

Solution :

The scores are entered in Table 2.5 and used for computing the mean, the deviations $(X - \bar{X})$ of the scores from the mean, and their respective squared values $(X - \bar{X})^2$.

Table 2.5. Table for computing variance and *SD* from aphid femur length scores, using the sum of squares ($n = 11$).

Scores (X)	$X - \bar{X}$	$(X - \bar{X})^2$
32	−4	16
38	+2	4
40	+4	16
36	0	0
34	−2	4
41	+5	25
40	+4	16
37	+1	1
34	−2	4
31	−5	25
33	−3	9
Σ 396 (ΣX)	–	120

$$\bar{X} = \frac{\Sigma X}{n} = \frac{396}{11} = 36 \times 10^{-2} \text{ mm};$$

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{120}{11-1} = 12.0 \times 10^{-2} \text{ mm}^2;$$

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{120}{11-1}} = 3.464 \cong 3.46 \times 10^{-2} \text{ mm}.$$

An alternative method **:**

To compute variance and *SD* using the mean and the sqared raw scores, the scores are entered in Table 2.6 and squared.

Table 2.6. Table for computing variance and *SD* from aphid femur length scores, using $\bar{X}$ and $X^2 (n = 11)$.

X	X^2
32	1024
38	1444
40	1600
36	1296
34	1156
41	1681
40	1600
37	1369
34	1156
31	961
33	1089
Σ 396 (ΣX)	14376 (ΣX^2)

$$\bar{X} = \frac{\Sigma X}{n} = \frac{396}{11} = 36 \times 10^{-2} \text{ mm};$$

$$s^2 = \frac{\Sigma X^2 - n\bar{X}^2}{n-1} = \frac{14376 - 11 \times (36)^2}{11-1} = 12.0 \times 10^{-2} \text{ mm}^2;$$

$$s = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{14376 - 11 \times (36)^2}{11-1}} = 3.464 \cong 3.46 \times 10^{-2} \text{ mm}.$$

***Example* 2.3.2.**

Compute the variance and the unbiased *SD* of the following interorbital width scores (mm) of a sample of pigeons.

12.1, 11.9, 11.4, 12.1, 11.4, 10.9, 12.5, 12.7, 10.9, 12.7, 12.5, 12.9.

Solution :

The scores are entered in Table 2.7 for working out the mean ($\bar{X}$) and the sum of squares, $\Sigma(X-\bar{X})^2$.

Table 2.7. Table for computing variance and *SD* of interorbital widths, using the sum of squares ($n = 12$).

	Scores (X)	$X-\bar{X}$	$(X-\bar{X})^2$
	12.1	+ 0.1	0.01
	11.9	− 0.1	0.01
	11.4	− 0.6	0.36
	12.1	+ 0.1	0.01
	11.4	− 0.6	0.36
	10.9	− 1.1	1.21
	12.5	+ 0.5	0.25
	12.7	+ 0.7	0.49
	10.9	− 1.1	1.21
	12.7	+ 0.7	0.49
	12.5	+ 0.5	0.25
	12.9	+ 0.9	0.81
Σ	144.0	–	5.46

$$\bar{X} = \frac{\Sigma X}{n} = \frac{144.0}{12} = 12.0 \text{ mm.}$$

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} = \frac{5.46}{12-1} = 0.496 \cong 0.50 \text{ mm}^2.$$

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{5.46}{12-1}} = 0.704 \cong 0.70 \text{ mm.}$$

An alternative method :

For computing variance and *SD* from raw scores (X) directly, the scores are entered in Table 2.8. Each score is squared, the squared scores (X^2) are entered in the table, and both scores and squared scores are totalled for further computation.

Table 2.8. Table for computing variance and *SD* directly from interorbital width scores (n = 12).

Scores (X)	X^2
12.1	146.41
11.9	141.61
11.4	129.96
12.1	146.41
11.4	129.96
10.9	118.81
12.5	156.25
12.7	161.29
10.9	118.81
12.7	161.29
12.5	156.25
12.9	166.41
Σ 144.0 (ΣX)	1733.46 (ΣX^2)

$$s^2 = \frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)} = \frac{12 \times 1733.46 - (144.0)^2}{12(12-1)} = 0.496 \cong 0.50 \text{ mm}^2.$$

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\frac{12 \times 1733.46 - (144.0)^2}{12(12-1)}} = 0.704 \cong 0.70 \text{ mm.}$$

***Example* 2.3.3.**

Compute the variance and the unbiased *SD* of cockroach winglength scores (mm) of the frequency distribution of Table 1.1.

Solution :

The class intervals, midpoints (X_c) and frequencies (f) given in Table 1.1 are entered in the first three columns of Table 2.9.

Table 2.9. Table for computing variance and *SD* from a cockroach winglength distribution ($n = 30$).

Class intervals	X_c	f	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
20 - 23	21.5	2	43.0	– 12	144	288
24 - 27	25.5	5	127.5	– 8	64	320
28 - 31	29.5	4	118.0	– 4	16	64
32 - 35	33.5	7	234.5	0	0	0
36 - 39	37.5	6	225.0	+ 4	16	96
40 - 43	41.5	4	166.0	+ 8	64	256
44 - 47	45.5	2	91.0	+ 12	144	288
Σ	–	30	1005.0	–	–	1312

The mean is first worked out from the sum of the products of f and X_c of all the class intervals. The difference between each X_c score and $\bar{X}$ is worked and squared, and each such squared difference is multiplied with the f of that class interval. The sum of the latter products, viz.,

$\Sigma f(X_C - \bar{X})^2$, is used as the sum of squares in computing variance and unbiased *SD*.

$$\bar{X} = \frac{\Sigma f X_C}{n} = \frac{1005.0}{30} = 33.5 \text{ mm.}$$

$$s^2 = \frac{\Sigma f(X_C - \bar{X})^2}{n-1} = \frac{1312}{30-1} = 45.241 \cong 45.24 \text{ mm}^2.$$

$$s = \sqrt{\frac{\Sigma f(X_C - \bar{X})^2}{n-1}} = \sqrt{\frac{1312}{30-1}} = 6.726 \cong 6.73 \text{ mm}$$

2.3.4. *Coefficient of variation*

This relative measure of dispersion expresses the standard deviation as a percentage of the mean and is, therefore, independent of the unit of the scores. You may also realize that the coefficient of variation (*CV*) makes the *SD* independent of the magnitude of its mean. These make the *CV* very suitable for comparing the variabilities of scores of two variables in two different units (e.g., winglength scores in cm and body weight scores in kg of a sample of hens), or of the same variable in two samples with vastly different magnitudes of their means (e.g., tiger tail lengths and mouse tail lengths, both in cm).

$$CV = \frac{s}{\bar{X}} \times 100.$$

***Example* 2.3.4.**

In a sample of lobsters, the mean Ca^{2+} concentrations (mg per kg of body fluids) were found to be 15.7 (*SD* 5.23) and 3.65 (*SD* 0.82) in respectively the extracellular and intracellular fluids. Compare the variabilities of Ca^{2+} concentrations in the two fluids.

Solution :

(a) Extracellular fluid :

$\bar{X}_1 = 15.7$ mg; $s_1 = 5.23$ mg.; $CV_1 = \frac{s_1}{\bar{X}_1} \times 100 = \frac{5.23}{15.7} \times 100$

$= 33.31\%$

(b) Intracellular fluid :

$\bar{X}_2 = 3.65$ mg; s_2=0.82 mg.; $CV_2 = \frac{s_2}{\bar{X}_2} \times 100 = \frac{0.82}{3.65} \times 100$

$= 22.47\%$

$\therefore \frac{CV_1}{CV_2} = \frac{33.31}{22.47} = 1.48.$

Thus, the variability of Ca^{2+} concentration in the extracellular fluid is about 1.48 times that in the intracellular fluid in the given sample.

Example **2.3.5.**

Compare the variabilities of vital capacity scores (litres) and body weight scores (kg) from the following statistics in a sample of humans.

Vital capacity (X) : mean = 5.3 L; SD = 1.45 L.

Body weight (Y) : mean = 54.5 kg; SD = 7.55 kg.

Solution :

(a) Vital capacity :

$\bar{X} = 5.3$ L; s_x = 1.45 L; $CV_x = \frac{s_x}{\bar{X}} \times 100 = \frac{1.45}{5.3} \times 100 = 27.36\%$

(b) Body weight :

$\bar{Y} = 54.5$ kg; s_Y = 7.55 kg; $CV_Y = \frac{s_y}{\bar{Y}} \times 100 = \frac{7.55}{54.5} \times 100 = 13.85\%$

$\therefore \frac{CV_x}{CV_Y} = \frac{27.36}{13.85} = 1.98.$

Thus, the variability of vital capacity scores is about 1.98 times that of body weight scores in this sample.

2.3.5. *Quartile deviation*

Quartile deviation (Q) is an absolute measure of dispersion, given by half the difference between the third and first quartiles (Q_3 and Q_1) of a frequency distribution. You may recall that Q_3 and Q_1 are the scores below which lie respectively 75% and 25% of the scores of a sample. So, you can see that Q also is in the same unit as the scores of the distribution.

$$Q = \frac{Q_3 - Q_1}{2}$$

As a measure of dispersion, Q is totally restricted to the central part of the distribution about Q_2 or median, and indicates nothing about the dispersion of scores beyond the third and first quartiles. Higher and smaller values of Q indicate respectively wider and narrower scatterings of scores in the central part, and consequent flattening and peakedness, respectively, of the distribution. This makes Q a convenient measure of the peakedness or *kurtosis* of the distribution. In a symmetric distribution of frequencies, Q gives the range covering exactly one-fourth of all the scores on each side of Q_2, and is identical with the distance between Q_1 and Q_2 in the low-value half of the distribution as also with the distance from Q_2 to Q_3 in the high-value half.

***Example* 2.3.6.**

Compute the quartile deviation for the frequency distribution of Table 2.4, using the percentiles worked out for that distribution in Example 2.2.4.

Solution :

As computed in Example 2.2.4., the required percentiles are quoted below.

P_{50} or Mdn = 33.8 mm; P_{25} or Q_1 = 28.0 mm; P_{75} or Q_3 = 38.5 mm.

$$\therefore Q = \frac{Q_3 - Q_1}{2} = \frac{38.5 - 28.0}{2} = 5.25 \text{ mm.}$$

2.3.6. *Kurtosis*

Kurtosis is the departure of the peakedness of a frequency distribution from that of a particular form of unimodal and symmetrical distribution called the *normal distribution* computed theoretically using the *Gaussian equation*. You will learn many details about the normal distribution in Module 35. It, however, suffices for the present to mention that the medium degree of peakedness of the normal distribution is called *mesokurtosis*, and is taken as a model for describing the peakedness of other distributions. *Leptokurtosis* is the peakedness of any distribution which has relatively more scores over a narrow middle range around its mean and also at its two ends, but less scores in the intervening regions, compared to the respective parts of a normal distribution having the same mean and variance. This gives the leptokurtic distributions, such as the *t* distributions of small samples, sharper, narrower and higher peaks and thicker tails than the corresponding normal distributions. *Platykurtosis* is the peakedness of any distribution which, in comparison to the normal distribution, has relatively less scores in the middle part around its mean and at its tails, but more scores in the intervening regions, so that its central part is broader and flatter and its tails thinner (Fig. 2.1.).

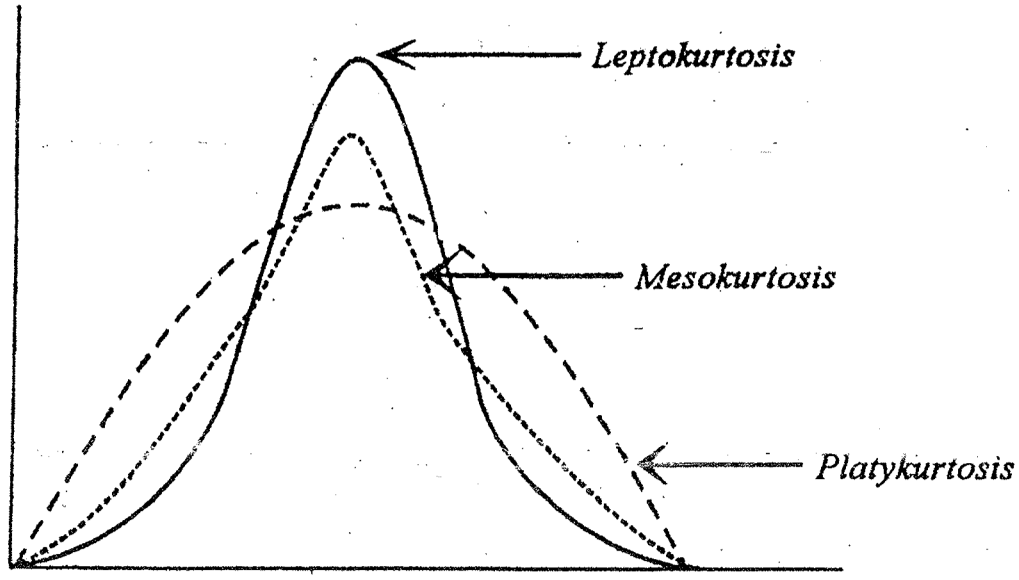


Fig. 2.1. Distributions with different forms of kurtosis.

You are aware that the quartile deviation (Q) is a measure of the range for 25% of the scores on each side of the *Mdn*. So, it can be used in working out the *percentile coefficient of kurtosis* (κ) as a measure of kurtosis. Mesokurtosis, leptokurtosis and platykurtosis are indicated if the value of κ respectively equals, falls below and exceeds 0.263. Coefficient of kurtosis is independent of the unit of raw scores.

$$\kappa = \frac{Q}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}.$$

***Example* 2.3.7.**

Work out the percentile coefficient of kurtosis for the frequency distribution of Table 2.4 (cockroach winglength data), using the following percentiles.

P_{10} = 24.3 mm. P_{25} or Q_1 = 28.0 mm.

P_{75} or Q_3 = 38.5 mm. P_{90} = 42.5 mm.

***Solution* :**

$$\kappa = \frac{Q}{P_{90} - P_{10}} = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{38.5 - 28.0}{2(42.5 - 24.3)} = 0.288.$$

Because the computed κ exceeds 0.263, the cockroach winglength distribution of Table 2.4. is slightly *platykurtic*.

2.3.7. *Skewness*

Skewness of a distribution is its departure from bilateral symmetry. A skewed distribution is an asymmetric distribution with more extreme scores at one of its ends to make that tail more drawn out (skewed) than its other tail which is shorter and thicker because of a denser concentration of the scores in that half of the distribution, closer to the central value. The distribution is called positively or negatively skewed according as whether the high-value (right) tail or the low-value (left) tail is more drawn out and pointed.

In symmetric distributions such as normal and *t* distributions, mean ($\bar{X}$), median (*Mdn*) and mode (M_0) coincide with the peak of the distribution; but in any asymmetric distribution, although M_0 is still the score at the peak, *Mdn* is deflected from it towards the skewed tail while $\bar{X}$ is deflected towards that tail even further. The higher the skewness, the greater are such deflections. While $\bar{X} < Mdn < M_0$ in negatively skewed distributions, $\bar{X} > Mdn > M_0$ in positively skewed distributions (Fig: 2.2). You may infer from these properties that the *Mdn*, being less affected by skewness than $\bar{X}$, is more dependable than the latter as a central value of asymmetric distributions, and that the difference between $\bar{X}$ and either M_0 or *Mdn* may be used to estimate the skewness. While Q_1 and Q_3 are equidistant from Q_2 in symmetric distributions, Q_1 is deflected towards the skewed tail of asymmetric distributions making (Q_2-Q_1) higher or lower than (Q_3-Q_2) according as the distribution is negatively or positively skewed, respectively.

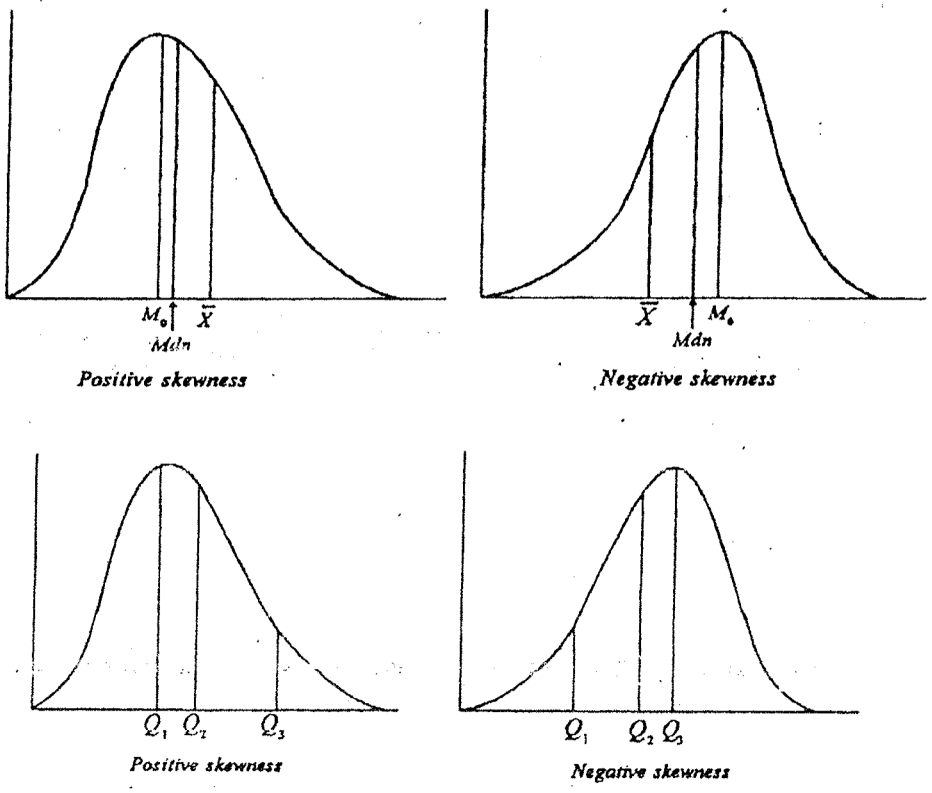


Fig. 2.2. Deflections of statistics in skewed distributions.

Skewness of a distribution is worked out in several ways. While a zero value of any of the skewness coefficients indicates a symmetric distribution, positive and negative values indicate respectively positive and negative skewness, and their magnitudes indicate the degrees of skewness. Skewness coefficients are independent of the units of raw scores.

(a) Pearson's coefficients of skewness :

$$Sk = \frac{\bar{X} - M_0}{s}; \qquad Sk = \frac{3(\bar{X} - Mdn)}{s}.$$

(b) Quartile coefficient of skewness:

$$Sk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{2Q}.$$

***Example* 2.3.8.**

Work out the coefficients of skewness of the frequency distribution of cockroach winglength scores of Table 2.4, using the following statistics.

$\bar{X} = 33.5$ mm. $M_0 = 34.5$ mm. $Mdn = 33.8$ mm.

$Q_1 = 28.0$ mm. $Q_3 = 38.5$ mm. $s = 6.73$ mm.

***Solution* :**

(a) $Sk = \frac{\bar{X} - M_0}{s} = \frac{33.5 - 34.5}{6.73} = -0.149.$

(b) $Sk = \frac{3(\bar{X} - Mdn)}{s} = \frac{3(33.5 - 33.8)}{6.73} = -0.134.$

(c) $Sk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{(38.5 - 33.8) - (33.8 - 28.0)}{38.5 - 28.0} = -0.105.$

So, the cockroach winglength distribution suffers from a negative skewness.

2.4. SAMPLING ERRORS AND THEIR MEASURES

You are aware that most scientific investigations are done with samples instead of entire populations, and that the population parameters are consequently not directly or precisely known - instead, they are estimated from the sample statistics. But the statistics of different samples, drawn from the same population, mostly differ from sample to sample. It, therefore, follows that most of these sample statistics, though used as estimates of the population parameter, miss the latter. This difference between a sample statistic and the corresponding population parameter is called the *sampling error* of that statistic. Because of different sampling errors of the statistics of different samples from the same population, the sample statistics remain dispersed around their population parameter in the form of a frequency distribution called the *sampling distribution* of the relevant statistic in the corresponding population. The sampling errors of a statistic are estimated by the *standard error* of the latter. You will learn about these in this section.

2.4.1. *Sampling errors*

The sampling error of a statistic is its difference from the corresponding parameter, for which it should serve as an estimate. The sampling error may be a positive or negative number or, in some cases, even zero according as the statistic is higher than, lower than or equal to the parameter. The sampling error varies from sample to sample, drawn from the same population, because random sampling chooses different individuals or cases for different samples, error in testing and measurement varies from sample to sample, and even if the same individual is included in two consecutive samples, his scores may differ in the two cases because of temporal and other changes. Even though a sample statistic differs from sample to sample, the mean of such a statistic of an infinite number of unbiased and representative samples from a population should coincide with the corresponding parameter; any difference between such a mean of sample statistics and the population parameter indicates *biased* samples.

Because sample statistics suffer from sampling errors, a difference between the statistics of two samples, e.g., the difference between their means $(\bar{X}_1 - \bar{X}_2)$, may suffer from sampling error to

make it differ from the difference between the corresponding parameters such as the difference between the population means $(\mu_1 - \mu_2)$. Such a sampling error is called *the sampling error of the difference between the statistics*. Even if two samples are from the same population, the difference between their statistics (e.g., $\bar{X}_1 - \bar{X}_2$) may still suffer from sampling error to make that difference differ from 0, though their parameters (e.g., μ_1 and μ_2) are identical here and have a zero difference only. You will understand the significance and application of these concepts as we proceed further with the subject.

In the absence of precise knowledge about the population parameter in most cases, we can only work out an estimate of the sampling error in the form of the *standard error* (*SE*).

2.4.2. *Sampling distributions*

You should be aware by now that most of the samples drawn from a population differ from each other with respect to any statistic (e.g., the mean) because of different magnitudes and algebraic signs of the sampling errors of their statistics. For example, some samples have their means higher than the parametric mean (μ) by different amounts, some of the sample means ($\bar{X}$) may coincide with μ, while the rest of the samples have their means lower than μ by different amounts. Because of such differences in sampling errors from sample to sample, the sample statistics such as sample means and sample standard deviations would be distributed in the form of frequency distributions around their respective parameters. Such a distribution of sample statistics is called its *sampling distribution*, e.g., the sampling distributions of means and of standard deviations. Such sampling distributions may also be worked out theoretically with the help of laws of probability.

You have also been told about how the differences between sample statistics, e.g., differences between sample means $(\bar{X}_1 - \bar{X}_2)$, suffer from sampling errors. Variations in the magnitude and algebraic signs of such sampling errors from pair to pair of samples cause the differences between those statistics, e.g., $\bar{X}_1 - \bar{X}_2$, to be distributed in the form of a frequency distribution - around 0 if the samples belong to the same population, or around the difference between the respective parameters, e.g., $\mu_1 - \mu_2$, if the samples of each pair come from two different populations. Such a distribution, worked out either experimentally or theoretically, is called a *sampling distribution of*

differences between statistics; e.g., sampling distributions of $(\bar{X}_1 - \bar{X}_2)$ and of $(s_1 - s_2)$.

The dispersions of statistics, or of their differences, in the respective sampling distributions may be estimated by *standard errors.*

2.4.3. *Standard errors*

Standard errors (*SE*) belong to the class of sampling statistics. They are used in estimating sampling errors of different types of statistics and of the differences between them, such as the standard errors of means $(s_{\bar{x}})$, standard deviations (s_s), proportions (s_p), quartile deviations (s_Q), differences between means $(s_{\bar{X}_1 - \bar{X}_2})$ and differences between proportions $(s_{p_1 - p_2})$. Their units are the same as those of the corresponding statistics. Just as standard deviation is a measure of dispersions of raw scores in a frequency distribution of the latter about their sample mean, the standard error is a measure of dispersions of a particular sample statistic in a frequency distribution of the latter about the corresponding parameter. Just as standard deviation estimates the average *error term* $(X - \bar{X})$ which is the average deviation of scores of a sample from their mean, the standard error estimates the average *sampling error*, i.e., the average deviation of a statistic from its population parameter. Conventionally, however, the term 'standard deviation' without any qualification is used to mean the measure of dispersion of scores about their mean while the term standard error is used for the measure of dispersion of any statistic about the corresponding parameter; because the *SE* of means is the most frequently used standard error, the term 'standard error' without any qualification is conventionally used to mean the of *SE* of means.

You should not apprehend that for working out an *SE* of any statistic, you would need to draw samples repeatedly from a population and to work out the values of that statistic in the respective samples. On the contrary, the *SE* of any statistic can be worked out from a *single sample* as such an *estimated SE* of that statistic that could be expected if a large number of samples were drawn to work it out from their statistics for the relevant population. However, the *SE* decreases with the rise in sample size.

Standard errors find applications in finding the dependability of the respective statistics, in

working out such confidence intervals from a sample statistic which has a given probability of including the corresponding parameter, in finding the probability of occurrence of the differences between the statistics of two samples by mere chances of random sampling, in estimating whether or not the experimental results are significant for making inferences, and also in estimating the variability of errors from sample to sample in measuring the given variable.

2.4.4. *SE of the mean*

The *SE* of the mean ($s_{\bar{X}}$), often referred to as the 'standard error' (*SE*) only, is the sampling statistic measuring the dispersions of means of different samples, drawn from a population, in their frequency distribution about the population mean. In other words, it is the standard deviation of the sampling distribution of means, estimating their average sampling error. It has the same unit as that of the sample means and so, of the raw scores of the samples. You have already been told that the *estimated SE* ($s_{\bar{X}}$), computed from a single sample, may be used in working out confidence intervals of the population mean (μ) from the sample mean ($\bar{X}$), in estimating the variability of errors from sample to sample in measuring the given variable, and in finding the significance of experimental results. It should always be worked out before the sample mean can be put to use for any such work. Basically, $s_{\bar{X}}$ is computed using the sample *SD* and the sample size (n), but the computational formulae vary with the nature of the population, finite or infinite, and the method of sampling, viz., simple random, stratified random, incidental, with replacement and without replacement. The estimated $SE(s_{\bar{X}})$ is taken to be an estimate of the *true SE* of the population ($\sigma_{\bar{X}}$) which may be computed from the theoretically computed *SD* (σ). For a population with its size (N) *infinitely large* ($N = \infty$),

$s_{\bar{X}} = \frac{s}{\sqrt{n}}$, if s is the *unbiased SD*, and

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$, using theoretically computed *SD*.

But for a *finite population* ($N < \infty$) and *sampling without replacement*,

$$s_{\bar{X}} = \frac{s\sqrt{N-n}}{\sqrt{n(N-1)}}; \qquad \sigma_{\bar{X}} = \frac{\sigma\sqrt{N-n}}{\sqrt{n(N-1)}}.$$

If the *SD* is not the unbiased one and has been computed with n instead of $(n-1)$, the formula for $s_{\bar{X}}$ should always be modified by using $(n-1)$ instead of n in the denominator. Thus, in such cases,

$s_{\bar{X}} = \frac{s}{\sqrt{n-1}}$, where $N = \infty$, and

$s_{\bar{X}} = \frac{s\sqrt{N-n}}{\sqrt{(n-1)(N-1)}}$, where $N < \infty$, and sampling is done without replacement.

The *SE of difference between means* $\left(s_{\bar{X}_1-\bar{X}_2}\right)$ is a measure of the average sampling error of the differences between means of each pair of samples drawn either from the same population or from two different populations. In other words, it is the standard deviation of the sampling distribution of such differences about their *mean difference* which is either the difference $(\mu_1 - \mu_2)$ between the parametric means if the samples come from separate populations, or zero if the samples belong to the same population $(\mu_1 - \mu_2 = 0)$. It thus estimates the dispersions of the differences between the sample means from the mean difference. The *SE* of difference between means finds important applications in testing experimental hypothesis from the difference between the means of two samples under different levels of experimental treatment. You will learn the details about such uses in a subsequent module. The *SE* of difference between means is in the same unit as the corresponding sample means and raw scores.

While the *SE* of the difference between two means ($\bar{X}_1$ and $\bar{X}_2$) is worked out experimentally using the standard errors of the two sample means, viz., $s_{\bar{X}_1}$ and $s_{\bar{X}_2}$, the *true or parametric SE of the difference*, $\sigma_{\bar{X}_1-\bar{X}_2}$, may be computed theoretically from the true standard errors ($\sigma_{\bar{X}_1}$ and $\sigma_{\bar{X}_2}$). Where s_1 and s_2 are the unbiased standard deviations of the respective samples of sizes n_1 and n_2,

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}}; \qquad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}};$$

$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(s_{\bar{X}_1}\right)^2 + \left(s_{\bar{X}_2}\right)^2}$, and similarly,

$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(\sigma_{\bar{X}_1}\right)^2 + \left(\sigma_{\bar{X}_2}\right)^2}$, as the *true SE* of difference.

Example **2.4.1.**

Work out the *SE* of difference between the mean Ca^{2+} concentrations (mg/kg of body fluids) in the extracellular fluid of two samples of lobsters, using the following sampling statistics.

Sample 1 : mean $\left(\bar{X}_1\right) = 15.7$ mg; unbiased *SD* $(s_1) = 4.73$ mg;
sample size $(n_1) = 25$.

Sample 2 : mean $\left(\bar{X}_2\right) = 17.8$ mg; unbiased *SD* $(s_2) = 5.45$ mg;
sample size $(n_2) = 36$.

Solution :

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{4.73}{\sqrt{25}} = 0.946\,\text{mg}.$$

$$s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{5.45}{\sqrt{36}} = 0.908\,\text{mg}.$$

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(s_{\bar{X}_1}\right)^2 + \left(s_{\bar{X}_2}\right)^2} = \sqrt{(0.946)^2 + (0.908)^2} = 1.311\,\text{mg}.$$

2.4.5. *Degrees of freedom*

You may have already noticed that in computing some sample statistics such as variance and standard deviation, one or more statistics computed earlier for the relevant sample have to be used as estimates of the respective population parameters. For example, while computing the standard deviation of a sample, the sample mean computed for the same scores have to be used as an estimate of the population mean. Because a parameter should not change, the sample statistic used as its estimate must also be kept unaltered during such work. This restricts the freedom of the scores of a sample for changes so that each sample statistic used as an estimate of a parameter does

not get altered due to changes of the scores. The *degree of freedom* (*df*) of a statistic is an expression of this restriction imposed on the changes of scores of the sample during its computation using another statistic computed earlier. It may be defined as that number of scores of a sample, the changes of which in magnitude as well as algebraic sign would not affect the values of the statistics being used as estimates of parameters in the computation of the statistic under reference.

In the case of the computation of *SD* using the sample mean, except any one of the *n* number of sample scores, all other scores may be freely altered in magnitude and algebraic sign; but the remaining one score, whichever it may be, is no longer free to change in all possible ways - it has to change only to such a value as would counteract and balance the changes of all other scores if the mean has to remain unchanged. So, the degrees of freedom or the number of scores free to change amount here to $(n - 1)$.

The *SE* of difference between two sample means, on the contrary, is computed using the two sample standard deviations, each of which is in turn computed using its sample mean as an estimate of the corresponding parametric mean. Thus, to keep the mean of each of the two samples unaltered, one score of each sample loses its freedom for change. This makes the *df* of $s_{\bar{X}_1-\bar{X}_2}$ amount to : $(n_1-1)+(n_2-1) = n_1+n_2-2$, where n_1 and n_2 are the total numbers of scores of the respective samples.

Thus, one *df* is lost for each precomputed statistic being used in the computation as an estimate of a parameter. If k number of precomputed statistics are so used while computing a statistic for a sample of size n, the *df* of the statistic being computed would be given by : $df = n-k$.

2.5. SUMMARY

Descriptive statistics have been classified with examples of each class. Properties and computations of mean, median and mode, as measures of central tendency, have been described. Percentiles and quartiles have been defined, their computations have been mentioned, and their uses in computing some measures of dispersion have also been described. Measures of dispersion have been classified into absolute and relative measures. Properties, computations and uses of range, variance, standard deviation and quartile deviation as absolute measures of dispersion and of coefficient of variation as a relative measure have been discussed.

Kurtosis and skewness of frequency distributions have been explained and their measures have been desqribed.

Sampling errors of sample statistics and sampling distributions of the latter, due to sampling errors, have been explained. Standard errors of statistics for measuring their sampling errors have been described with particular reference to the sample means and their differences.

Degrees of freedom of statistics have been explained with examples.

2.6. GLOSSARY

Central tendency : Tendency of scores to gather around a central point on their scale. Measures of central tendency are the central values.

Central value : A statistic of location showing the location of a central score of a frequency distribution on the scale of its scores.

Coefficient of variation (*CV*) : The standard deviation expressed as a percentage of the mean.

Degrees of freedom (*df*) : Number of scores in a sample or of classes in a distribution that can be changed without altering already computed statistics used as estimates of parameters in the computation of a given statistic.

Dispersion : Scattering of scores around a central value.

Kurtosis : Peakedness of a distribution compared to that of a normal distribution .

Mean ($\bar{X}$) : The arithmetic average of the scores of the sample.

Mean square (*MS*) : Same as variance.

Measures of dispersion : Statistics such as variance and standard deviation, measuring the dispersions of scores around a central value.

Median (*Mdn*) : The score below which lies exactly 50% of the all the scores of the sample.

Mode (M_0) : The score occurring most frequently in the sample.

Percentile : Each percentile is that score below which lies a given percentage of all the scores of a sample.

Quartile : Each quartile is that score below which occurs a given number of quarters of all

the scores.

Quartile deviation (Q) : Half of the difference between the third and first quartiles.

Range : The interval covering all the scores of the sample from the lowest to the highest score.

Sampling distribution : Frequency distribution of a statistic of a number of samples drawn from the same population, around the population parameter.

Sampling error : The difference between a sample statistic and the corresponding population parameter.

Skewness : Bilateral asymmetry of a distribution with one of its tails more drawn out than the other.

Standard deviation (SD or s) : The square root of the arithmetic average of the squared deviatons of scores from the mean.

Standard error (SE) : The measure of the sampling error of a statistic.

Statistics of dispersion : Same as measures of dispersion.

Statistics of location : Statistics such as mean and median which show the location of a given point of the frequency distribution on the scale of its scores.

Sum of squares (SS) : The sum of the squared deviations of scores from the mean.

Variability : Tendency of scores for dispersion around a central value of the sample.

Variance (s^2) : The arithmetic average of the squared deviations of scores from the mean.

2.7. QUESTIONS

1. (a) Classify descriptive statistics into its major classes and their respective subclasses, defining and giving examples at each stage.

   (b) Define and explain the degrees of freedom of a statistic, citing examples.

2. (a) Define skewness and kurtosis. Explain how they affect a frequency distribution and its properties. Give the computational formulae for measuring skewness and kurtosis of a distribution.

(b) Work out the skewness and kurtosis of the frequency distribution of interorbital widths (in mm) in a sample of doves using the following statistics of the sample.

P_{10} = 12.25 mm. P_{50} = 14.88 mm. P_{75} = 15.81 mm.

P_{90} = 16.38 mm. P_{25} = 13.45 mm.

3. (a) Give the definitions, properties, computational formulae and uses of an absolute measure of dispersion, and of a relative measure of dispersion derived from it.

(b) Compute the *SE* of difference between mean blood sugars (in mg per dl of blood) of two samples of Bengal goats, using the following statistics. Also mention the degrees of freedom of the computed statistic and of the statistics involved in its computation, stating the reasons for supporting your conclusions.

Sample 1: Mean = 75.2 mg. *SD* = 13.90 mg. Sample size : 36.

Sample 2 : Mean = 80.5 mg. *SD* = 15.23 mg. Sample size : 49.

4. (a) Define median and mode and describe their properties. How are they affected by more extreme scores in one tail of the distribution than in the other?

(b) Compute the mode, median, quartile deviation and skewness of the following distribution of Ca^{2+} concentrations (in mg/ kg of body fluid) in the extracullular fluid of a sample of marine crabs.

Class intervals :	8.5-11.5	11.5-14.5	14.5-17.5	17.5-20.5	20.5-23.5
Frequencies :	4	10	16	7	3

5. (a) What do you mean by sampling errors of a statistic? Discuss how the sampling errors give rise to sampling distributions of a statistic.

(b) Explain what is meant by the *SE* of a statistic and its uses. Write about the *SE* of the mean and the *SE* of differences between means, mentioning their computational formulae.

(c) Compute the *SE* of difference between the mean wing-lengths (mm) of two samples of locusts using the following statistics, and work out its degrees of freedom.

Sample 1 : Mean = 35.7 m. *SD* = 10.25 m. Sample size : 30.

Sample 2 : Mean = 32.8 m. *SD* = 7.88 m. Sample size : 28.

6. (a) Describe the properties of mean and how it can be computed from ungrouped raw scores and grouped scores of a frequency distribution, respectively.

(b) How is the mean used in computing the standard deviation, the coefficient of variation, the variance and the skewness? What do you mean by the sum of squares? Mention one of its uses.

7. (a) Compute the mean, *SD*, coefficient of variation and *SE* of the mean of the following winglength (mm) data of a sample of cockroaches.

Class intervals (in score limits)	18-22	23-27	28-32	33-37	38-42
Frequencies :	6	8	15	10	6

(b) Compute the mean, median, skewness and *SE* of the mean of the following femur length scores ($mm \times 10^{-2}$) of a sample of aphids.

29, 33, 40, 41, 39, 33, 28, 30, 33, 30, 28, 31, 35, 40, 39.

BIBLIOGRAPHY FOR FURTHER READING

1. Das, D. and Das, A., *Statistics in Biology and Psychology,* 4th edition, Academic Publishers, Kolkata.

2. Sokal, R.R. and Rohlf, F.J., *Introduction to Biostatistics,* 2nd edition, W.H. Freeman & Co., San Francisco.

3. Sokal, R.R. and Rohlf, F.J., *Biometry*, W.H. Freeman & Co., San Francisco.

- - - - -

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-B Unit-I

Module No. - 35

UNIT 1 : PROBABILITY DISTRIBUTIONS

1.0. OBJECTIVES

It is intended that after going through this unit, you should be able to do the following :

* understand and explain probability and probability distributions,
* understand and describe binomial probability distributions, their properties, assumptions and uses in biostatistics,
* understand and describe Poisson probability distributions for rare events, their assumptions, properties and uses in biostatistics,
* define normal probability distributions and unit normal curve, and give their equations,
* describe the properties of normal probability distribution and mention its applications,
* define and explain standard scores such as z and t scores, and their distributions as well as uses,
* define and explain Student's t distributions for small samples, drawn from a normally distributed population, and give their equations, and
* describe the properties and uses of t distributions.

1.1. PROBABILITY

You should realize that right from the drawing of unbiased samples from a population for scientific investigations, planning experimental designs, and even upto the interpretation of significance of the observed results of an experiment, statistical methods depend on probabilities and the laws of probability. *Probability (P)* is the expected relative frequency (f/n) of any score, phenomenon or event when the total number (n) of all scores, phenomena or events is so large as to approach infinity : $P = f/n$. For example, if a sample of 15000 insects of a particular species includes 150 insects with winglength score of 6.7 mm, the probability of occurrence

of that winglength score may amount to :

$$P = \frac{f}{n} = \frac{150}{15000} = 0.01.$$

However, the smaller is the total number (n) of cases or observations under consideration, the wider the difference between the relative frequency of the given score or event in it and the probability or *limiting* relative frequency of such score or event in an infinitely large number of cases or observations.

1.2. PROBABILITY DISTRIBUTIONS

In Unit 1 of Module 34, you have already learnt about the frequency distribution of scores in a sample. If you plot the relative frequencies (f/n) or *probabilities* of the scores of a very large sample ($n \cong \infty$) against the respective scores, you get the probability distribution of those scores. In Table 1.1, for example, the grouped frequency distribution of winglength scores in a large sample of cockroaches is converted to a probability distribution of those scores.

Table 1.1. A probability distribution of cockroach winglength scores (mm) using their frequency distribution in a large sample ($n = 280$).

Class intervals (mm)	f	$P = f/n$
17-20	17	0.061
21-24	54	0.193
25-28	139	0.496
29-32	48	0.171
33-36	22	0.079
Total	**280(n)**	**1.000**

As has been cited in Table 1.1, the observed frequencies of scores of a variable obtained from a large sample may be used in forming an *experimental probability distribution* of such scores. On the other hand, you may also work out a *theoretical probability distribution* using the laws of probability and specific models of distribution. In both cases, you may present the probability distribution graphically by plotting the probability of each score against that score -- for this, the probabilities and the scores are scaled along the Y- and X-axes, respectively. You will shortly learn about two probability distributions, viz., binomial distribution and Poisson distribution, which may be theoretically computed using the binomial equation and the Poisson equation, respectively. They are also *discrete probability distributions* because their X- axis scales for the whole numbers of cases or events have genuine gaps. Two other theoretical probability distributions to be described shortly, viz., normal distribution and t distribution, are based on the Gaussian equation and Gosset's

equation, respectively; but these two are *continuous probability distributions* having no genuine gap in the X-axis scale for the scores of a continuous measurement variable.

1.3. BINOMIAL DISTRIBUTION

With respect to some variables, a population may be divided into only two classes. Such variables are called *dichotomous variables.* Sex (male-female), pregnant-nonpregnant, HIV positive-HIV negative, Rh positive-Rh negative and living-dead are examples of such variables. In samples, each having a total of n number of cases or animals and drawn from a population with p and q proportions of cases in the two classes of the dichotomous variable, probabilities of random occurrences of various combinations of the two types of cases or animals are given by the successive terms of the following *binomial expansion.*

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1\times 2}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{1\times 2\times 3}p^{n-3}q^3 + \frac{n(n-1)(n-2)(n-3)}{1\times 2\times 3\times 4}p^{n-4}q^4 + \ldots\ldots$$

$$+\frac{n(n-1)(n-2)(n-3)\ldots\times 3\times 2}{1\times 2\times 3\times 4\times \ldots\ldots\ldots (n-1)}pq^{n-1} + q^n.$$

In a sample of size n, the probability (P) of random occurrence of a given X number of cases of any of the two classes is computed by working out only that particular term of the binomial expansion which has X as the power of the proportion (p) of that class, and the number (n–X) of the cases of the other class as the power of the proportion (q) of the latter. The given X number of cases has occurred due to random sampling so long as the computed P exceeds a chosen level of significance (α) - see Section 2.3.

Probabilities of various combinations of the two types of cases, occurring in the samples on repeated trials, form the *binomial probability distribution.*

Properties of the binomial probability distribution are as follows :

(*a*) Each event or animal of either class occurs *at random* in the sample. In other words, its occurrence is not influenced by either the occurrence or the absence of any other event or animal in the sample.

(*b*) Because the events or animals must occur in whole numbers only and never in fractional numbers, there are genuine gaps between the numbers of events or animals occurring in the sample. So, it is a *discrete probability distribution.*

(*c*) The *theoretical* binomial distribution can be worked out using the binomial expansion.

(*d*) It is a probability distribution of the events of a *dichotomous variable* with neither of its classes having a very low or near-zero proportion in the population.

(*e*) The distribution for the events of a given class, having the proportion p, has the following mean (μ) and

SD (σ) : $\mu = np$; $\sigma = \sqrt{npq}$, where q is the proportion of the other class.

(*f*) The distribution of a given class with proportion p is *skewed positively* if p is less than 0.50, and *negatively* if p exceeds 0.50, but has *no skewness* if p amounts to 0.50.

(*g*) If the proportion p of a given class falls outside the range 0.21-0.79 in the population, its distribution is *leptokurtic*; otherwise its distribution is *platykurtic*.

Some of these properties are evident in a *bar diagram* representing the binomial probability distribution of one of the two classes, having a proportion p in the population (Fig. 1.1). The ordinate and abscissa are scaled in the probabilities and the number of events respectively, with the bars having bases of equal widths and small intervening gaps separating the consecutive bars.

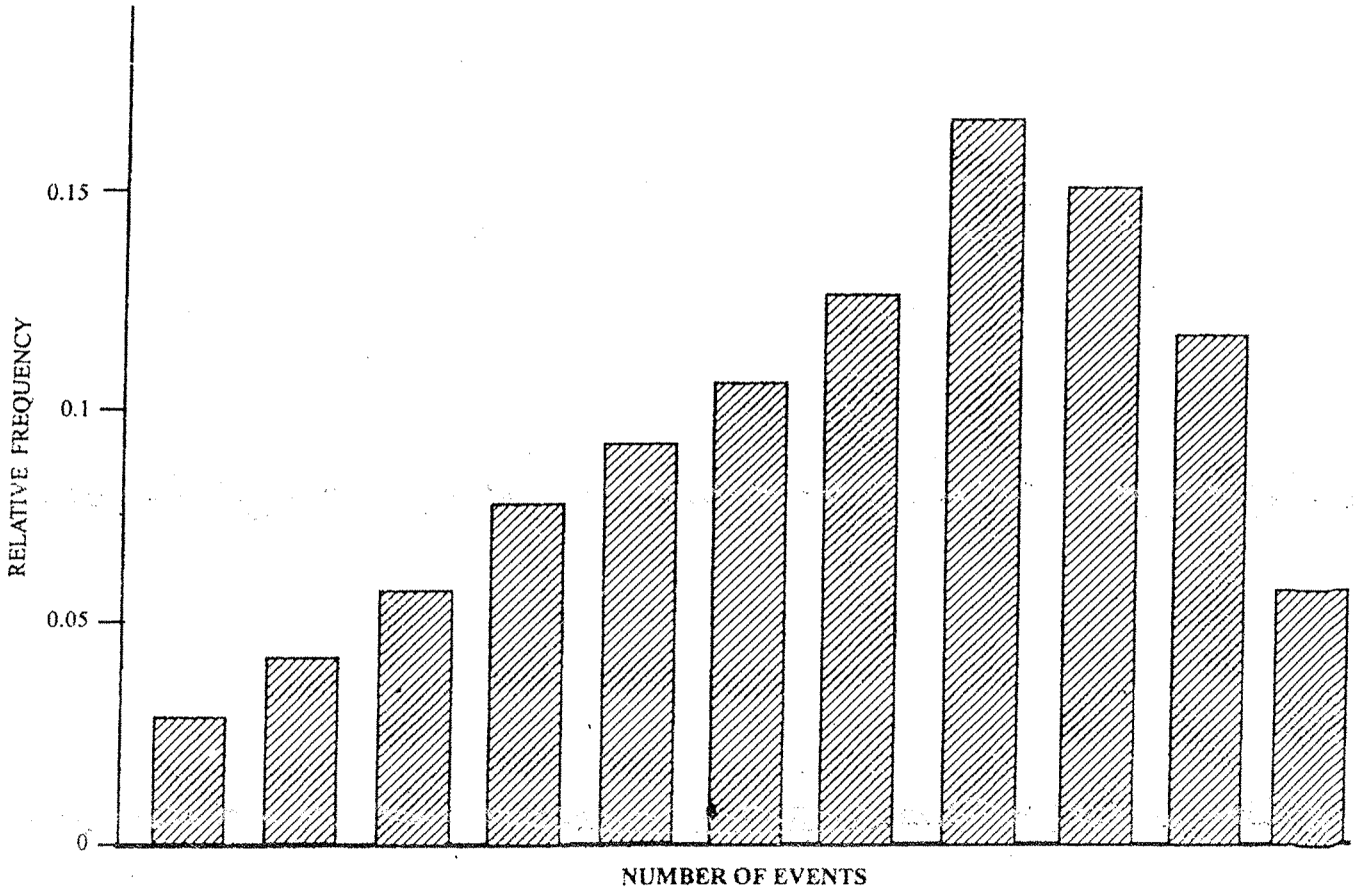


Fig. 1.1 : Bar diagram of the binomial probability distribution of one of the two classes. p = 0.60; n = 20.

The following *assumptions* should be justifiable if the binomial distribution is to be used for finding the probability of random occurrence of a given combination of cases of two classes of a variable in a sample:

(*a*) There should be two classes in the population, corresponding to the classes of the dichotomous variable.

(*b*) The proportion of the population in either of the classes should be precisely known and should not be too low or near-zero.

(*c*) Each event or case should occur at random in the sample and free from any influence of the other events or cases.

(*d*) The mean, *SD* and coefficient of dispersion (*CD*) of the distribution of one of the classes, which has the proportion *p* in the population, should be given by the following:

$$\mu = np;\ \sigma = \sqrt{npq}\ ;\ CD = \frac{\sigma^2}{\mu} = q\ .$$

The binomial distribution is often used in Zoology to work out the probability of random occurrence of the sex-ratio found in a sample of animals drawn from a population with known sex-ratio.

Example 1.3.1.

What is the probability of random occurrence of 5 male animals in a sample of 10 animals from a population with the known male-female ratio of 11:9 ? Interpret your result. (α=0.05).

Solution :

Proportions in population : (a) male animals $(p) = \frac{11}{11+9} = 0.55$;

(b) female animals $(q) = 1 - p = 1 - 0.55 = 0.45$.

Combination of animals of two classes :

(a) male animals : 5;

(b) sample size (n) : 10;

(c) female animals : $n - 5 = 10 - 5 = 5$.

Probabilities of various combinations of males and females are given by the following binomial expansion.

$$(p+q)^n = (0.55+0.45)^{10}$$

$$= (0.55)^{10} + 10\times(0.55)^9\times 0.45 + \frac{10\times 9}{1\times 2}\times(0.55)^8\times(0.45)^2 + \frac{10\times 9\times 8}{1\times 2\times 3}\times(0.55)^7\times(0.45)^3$$

$$+\frac{10\times 9\times 8\times 7}{1\times 2\times 3\times 4}\times(0.55)^6\times(0.45)^4 + \frac{10\times 9\times 8\times 7\times 6}{1\times 2\times 3\times 4\times 5}\times(0.55)^5\times(0.45)^5 + \ldots\ldots$$

$$+\frac{10\times 9\times 8\times 7\times 6\ldots\times 3\times 2}{1\times 2\times 3\times\ldots\times 9}\times 0.55\times(0.45)^9 + (0.45)^{10}.$$

So, the probability of occurrence of 5 males, i.e., $P(5)$, or that of the combination of 5 males and 5 females, is given by the sixth term of the expansion with both p and q raised to the power 5.

$$P(5) = \frac{10\times 9\times 8\times 7\times 6}{1\times 2\times 3\times 4\times 5}\times(0.55)^5\times(0.45)^5$$

$$= 0.234.$$

$\alpha = 0.05.$

Because the computed P of 0.234 exceeds the chosen α of 0.05, it is inferred that the given number of 5 males occurred in the sample due merely to random sampling ($P>0.05$).

1.4. POISSON DISTRIBUTION

This is the probability distribution of animals or cases belonging to the rare class of a dichotomous variable. As in case of the binomial distribution, here also the population is divided into two classes corresponding to the two classes of the dichotomous variable. But here the proportion of the population in one of the classes is very low, nearing 0.00, while that in the other class is very high and close to almost 1.00. In samples, each having a total of n number of cases or animals and drawn from a population with p proportion of cases in the rare class, probabilities of random occurrences of various numbers (X) of animals or events of only the rare class -- not of the other and larger class - are given by the series of terms of the *Poisson distribution* on the basis of the mean (μ) of the distribution, the base (e) of natural logarithm, and the respective X values such as 0, 1, 2, 3, ...n, etc. Thus, μ of the distribution and $P(X)$ or probability of random occurrence of X number of rare events are given by :

$$\mu = np;\ P(X) = \frac{\mu^X}{X!e^{\mu}}$$

So, for 0, 1, 2, 3, ... or n number of rare events, the Poisson probabilities are given respectively by the following terms.

$$\frac{\mu^0}{e^{\mu}} \text{ or } \frac{1}{e^{\mu}}; \frac{\mu}{1!e^{\mu}}; \frac{\mu^2}{2!e^{\mu}}; \frac{\mu^3}{3!e^{\mu}}; \frac{\mu^4}{4!e^{\mu}}; \ldots\ldots \frac{\mu^n}{n!e^{\mu}}.$$

Properties of the Poisson distribution are summarized below.

(*a*) It is the probability distribution of events or animals belonging to the *rare class* of a dichotomous variable, not of the cases belonging to the larger class.

(*b*) Because the events or animals of the rare class must occur in whole numbers only and never in fractional numbers, there are genuine gaps between those numbers occurring in the sample. So, it is a *discrete probability distribution.*

(*c*) Each rare event or animal occurs *at random* in the sample, without being influenced by the occurrence or non-occurrence of other rare cases.

(*d*) The proportion (p) of the population in the rare class is very low, almost near zero, while the mean (μ) and the variance (σ^2) of the distribution of the rare events are identical in value which does not exceed 5.

$\mu = \sigma^2 = np.$

(*e*) Poisson distributions are progressively more *leptokurtic* with the fall in value of μ.

(*f*) Poisson distributions are progressively more *positively skewed* with the fall in value of μ. If μ is too low, the distribution loses its negative tail altogether.

The following *assumptions* should be justifiable if the Poisson distribution is to be used for finding the probability of random occurrence of a given number of events of the rare class in a sample.

(*a*) The proportion of the population in the rare class is very low, almost approaching zero.

(*b*) Each rare event occurs at random in the sample and is not influenced by other such events.

(*c*) The mean and variance of the distribution of rare events are identical and do not exceed 5.

Example 1.4.1.

Work out the probability of random occurrence of 4 animals carrying a particular mutant gene in a sample of 325 animals, drawn from a population known to have 46 such mutant animals per 10000.

Solution :

Where p is the proportion of mutant animals in the population, n is the sample size, X is the number of mutant animals in the sample, μ is the mean of the Poisson distribution of mutant animals, and e is the base of natural logarithm and amounts to 2.7183.

$$p = \frac{46}{10000} = 0.0046;\ n = 325;\ X = 4;$$

$$\mu = np = 325 \times 0.0046 = 1.495;$$

$$P(X) = \frac{\mu^X}{X! e^{\mu}}$$

$$\text{or, } P(4) = \frac{(1.495)^4}{4!(2.7183)^{1.495}} = 0.047.$$

1.5. STANDARD SCORES

For the sake of comparability and better interpretation, raw scores of an experiment, their differences, and their deviations from their means are often transformed into some scores having standard reference values. Such transformed scores are called *standard scores* such as z scores, T scores and C scores. Of them z scores are examples of *linearly transformed scores*, resulting from the changes in unit, mean, *SD* and zero points of the scale of original raw scores, but retaining the relative magnitudes of the original scores as well as the form and properties of their distributions. But T scores and C scores are examples of *nonlinearly transformed scores* which differ from original raw scores in unit, mean, *SD* as well as the form and properties of the distribution.

The *normal probability distribution*, about which you will read in the very next section, is actually the probability distribution of standard z scores, obtained by the *linear transformations* of raw scores of many continuous measurement variables. Each raw score may be transformed into the z score by dividing its deviation from the sample mean by the *SD*; each statistic (e.g., the sample mean) may be transformed into the z score by dividing its deviation from the corresponding parameter (e.g., the population mean) by the *SE* of that statistic (i.e., the *SD* of its sampling distribution); each difference between two sample means

may be similarly converted to the z score by dividing that difference by the *SE* of that difference. So, the z scores are in *SD* units (σ units).

$$\text{Thus, } z = \frac{X - \bar{X}}{s}; \quad z = \frac{\bar{X} - \mu}{s_{\bar{X}}}; \quad z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

The *Student's t distribution*, to be described in a subsequent section of this Module, is actually the probability distribution of t scores which are also linearly transformed scores for use as standard scores for *small samples* (n<30). The t scores, also given in σ units, are computed identically as the z scores, from the raw scores, their statistics like the sample mean ($\bar{X}$), and the difference between two sample means, when the sample has a small size (n<30). For small samples,

$$t = \frac{X - \bar{X}}{s}; \quad t = \frac{\bar{X} - \mu}{s_{\bar{X}}}; \quad t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}.$$

1.6. NORMAL DISTRIBUTIONS

Normal distributions are *theoretical probability distributions* of the scores of many *continuous variables* in a population or in a large sample (n>30) from such a population, provided those scores depend upon the independent effects of almost an infinite number of other variables which are independent of each other. Basically, they are probability distributions of the standard z scores obtained by *linear transformations* of raw scores, their sample means, and the differences between two such sample means. The probability of random occurrence of each such z score is given by the corresponding ordinate Y of the normal distribution (Fig. 1.2) in terms of the following *Gaussian equation,* provided the raw scores and so, their respective z scores are normally distributed in the population or in a large sample drawn from such a population.

$$Y = \frac{n}{\sigma\sqrt{2\pi}} e^{-z^2/2}.$$

By working out the Y scores for different z scores, scaling them along the Y- and X- axis respectively, and plotting the Y scores against the respective z scores, a unimodal, bilaterally symmetrical bell-shaped *normal distribution curve* is graphically drawn, its precise form and size varying with the sixe (n), mean, *SD* (σ) and the class interval length (i) of the sample. To avoid such variations in normal curves from sample to sample, a *unit normal curve* is plotted as a standard reference curve, using the following form of the Gaussian equation where n, σ and i are all taken as 1.00 only (Fig. 1.2).

$$Y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

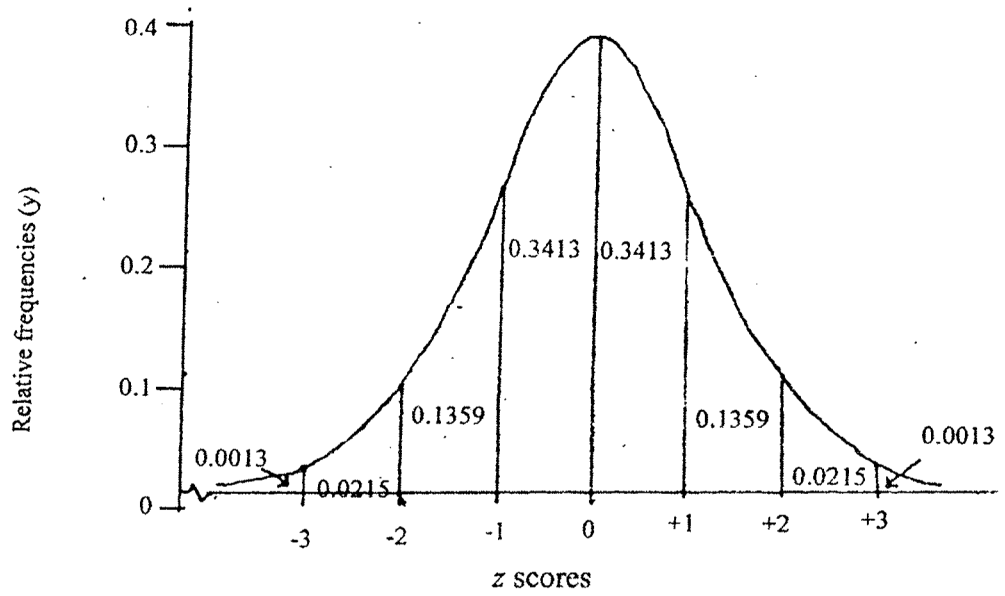


Fig. 1.2. : A unit normal curve showing its fractional areas between different ordinates.

Properties of the normal probability distribution (the unit normal curve) are summarized below.

(*a*) It is computed as a *theoretical probability distribution*, using the Gaussian equation cited above.

(*b*) Because the scores of only continuous measurement variables are transformed into *z* scores for the normal curve, there is no genuine gap between the *z* scores on the X axis. So, it is a *continuous probability distribution.*

(*c*) It is perfectly symmetrical bilaterally and possesses *no skewness*.

(*d*) It is a *mesokurtic* distribution with a medium degree of kurtosis -- neither too peaked, nor too flat.

(*e*) It is a *unimodal* distribution with its mean, median and mode identical with each other and corresponding to the highest ordinate at the centre. The *z* score for each of them amounts to 0 while the *SD* of the unit normal curve is 1.00.

$$z = \frac{\mu - \mu}{\sigma} = 0.$$

(*f*) It is an *asymptotic* distribution with its two tails extending to -∞ and +∞, respectively.

(*g*) Its *highest ordinate* corresponds to the central *z* score of 0.00 and measures 0.3989 for a unit normal curve.

(*h*) The *area* under the normal curve corresponds to the sample size (*n*) and is taken as 1.0000 for the unit normal curve. Because of its perfect bilateral symmetry, each half of the normal curve, on either side of the highest ordinate at the *z* score of 0.00, covers a fractional area corresponding to $\frac{1}{2}n$ and is taken as 0.5000 for a unit normal curve. The bilateral symmetry also ensures that the fractional area between two given *z* scores in the positive half of the normal curve equals the fractional area between the corresponding *z* scores in the negative half. This is shown in Fig. 1.2.

(*i*) The probability of random occurrence, in a sample, of a particular *z* score and all others beyond it in any tail is given by the fractional area in that tail of the unit normal curve beyond the particular *z* score. The *one-tail probability* can be worked out by subtracting the fractional area from μ (z=0.00) to the given *z* score, from 0.5000 which is the total fractional area in either half of the unit normal curve. Thus, the 1-tail probability of a *z* score of +1.00 may be worked out as follows, using the fractional areas of Fig. 1.2.

P = 0.5000 - (fractional area from μ to the *z* score of +1.00)

= 0.5000 - 0.3413 = 0.1587.

(*j*) The *two-tail probability* of random occurrence of both positive and negative *z* scores of a particular value and of all others beyond the latter in the two tails is given by the total fractional area in both the tails of the unit normal curve beyond respectively the positive and the negative values of that *z* score. Because of bilateral symmetry of the normal curve, it can be worked out as follows using the fractional area from μ to the given *z* in any one tail. Thus, the 2-tail probability of the *z* score of ±1.00 would come to :

P = 2[0.5000 - (fractional area from μ to the *z* score of 1.00)]

= 2 [0.5000 - 0.3413] = 0.3174.

(*k*) If large samples are drawn from any normally distributed population, their means ($\bar{X}$) form a *normal sampling distribution* around the corresponding population mean. Also, the differences between such sample means form a similar normal sampling distribution. Hence, the probability of random occurrence of a given difference ($\bar{X}_1 - \bar{X}_2$) between two sample means may be obtained from the fractional area in the tail(s) beyond the *z* score of that difference in a unit normal curve.

The following *assumptions* should be justifiable if z scores and the unit normal curve areas are to be used for working out the probability of random occurrence of an observed difference $(\bar{X}_1 - \bar{X}_2)$ between two sample means.

(*a*) The scores whose means are being compared should belong to a continuous measurement variable.

(*b*) The scores of that variable should be normally distributed in the relevant population.

(*c*) Each score should occur at random in the sample, beyond any influence from the occurrence or absence of any other score.

1.7. STUDENT'S *t* DISTRIBUTION

You have already learnt about the *small sample statistic,* called the *t* score, which is obtained by the linear transformation of raw scores of small samples and their statistics like the mean, the difference between means and the correlation coefficient, basically in the same way as the *z* scores for large samples. The *t* distributions are the probability distributions of such *t* scores and are used in finding the probability of random occurrence of any given *t* score as well as the raw score or statistic which has been transformed into the latter. For plotting any *t* distribution curve, the probabilities -- in the forms of ordinates or *Y* scores of that curve -- are computed for different *t* scores having given *degrees of freedom* (*df*), with the help of complex mathematical equations formulated by W.S. Gossett (pseudonym "Student").

$$Y = \frac{\left[\frac{df-1}{2}\right]!}{\left[\frac{df-2}{2}\right]!} \times \frac{1}{\sqrt{\pi df}} \Bigg/ \left[1+\frac{t^2}{df}\right]^{\frac{df+1}{2}}.$$

The *properties* of *t* distributions are summarized below.

(*a*) The *t* scores, computed from the scores of only continuous measurement variables, have no genuine intervening gaps in the X axis of the *t* distribution curve. So, *t* distributions are *continuous probability distributions.*

(*b*) The *t* distributions may be plotted as *theoretical probability distributions,* using the Gossett equation.

(*c*) They are innumerable in number, each being the probability distribution of the *t* scores having a

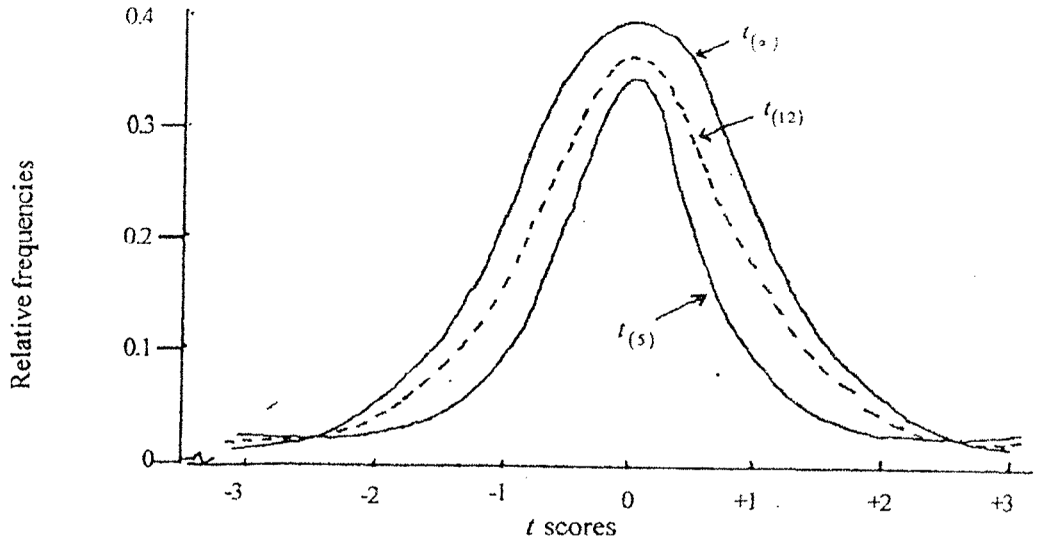


Fig. 1.3: A number of t distributions.

specific *df* only. So, the probability of random occurrence of any *t* score with a given *df* should be found out with reference to the particular *t* distribution with the same *df*. The *1-tail* and *2-tail probabilities* of a given *t* score are given by the fractional areas beyond that score in a single tail or both tails, respectively, in the *t* distribution curve with the same *df*.

(*d*) Each *t* distribution is *unimodal* with its mean, median and mode identical with the *t* score of 0.00 for the mean.

(*e*) Each *t* distribution is *asymptotic* with its two tails reaching the negative and positive infinities respectively.

(*f*) The *t* distributions are *leptokurtic*, but progressively less so with the rise in *df*, so that the *t* distribution for ∞ degrees of freedom is the same as the mesokurtic normal distribution.

(*g*) The *t* distributions have *no skewness* so that the fractional area beyond a given *t* score or between any two given *t* scores is identical in both the halves of a distribution.

The following *assumptions* should be justifiable if *t* scores and *t* distributions are to be used for working out the probability of random occurrence of an observed difference $(\bar{X}_1 - \bar{X}_2)$ between two sample means.

(*a*) The scores whose means are being compared should belong to a continuous measurement variable.

(*b*) The scores of that variable should be normally distributed in the population and should form a *t* distribution in a small sample.

(*c*) Each score should occur at random in the sample, beyond any influence from any other score.

1.8. SUMMARY

Probabilities and different types of probability distributions have been defined and explained.

Binomial distribution is a discrete probability distribution giving the probabilities of different combinations of numbers of events of two classes of a dichotomous variable. Its properties, assumptions and computational formula have been described.

Poisson distribution is a discrete probability distribution giving the probabilities of numbers of events of the rare class of a dichotomous variable. Its properties, assumptions and computational formula have been described.

Linearly and nonlinearly transformed standard scores have been defined with examples.

Normal distribution, a continuous probability distribution based on the Gaussian equation, and the unit normal curve have been defined and explained. Properties, assumptions and applications of normal distribution have been described.

Student's *t* distributions, continuous but *df*-dependent, are based on Gossett's equation. Their properties, assumptions and applications have been described.

1.9. GLOSSARY

Binomial distribution : A theoretical probability distribution of events of a dichotomous variable, plotted using the binomial equation.

Continuous probability distribution : A probability distribution of the scores of continuous measurement variables.

Discrete probability distribution : A probability distribution of the scores of discontinuous measurement variables.

Gaussian equation : The mathematical equation of Karl Gauss for expressing probabilities of scores as functions of their standard *z* scores.

Leptokurtic distribution: A distribution more peaked than the normal distribution.

Mesokurtic distribution: A distribution with a medium degree of peakedness like that of the normal distribution.

Negatively skewed distribution : An asymmetric distribution with its negative tail longer than the positive tail.

Normal distribution : A theoretical probability distribution of the scores of continuous measurement variables, plotted using the Gaussian equation.

Platykurtic distribution: A distribution flatter than the mesokurtic normal distribution.

Poisson distribution : A theoretical probability distribution of rare events, based on the Poisson equation.

Positively skewed distribution : An asymmetric distribution with a longer positive tail than the negative tail.

Probability : The limiting relative frequency of occurrence of an event or a score in a very large or infinite number of observations.

Probability distribution : Distribution of probabilities of occurrences of different scores or events in a sample or population.

Standard score : A score such as *z* score, computed by either linear or nonlinear transformation of a raw score or a statistic, for use as a comparable and standard reference score.

t distribution : A continuous but *df*-dependent probability distribution, plotted theoretically using Gossett's equation.

t score : A sampling statistic in *SD* (σ) units, which is computed by the linear transformation of a raw score, a mean, a difference between means or a correlation coefficient of a sample for use as a standard score.

Theoretical probability distribution : A theoretically computed probability distribution of scores or events of a variable in a sample or population, in terms of a mathematical equation.

Unit normal curve : A standard normal distribution curve, plotted by taking the sample size, *SD* and class interval length as 1.00 each.

z score : A standard score in *SD* (σ) units, into which a raw score, a statistic or a difference between scores or statistics of a large sample is linearly transformed.

1.10. QUESTIONS

1. (a) Explain what you mean by probability.

   (b) Define probability distributions and their different types with suitable examples.

   (c) Define standard scores and describe briefly their types according to their computational methods, with examples.

(d) Change the following frequency distribution of interorbital widths (in mm) of a sample of pigeons into a probability distribution.

Class intervals :	10.6 - 11.0	11.1-11.5	11.6-12.0	12.1-12.5	12.6-13.0	13.1-13.5
Frequencies :	12	28	46	32	15	7

2. (a) Define a binomial probability distribution. Explain how the terms of binomial expansion may be used in working out the probabilities of different combinations of the numbers of two types of events of a dichotomous variable.

(b) Describe the properties of the binomial distribution.

(c) State the assumptions underlying the application of the binomial distribution in biostatistics.

(d) Work out and interpret the probability of random occurrence of 3 male animals in a sample of 15 animals from a population having the male-female sex-ratio of 52:48, using the binomial expansion ($\alpha = 0.01$)..

3. (a) What is a Poisson probability distribution? Explain how the terms of the Poisson equation may give the probabilities of different numbers of events of the rare class of a dichotomous variable in a sample.

(b) Discuss the properties of the Poisson distribution.

(c) Write the assumptions to be fulfilled if you have to use the Poisson distribution for finding the probability of random occurrence of a given number of rare events in a sample.

(d) Compute the probability of random occurrence of 7 mutant animals in a sample of 343 animals drawn from a population in which 104 mutant animals occur per 10000.

4. (a) Write the mathematical equation for the unit normal curve and explain why and how you may change the Gaussian equation to that for the unit normal curve.

(b) Discuss the properties of normal probability distribution.

(c) Give the assumptions for using the unit normal curve in finding the probability of random occurrence of a score in a sample.

5. (a) Name the standard scores whose probability of occurrence is given by the normal distribution and the t distribution, respectively. Write the mathematical equations for a normal distribution and the unit normal curve, respectively.

(b) Explain how you would find the one-tail and the two-tail probabilities of a given z score.

(c) Discuss the similarities and differences between the properties of t distributions and normal distributions.

6. (a) Name the small sample statistic whose probability distribution may be theoretically worked out using Gossett's equation. Write the formulae for transforming a raw score, a sample mean and a difference between two sample means, respectively, into that statistic and mention its unit.

(b) Describe the assumptions for using the probability distribution of the small sample statistic, mentioned above, in finding probabilities of random occurrence of a difference between two sample means

(c) Discuss the properties of that probability distribution.

UNIT 2: TESTING OF HYPOTHESIS

2.0. OBJECTIVES

It is intended that after going through this unit, you should be able to do the following :

* understand and explain what is meant by the significance of the findings of an experiment,
* describe the null hypothesis and its importance in finding the significance of experimental observations,
* understand and explain the levels of significance and critical scores,
* realize that the rejection or retention of null hypothesis on the basis of a chosen level of significance leads to probabilities of errors of inference,
* distinguish between one-tail and two-tail tests and their uses in making the required inferences,
* understand the assumptions for significance tests using *z* scores and *t* scores, respectively,
* compute *z* scores from the differences between means of large samples and find the significance of such differences by referring the *z* scores to the unit normal curve table,
* compute *t* scores from the differences between means of both small and large samples and find the significance of such differences by referring the *t* scores to the critical *t* score table,
* understand how nonparametric chi square scores can be used in testing of hypothesis by the analysis of experimentally observed frequency distributions,
* work out and use the chi square score from the observed frequency distribution to find if the latter fits significantly with a proposed distribution, and
* use the computed chi square score to test whether or not there is a significant association between two variables.

2.1. SIGNIFICANCE OF EXPERIMENTAL RESULTS

You should be aware by now that all scientific investigations including experiments are performed using samples drawn from the population, and almost never by using the entire population. Because of variations from sample to sample from the same population, experimental observations also show variations. There is, thus, always a probability that the observed results may have been obtained simply because of the

random choice and use of a particular sample, and may not coincide with the findings from the entire population if the latter could be used for the investigation. But all our investigations aim at generalizing the findings in a sample for the entire relevant population. It thus becomes necessary to find statistically the probability of the experimental findings of a sample being true for the relevant population; in other words, statistical tests become unavoidable to find whether or not the experimental observations are meaningful or *significant* for the population and have only a low probability of being due to the accidental use of a particular sample instead of the entire population.

Consider an experiment to explore whether or not the serum Ca^{2+} is changed in a species of insects on administration of calcitonin, by finding the mean serum Ca^{2+} in two groups of insects from that population after administration of two levels (or doses) of calcitonin to the two groups. In each such experiment, the variable whose changes are being explored is called the *dependent variable* while the variable(s) whose effects on the dependent variable are under investigation is/are called the *independent variable(s)*; in the cited experiment, serum Ca^{2+} and calcitonin may be called the dependent variable and the independent variable, respectively. Now, even if the two group means are found to differ, there may be two possibilities. *First,* there has been no real change in the mean Ca^{2+} due to the effect of calcitonin and the observed difference may be explained away by the varying *sampling errors* of the means, which cause samples from the same population to form a sampling distribution around the population mean. *Secondly,* the mean serum Ca^{2+} may have actually changed due to the effects of calcitonin so that their difference cannot be explained away by the varying sampling errors of the sample means from the same population. If the first alternative is correct, then the two samples continue to belong to the same population and their means continue to be the estimates of the same population mean. But if the second alternative is correct, then the two samples no longer belong to the same population after being treated with two different levels of the independent variable; so, their means no longer represent the estimates of the parametric mean of the same population, and their difference is genuine and *not* explained away by the sampling errors of the means. It is, however, impossible to know for certain which one of these two possibilities is correct in any experiment. However, a *significance test* may be statistically performed with the experimental results to work out the probability of the observed results being genuine or not. This is a test for the significance or meaningfulness of the observed results; only if the probability of the results arising by mere chance, because of the accidental choice of particular sample for the experiment, can be judged to be too low, observed results may be considered significant or meaningful, which means that they have a high probability of being genuine enough for generalization over the entire population. On

the countrary, if the statistical test for significance yields too high a probability of the observed results occurring by chance due to the accidental choice of the particular sample, the results are not considered meaningful or significant and are not worthy of generalization over the relevant population. This basic procedure you will find repeated as a follow-up of all scientific investigations for making inferences about the experimental results.

2.2. NULL HYPOTHESIS

It is this hypothesis whose correctness is tested in all significance tests for making inferences. It is evident to you by now that each experiment or investigation is performed to substantiate a hypothesis. The latter is called the *alternative hypothesis* (H_a) for reasons you will learn below. However, practically all experiments or investigations, you know, are conducted using random samples chosen from the population, and not by subjecting the entire population to such studies. There is, therefore, always a probability that the observed results have arisen by chance because of the use of a randomly chosen sample, instead of the entire population, in the experiment. It is this probability which the investigator has to weigh while making inferences from his experiment. And it is this probability of the observed results happening by chance, due to the use of a random sample, which is taken care of by the null hypothesis.

The *null hypothesis* (H_0) always contests the alternative hypothesis (H_a) which is sought to be substantiated by the relevant experiment, and thereby proposes to negate or nullify the latter hypothesis. The H_0 always contends that the observed result of the experiment, viz., the observed difference between two group means, or the observed correlation between two or more variables, or the goodness of fit between the observed and the proposed frequency distributions, is not significant. It considers that the observed results have been obtained simply because of the random choice of the particular sample and would have been different if the entire population or even a second random sample would have been used for the experiment. In other words, the H_0 proposes that the findings may be explained away by the sampling errors and are thus not meaningful or significant.

So, after each experiment, the probability of this H_0 being correct has to be worked out statistically. Only if this probability (P) is found to be too low, the H_0 is rejected, the H_a is retained instead, and the observed results of the experiment are considered significant. On the contrary, if the P of the correctness of H_0 is found to be too high, the H_0 is retained and the H_a is rejected instead, leading to the inference that the observed results are not significant. Because in this way, the hypothesis (H_a) underlying the experiment is retained or rejected depending on

the probability (P) of the H_0 being correct, the experimental hypothesis has been called the alternative hypothesis.

Because the H_0 always contests the H_a, its form depends on the H_a sought to be substantiated in the experiment. If the H_a proposes that there is a significant difference between two group means in the given experiment, the H_0 contends that there is no significant difference between the means and the observed difference has occurred simply due to the choice and use of the random sample in the experiment. If the H_a proposes that the frequency distribution observed in the sample has a significant goodness of fit with a proposed distribution (e.g., normal distribution, Mendelian distribution, binomial distribution, etc.), the H_0 contends that there is no significant goodness of fit between the two distributions. If the H_a proposes that there is a significant correlation between two variables, the H_0 contests it by proposing that there is no significant correlation between the variables. Because the H_0 tries to negate and nullify what is proposed by the H_a, it has been called the null hypothesis.

2.3. LEVELS OF SIGNIFICANCE AND ERRORS OF INFERENCE

The question which should naturally occur to you is, when to consider the probability (P) of the H_0 being correct to be too low, justifying its rejection, and when to consider the P to be so high that the H_0 has to be retained. For this, the investigator has to use a level of significance. The *level of significance* (α) is that level of probability upto which the computed probability (P) of correctness of the H_0 may be considered too low and the H_0 may be rejected; in other words, the H_0 is rejected and the observed results are considered significant so long as P is either lower than or equal to the level of significance ($P \leq \alpha$). On the contrary, if the worked out probability (P) of correctness of the H_0 exceeds the chosen level of significance, P is considered too high to justify the rejection of the H_0; so, if P is higher than the chosen α ($P > \alpha$), the H_0 is retained and the observed results of the experiment are considered not significant.

However, there is always a chance of the inference going wrong, because it is made on the basis of the level of significance which is only a level of probability of the H_0 being correct. If the chosen level of significance has led to an inference that the observed results are significant, there is a probability, howsoever small, that the H_0 was actually correct in this particular case and has been wrongly rejected. This type of error, called the *type I error* of inference, has a probability (α) identical with that of the level of significance used in rejecting the H_0. In other words, it is the probability of accepting the observed results as significant though they are really not significant and would not have been obtained if the entire population were used instead of the given sample. The higher the level

of significance used in drawing the inference, the greater is the probability of the type I error. So, the probability of type I error may be reduced by lowering the level of significance for making the inference. But this would cause a corresponding rise in the probability of an opposite error of inference, viz., the *type II error*, which results from the acceptance of the H_0 and the consequent rejection of the observed results as not significant, although the H_0 was not correct in this case and the observed results were genuine and would have been obtained even on using the entire population instead of the given sample. The probability (β) of the type II error may approach ($1-\alpha$) where α is the chosen significance level as well as the probability of type I error. Thus, if the H_0 is rejected using the 0.05 level of significance, the probability (α) of the type I error in that inference amounts to 0.05. On the contrary, if the H_0 is accepted using the 0.05 level of significance, the probability (β) of the type II error in that inference may be as high as ($1-\alpha$) or 0.95.

More importance is usually given to the type I error than to the type II error while making inferences, so that the probability of wrong acceptance of the experimental results as significant may be minimized. For most biological experiments, the level of significance does not exceed 0.05 so that the H_0 is not rejected if its probability (P) exceeds 0.05 ($P>0.05$). However, for minimizing the probability of the type I error of inference, the level of significance may be lowered below 0.05; 0.02, 0.01, 0.005 or 0.001 significance level (α) may be chosen for a progressively lower probability of type I error in the acceptance of the experimental results as significant.

2.4. CRITICAL SCORES

The probability of random occurrence of a computed statistic (e.g., the *t* score computed from an observed difference between group means or from a correlation coefficient between two variables, the chi square (χ^2) from the difference between two frequency distributions, and the *F* ratio of two variances) is very often found by comparing the computed statistic with its scores beyond which lie fractional areas corresponding to respective significance levels, in the tails of its probability distribution. Any particular score, beyond which the fractional area in its probability distribution corresponds to a chosen significance level (α), is called the *critical score* for that α. Except the unit normal curve for *z* score distribution, the probability distributions of other such scores as the *t* scores, chi squares and *F* ratios vary according to the *df* of those scores. So, their critical score for any given significance level varies with the *df* of the relevant score. Moreover the critical scores of a statistic for a given significance level differ according to whether it is a one-tail test or a two-tail test (see Section 2.5 below). This is

because the one-tail significance level (*1-tail* α) is the fractional area beyond the relevant critical score in a single tail of its probability distribution while the two-tail significance level (*2-tail* α) is the sum of the fractional areas beyond the relevant critical score in both the tails of the distribution. Evidently, for any given *df*, the 2-tail critical score for a particular α would be higher than the 1-tail critical score for the same α (see Table 2.1).

Table 2.1 : Some critical scores.

		$\alpha = 0.05$		$\alpha = 0.01$	
Statistic	*df*	2-tail	1-tail	2-tail	1-tail
z	-	1.960	1.645	2.575	2.327
t	10	2.228	1.812	3.169	2.764
t	30	2.042	1.697	2.750	2.457
t	∞	1.960	1.645	2.576	2.326
χ^2	1	3.84	2.71	6.64	5.41
χ^2	2	5.99	4.60	9.21	7.82
χ^2	3	7.82	6.25	11.34	9.84
χ^2	4	9.49	7.78	13.28	11.67

After working out a *t* score, χ^2 or *F* and its *df* from the observed data, it is compared with the corresponding critical score with the same *df* and for the chosen significance level (α). If the computed score equals or exceeds the critical score, the probability *P* of the H_0 being correct is respectively equal to the α ($P=\alpha$) or lower than it ($P<\alpha$); in both cases *P* may be considered too low and the H_0 may be rejected, inferring that the observed results are significant. The probability of type I error in this inference is also mentioned as ($P=\alpha$) or ($P<\alpha$) giving the value of the α used. If, on the contrary, the computed *t*, χ^2 or *F* score is lower than the corresponding critical score, *P* exceeds α ($P>\alpha$) and is considered too high; so, the H_0 cannot be rejected and it is inferred that observed results are not significant.

The critical scores are given, mentioning their respective *df* and significance levels (α). For example, the 2-tail and 1-tail critical *t* scores (df=30) for an α of 0.05, are given as follows (see Table 2.1).

Critical 2-tail $t_{\alpha(df)}$: $t_{0.05\,(30)} = 2.042$.

Critical 1-tail $t_{\alpha(df)}$: $t_{0.05\,(30)} = 1.697$.

If the t score computed from the difference between two group means amounts to 2.014 and has the df of ∞, it is compared for a 2-tail significance test ($\alpha = 0.05$) with the critical 2-tail critical $t_{0.05\,(\infty)}$:

Computed $t = 2.014$; critical $t_{0.05\,(\infty)} = 1.960$ (see Table 2.1).

Because this computed t score exceeds the critical t, the probability P of the H_0 being correct is lower than α ($P<0.05$) and is, therefore, considered to be too low. So, the H_0 may be rejected and it may be inferred that there is a significant difference between the group means ($P<0.05$).

2.5. TWO-TAIL AND ONE-TAIL TESTS

For an experiment conducted to test the H_a that there is a significant difference between two group means irrespective of the algebraic sign of such difference, the z or t score computed from the latter may be positive as well as negative and its probability of random or chance occurrence is given by the fractional areas beyond it in both tails of its probability distribution. So, the test for significance of such a difference between means is called a *two-tail test*. The H_0 of such a two-tail test contends that there is no significant difference between the means, irrespective of the algebraic sign of such a difference. So, the probability P of the correctness of such an H_0 has to be determined using the two-tail critical z or t score for the chosen significance level (α), the fractional area for which is divided into identical halves ($\alpha/2$ each) in two tails of the probability distribution of t or z. If the P is found to be equal to or lower than the chosen α in such a two-tail test ($P \leq \alpha$), it is inferred that the observed difference between means is significant, irrespective of its algebraic sign.

On the contrary, the H_a underlying some experiments proposes that one group mean is significantly higher (or lower) than the other so that their difference bears a particular algebraic sign, either positive or negative only. The z or t score computed from that difference would bear only that particular algebraic sign and not the other. So, the probability of chance occurrence of such a z or t is given by the fractional area beyond it in a single tail of its probability distribution and not in both tails. The test for significance of such

a difference, bearing a particular algebraic sign only, is therefore called a *one tail* test. The H_0 of such a one-tail test contends that one group mean is not significantly higher (or lower) than the other. The probability of its correctness (P) has to be found out using the one-tail critical t or z score for the chosen significance level (α), the entire fractional area for which is located in a single tail of the probability distribution of t or z. If P is found to be equal to or lower than the chosen α, it is considered too low ($P \leq \alpha$), the H_0 cannot be retained and it is inferred that one of the group means is significantly higher (or lower) than the other. But if P exceeds the α, it is considered too high, the H_0 cannot be rejected and it is inferred that one group mean is not significantly higher (or lower) than the other ($P > \alpha$).

You should recall that for both types of tests, P is found out by comparing the computed score (e.g., the computed t) with the critical score for the chosen level of significance (α). If the computed score equals the critical score, P equals α; if the computed score exceeds the critical score, P is lower than α. In both cases, the H_0 may be rejected. But if the computed score is lower than the critical score, P exceeds α and the H_0 is retained. [See Section 2.4].

2.6. SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS USING z SCORES

In an experiment exploring the possible change of a variable (*dependent variable*) due to the effect of another variable (*independent variable*), often two large ($n > 30$), separate and unmatched groups of animals, sampled from a population, are exposed to or administered with two different levels (doses, amounts, qualities, etc.) of the independent variable and the scores of the dependent variable are measured in each group after such treatment (*large independent group experiment*). The difference between the two group means is then computed and subjected to a significance test. In case of a two-tail test, the H_a proposes a significant difference between the means while the H_0 contests it by contending that the group means have no significant difference; in a one-tail test, the H_a proposes that one group mean is significantly higher (or lower) than the other while the H_0 contends that one of the means is not significantly higher (or lower) than the other (see Sections 2.2 to 2.5). In both cases, the probability of the H_0 being correct is found out by converting the difference between means to a z score and finding the probability of its random occurrence as per the H_0, using the unit normal curve. If this probability (P) is found to be either equal to or lower than the

level of significance (α), the H_0 is rejected ($P \leq \alpha$); this leads to the inference in the two-tail test that there is a significant difference between the group means, and in the one-tail test, that one group mean is significantly higher (or lower) than the other. But if $P > \alpha$, the H_0 is retained -- it is then inferred in a two-tail test that there is no significant difference between the group means, and in a one-tail test, that one group mean is not significantly higher (or lower) than the other. For such tests in zoological experiments, α is never fixed above 0.05.

If you propose to use *z* scores and the unit normal curve for such tests, the following ***assumptions*** must be justifiable before you can apply them, viz., that the dependent variable is a ***continuous measurement variable***, having a ***normal distribution*** in the population, that the sample size is ***large***, and that each individual animal has been chosen for the sample ***at random***. You cannot use *z* scores and the normal curve, if any of these assumptions cannot be made justifiably.

Procedure :

First, work out the group means ($\bar{X}_1$ and $\bar{X}_2$), *SD*s (s_1 and s_2) and the *SE*s of the means ($s_{\bar{X}_1}$ and $s_{\bar{X}_2}$) of the scores (X_1 and X_2) of the respective groups, having n_1 and n_2 as their respective sizes. Use the *SE*s in computing the *SE* of the difference between means ($s_{\bar{X}_1 - \bar{X}_2}$). Then, use the latter in transforming the difference between the means ($\bar{X}_1 - \bar{X}_2$) to the *z* score.

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1};\quad \bar{X}_2 = \frac{\Sigma X_2}{n_2};\quad s_1 = \sqrt{\frac{\Sigma (X_1 - \bar{X}_1)^2}{n_1 - 1}};\quad s_2 = \sqrt{\frac{\Sigma (X_2 - \bar{X}_2)^2}{n_2 - 1}};$$

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}};\quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}};\quad s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(s_{\bar{X}_1}\right)^2 + \left(s_{\bar{X}_2}\right)^2};$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}.$$

Next, note from the unit normal curve table the *area* of that curve from its mean μ ($z = 0.00$) to the computed *z* score, and use it in working out the probability (*P*) of the H_0 being correct.

(a) For a two-tail test :

$P = 2$ [0.5000 - (unit normal curve area corresponding to the computed z)].

If P is either lower than or equal to the chosen α (0.05 or lower), i.e., if $P \leq \alpha$, the H_0 is rejected and it is inferred that there is a significant difference between the group means.

(b) For a one-tail test :

$P = 0.5000$ - (unit normal curve area corresponding to the computed z).

If P is either lower than or equal to the chosen α (0.05 or lower), i.e., if $P \leq \alpha$, the H_0 is rejected and one group mean is significantly higher (or lower) than the other.

In both types of test, the H_0 is retained and there is no significant difference when P exceeds either a chosen α or the α of 0.05 ($P > \alpha$).

Example 2.6.1.

Find whether or not there is a significant difference between the mean interorbital widths (mm) of the following two groups of pigeons, choosing the relevant area corresponding to the computed z score ($\alpha = 0.05$).

Interorbital widths (mm) :

Group 1 (X_1) : 11.4, 11.9, 12.1, 11.4, 12.1, 11.9, 11.4, 11.2, 12.5, 12.1, 13.5, 12.2, 12.0, 12.6, 11.4, 12.2, 11.3, 12.7, 12.3, 13.3, 12.8, 12.5, 12.5, 11.9, 11.8, 11.5, 12.6, 12.0, 11.7, 12.2.

Group 2 (X_2) : 10.8, 11.3, 10.8, 11.5, 12.4, 11.6, 11.8, 11.2, 11.5, 12.1, 12.4, 11.6, 12.0, 12.5, 11.7, 12.3, 12.2, 11.8, 12.3, 10.8, 11.8, 12.8, 11.8, 12.2, 11.8.

z scores :	1.98	1.99	2.00	2.01	2.02
Areas of unit normal curve :	0.4761	0.4767	0.4772	0.4778	0.4783

Solution :

A two-tail z test is undertaken to find the probability (P) of correctness of the H_0 that there is no significant difference between the group means. For this, the scores of the two groups are entered in the first two columns of Table 2.2, and used for further computations.

Table 2.2 : Table for computing the means and *SD*s of the interorbital width data (n_1=30, n_2=25).

X_1	X_2	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
11.4	10.8	-0.7	0.49	-1.0	1.00
11.9	11.3	-0.2	0.04	-0.5	0.25
12.1	10.8	0	0	-1.0	1.00
11.4	11.5	-0.7	0.49	-0.3	0.09
12.1	12.4	0	0	+ 0.6	0.36
11.9	11.6	-0.2	0.04	- 0.2	0.04
11.4	11.8	-0.7	0.49	0	0
11.2	11.2	- 0.9	0.81	-0.6	0.36
12.5	11.5	+0.4	0.16	-0.3	0.09
12.1	12.1	0	0	+0.3	0.09
13.5	12.4	+1.4	1.96	+0.6	0.36
12.2	11.6	+0.1	0.01	-0.2	0.04
12.0	12.0	-0.1	0.01	+0.2	0.04
12.6	12.5	+0.5	0.25	+0.7	0.49
11.4	11.7	-0.7	0.49	-0.1	0.01
12.2	12.3	+0.1	0.01	+0.5	0.25
11.3	12.2	-0.8	0.64	+0.4	0.16
12.7	11.8	+0.6	0.36	0	0
12.3	12.3	+0.2	0.04	+0.5	0.25
13.3	10.8	+1.2	1.44	-1.0	1.00

Table 2.2 (Continued)

X_1	X_2	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
12.8	11.8	+0.7	0.49	0	0
12.5	12.8	+0.4	0.16	+1.0	1.00
12.5	11.8	+0.4	0.16	0	0
11.9	12.2	-0.2	0.04	+0.4	0.16
11.8	11.8	-0.3	0.09	0	0
11.5	-	-0.6	0.36	-	-
12.6	-	+0.5	0.25	-	-
12.0	-	-0.1	0.01	-	-
11.7	-	-0.4	0.16	-	-
12.2	-	+0.1	0.01	-	-
Σ 363.0	295.0	-	9.46	-	7.04
(ΣX_1)	(ΣX_2)		$\Sigma (X_1 - \bar{X}_1)^2$		$\Sigma (X_2 - \bar{X}_2)^2$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{363}{30} = 12.1 \text{ mm.} \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{295}{25} = 11.8 \text{ mm.}$$

$$s_1 = \sqrt{\frac{\Sigma (X_1 - \bar{X}_1)^2}{n_1 - 1}} = \sqrt{\frac{9.46}{30 - 1}} = 0.57 \text{ mm.} \quad s_2 = \sqrt{\frac{\Sigma (X_2 - \bar{X}_2)^2}{n_2 - 1}} = \sqrt{\frac{7.04}{25 - 1}} = 0.54 \text{ mm.}$$

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{0.57}{\sqrt{30}} = 0.104 \text{ mm.} \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{0.54}{\sqrt{25}} = 0.108 \text{ mm.}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2} = \sqrt{(0.104)^2 + (0.108)^2} = 0.150 \text{ mm.}$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{12.1 - 11.8}{0.150} = 2.00.$$

Area of unit normal curve corresponding to the z score of 2.00 amounts to 0.4772.

$\therefore P$ = 2 [0.5000 - (Area of unit normal curve corresponding to the computed z)]

= 2[0.5000 - 0.4772] ≅ 0.046.

$\alpha = 0.05$

As the probability P for the correctness of the H_0 is lower than 0.05, the H_0 is rejected and it is inferred that the observed difference between the group means is significant ($P<0.05$).

2.7. SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS USING t SCORES

An observed difference between two group means in an experiment can also be tested for its significance by transforming that difference to the Student's t score, working out the df of the computed t, and then comparing the computed t with the critical t scores from a critical t score table for finding the probability P of the H_0 being correct. You may recall that t distributions differ from each other according to the degrees of freedom of the t scores. Hence, the computed t must be compared with only such critical t scores for different significance levels (α) as have the df of the computed t. Moreover, either one-tail or two-tail critical t scores should be used for such comparison, according as a one-tail or a two-tail test is being done (see Sections 2.4 and 2.5).

In case of a *two-tail t test*, the H_a proposes a significant difference between the group means while the H_0 contests it by contending that there is no significant difference between the means. In a *one-tail t test*, the H_a proposes that one group mean is significantly higher (or lower) than the other while the H_0 contends that one group mean is not significantly higher (or lower) than the other. In both types of test, if the computed t is found to be either greater than or equal to the critical t with the same df and for a chosen level of α, i.e., if $t \geq t_\alpha$, the probability P of the correctness of the H_0 is respectively lower than and equal to that significance level. The H_0 may then be rejected. This leads to the inference in a two-tail t test that the observed difference between the means is significant ($P<\alpha$ or $P=\alpha$ as the case may be); for a one-tail t test, the inference should be that one of the group means is significantly higher (or lower) than the other. But if the computed t is found to be lower than the critical t ($t<t_\alpha$), the probability P of correctness of the H_0 exceeds the chosen α ($P>\alpha$); in a two-tail test, this leads to the inference that the observed difference between means is not significant while in a one-tail test, it is inferred that one group mean is not significantly higher (or lower) than the other. For zoological experiments, α is not fixed above 0.05.

Although Student's t is a small-sample statistic and t tests were originally proposed for use in small samples ($n<30$), t tests can also be used for large samples, even where $n = \infty$, because the leptokurtic t distributions progressively approach the mesokurtic normal distribution with the rise in df, and coincide with the normal distribution at ∞ degrees of freedom. This will be evident to you if you note that the critical z scores for different levels of significance are practically identical with the critical t scores ($df = \infty$) for those levels (Table 2.1).

Only if the following *assumptions* can be made justifiably, the *t* tests can be applied, viz., that the dependent variable of the experiment is a *continuous measurement variable* with its scores *distributed normally* in the population, and that each animal occurs in the sample *at ramdom* only.

Different types of *t* tests are used for different types of experiments.

(a) *t test for large independent groups :*

For *t* tests for significance of difference between the means of two large independent groups, each exceeding 30 in size, the difference ($\overline{X}_1 - \overline{X}_2$) between the means is transformed into *t*, using the *SE* of their difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$ computed from the *SD*s (s_1 and s_2) and the sizes (n_1 and n_2) of the groups, and the *df* of the *t* score is worked out.

$$\overline{X}_1 = \frac{\Sigma X_1}{n_1};\ \overline{X}_2 = \frac{\Sigma X_2}{n_2};\ s_1 = \sqrt{\frac{\Sigma\left(X_1 - \overline{X}_1\right)^2}{n_1 - 1}};\ s_2 = \sqrt{\frac{\Sigma\left(X_2 - \overline{X}_2\right)^2}{n_2 - 1}};$$

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}};\quad t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1-\overline{X}_2}};\quad df = n_1 + n_2 - 2.$$

The computed *t* is then compared with the critical 2-tail or 1-tail *t* scores, according as a two-tail or a one-tail test is intended. Where the computed *t* exceeds or equals the critical *t* (t_α) of any significance level not higher than 0.05, i.e., $t \geq t_\alpha$ where $\alpha \leq 0.05$, the probability *P* of the correctness of H_0 is respectively lower than or equal to that $\alpha (P \leq \alpha)$ and H_0 may consequently be rejected. The observed results are then considered significant -- for a two-tail test the observed difference between the means is significant while for a one-tail test, one group mean is significantly higher (or lower) than the other. But if the computed *t* is lower than the critical *t* for the chosen significance level or the 0.05 level of $\alpha (t < t_\alpha)$, *P* exceeds α, the H_0 cannot be rejected and the observed results are not significant ($P > \alpha$).

Example 2.7.1.

Apply *t* test to the pigeon interorbital width score (mm) data of Example 2.6.1 and Table 2.2, to test whether or not the two group means differ significantly ($\alpha = 0.01$), choosing the correct critical t_α from the following two-tail critical *t* scores.

$t_{0.01(55)} = 2.668$; $t_{0.05(55)} = 2.004$; $t_{0.01(54)} = 2.670$;

$t_{0.05(54)} = 2.005$; $t_{0.01(53)} = 2.672$; $t_{0.05(53)} = 2.006$.

Solution :

You should refer to Table 2.2 for the data and the relevant computations. Depending on these computations whose results are given there (Example 2.6.1),

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{363}{30} = 12.1 \text{ mm.} \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{295}{25} = 11.8 \text{ mm.}$$

$$s_1 = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2}{n_1 - 1}} = \sqrt{\frac{9.46}{30-1}} = 0.57 \text{ mm.} \quad s_2 = \sqrt{\frac{\Sigma(X_2 - \bar{X}_2)^2}{n_2 - 1}} = \sqrt{\frac{7.04}{25-1}} = 0.54 \text{ mm.}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}} = \sqrt{\frac{(0.57)^2}{25} + \frac{(0.54)^2}{30}} = 0.151.$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{12.1 - 11.8}{0.151} = 1.987. \quad df = n_1 + n_1 - 2 = 30 + 25 - 2 = 53.$$

Critical $t_{0.01(53)} = 2.672$.

Because the computed t is lower than the critical $t_{0.01}$ ($t < t_{0.01}$), the probability P of the H_0 being correct exceeds the chosen α of 0.01 ($P > 0.01$). So, the H_0 cannot be rejected. Hence, it is inferred that the observed difference between the group means is not significant ($P > 0.01$).

(b) *t test for small independent groups :*

Where the groups are independent of each other but small, each less than 30 in size, the *SE* of difference between the means is computed using a pooled *SD* ($\hat{s}$) of both the groups, instead of their individual *SD* scores. Otherwise, the procedure is similar to that for large independent groups.

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1}; \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2}; \quad \hat{s} = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}};$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2} = \hat{s}\sqrt{\frac{n_1 + n_2}{n_1 n_2}};$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}; \quad df = n_1 + n_2 - 2.$$

The computed t is compared to either 1-tail or 2-tail critical t scores of the same df, according as one-tail or two-tail test is intended, and inferences are made as given for large independent groups (see above).

Example 2.7.2.

Find whether or not there is a significant difference between the mean interorbital widths (mm) of two groups of pigeons, consisting of 10 and 8 animals respectively, using the correct critical t score from those given below. (α=0.05).

Group 1 (X_1) : 11.9, 11.4, 11.9, 11.4, 11.2, 12.2, 12.6, 12.2, 12.7, 12.5.

Group 2 (X_2) : 10.8, 10.9, 10.5, 11.0, 10.6, 10.5, 10.8, 10.5.

$t_{0.01(18)} = 2.878$; $t_{0.05(18)} = 2.101$; $t_{0.01(17)} = 2.898$; $t_{0.05(17)} = 2.110$; $t_{0.01(16)} = 2.921$; $t_{0.05(16)} = 2.120$.

Solution :

The scores of the two groups are entered in the first two columns of Table 2.3 and used for further computations.

Table 2.3: Table for computing means and pooled *SD* of interorbital width data. (n_1=10, n_2=8).

X_1	X_2	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
11.9	10.8	-0.1	0.01	+0.1	0.01
11.4	10.9	-0.6	0.36	+0.2	0.04
11.9	10.5	-0.1	0.01	-0.2	0.04
11.4	11.0	-0.6	0.36	+0.3	0.09
11.2	10.6	-0.8	0.64	-0.1	0.01
12.2	10.5	+0.2	0.04	-0.2	0.04
12.6	10.8	+0.6	0.36	+0.1	0.01
12.2	10.5	+0.2	0.04	-0.2	0.04

Table 2.3 (Continued)

12.7	-	+0.7	0.49	-	-
12.5	-	+0.5	0.25	-	-
Σ 120.0	85.6	-	2.56	-	0.28
(ΣX_1)	(ΣX_2)		$\Sigma (X_1 - \bar{X}_1)^2$		$\Sigma (X_2 - \bar{X}_2)^2$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{120}{10} = 12.0\,\text{mm}. \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{85.6}{8} = 10.7\,\text{mm}.$$

$$\hat{s} = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{2.56 + 0.28}{10 + 8 - 2}} = 0.42\,\text{mm}.$$

$$s_{\bar{X}_1 - \bar{X}_2} = \hat{s}\sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 0.42\sqrt{\frac{10 + 8}{10 \text{x} 8}} = 0.199;$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{12.0 - 10.7}{0.199} = 6.533; \; df = n_1 + n_2 - 2 = 10 + 8 - 2 = 16.$$

Critical $t_{0.05(16)} = 2.120$.

As the computed t is higher than the critical t for the chosen level of α, $P<\alpha$. Hence, the H_0 is rejected and it is inferred that the observed difference between the means is significant ($P<0.05$).

(c) *t test for small single-group experiments :*

In a *single-group experiment,* the same group of individuals, drawn at random from the population, is subjected to the treatment with successive levels of the independent variable and the subsequent scores of the dependent variable are measured to explore the possible changes of the latter due to the independent variable. In such an experiment, each individual yields two scores of the dependent variable, one after being subjected to each level of treatment with the independent variable. Because of such *paired scores* of each individual of the group, the two sets of scores, obtained from the same group while functioning as the control group and the experimental group respectively, bear a correlation between each other. This correlation between the paired scores may affect the observed difference between the group means and should be taken into consideration while testing the significance of that difference. But because of considerable errors in correlation coefficient computed from small groups ($n<30$), a t test is performed using the differences between the paired scores, but avoiding the use of the correlation coefficient.

In this *t test by the difference method*, the difference (D) between the scores (X_1 and X_2) of each pair is worked out and all such D scores are used in computing the mean difference ($\overline{D}$). The latter is then used in the computations of the standard deviation (s_D) and the standard error ($s_{\overline{D}}$) of the mean difference. The mean difference $\overline{D}$ is then transformed into the corresponding t score. Where n is the group size, i.e., the number of individuals in the group,

$$D = X_1 - X_2;\ \overline{D} = \frac{\Sigma D}{n};\ s_D = \sqrt{\frac{\Sigma(D-\overline{D})^2}{n-1}};\ s_{\overline{D}} = \frac{s_D}{\sqrt{n}};\ t = \frac{\overline{D}}{s_{\overline{D}}};\ df = n-1.$$

The computed t score is then compared with the critical 2-tail or 1-tail t scores, according as a two-tail or a one-tail test is intended. Where the computed t exceeds or equals the critical t (t_α) of any significance level not higher than 0.05, i.e., $t \geq t_\alpha$ where $\alpha \leq 0.05$, the probability P of the H_0 being correct is respectively lower than and equal to that significance level. The H_0 may then be rejected. This leads to the inference in a 2-tail test that the observed difference between the means of the two sets of paired scores is significant ($P \leq \alpha$); for a one-tail test, the inference should be that one of the means is significantly higher (or lower) than the other. But if the computed t is lower than the critical t ($t < t_\alpha$), the probability P of the H_0 exceeds the chosen α ($P > \alpha$); in a two-tail test, this leads to the inference that the observed difference is not significant while in a one-tail test, it is inferred that the mean of one set of paired scores is not significantly higher (or lower) than that of the other set.

Example 2.7.3.

Using the following data, find whether or not the tracheal ventilations of a group of water beetles are significantly higher before exposure to a pesticide than after exposure to it. (α=0.01).

Beetle No. :	1	2	3	4	5	6	7	8
Tracheal ventilations (ml/min) :								
(a) before exposure (X_1)	85.2	76.3	88.0	80.3	79.5	89.6	79.1	80.5
(b) after exposure (X_2)	81.4	70.8	84.7	76.7	74.0	82.8	77.3	78.8

Critical t scores :

2-tail t_α: $t_{0.05(7)} = 2.365$; $t_{0.01(7)} = 3.499$; $t_{0.05(6)} = 2.447$; $t_{0.01(6)} = 3.707$.

1-tail t_α: $t_{0.05(7)} = 1.895$; $t_{0.01(7)} = 2.998$; $t_{0.05(6)} = 1.943$; $t_{0.01(6)} = 3.143$.

Solution :

A one-tail t test by the difference method has to be performed in this case. For this, the scores of the two sets are entered in Table 2.4 and used for subsequent computations.

Table 2.4 : Table for computing mean difference and *SD* of the differences ($n = 8$).

Ind. No	X_1	X_2	$D = X_1 - X_2$	$D - \overline{D}$	$(D - \overline{D})^2$
1	85.2	81.4	+3.8	-0.2	0.04
2	76.3	70.8	+5.5	+1.5	2.25
3	88.0	84.7	+3.3	-0.7	0.49
4	80.3	76.7	+3.6	-0.4	0.16
5	79.5	74.0	+5.5	+1.5	2.25
6	89.6	82.8	+6.8	+2.8	7.84
7	79.1	77.3	+1.8	-2.2	4.84
8	80.5	78.8	+1.7	-2.3	5.29
Σ	658.5	626.5	32.0	-	23.16

$$\overline{X}_1 = \frac{\Sigma X_1}{n} = \frac{658.5}{8} = 82.3 \text{ ml}; \quad \overline{X}_2 = \frac{\Sigma X_2}{n} = \frac{626.5}{8} = 78.3 \text{ ml};$$

$$\overline{D} = \frac{\Sigma D}{n} = \frac{32.0}{8} = 4.0 \text{ ml}; \quad s_D = \sqrt{\frac{\Sigma (D - \overline{D})^2}{n-1}} = \sqrt{\frac{23.16}{8-1}} = 1.82 \text{ ml};$$

$$s_{\overline{D}} = \frac{s_D}{\sqrt{n}} = \frac{1.82}{\sqrt{8}} = 0.643 \text{ ml}; \quad t = \frac{\overline{D}}{s_{\overline{D}}} = \frac{4.0}{0.643} = 6.221;$$

$df = n - 1 = 8 - 1 = 7$. 1-tail critical $t_{0.01(7)} = 2.998$.

Because the computed t is higher than the critical $t_{0.01}$ score, the probability P of the correctness of H_0 is lower than the chosen α of 0.01. Hence, P is considered too low, the H_0 is consequently rejected, and it is inferred that the mean tracheal ventilation before exposure to the pesticide is significantly higher than that after the exposure ($P<0.01$).

2.8. CHI SQUARE TESTS

These are nonparametric tests of hypothesis. A *nonparametric statistic* is one, in whose computation no statistic computed earlier has been used as an estimate of the corresponding parameter. Nor does its significant test involve the use of probability distributions of parametric statistics, such as the normal and t distributions. So, in contrast to *parametric statistics* like z and t which are computed using other statistics as estimates of parameters and are interpreted with reference to normal and t distributions, applications of nonparametric statistics like the chi square do not require the *assumptions* of the continuous nature and the normality of distribution of the variable to be justifiable. So, nonparametric statistics like chi squares can be used for continuous, discontinuous and ordinal variables and even to small samples. In fact, chi square tests can also be used for qualitative or nominal variables. But for using the chi square, it should be justifiable to assume that each individual has been chosen at random for the sample and the choice is not affected by the choice or rejection of any other individual.

The significance of a computed chi square (χ^2) is found out by comparing it with the critical chi square (χ^2_α) score beyond which lies the fractional area equivalent to the level of significance (α) in *only the right tail* of the relevant chi square distribution. However, as such distributions vary with the degrees of freedom of chi squares, the critical χ^2_α scores to be compared with the computed χ^2 must have the same *df* as the latter. Some other nonparametric statistics such as the log likelihood ratio statistic G, the contingency coefficient and Yule's phi coefficient are also interpreted using the chi square distribution.

Chi square is often used in analyzing the frequency distribution observed in a sample to find either (a) whether or not the observed distribution fits or corresponds significantly with a proposed distribution like the Mendelian 9:3:3:1 distribution, the binomial distribution and the normal distribution, or (b) whether or not two given variables are significantly associated with one another. Accordingly, two different chi square tests, viz., a test for goodness of fit and a test of independence, are respectively used. The computed χ^2 is considered significant, if the computed χ^2 either exceeds or equals the critical χ^2_α. On the contrary, if the computed χ^2 is lower than the critical χ^2_α, it is considered not significant.

(a) *Chi square test for goodness of fit* :

A test for goodness of fit is intended to find whether or not a frequency distribution observed in the sample fits significantly with a specific proposed distribution. In this test, the observed frequencies (f_0) of cases in the distribution found in the sample are tabulated according to its respective classes. The corresponding expected frequencies (f_e) of those classes are worked out on the basis of the proposed distribution. The difference (f_0-f_e) between the observed and expected frequencies is worked out for each class and then used in computing the chi square.

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}.$$

The *df* of the computed χ^2 is given by the difference between the total number (k) of classes in the frequency distribution and the number of classes which lose their freedom for change to keep the sample size and the parameters of the proposed distribution constant. Thus, the computed χ^2 has the *df* of k-1, k-2 or k-3, according as the proposed distribution used in its computation consists of a Mendelian distribution, a binomial distribution or a normal distribution. If the computed χ^2 has the *df* of 1 only and the f_e of one or more classes is lower than 5, *Yates' correction* is applied to the (f_0-f_e) differences of *all the classes*, reducing each absolute difference without its algebraic sign, i.e., each $|f_0-f_e|$, by 0.5. Thus, for each class,

corrected $(f_0-f_e) = |f_0-f_e| - 0.5$,

$$\chi^2 = \Sigma \frac{(|f_0 - f_e| - 0.5)^2}{f_e}$$

The H_a proposes that there is a significant goodness of fit between the observed and proposed distributions. In other words, according to the H_a, there is no significant difference between the distributions so that the χ^2 computed from such difference is not significant. The H_0 contends, on the contrary, that there is no significant goodness of fit between the distributions. In effect, the H_0 proposes that there is a significant difference between the distributions and the χ^2 computed from such difference is also significant. To test this, the computed χ^2 is compared with the critical χ^2_α of a chosen significance level. If the computed χ^2 is found to be lower than the critical χ^2_α, it is considered not significant ($P>\alpha$); it is then inferred that the difference between the two distributions is not significant so that there is a significant goodness of fit between them. But if the computed χ^2 equals or exceeds the critical χ^2_α, it is considered significant ($P\leq\alpha$); it is then inferred that there is a significant difference between the distributions so that there is no significant goodnes of fit between them.

Example 2.8.1.

Find whether or not the following phenotype distribution in a sample of 96 flies from a particular housefly population has a goodness of fit with the Mendelian 9:3:3:1 distribution. (α=0.05).

Phenotypes :	Grey body red eye	Black body red eye	Grey body scarlet eye	Black body scarlet eye
Frequencies :	60	16	15	5

Critical χ^2_α: $\chi^2_{0.05(1)} = 3.84$; $\chi^2_{0.05(2)} = 5.99$; $\chi^2_{0.05(3)} = 7.82$; $\chi^2_{0.05(4)} = 9.49$.

Solution :

The obseved frequencies (f_0) of different phenotypes are entered in Table 2.5. The expected frequencies (f_e) are computed using the sample size (n) and the proportionate frequencies of phenotypes as proposed by the Mendelian 9:3:3:1 distribution. Thus, for the grey-body red-eye phenotype having a proportionate frequency of 9/16, f_e=96×9/16=54. The (f_0-f_e) of each phenotype is thus worked out and used in computing the chi square (Table 2.5).

Table 2.5. Table for computing chi square for a test for goodness of fit. (n=96)

Phenotypes	f_0	f_e	f_0-f_e	$(f_0-f_e)^2$	$(f_0-f_e)^2/f_e$
Grey red	60	54	+6	36	0.67
Black red	16	18	-2	4	0.22
Grey scarlet	15	18	-3	9	0.50
Black scarlet	5	6	-1	1	0.17
Total	96 (n)	96	-	-	1.56

$$\chi^2 = \Sigma\frac{(f_0 - f_e)^2}{f_e} = 1.56 \text{ (from Table 2.5).}$$

$df = k\text{-}1 = 4 - 1 = 3$.

Critical $\chi^2_{0.05(3)} = 7.82$.

The computed χ^2 is found to be lower than the critical $\chi^2_{0.05}$. So, the computed χ^2 is not significant (P>0.05). Hence, there is no significant difference between the two distributions. In other words, the observed

distribution has a significant goodness of fit with the proposed Mendelian distribution.

(b) *Chi square test of independence :*

This test is used to find whether there is a significant association between two given variables, or they are independent of each other. The H_a proposes here that there is a significant association between the variables; the H_0 contests the H_a by proposing that there is no significant association between the variables, i.e., the variables are independent of each other.

The data consist of observed frequencies (f_0) of cases in each class corresponding to a given combination of the two variables. The f_0 of each such class is entered in the relevant cell of a contingency table in which the classes of one variable run along its c number of columns and those of the other variable along its r number of rows. The df of the χ^2 is given by : $(r\text{-}1)(c\text{-}1)$. The total (f_r) of the f_0 scores of each row and the total (f_c) of the f_0 scores of each column are computed and entered in the marginal column for f_r and the marginal row for f_c respectively. [See Table 2.6.]

The expected frequencies (f_e), corresponding to the f_0 scores of only that number of cells as is given by the df, are computed in terms of the H_0, using the sample size (n) and the f_r and f_c scores for the respective cells : $f_e = f_r f_c / n$. The f_e scores of the remaining cells are obtained as the differences between the already obtained f_e scores and either the f_r or the f_c for the respective cells.

The difference ($f_o - f_e$) between each f_0 and the corresponding f_e is then worked out and used along with the same f_e in computing the chi square. (See Example 2.8.2).

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}.$$

Where the computed χ^2 has the df of 1 and one or more f_e scores are lower than 5, each ($f_0 - f_e$) difference is first corrected by *Yates' correction* in the same way as in similar cases in χ^2 test for goodness of fit. The corrected ($f_0 - f_e$) differences are then used in computing χ^2.

If the computed χ^2 either exceeds or equals the critical χ^2_α for a chosen level of α which is not above 0.05, the computed χ^2 is considered significant, the H_0 is rejected and it is inferred that there is a significant association between the given variables ($P \leq \alpha$). But if the computed χ^2 is lower than the critical χ^2_α, it is not significant and there is no significant association between the variables ($P > \alpha$).

Example 2.8.2.

In a sample of 50 persons of Polynesian race, 12, 18, 11 and 9 persons had respectively A, B, AB and O blood groups. Out of a sample of 50 persons of Mongolian race, on the contrary, 18, 12, 9 and 11 persons had those respective blood groups. Find whether or not there is a significant association between blood group and race, choosing the appropriate critical χ^2_α from among the following. ($\alpha = 0.05$).

Critical χ^2_α: $\chi^2_{0.05(1)} = 3.84$; $\chi^2_{0.05(2)} = 5.99$; $\chi^2_{0.05(3)} = 7.82$; $\chi^2_{0.05(4)} = 9.49$; $\chi^2_{0.05(10)} = 18.31$.

Solution :

The observed frequencies (f_0) are entered in the relevant cells of a contingency table (Table 2.6). The sum of the f_0 scores of each row and that of the f_0 scores of each column are entered in the f_r column and the f_c row of the table, respectively. The *df* of the chi square is computed as follows, using the r number of rows and c number of columns containing the f_0 scores.

$r = 2$; $c=4$; $df = (r-1)(c-1) = (2-1)(4-1) = 3$.

Table 2.6 : Contingency table for association between blood groups and races.

Races	A group		B group		AB group		O group		Total
	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	(f_r)
Polynesian	12	15	18	15	11	10	9	10	50
Mongolian	18	15	12	15	9	10	11	10	50
Total (f_c)	30	30	30	30	20	20	20	20	100 (n)

As per the computed *df* of 3, the f_e of the following 3 cells are computed in terms of the H_o, using the respective f_r and f_c scores and the sample size (n), and entered in Table 2.6.

(a) Polynesians with A group : $f_e = \frac{f_r f_c}{n} = \frac{50 \times 30}{100} = 15$.

(b) Mongolians with B group : $f_e = \frac{f_r f_c}{n} = \frac{50 \times 30}{100} = 15$.

(c) Mongolians with O group : $f_e = \frac{f_r f_c}{n} = \frac{50 \times 20}{100} = 10$.

The remaining f_e scores are obtained as the differences between the f_r or f_c scores and the already computed f_e scores. For example, for Polynesians with B group,

f_e = (f_c of B group) - (f_e of Mongolians with B group)

= 30 - 15 = 15.

The χ^2 is then computed using each $(f_o - f_e)$ difference and the corresponding f_e

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = \frac{(12-15)^2}{15} + \frac{(18-15)^2}{15} + \frac{(11-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(18-15)^2}{15}$$

$$+ \frac{(12-15)^2}{15} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} = 2.80.$$

df=3. Critical $\chi^2_{0.05(3)} = 7.82$.

Because the computed χ^2 is lower than the critical $\chi^2_{0.05}$, the probability P of the correctness of the H_0 is higher than 0.05. So, the H_0 cannot be rejected, the computed χ^2 is considered not significant, and it is inferred that there is no significant association between blood groups and the races (P>0.05).

2.9. SUMMARY

In this unit, it has been explained how the significance of experimental results may be explored. In each experiment, the null hypothesis contests the experimental or alternative hypothesis by contending that the observed results are not significant, having arisen by chance due to the use of a particular random sample. Only if the probability of correctness of the null hypothesis, worked out in a significance test, does not exceed a chosen maximum limit, called the significance level, the null hypothesis is rejected and the observed results are considered significant. Critical t scores for given significance levels have such fractional areas beyond them in their probability distributions as equal the respective significance levels. An observed result is considered significant only if the t score computed from it either exceeds or equals the critical t score for the relevant significance level. A two-tail t test is done, using two-tail critical t scores, for finding whether or not two group means differ significantly; a one-tail t test uses one-tail critical t scores to find whether or not one group mean is significantly higher than the other.

The difference between the means of two large independent groups, drawn from a normally distributed population, is transformed into z score and the probability of its random occurrence is obtained as the fractional area of the unit normal curve in its tail(s) beyond the computed z. Only if this probability does not

exceed 0.05, the difference is considered significant. For small groups, the t score computed from the difference between their means is compared with critical t_α for a significance level not exceeding 0.05. For small independent groups, the t score is computed using the pooled SD of both the groups. In small single-group experiments, t is computed from the differences between the paired scores of individuals to avoid the need for computing the correlation coefficient between them.

Chi square tests are nonparametric analysis of frequencies for testing either the goodness of fit between two distributions or the association between two variables. The computed χ^2 is compared with the critical χ^2 scores for chosen significance levels to find its significance for making the inference.

2.10. GLOSSARY

Alternative hypothesis (H_a): The hypothesis, sought to be proved by the undertaken experiment and contested by the null hypothesis.

Chi square (χ^2) : A nonparametric statistic for the analysis of frequencies in testing either the goodness of fit between two frequency distributions or the association between two variables.

Critical score : A specific score of a statistic such as z score, t score or chi square score, beyond which lies that fractional area in the tail of its probability distribution, which corresponds to a given level of significance.

Dependent variable : That variable in an experiment, whose changes on exposure to independent variable(s) are being explored.

Error of inference, type I : Wrong inference made by rejecting a correct null hypothesis, its probability equalling the level of significance used.

Error of inference, type II : Wrong rejection of significant observations as not significant, made by the acceptance of a wrong null hypothesis.

Goodness of fit test : Chi square test for finding whether or not two frequency distributions fit significantly with each other.

Independence test : Chi square test for finding whether or not two variables are independent of each other.

Independent group experiment : An experiment using separate and unmatched groups of animals for treatment with different levels of independent variable(s).

Independent variable : The variable whose possible effects on the dependent variable are studied in an

experiment.

Level of significance (α) : That highest level of the probability of the null hypothesis being correct, upto which the null hypothesis is rejected and the experimental results are considered significant.

Nonparametric statistic : A statistic like chi square or Spearman's rho, in whose computation no other statistic is used as an estimate of any population parameter.

Null hypothesis (H_0) : The hypothesis which contests the alternative hypothesis by contending that the observed experimental results are not significant and have arisen by chance only.

One-tail test : A test about whether or not one group mean is significantly higher (or lower) than another, in an experiment.

Parametric statistic : A statistic like *z, t, SD* and *SE,* whose computation needs the use of some other statistic as an estimate of a population parameter.

Single-group experiment : An experiment using the same group of animals for successive applications of different levels of the independent variable.

t test : A test for significance of a difference, a correlation coefficient, etc., by comparing their *t* scores with critical *t* scores.

Two-tail test : A test about whether or not the difference between two group means, irrespective of its algebraic sign, is significant.

Yates' correction : A correction needed if an expected frequency is too low in a chi square test.

2.11. QUESTIONS

1. (a) Describe the two contesting hypotheses which are tested for the significance of experimental results, citing examples.

   (b) Discuss the importance of levels of significance and critical scores in the testing of hypothesis.

   (c) Explain one-tail and two-tail tests of significance using examples.

2. (a) Explain what you mean by single-group and independent group experiments.

   (b) Write about the null hypothesis for a two-tail *t* test.

   (c) Describe the assumptions for the *t* test.

(d) Compute Student's t using the pooled SD of the two small independent groups whose scores are cited below, to test whether the two group means differ significantly, choosing the correct critical t score ($\alpha = 0.05$).

Cockroach winglength scores (mm) :

Group 1 : 32, 34, 37, 38, 39, 40, 41, 42,

Group 2 : 20, 22, 26, 28, 30, 31, 32.

2-tail critical t scores : $t_{0.01(15)} = 2.947$; $t_{0.05(14)} = 2.145$; $t_{0.01(13)} = 3.012$; $t_{0.05(15)} = 2.131$;

$t_{0.05(13)} = 2.160$; $t_{0.01(14)} = 2.977$.

1-tail critical t scores : $t_{0.01(15)} = 2.602$; $t_{0.05(14)} = 1.761$; $t_{0.01(13)} = 2.650$; $t_{0.05(15)} = 1.753$;

$t_{0.05(13)} = 1.771$; $t_{0.01(14)} = 2.624$.

3. (a) Describe the null hypothesis for a one-tail test of significance of the difference between two group means.

(b) Write about the assumptions for significance tests using z scores.

(c) Describe, mentioning computational formulae, how you would compute the z score and test the significance of the difference between two group means in a two-tail test.

(d) Compute the z score of the difference between the group means of the following cockroach winglength scores (mm) and test whether or not the Group 1 mean is significantly higher than the Group 2 mean, choosing the area of the unit normal curve from its μ to the computed z score ($\alpha = 0.01$).

Group 1 : 25, 27, 27, 29, 34, 34, 35, 37, 38, 38, 40, 40, 43, 43.

Group 2 : 20, 22, 26, 28, 30, 32, 33, 33, 38, 38, 42, 42.

z scores :	1.17	1.18	1.19	1.20	1.21
Areas :	0.3790	0.3810	0.3830	0.3849	0.3869

4. (a) Describe the null hypothesis for a one -tail t test.

(b) Describe, mentioning computational formulae, how you would compute Student's t by the difference

method and use it to find whether or not one group mean is significantly lower than the other, in an experiment using a small single group.

(c) Using the following data obtained from a sample of 7 lobsters, work out the *t* test by difference method to find whether or not the mean Ca^{2+} concentration (mg per kg of body-fluids) changes significantly on administration of calcitonin to them ($\alpha = 0.01$).

Animal No.	1	2	3	4.	5	6	7
Body-fluid Ca^{2+}							
(i) before calcitonin :	10.2	14.6	13.0	15.6	11.4	14.7	12.9
(ii) after calcitonin :	13.3	15.2	14.3	16.8	13.0	15.3	14.0

Critical t scores (one-tail) : $t_{0.01(12)} = 2.681$; $t_{0.01(6)} = 3.143$;

Critical t scores (two-tail) : $t_{0.01(12)} = 3.055$; $t_{0.01(6)} = 3.707$.

5. (a) Explain what is meant by a chi square test for goodness of fit.

(b) Describe how you would work out a chi square test for goodness of fit, mentioning the computational formula and the test for significance of the computed statistic.

(c) When and how would you apply Yates' correction in a chi square test?

(d) How would you find the *df* for the computed chi square for goodness of fit with a binomial distribution, a normal distribution and a Mendelian 9:3:3:1 phenotype distribution, respectively?

(e) Find whether or not the following observed distribution of phenotypes in a sample of 384 Drosophila flies has a significant goodness of fit with the proposed Mendelian 9:3:3:1 distribution ($\alpha = 0.05$).

Phenotypes :	AB	Ab	aB	ab	*Total*
Number of animals :	232	76	58	18	384

Critical χ^2 *values* : $\chi^2_{0.05(2)} = 5.99$; $\chi^2_{0.05(3)} = 7.82$; $\chi^2_{0.05(1)} = 3.84$; $\chi^2_{0.05(4)} = 9.49$.

6. (a) Explain what you mean by parametric statistics and nonparametric statistics, with examples.

   (b) How would you determine the degree of freedom of the computed chi square in a test of independence?

   (c) Describe how you would compute the chi square and test its significance for testing the presence or absence of association between two variables.

   (d) Out of 20 diabetics, 10 were hypertensive while the others had normal BP; out of 30 nondiabetics, 10 were hypertensive while the remaining 20 had normal BP. Use the chi square test to find if there is a significant association between diabetes and hypertension ($\alpha = 0.05$).

Critical χ^2 values : $\chi^2_{0.05(1)} = 3.84$; $\chi^2_{0.05(2)} = 5.99$; $\chi^2_{0.05(3)} = 7.82$; $\chi^2_{0.05(4)} = 9.49$.

BIBLIOGRAPHY FOR FURTHER READING

1. Das, D. and Das, A., *Statistics in Biology and Psychology,* 4th edition, Academic Publishers, Kolkata.

2. Sokal, R.R. and Rohlf, F.J., *Introduction to Biostatistics,* 2nd edition, W.H. Freeman & Co., San Francisco.

3. Sokal, R.R. and Rohlf, F.J., *Biometry*, W.H. Freeman & Co., San Francisco.

VIDYASAGAR UNIVERSITY

DIRECTORATE OF DISTANCE EDUCATION
MIDNAPORE - 721 102

M.Sc. in Zoology

Part-I

Paper-III Group-B Unit-II

Module No. - 36

Unit - I : CORRELATION AND REGRESSION

1.0 OBJECTIVES :

After studying this unit, you should be able to do the following –

* understand what is correlation and describe its different types,
* understand the assumptions for product-moment correlation and describe its properties,
* work out the product-moment *r* between two variables and find the significance of computed *r*,
* understand where Kendall's tau has to be computed instead of product-moment *r* for correlation,
* work out Kendall's tau between two variables and find its significance,
* understand what is regression and what are its different types and models,
* understand the assumptions for simple linear regression and describe its properties,
* work out the linear regression equation of one variable on another,
* understand when and how partial and multiple linear correlations are worked out, and
* understand when and how multiple linear regression equation is computed.

1.1 CORRELATION :

Correlation is the measure of the association or relation between the changes of two or more variables in the individuals of a population or a sample from that population. For example, a correlation between the scores of height and weight in the individuals of a sample assesses whether or not the changes in height are significantly associated with changes in weight and if so, to what

degree and in which direction - same or opposite. The statistic of correlation, known as the **correlation coefficient**, thus gives both the strength and the algebraic sign of the association between the changes of two or more variables. The correlation coefficient ranges from −1.00 to +1.00; the maximum negative correlation is indicated by a correlation coefficient of −1.00 while the highest possible positive correlation is indicated by a correlation coefficient of +1.00. Correlation is **positive** if a rise in one variable is associated with or accompanied by a rise in the other variable, while it is **negative** if a rise in one variable is associated with the fall in the other. Correlation is **linear** if the **scattergram** consisting of scattered points, obtained by plotting the scores of one of the variables against those of the other, conforms to a straight line; but it is **nonlinear** where the scattergram follows a nonlinear course such as elliptical or hyperbolic. Correlation is either **simple** or **multiple**, depending on whether two or more than two variables are correlated. Two methods of simple linear correlation are described in Sections 1.2 and 1.3 given below. However, presence of a significant correlation between two variables does not mean that the changes of either constitute the cause for the changes of the other. Moreover, the correlation coefficient, worked out and found significant for a particular section of the population or under a given set of conditions, may not hold good for other sections of the population or under other conditions. For example, a correlation coefficient between height and weight, worked out using a sample of newborn animals, may not hold good for full-grown animals of the same species.

1.2 PEARSON'S PRODUCT-MOMENT CORRELATION :

The product-moment correlation coefficient or **Pearson's r** is a **simple linear correlation coefficient**. It serves to determine the strength or magnitude as well as the direction or algebraic sign of the association between two continuous measurement variables, provided the relationship between their scores conforms to a straight line. Product-moment *r* cannot be used directly as multiple correlation between more than two variables, nor if one or both the variables is/are discontinuous, ordinal or qualitative.

The **assumptions** to be justifiable before you may use the product-moment *r* for correlation are as follows: that both variables give continuous metric data, that scores of both are normally or almost normally distributed in the population, that neither has a very skewed distribution, that the

paired scores of the two variables for each individual occur at random in the sample and beyond the influence of other such paired scores, and that there is a linear association between the variables.

As far as the **properties** of *r* are concerned, it ranges from –1.00 to +1.00, the two extreme scores being measures of maximum negative and positive correlations respectively, while a score of 0.00 signifies the absence of linear correlation; its score is not changed on adding or subtracting a constant number to or from all the scores of one or both variables, and also on multiplying or dividing all the scores by a constant number; it is given by those proportions of total variances of the two variables as are associated with each other, and consequently depends on the covariance of the variables, viz., **Cov(X, Y)**; the *r* scores of samples from the same population lie dispersed, due to their different sampling errors, in a sampling distribution around the parametric correlation coefficient (ρ) – this sampling distribution of *r* scores is either positively or negatively skewed according as ρ has a negative or positive value, but has no skewness when ρ amounts to zero; although *r* must be significant before a linear regression of one of the correlated variables may be worked out on the other variable, *r* has no direct predictive value, nor does it indicate that the changes of either of the correlated variables constitute the cause of the changes of the other.

Product-moment *r* may be computed by any of the following formulae :

(a) $$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}},$$

where $\sum X$ and $\sum Y$ are the respective sums of scores of the two variables, $\sum X^2$ and $\sum Y^2$ are the sums of squared scores of the respective variables, $\sum XY$ is the sum of the products of the X and Y scores of each individual, and n is the sample size, i.e., the number of individuals in the sample.

(b) $$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}},$$

where $(X - \bar{X})$ and $(Y - \bar{Y})$ are the differences of each X and each Y scores from their respective means, $\sum (X - \bar{X})(Y - \bar{Y})$ is the **sum of products**, i.e., the total of the products of $(X - \bar{X})$ and

$(Y-\bar{Y})$ of each individual in the sample, and $\sum(X-\bar{X})^2$ and $\sum(Y-\bar{Y})^2$ are the sums of squared deviations (sums of squares) of the X and Y scores from their respective sample means

(c) $$r=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)s_X s_Y},$$

where s_X and s_Y are unbiased standard deviations of X and Y scores, respectively, computed using $(n-1)$ instead of n in the denominator, as given below :

$$s_X=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}, \qquad s_Y=\sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}}$$

For correlations, the alternative hypothesis (H_a) proposes that there is a significant correlation between the two given variables. The **null hypothesis** (H_o) contests this H_a and proposes that there is no significant correlation between the given variables, and that the observed correlation coefficient would not have resulted if the investigations were done using the entire population. To find the probability P of this H_o being correct, the computed r is subjected to a significance test. For this, the computed r is transformed into the Student's t score using the SE of r, viz., s_r, and the computed t is compared with critical t scores for different levels of significance (α).

$$s_r=\sqrt{\frac{1-r^2}{n-2}}; \qquad t=\frac{r}{s_r}; \qquad df=n-2$$

Only if the computed t score is either higher than or equal to the critical t score (same df as of the computed t) for a chosen level of significance not higher than 0.05, the P of correctness of the H_o is considered too low $(P \le \alpha)$ and it is inferred that there is a significant correlation between the given variables. On the contrary, if the computed t is lower than the critical score for either a chosen α or the α of 0.05, then the H_o has high probability of being correct and it is, therefore, inferred that there is no significant correlation between the given variables ($P>\alpha$).

Example 1.2.1

Find whether or not there is a significant correlation (α = 0.01) between gill weight and body weight, using the following data from a sample of crabs and choosing the correct critical t score out of those quoted below.

Individual No.	1	2	3	4	5	6	7	8	9	10
Gill weight (g) :	0.10	0.18	0.20	0.23	0.12	0.16	0.39	0.34	0.14	0.34
Body weight (g) :	9.10	14.20	9.45	15.18	12.00	14.25	15.48	12.80	12.20	15.50

critical t scores : $t_{0.01(19)} = 2.861$; $t_{0.01(20)} = 2.845$; $t_{0.01(18)} = 2.878$; $t_{0.01(8)} = 3.355$.

Solution :

(a) Using the sum of products and the sums of squares :

The scores of gill weight (X) and body weight (Y) are entered in Table 1.1 for computing r using the sum of products.

Table 1.1 : Table for correlating gill weights and body weights using the sum of products. (n=10).

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
0.10	9.10	– 0.12	0.0144	– 3.92	15.3664	0.4704
0.18	14.20	– 0.04	0.0016	+1.18	1.3924	– 0.0472
0.20	9.45	– 0.02	0.0004	– 3.57	12.7449	0.0714
0.23	15.18	+0.01	0.0001	+2.16	4.6656	0.0216
0.12	12.00	– 0.10	0.0100	– 1.02	1.0404	0.1020
0.16	14.25	– 0.06	0.0036	+ 1.23	1.5129	– 0.0738
0.39	15.48	+ 0.17	0.0289	+ 2.46	6.0516	0.4182
0.34	12.80	+ 0.12	0.0144	– 0.22	0.0484	– 0.0264
0.14	12.20	– 0.08	0.0064	– 0.82	0.6724	0.0656
0.34	15.50	+ 0.12	0.0144	+ 2.48	6.1504	0.2976
Σ 2.20	130.16	–	0.0942	–	49.6454	1.2994

$$\bar{X} = \frac{\sum X}{n} = \frac{2.20}{10} = 0.22\text{g.} \qquad \bar{Y} = \frac{\sum Y}{n} = \frac{130.16}{10} = 13.02\text{g.}$$

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}} = \frac{1.2994}{\sqrt{0.0942 \times 49.6454}} = +0.60$$

(b) Using the standard deviations :

Table 1.1, used in the preceding method, is applicable for this method also. $\bar{X}$, $\bar{Y}$, $(X-\bar{X})^2$, $(Y-\bar{Y})^2$ and $(X-\bar{X})(Y-\bar{Y})$ are all computed as before. The standard deviations are then computed using $\sum(X-\bar{X})^2$ and $\sum(Y-\bar{Y})^2$ obtained from Table 1.1.

$$s_X = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{0.0942}{10-1}} = 0.102\text{g.} \qquad s_Y = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}} = \sqrt{\frac{49.6454}{10-1}} = 2.349\text{g}$$

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)s_x s_y} = \frac{1.2994}{(10-1)\times 0.102 \times 2.349} = +0.60$$

(c) Using the raw scores :

The scores of gill weight (X) and body weight (Y) are entered in Table 1.2 which is then used for further computations.

Table 1.2 : Table for correlating gill weights and body weights using raw scores. (n=10).

	X	Y	X^2	Y^2	XY
	0.10	9.10	0.0100	82.8100	0.9100
	0.18	14.20	0.0324	201.6400	2.5560
	0.20	9.45	0.0400	89.3025	1.8900
	0.23	15.18	0.0529	230.4324	3.4914
	0.12	12.00	0.0144	144.0000	1.4400
	0.16	14.25	0.0256	203.0625	2.2800
	0.39	15.48	0.1521	239.6304	6.0372
	0.34	12.80	0.1156	163.8400	4.3520
	0.14	12.20	0.0196	148.8400	1.7080
	0.34	15.50	0.1156	240.2500	5.2700
Σ	2.20	130.16	0.5782	1743.8078	29.9346

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}} = \frac{10 \times 29.9346 - 2.20 \times 130.16}{\sqrt{\left[10 \times 0.5782 - (2.20)^2\right]\left[10 \times 1743.8078 - (130.16)^2\right]}} = 0.60$$

The r, computed by any of the above methods, is transformed into t score using the $SE(s_r)$.

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.60)^2}{10-2}} = 0.283. \quad t = \frac{r}{s_r} = \frac{0.60}{0.283} = 2.120$$

$df = n - 2 = 10 - 2 = 8.$

Critical $t_{0.01(8)} = 3.355.$

As the computed t is lower than the critical t, $P > \alpha$. So there is no significant correlation between gill weight and body weight ($P > 0.01$).

Example 1.2.2

Work out the product-moment r between oxygen consumption (ml/minute) and tracheal ventilation (ml/minute) of the following sample of ten beetles, using the sum of products. Test the significance of the computed r, choosing the correct critical t score out of the critical t scores quoted below ($\alpha = 0.05$).

Individual No. :	1	2	3	4	5	6	7	8	9	10
O_2 consumption (X) :	3.7	2.7	3.4	2.5	2.4	3.9	3.3	2.3	3.0	2.8
Tracheal ventilation (Y) :	86.5	74.2	84.6	71.0	70.6	88.1	82.5	65.0	77.2	75.3

Critical t scores : $t_{0.05(20)} = 2.086$; $t_{0.05(19)} = 2.093$; $t_{0.05(18)} = 2.101$; $t_{0.05(10)} = 2.228$; $t_{0.05(9)} = 2.262$; $t_{0.05(8)} = 2.306$.

Solution :

The scores of O_2 consumption (X ml) and of tracheal ventilation (Y ml) are entered in Table 1.3 which is used for further computations.

Table 1.3 : Table for correlating oxygen consumption and tracheal ventilation using the sum of products ($n = 10$).

	X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
	3.7	86.5	+ 0.7	0.49	+ 9.0	81.00	6.30
	2.7	74.2	– 0.3	0.09	– 3.3	10.89	0.99
	3.4	84.6	+ 0.4	0.16	+ 7.1	50.41	2.84
	2.5	71.0	– 0.5	0.25	– 6.5	42.25	3.25
	2.4	70.6	– 0.6	0.36	– 6.9	47.61	4.14
	3.9	88.1	+ 0.9	0.81	+ 10.6	112.36	9.54
	3.3	82.5	+ 0.3	0.09	+ 5.0	25.00	1.50
	2.3	65.0	– 0.7	0.49	– 12.5	156.25	8.75
	3.0	77.2	0	0	– 0.3	0.09	0
	2.8	75.3	– 0.2	0.04	– 2.2	4.84	0.44
Σ	30.0	775.0	–	2.78	–	530.70	37.75

$$\bar{X} = \frac{\sum X}{n} = \frac{30}{10} = 3.0 \text{ ml.} \qquad \bar{Y} = \frac{\sum Y}{n} = \frac{775}{10} = 77.5 \text{ ml.}$$

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}} = \frac{37.75}{\sqrt{2.78 \times 530.70}} = +0.98.$$

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.98)^2}{10-2}} = 0.070. \qquad t = \frac{r}{s_r} = \frac{0.98}{0.070} = 14.000.$$

$df = n - 2 = 10 - 2 = 8$

Critical $t_{0.05(8)} = 2.306$.

As the computed t is higher than the critical t, $P<\alpha$. So, there is a significant correlation between oxygen consumption and tracheal ventilation ($P<0.05$).

1.3 KENDALL'S RANK CORRELATION:

This is a **nonparametric** method for finding the strength and the direction of linear correlation, even in quite small samples, between either the ranks of individuals with respect to ordinal variables (e.g., ferocity, prowess and trustworthiness) or the ranks allotted to the scores of discontinuous measurement variables (e.g., heart rate, litter size and respiratory rate) or the ranks given to even non-normally distributed scores of continuous measurement variables. But it cannot be used for correlation where one or both variables are nominal or qualitative variables (e.g., sex and race) nor for correlating more than two variables. Thus, the rank correlation coefficient, called Kendall's *tau* (τ), is a **rank statistic** for **simple linear correlations.** It is, however, less powerful than the product-moment *r*. Its values range from –1.00 to +1.00 — the two extreme scores indicating respectively the maximum negative and the maximum positive correlations. Absence of correlation is indicated by the zero value of *tau*. Kendall's *tau* is widely applicable because of its very few assumptions which are as follows: that there is a linear association between the variables, and that the paired ranks of each individual with respect to the two variables occur at random in the sample, unaffected by the ranks of any other individual in the sample. Kendall's *tau* should be computed only if the assumptions for product-moment *r* are not justifiable.

For computing *tau*, ranks are first given in an ascending order to the scores of each variable separately — this step is not required if the individuals of the sample already bear ranks with respect to an ordinal or ranked variable. If two or more scores of a variable are identical (*tied scores*), each of them is given an identical *average rank* which is the mean of the separate ranks they would have got if they were not identical. The next score following the tied set is given the same rank that it would have got if the preceding scores were not tied and held separate ranks. You should realize that *tau* as well as other rank statistics suffers from *errors* because (i) an average rank is given to all the scores of a tied set instead of their separate individual ranks, and (ii) the scores given successive ranks differ from each other by varying amounts.

After assigning the ranks, arrange in an ascending order along a column the ranks of that variable which has no tied scores, accompanying them by the respective paired ranks of the other

variable in another column beside the former column; if both or neither of the variables have/has tied scores, the ranks of any of the two may be arranged in the first column and the respective paired ranks of the other in the second column (Tables 1.4 and 1.5).

Each rank of the second column of the paired ranks is next taken in turn as the pivotal rank, and compared to the subsequent ranks of that column, counting each such subsequent rank as 1, 0.5 or 0 according as the latter is higher than, equal to or lower than the pivotal rank. The total of the counts of such subsequent ranks for all the pivotal ranks of the second column, viz., $\sum C_i$, is used in computing *tau* (τ). In this computation, $\sum T_X$ and $\sum T_Y$ are used as *correction terms* for the errors due to average ranks given to tied scores of the respective variables (X and Y); for each tied set of every variable, $T = t(t-1)$ where t is the number of tied scores in that set, while $\sum T$ is the sum of the T scores of all the tied sets of that variable. For a sample of size n,

$$\sum T_X = \sum[t_X(t_X-1)];\ \sum T_Y = \sum[t_Y(t_Y-1)];$$

$$\tau = \frac{4\sum C_i - n(n-1)}{\sqrt{[n(n-1)-\sum T_X][n(n-1)-\sum T_Y]}}.$$

Where $n > 10$, the computed *tau* (τ) is transformed into Student's t score using the SE (s_τ) of *tau*, and the computed t is next compared with the critical $t_{\alpha(df)}$ for the chosen α for inference.

$$s_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}};\ t = \frac{\tau}{s_\tau};\ df = \infty.$$

The computed *tau* and so, the correlation between the given variables may be considered significant only if the computed t either exceeds or equals the critical $t_{\alpha(\infty)}$ for the chosen significance level (α).

Example 1.3.1

Find whether or not there is a significant correlation ($\alpha = 0.01$) between respiratory rates (X per minute) and heart rates (Y per minute) using the following data from a sample of 12 Orang utans and choosing the correct critical t from those quoted below.

Individual :	1	2	3	4	5	6	7	8	9	10	11	12
X scores :	19	15	12	16	17	20	18	13	21	23	24	14
Y scores :	75	68	70	72	74	76	75	69	80	78	80	67

Critical t scores : $t_{0.01(11)}$=3.106; $t_{0.01(\infty)}$=2.576; $t_{0.01(10)}$=3.169.

Solution:

The data are entered in Table 1.4 for assigning ranks in ascending orders (R_X and R_Y) to the scores of X and Y.

Table 1.4 Ranking of respiratory rate (X) and heart rate (Y) scores.

Respiratory rates (X)	R_X	Heart rate (Y)	R_Y
19	8	75	7.5
15	4	68	2
12	1	70	4
16	5	72	5
17	6	74	6
20	9	76	9
18	7	75	7.5
13	2	69	3
21	10	80	11.5
23	11	78	10
24	12	80	11.5
14	3	67	1

As there is no tied scores of variable X, its ranks (R_X) are arranged in ascending order in the first column of Table 1.5, entering the paired rank of each score of variable Y in the second column of that Table. Using each rank of the second column (R_Y) as the pivotal rank in turn, the subsequent ranks of that column are successively counted as 1, 0.5 or 0 as the case may be. These counts are entered and totalled in Table 1.5.

Table 1.5. Count of ranks for correlating respiratory and heart rates.

R_X	R_Y	Count of subsequent ranks	Total (C_i)
1	4	0 + 0 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	8
2	3	0 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	8
3	1	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	9
4	2	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	8
5	5	1 + 1 + 1 + 1 + 1 + 1 + 1	7
6	6	1 + 1 + 1 + 1 + 1 + 1	6
7	7.5	0.5 + 1 + 1 + 1 + 1	4.5
8	7.5	1 + 1 + 1 + 1	4
9	9	1 + 1 + 1	3
10	11.5	0 + 0.5	0.5
11	10	1	1
12	11.5		-
Σ			59.0

$n = 12.\ \sum T_X = \sum\left[t_X(t_X - 1)\right] = 0.\ \sum T_Y = \sum\left[t_Y(t_Y - 1)\right] = 2(2-1) + 2(2-1) = 4.$

$$\tau = \frac{4\sum C_i - n(n-1)}{\sqrt{\left[n(n-1) - \sum T_X\right]\left[n(n-1) - \sum T_Y\right]}} = \frac{4 \times 59.0 - 12(12-1)}{\sqrt{\left[12(12-1) - 0\right]\left[12(12-1) - 4\right]}}$$

$= + 0.80.$

$$s_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}} = \sqrt{\frac{2(2 \times 12 + 5)}{9 \times 12(12-1)}} = 0.221.$$

$$t = \frac{\tau}{s_\tau} = \frac{0.80}{0.221} = 3.620. \quad df = \infty.$$

Critical $t_{0.01(\infty)} = 2.576$.

As the computed t is higher than the critical $t_{0.01}$, $P<0.01$. So, there is a significant correlation between respiratory rate and heart rate ($P<0.01$).

1.4 REGRESSION:

Regression is the statistical method for predicting the probable score of a variable in an individual from the known or measured score(s) of one or more variables in the same individual. The variable, whose score is predicted, is called the **dependent variable**; the variable(s), whose scores are used in working out the prediction, may be called the **independent variable(s)**. The dependent variable is always a single one while the number of independent variable(s) used for the prediction may be one or more; according to the number of independent variable(s), one or more, regression is respectively **simple** or **multiple**. Thus, a **simple regression** would be done for predicting the probable score of oxygen consumption (dependent variable) of an insect on the basis of the observed score of its tracheal ventilation (independent variable); but a **multiple regression** would predict the probable score of O_2 consumption on the scores of tracheal ventilation and blood O_2 tension (both independent variables) of the same insect. Again, the method of **linear regression** is required to predict the probable score of a dependent variable on the basis of the linear association between it and the independent variable(s); but a relevant method of **nonlinear regression** (e.g., quadratic and hyperbolic) would be needed in working out the probable score of dependent variable in an individual from the observed score(s) of the independent variable(s) if the relation between the dependent and independent variables is nonlinear.

Model I regression is used in predicting the probable score of the dependent variable where a strict control is maintained by the investigator on the doses, mode and timings of administration, etc., of the independent variable(s) so as to eliminate random errors or uncontrolled variations of the latter, enabling thereby the expression of not only an interdependence but also a cause-and-effect relation between the independent and dependent variables. For example, the simple regression of diastolic blood pressure on the administered doses of a vasodilator drug would be a model I regression. **Model II regression** is done to predict the probable score of the dependent variable on the score(s) of one or more independent variables whose random errors or variations cannot be eliminated because of no planned control of the doses, applications, etc., of the latter by the investigator. Thus the model II regression, though an expression of interdependence of the independent and dependent variables, cannot serve for a cause-and-effect relation between them.

An example is the model II regression of blood insulin concentration on the blood sugar level — the independent variable, blood sugar, is not "fixed" or deliberately controlled by the investigator and is free to vary at random during the experiment. **Model III regression,** always multiple, may be used in predicting the probable score of dependent variable on the scores of more than one independent variable, some uncontrolled and random and the others controlled and "fixed". Such a model is used for the regression of pulmonary ventilation on the scores of atmospheric relative humidity and the administered doses of ephedrine, the first being a random variable and the second a "fixed" one.

An important **property** of regression is that it holds good only within the limits of the particular section of the population and other conditions, in which it has been worked out – it **cannot be extended beyond those limits and conditions.** For example, a regression of wing length on trunk length, worked out in newly hatched cockroaches, may not hold good in aging cockroaches. Secondly, the score of the dependent variable, predicted by regression, is **a probable score only** and may differ from the actual scores of the dependent variable in relevant individuals by varying amounts. Thirdly, because of such differences between the predicted scores and the actual ones, the actual scores of the dependent variable in a number of individuals, having a given score of the independent variable, lie dispersed in the form of a **distribution** around the score, predicted by regression on the given score of the independent variable. Fourthly, being based on the interdependence or association of the dependent and independent variables, regression is meaningful only if those variables have a **significant correlation**. The lower this correlation, the higher would be the differences between the predicted scores and the actual scores of the dependent variable and so, the greater the **error of estimate** in the regression. Fifthly, regressions are worked out using **regression equations** based on the associations between the dependent and independent variables; the basic statistic for a regression is a **regression coefficient** which gives the average rate of change in the dependent variable score for unit change in the independent variable score. A **regression line** is drawn by plotting the predicted scores of the dependent variable against the corresponding scores of the independent variable and may be rectilinear or curved according to the form of association between the variables. Although regression of a variable Y on a correlated variable X as well as that of X on Y may be worked out, usually the regression of that

variable is done on the other, as is more difficult of the two to be measured directly and precisely.

1.5 SIMPLE LINEAR REGRESSION :

This is the regression of a dependent variable (Y) on a **single independent variable** (X) where their association can be described in terms of a straight line and they bear a significant **linear correlation** with each other. Only models I and II of regression can apply here, according as the independent variable does not or does suffer from random errors, respectively (see Section 1.4). For working out a simple linear regression, the **assumptions** which should be justifiable are: that the variables are continuous measurement variables or are based on such variables, that their scores are distributed normally or almost normally in the population, and that the variables have a significant linear correlation with each other.

Some **properties** of simple linear regression are as follows. The graphical plot of the predicted dependent variable scores $(\hat{Y})$ against the respective independent variable scores (X) conforms to a **straight line**. Secondly, the **regression equation** is an equation for a straight line, the slope of which is given by the linear regression coefficient. For the regression of the variable Y on the variable X, the predicted $\hat{y}$ score, corresponding to the independent variable score X, is given by:

$$\hat{Y} = a_{YX} + b_{YX}X.$$

Here b_{YX} is the **regression coefficient** of Y on X, giving the slope of the regression line of Y on X and being the average change of Y scores for a unit change of X scores, while a_{YX} is the Y-intercept of the regression line and is computed from the b_{YX} itself. The regression line of X on Y is similarly given by its regression coefficient b_{XY} and the corresponding a_{XY} ; where $\hat{X}$ is the predicted X score on the Y score,

$$\hat{X} = a_{XY} + b_{XY}Y.$$

Thirdly, the two regression lines, viz., those of Y on X and X on Y respectively, are separate from each other and intersect at only one point corresponding to $\overline{X}$ and $\overline{Y}$. But the angle between the two lines declines progressively with the rise in the value of r between the variables, and the

lines coincide when r reaches the highest value of either +1.00 or −1.00. Fourthly, because the predicted score is only a very likely score of the dependent variable for all individuals possessing the corresponding independent variable score, their actual dependent variable scores form a **normal distribution** around the relevant predicted score as the mean (μ), and all such distributions are **homoscedastic**, having homogeneous variances.

The simple linear regression of a variable Y on another variable X is computed as follows. using any alternative formula for b_{YX}.

$$b_{YX} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}, \quad or, \; b_{YX} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}, \; or, \; b_{YX} = r_{YX}\frac{s_Y}{s_X}$$

where r_{YX} is the correlation coefficient between the variables, and s_Y and s_X are the standard deviations of the respective variables. The computed b_{YX} is next used in computing a_{YX} and the regression equation.

$$a_{YX} = \bar{Y} - b_{YX}\bar{X}; \quad \hat{Y} = a_{YX} + b_{YX}X.$$

The *SE of estimate* (s_{YX}), which is an estimate of the differences between the predicted $\hat{y}$ scores and the corresponding actual Y scores of the dependent variable, is given by:

$$s_{YX} = s_Y\sqrt{1-r_{YX}^2}.$$

Example 1.5.1

Compute the linear regression equation of tracheal ventilation (Y ml/minute) on oxygen consumption (X ml/minute) using the data of Example 1.2.2, obtained from a sample of ten beetles.

Solution :

The regression coefficient (b_{YX}) may be computed by any of the following alternative methods.

(a) Using the sum of products :

The data are entered in Table 1.6 for computing b_{YX} using the sum of products, $\sum(X-\bar{X})(Y-\bar{Y})$.

Table 1.6 : Table for computing the regression coefficient of tracheal ventilation (Y) on oxygen consumption (X) using the sum of products. (n = 10).

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
3.7	86.5	+0.7	+9.0	0.49	6.30
2.7	74.2	–0.3	–3.3	0.09	0.99
3.4	84.6	+0.4	+7.1	0.16	2.84
2.5	71.0	–0.5	–6.5	0.25	3.25
2.4	70.6	–0.6	–6.9	0.36	4.14
3.9	88.1	+0.9	+10.6	0.81	9.54
3.3	82.5	+0.3	+5.0	0.09	1.50
2.3	65.0	–0.7	–12.5	0.49	8.75
3.0	77.2	0	–0.3	0	0
2.8	75.3	–0.2	–2.2	0.04	0.44
Σ 30.0	775.0	-	-	2.78	37.75

$$\bar{X}=\frac{\sum X}{n}=\frac{30.0}{10}=3.0\text{ml.} \qquad \bar{Y}=\frac{\sum Y}{n}=\frac{775.0}{10}=77.5\text{ml.}$$

$$b_{YX}=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}=\frac{37.75}{2.78}=13.58$$

(b) Using the product-moment r and standard deviations :

The data are entered in Table 1.7 for computing r_{YX} and the standard deviations to be used in computing b_{YX}.

$$\bar{X}=\frac{\sum X}{n}=\frac{30.0}{10}=3.0\text{ml.} \qquad \bar{Y}=\frac{\sum Y}{n}=\frac{775.0}{10}=77.5\text{ml.}$$

$$s_Y=\sqrt{\frac{\sum(Y-\bar{Y})^2}{n-1}}=\sqrt{\frac{530.70}{10-1}}=7.68\text{ml.} \qquad s_X=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{2.78}{10-1}}=0.56\text{ml.}$$

$$r_{YX}=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)s_X s_Y}=\frac{37.75}{(10-1)\times 0.56\times 7.68}=+0.98,$$

$$b_{YX} = r_{YX}\frac{s_Y}{s_X} = 0.98 \times \frac{7.68}{0.56} = 13.44.$$

Table 1.7 : Table for computing the regression coefficient of tracheal ventilation (Y) on O_2 consumption (X), using product-moment r and standard deviations ($n = 10$).

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
3.7	86.5	+ 0.7	0.49	+ 9.0	81.00	6.30
2.7	74.2	– 0.3	0.09	– 3.3	10.89	0.99
3.4	84.6	+ 0.4	0.16	+ 7.1	50.41	2.84
2.5	71.0	– 0.5	0.25	– 6.5	42.25	3.25
2.4	70.6	– 0.6	0.36	– 6.9	47.61	4.14
3.9	88.1	+ 0.9	0.81	+ 10.6	112.36	9.54
3.3	82.5	+ 0.3	0.09	+ 5.0	25.00	1.50
2.3	65.0	– 0.7	0.49	– 12.5	156.25	8.75
3.0	77.2	0	0	– 0.3	0.09	0
2.8	75.3	– 0.2	0.04	– 2.2	4.84	0.44
Σ 30.0	775.0	–	2.78	–	530.70	37.75

(c) Using raw scores:

The data are entered in Table 1.8 for computing b_{YX} from raw scores.

Table 1.8 : Table for computing b_{YX} from raw scores ($n = 10$).

	X	Y	XY	X^2
	3.7	86.5	320.05	13.69
	2.7	74.2	200.34	7.29
	3.4	84.6	287.64	11.56
	2.5	71.0	177.50	6.25
	2.4	70.6	169.44	5.76
	3.9	88.1	343.59	15.21
	3.3	82.5	272.25	10.89
	2.3	65.0	149.50	5.29
	3.0	77.2	231.60	9.00
	2.8	75.3	210.84	7.84
Σ	30.0	775.0	2362.75	92.78

$$b_{YX} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = \frac{10 \times 2362.75 - 30.0 \times 775.0}{10 \times 92.78 - (30.0)^2} = 13.58.$$

The a_{YX}, the regression equation and the $SE(s_{YX})$ of estimate are computed, using the b_{YX} computed by any of the preceding alternative methods.

$$a_{YX} = \overline{Y} - b_{YX}\overline{X} = 77.5 - 13.58 \times 3.0 = 36.76, \quad \hat{Y} = a_{YX} + b_{YX}X = 36.76 + 13.58X.$$

$$s_{YX} = s_Y\sqrt{1 - r_{YX}^2} = 7.68\sqrt{1 - (0.98)^2} = 1.528.$$

1.6 PARTIAL AND MULTIPLE CORRELATIONS :

Correlations involving more than two variables include partial and multiple correlations.

Partial correlation

The simple linear product-moment correlation coefficient (r_{12}) between two variables (X_1 and X_2) is partly due to the correlations of each of them to other variables such as r_{13} between X_1 and variable X_3, and r_{23} between X_2 and X_3. To estimate the correlation between the two desired variables (X_1 and X_2) eliminating the effects of the correlation between either of them and the other variable X_3, the *partial correlation* ($r_{12.3}$) between the desired variables (X_1 and X_2) is worked out, eliminating or partialling out the effects of the other variable X_3 on both of them. This is known as the *first order partial correlation* between the given variables (say, $r_{12.3}$ between X_1 and X_2) partialling out the effect of a single other variable X_3. Similarly, second, third or still higher orders of partial r can be worked out between two variables (e.g. $r_{12.34}$ or $r_{12.345}$, etc.) by eliminating the effects of respectively 2, 3 or more of other variables on both X_1 and X_2.

Partial r can be worked out and used only if the following *assumptions* are justifiable: that all the relevant variables are continuous variables having either normal or almost normal distributions, that there are linear associations between them, and that their paired scores occur randomly in the sample according to the laws of probability.

The first order partial r between X_1 and X_2, partialling out X_3, is computed using the product-moment r scores (r_{12}, r_{13} and r_{23}) between the respective variables. The computed $r_{12.3}$ is next transformed into Student's t, using the SE of $r_{12.3}$ $\left(s_{r_{12.3}}\right)$. The computed t is finally compared with the appropriate critical $t_{\alpha(df)}$ for finding the probability (P) of correctness of the H_0 that the computed $r_{12.3}$ has resulted only due to the chances of random sampling.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left[1 - r_{13}^2\right]\left[1 - r_{23}^2\right]}};\quad s_{r_{12.3}} = \sqrt{\frac{1 - r_{12.3}^2}{n-3}};\quad t = \frac{r_{12.3}}{s_{r_{12.3}}};\ df = n - 3.$$

The computed $r_{12.3}$ is considered significant only if the computed t either exceeds or equals the critical t for a chosen level of significance (α), but is not significant if the critical t is higher than the computed t.

Multiple correlation

If more than two variables (say X_1, X_2 and X_3) are significantly correlated to each other, multiple correlation ($R_{1.23}$) can be worked out between any one of them (say, X_1) and the combination of the others (say, X_2 and X_3). In such a case, X_1 would be called the *criterion* while X_2 and X_3 would be the *predictors*. Multiple correlation may be *linear* if the variables are linearly correlated with each other, and *nonlinear* if their relations are not linear, like hyperbolic or sigmoid.

Following *assumptions* should be justifiable for working out multiple linear correlations: that the variables are continuous variables with normal or near-normal distributions of their scores, that the scores of these variables are linearly correlated, and that their paired scores occur in the sample at random according to the laws of probability.

Using β_2 and β_3 coefficients which are proportions of the variance of X_1 associated with the variances of the respective predictors X_2 and X_3, the multiple linear correlation coefficient ($R_{1.23}$) of the criterion (X_1) with two predictors (X_2 and X_3) is worked out and tested for its significance.

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2};\quad \beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2};\quad R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}};$$

$$s_{R_{1.23}} = \frac{1}{\sqrt{n-3}};\quad t = \frac{R_{1.23}}{s_{R_{1.23}}};\quad df = n - 3$$

The $R_{1.23}$ is significant only if the t, computed from it as above, either exceeds or equals the critical $t_{\alpha(df)}$ for a chosen α, but not significant if the critical t score exceeds the computed t.

Example 1.6.1

The product-moment r scores between gill weights (X_1 gm), body weights (X_2 gm) and body lengths (X_3 cm) were found to be as follows in a sample of 33 salmons: $r_{12} = +0.65$; $r_{13} = +0.25$; $r_{23} = +0.32$. Find (a) if there is a significant partial linear correlation between gill weights (X_1) and body

weights (X_2), partialling out body lengths (X_3), and (b) if there is a significant multiple correlation between X_1 and the combination of X_2 and X_3. ($\alpha = 0.01$). Critical $t_{0.01(32)}$=2.738; $t_{0.01(30)}$=2.750.

Solution :

(a) Partial correlation:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left[1-r_{13}^2\right]\left[1-r_{23}^2\right]}} = \frac{0.65 - 0.25 \times 0.32}{\sqrt{\left[1-0.25^2\right]\left[1-0.32^2\right]}} = 0.62$$

$$s_{r_{12.3}} = \sqrt{\frac{1-r_{12.3}^2}{n-3}} = \sqrt{\frac{1-0.62^2}{33-3}} = 0.143;$$

$$t = \frac{r_{12.3}}{s_{r_{12.3}}} = \frac{0.62}{0.143} = 4.336; \quad df = n-3 = 33-3 = 30$$

Critical $t_{0.01(30)}$=2.750. Computed t > Critical $t_{0.01}$; $\therefore$ P<0.01.

Hence, there is a significant partial correlation between gill weights and bodyweights (P<0.01).

(b) Multiple correlation :

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1-r_{23}^2} = \frac{0.65 - 0.25 \times 0.32}{1-0.32^2} = 0.635;$$

$$\beta_3 = \frac{r_{13} - r_{12}r_{23}}{1-r_{23}^2} = \frac{0.25 - 0.65 \times 0.32}{1-0.32^2} = 0.047;$$

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}} = \sqrt{0.635 \times 0.65 + 0.047 \times 0.25} = +0.65.$$

$$s_{R_{1.23}} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{33-3}} = 0.183; \quad t = \frac{R_{1.23}}{s_{R_{1.23}}} = \frac{0.65}{0.183} = 3.552;$$

$df = n - 3 = 33 - 3 = 30$. Critical $t_{0.01(30)} = 2.750$.

Computed t > Critical $t_{0.01}$; $\therefore$ $P < 0.01$.

Hence, there is a significant multiple correlation between gill weights and the combination of body weights and body lengths (P<0.01).

1.7 MULTIPLE REGRESSION :

This is the statistical prediction of the likely score of a variable, also called *criterion* (say, X_1) of an individual from the measured scores of two or more other variables, also called *predictors* (say, X_2 and X_3), of the same individual, provided the variables are correlated with each other.

Multiple regression may be linear if the variables are linearly correlated, and nonlinear if their relations are nonlinear.

Assumptions for multiple linear regressions are as follows : that the variables are continuous variables with normal or almost normal distributions, that they possess linear relations with each other, and that their paired scores occur at random in the sample according to the laws of probability.

For working out a three-variable multiple linear regression of a criterion (X_1) on two predictors (X_2 and X_3), *partial regression coefficients* ($b_{12.3}$ and $b_{13.2}$) are computed using the beta coefficients (β_2 and β_3) and the standard deviations (s_1, s_2 and s_3) of the respective variables; $b_{12.3}$ and $b_{13.2}$ are the average changes of the criterion (X_1) for unit changes of the predictors X_2 and X_3, respectively, when the respective other predictor is kept constant. The vertical intercept ($a_{1.23}$) of the multiple regression line is in turn worked out using $b_{12.3}$, $b_{13.2}$ and the means of the three variables ($\overline{X}_1$, $\overline{X}_2$ and $\overline{X}_3$).

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}; \quad \beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}; \quad b_{12.3} = \beta_2 \times \frac{s_1}{s_2}; \quad b_{13.2} = \beta_3 \times \frac{s_1}{s_3};$$

$$a_{1.23} = \overline{X}_1 - b_{12.3}X_2 - b_{13.2}X_3 .$$

The multiple linear regression equation and the SE of estimate ($s_{1.23}$) are then worked out as follows.

$$\hat{X}_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3; \quad s_{1.23} = s_1\sqrt{1 - R_{1.23}^2} .$$

Example 1.7.1

Work out the multiple linear regression equation of glomerular filtration rate (X_1 ml/min) on glomerular blood pressure (X_2 mm Hg) and plasma protein osmotic pressure (X_3 mm Hg), using the following data from a sample of 52 langurs.

$\overline{X}_1 = 120$; $s_1 = 18.5$; $\overline{X}_2 = 55$; $s_2 = 10.1$; $\overline{X}_3 = 25$; $s_3 = 4.2$.

$r_{12} = +0.72$; $r_{13} = -0.26$; $r_{23} = +0.18$.

Solution:

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} = \frac{0.72 - (-0.26) \times 0.18}{1 - 0.18^2} = 0.792.$$

$$\beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} = \frac{-0.26 - 0.72 \times 0.18}{1 - 0.18^2} = -0.403.$$

$$b_{12.3} = \beta_2 \times \frac{s_1}{s_2} = 0.792 \times \frac{18.5}{10.1} = 1.45.$$

$$b_{13.2} = \beta_3 \times \frac{s_1}{s_3} = -0.403 \times \frac{18.5}{4.2} = -1.78.$$

$$a_{1.23} = \bar{X}_1 - b_{12.3}\bar{X}_2 - b_{13.2}\bar{X}_3$$

$$= 120 - 1.45 \times 55 - (-1.78) \times 25 = 84.75.$$

$$\hat{X}_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

or, $\hat{X}_1 = 84.75 + 1.45X_2 - 1.78X_3$.

1.8 SUMMARY :

Correlation, the quantitative measure of association between two or more variables, may be simple or multiple, linear or nonlinear, and positive or negative. Product-moment *r* is a simple linear correlation coefficient between two normally or almost normally distributed continuous measurement variables. The sample *r* scores form a sampling distribution around the population ρ, which is either positively or negatively skewed except when ρ is 0.00. Product-moment *r* may be computed either from raw scores directly, or using the sum of products. The significance of *r* is tested by transforming it to *t* score. Kendall's *tau* is a nonparametric simple linear correlation coefficient, computed from the ranks of the individuals in the two variables and tested for significance by conversion to *t* score.

Regression is the statistical prediction of the score of a dependent variable (criterion) from the score(s) of independent variable(s) or predictor(s). Regression may be simple or multiple, and linear or nonlinear. Different models of regression are used according to the natures of the independent variables. The regression coefficient for a simple linear regression may be computed either directly from raw scores, or using the product-moment *r* between the relevant variables, or from their sum of products. It can be worked out when the variables are significantly correlated linearly and are continuous measurement variables with normal or almost normal distributions.

Partial correlation is worked out between two variables, eliminating the effects of one or more other variables on them. Multiple linear correlation is the linear correlation between a variable (criterion) and a combination of two or more other variables (predictors). Both these types of linear correlations are worked out basically using the product-moment *r* scores between the criterion and the predictors.

Multiple linear regression is used for statistical prediction of the score of a criterion in an individual from the measured scores of two or more predictors in the latter. It is also worked out from the product-moment *r* scores between the criterion and the predictors.

1.9 GLOSSARY :

Correlation : Magnitude and direction of association between variables.

Correlation coefficient : Statistic for magnitude and sign of correlation.

Criterion : The variable whose scores are predicted by regression using the scores of one or more other variables in the same individual.

Dependent variable : Same as criterion.

Errors of estimate : Estimate of differences between the predicted and actual scores of the criterion in the individuals of a sample.

Homoscedasticity : Homogeneity of variances.

Independent variable : Variable(s) whose observed scores are used in predicting the scores of a criterion in regression.

Kendall's tau : Nonparametric linear correlation coefficient between the ranks of two variables.

Linear correlation : Correlation between variables, the association between which conforms to a straight line.

Linear regression : Regression where the criterion and the predictor(s) are linearly correlated.

Models of regression : Methods of regression according to the "fixed" or random nature of the predictor(s).

Multiple correlation : Correlation between one variable and a combination of two or more others.

Multiple regression : Regression of a criterion on two or more predictors.

Negative correlation : Correlation where the rise of one variable is associated with the fall of the other.

Nonlinear correlation : Correlation between variables which are associated nonlinearly with each other.

Nonlinear regression : Regression of a criterion on predictors correlated nonlinearly with it.

Positive correlation : Correlation where the rise of one variable is associated with the rise of the other.

Predictor : Same as independent variable.

Product-moment *r* : Linear correlation coefficient between two continuous and normally or near-normally distributed variables.

Regression : Statistical prediction of the score of a variable from the observed score(s) of one or more other variables.

Regression coefficient : The prediction statistic giving the slope of the regression line.

Regression line : The line drawn by plotting the predicted scores of the dependent variable against the corresponding observed scores of the independent variable.

Scattergram : Scattered points by plotting the scores of one variable against those of another in the respective individuals.

Simple correlation : Correlation between two variables.

Simple regression : Regression of one variable on a single other variable.

Sum of products : The sum of the products of differences between the scores of each individual and their respective means in two variables.

1.10 QUESTIONS :

1. (a) Define correlation and correlation coefficient. Explain different types of correlation.
   (b) Discuss the properties and assumptions of product-moment correlation coefficient.
   (c) Compute product-moment correlation coefficient between the following scores of total lengths (X) of aphid stem mothers and the thorax lengths (Y) of their respective parthenogenetic offsprings in a sample of ten aphid stem mothers, and find its significance ($\alpha = 0.05$).

		1	2	3	4	5	6	7	8	9	10
X	:	8.6	6.5	8.0	8.6	11.0	7.3	10.3	6.5	6.5	10.0
Y	:	5.75	4.80	5.60	6.15	6.20	5.72	5.92	4.30	4.80	6.00

Critical t scores : $t_{0.05(10)} = 2.228$; $t_{0.05(9)} = 2.262$; $t_{0.05(18)} = 2.101$; $t_{0.05(8)} = 2.306$; $t_{0.05(19)} = 2.093$.

2. (a) What is Kendall's *tau*? When would you compute it?

   (b) Describe the assumptions and methods of computation of Kendall's *tau*. How do you test the significance of the computed *tau*?

   (c) Compute Kendall's *tau* using the data given in question 1(c) above, and test its significance ($\alpha = 0.05$).

3. (a) Which correlation coefficient would you compute between litter size and mother animal's age in a group of animals? Justify you choice. Mention the assumptions for your chosen correlation coefficient.

   (b) Describe how you would compute the correlation coefficient chosen for the question 3(a) and test the significance of the computed statistic, mentioning the formulae used.

   (c) What are the sources of error in that computation? How are such errors minimized?

4. (a) What is the sum of products? Describe its use in computing product-moment r and simple linear regression, mentioning the relevant formulae.

   (b) Define scattergram and standard error of estimate. Mention their relevance in correlation and prediction statistics.

   (c) Work out the regression equation of thoracic lengths of parthenogenetic offsprings on the lengths of their respective aphid stem mothers, using the data given in question 1(c) above and with two alternative formulae based on the product-moment r and the sum of products, respectively.

5. (a) What is linear regression? Describe different models of regression, citing examples.

   (b) Discuss the properties and assumptions of simple linear regression.

   (c) Compute the linear regression equation and the error of estimate for the regression of serum cholesterol (mg/dl) on blood sugar (mg/dl), using the following data in a sample of 10 individuals.

Individual No.	:	1	2	3	4	5	6	7	8	9	10
Blood sugar (X)	:	110	90	85	205	116	140	94	120	200	220
Serum cholesterol (Y)	:	170	164	158	230	177	210	165	180	222	244

6. (a) Compare the assumptions for product-moment r and Kendall's *tau*.

   (b) Compute product-moment r between serum cholesterol and blood sugar, using the data given in question 5(c), and find its significance ($\alpha = 0.05$). Critical t scores: $t_{0.05(10)}$=2.228; $t_{0.05(9)}$=2.262; $t_{0.05(19)}$=2.093; $t_{0.05(8)}$=2.306; $t_{0.05(18)}$=2.101.

   (c) Use the same data to compute Kendall's *tau* between the given variables and test its significance ($\alpha = 0.05$).

BIBLIOGRAPHY FOR FURTHER READING :

1. Das D. and Das A., *Statistics in Biology and Psychology*, 4th edition, Academic Publishers, Kolkata.
2. Sokal, R.R. and Rohlf, F.J., *Introduction to Biostatistics*, 2nd edition, W.H. Freeman & Co., San Francisco.
3. Sokal, R.R. and Rohlf, F.J., *Biometry*, W.H. Freeman & Co. San Francisco.
4. Bliss, C., *Statistics in Biology*, Vols. I-II, McGraw-Hill, New York.
5. Ipsen, J. and Feigl, P., *Bancroft's Introduction to Biostatistics*, Harper and Row, New York.
6. Armitage, P. and Berry, G., *Statistical Methods in Medical Research*, Blackwell, Oxford.

Unit - 2 : ANALYSIS OF VARIANCES

2.0 OBJECTIVES :

After studying this unit, it should be possible for you to :

* realize the importance of variances in statistics,
* understand and explain the application of analysis of variances in the testing of hypothesis,
* identify one-way and higher orders as well as models of anova and work out the anova accordingly,
* understand and explain the assumptions for anova,
* work out the variance ratio and find its significance in the preliminary F test for one-way anova,
* follow up a significant variance ratio by multiple comparison tests between group means in pairs, and
* work out the Kruskal-Wallis one-way anova and the multiple comparison Mann-Whitney U test, and interpret them.

2.1 ANOVA :

Analysis of variances or anova consists of very powerful methods for the testing of hypothesis. Unlike t tests, anova explores the significance of differences between variances of different groups, instead of testing directly the differences between group means. It has much more rigorous assumptions than the t test and can be used simultaneously for two or more groups.

Variance (s^2) is an **absolute measure of dispersion** of raw scores around the sample (group) mean, the dispersion of the scores resulting from their varying differences (**error terms**) from the mean. You have learnt in Section 2.3 of Module 34 that it is generally computed as the mean of the squared deviations of raw scores from the group mean, using the **sum of squares** (SS) as the numerator and the df as the denominator – hence its other name the **mean square** (MS) and its unit identical with the squared unit of the raw scores.

$$s^2 \text{ or } MS = \frac{SS}{df} = \frac{\sum (X - \bar{X})^2}{n-1}$$

Variances of different samples or groups drawn from the same population are estimates of that population variance (σ^2), but still differ from each other simply due to random variations of scores of different samples. So, the difference between variances of two such samples from the same population is not genuine and can be explained away by their respective sampling errors owing to random differences between the sample scores. But variances of two samples or groups, belonging to two different populations, are estimates of the respective population variances; the difference between two such sample variances is an estimate of the difference between their respective population variances, cannot be explained away by their sampling errors, and may be considered significant.

In an experiment, different groups are exposed to an independent variable — each group to a different level of the latter — and the dependent variable is measured in the groups after such exposure. The variance of scores **within the groups**, each exposed only to a particular level of the independent variable, is the measure of the random variations of scores and their consequent **error terms**, viz., $X - \bar{X}$, and may be called the **error variance**. On the contrary, the variance of scores, belonging to different groups and known as the **between-groups variance**, may result from two causes, viz., (i) the random variations of the scores as give the **within-groups variance**, and (ii) in case the independent variable has produced changes in the dependent variable, an added variance between the groups because of different variations of scores of different groups due to exposure to different levels of the independent variable – this **added variance** would be absent in the between-groups variance and the latter would only be due to the same random variations of scores as are measured by the within-groups variance, when the independent variable has produced no change in the dependent one.

Anova is used to partition the total variance (s_t^2) of the dependent variable scores into their between-groups variance (s_b^2) and within-groups variance (s_w^2), and to explore whether or not the computed s_b^2 is significantly higher than the computed s_w^2; the **variance ratio** (F ratio) between s_b^2 and s_w^2 is computed in anova and its significance is tested with reference to critical F_α scores — an F ratio found significant indicates that the s_b^2 is significantly higher than s_w^2 and has a significant added

variance so that it may be inferred that the levels of independent variable have produced significant differences in the dependent variable scores of different groups ($P \leq \alpha$).

Anova would be a **one-way anova** when there is a single independent variable involved, a **two-way anova** where two independent variables are involved, a **three-way anova** for an experiment having three independent variables, and so on. Thus, a one-way anova has to be worked out for the data of an experiment studying the effects of three levels of a pesticide (independent variable) on the oxygen consumption (dependent variable) in a sample of insects; a two-way anova is needed for the findings of an experiment about the effects of different levels of combinations of a pesticide and an insect hormone, on the oxygen consumption in a sample of insects.

Different models of anova have to be used according to the nature(s) of the independent variable(s) used in the relevant experiment. **Model I anova** is applied to the dependent variable scores where the independent variable(s) is/are **"fixed" experimental treatment(s)**, controlled by the investigator and consequently, not suffering from random errors. **Model II anova** is meant for the observed results of an experiment using **random** independent variable(s), not controlled by the investigator and suffering from random errors. **Model III anova** is a multi-way anova used for the data of an experiment with two or more independent variables, some "fixed" and other(s) random. Examples: model I anova between oxygen consumption and controlled ("fixed") levels of a pesticide in a sample of insects; model II anova between oxygen consumption and uncontrolled levels of serum concentration of an insect hormone in an insect sample; model III anova between oxygen consumption and combinations of the "fixed" levels of a pesticide and the uncontrolled serum concentrations of an insect hormone in a sample of insects.

Assumptions to be satisfied, if anova has to be applied to the data, are as follows: that the dependent variable scores and so their error terms $(X - \bar{X})$ are **normally distributed**, that they occur **at random** and **independent of each other** in the groups, that the variances within the groups are almost similar (**homoscedasticity** or homogeneity of variances), and that the total variation of the variable results from the **additivity** of different components of that variation due to different factors.

2.2 ONE-WAY ANOVA

One-way anova is applied to the scores of a dependent variable, obtained in an experiment where two or more levels of a single independent variable have been applied to as many groups of individuals. It is meant for exploring whether the exposure to a single independent variable has led to significant differences in variances between the different groups. One-way anova may be either **model I anova** when the indepenent variable is deliberately "fixed" or controlled by the investigator and not suffering from random errors, or **model II anova** when the independent variable is not under the deliberate control by the investigator and not free from random errors.

For a one-way anova, the **total sum of squares** (SS_t) has to be partitioned into two component sums of squares, viz., the within-groups sum of squares (SS_w) and the between-groups sum of squares (SS_b). SS_t is the sum of squared differences of the raw scores of all the k number of groups from the grand mean $(\bar{X})$ of all those groups combined. When X_1, X_2, etc., are the scores of the respective groups of sizes n_1, n_2, etc., and $\bar{X}_1$, $\bar{X}_2$, etc., the respective group means, SS_t can be computed either from the differences of raw scores from $(\bar{X})$ or directly from the raw scores:

$$\bar{X} = \frac{\sum X_1 + \sum X_2 + \ldots\ldots + \sum X_k}{n_1 + n_2 + \ldots\ldots + n_k}, \text{ or, } \bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \ldots\ldots + n_k\bar{X}_k}{n_1 + n_2 + \ldots\ldots + n_k}.$$

$$SS_t = \sum(X_1 - \bar{X})^2 + \sum(X_2 - \bar{X})^2 + \ldots\ldots + \sum(X_k - \bar{X})^2, \text{ or,}$$

$$SS_t = \sum X_1^2 + \sum X_2^2 + \ldots\ldots + \sum X_k^2 - \frac{\left(\sum X_1 + \sum X_2 + \ldots\ldots + \sum X_k\right)^2}{n_1 + n_2 + \ldots\ldots + n_k}.$$

$$df_t = (n_1 + n_2 + \ldots\ldots + n_k) - 1 = N - 1.$$

The **between-groups sum of squares** (SS_b) is the sum of the groupsize-weighted squared differences of the group means from the grand mean. It can be computed either using the k number of group means and the group sizes, or directly from raw scores.

$$SS_b = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + \ldots\ldots + n_k(\bar{X}_k - \bar{X})^2, \text{ or,}$$

$$SS_b = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \ldots\ldots + \frac{\left(\sum X_k\right)^2}{n_k} - \frac{\left(\sum X_1 + \sum X_2 + \ldots\ldots + \sum X_k\right)^2}{n_1 + n_2 + \ldots\ldots + n_k}.$$

$$df_b = k - 1.$$

The **within-groups sum of squares** (SS_w) is the sum of squared differences of raw scores from their respective group means. It can be computed as the sum of the **squared error terms** of the raw scores. But because of the additivity of the sums of squares, SS_w is usually worked out as the difference between SS_t and SS_b.

$$SS_w = \sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2 + \ldots\ldots + \sum(X_k - \bar{X}_k)^2, \text{ or,}$$

$$SS_w = SS_t - SS_b .$$

$$df_w = (n_1 + n_2 + \ldots\ldots n_k) - k = N - k .$$

The **between-groups variance** (s_b^2) and the **within-groups variance** (s_w^2) are then computed by dividing the respective sums of squares with their degrees of freedom, and next used in working out Fisher's **variance ratio** (F ratio). The combination of df_b and df_w of s_b^2 and s_w^2 — these variances are also called the **greater** and the **lesser mean squares** respectively — constitutes the *df* of the computed F.

$$s_b^2 = \frac{SS_b}{df_b}; \quad s_w^2 = \frac{SS_w}{df_w}; \quad F = \frac{s_b^2}{s_w^2}.$$

$$df \text{ of } F: (df_b, df_w) \text{ or } (k-1,\ N-k).$$

The H_a proposes that s_b^2 has an additional variance component over and above the error variance measured by s_w^2. Contesting the H_a, the H_o proposes that there is no such added variance component in s_b^2 which is, therefore, simply another measure of the error variance measured by s_w^2. Thus, according to the H_o, the F ratio between s_b^2 and s_w^2 does not differ significantly from 1.00. This, in effect, means that the group means do not differ significantly and are estimates of the same population mean.

Sampling distributions of F scores vary according to their *df*, their positive skewness declining progressively with the rise in *df*. So, for finding the probability P of the correctness of H_o, the computed F score must be compared with the critical F scores having the same *df* as that of the computed one. Moreover, the computed F can never bear a negative sign because of being the ratio of two variances; so, its significance level (α) is given by the fractional area beyond the critical F_α in only the positive tail of the relevant F distribution.

If the computed F either equals or exceeds the critical F_α for a chosen significance level (α) or for the 0.05 level, the probability P of the H_0 being correct is considered too low ($P \leq \alpha$) and the H_0 may be rejected. It is then inferred, in case of a model I anova, that there is a significant **added treatment component** in the between-groups variance, which is absent in the within-groups variance and has resulted from the exposure of the groups to different levels of the "fixed" experimental treatment applied as the independent variable; the group means, therefore, are considered as significantly different from each other ($P \leq \alpha$). But a similar finding in a model II anova leads to the inference that there is a significant **added variance component** (s_a^2) in the between-groups variance owing to the uncontrolled effects of different levels of the randomly varying independent variable, consequently causing a significant difference between the group means $(P \leq \alpha)$.

If more than two groups have been used for as many levels of the independent variable, a significant F ratio in a model I anova should be followed by **multiple comparison tests** (see Section 2.3) between the group means of each chosen pair of groups to find which of the group means differ significantly from each other. However, multiple comparison test is not required if only two groups have been used in the experiment because a significant F ratio itself indicates a significant difference between two such group means.

A significant F ratio in a model II anova should be followed by the computation of the magnitude of the **added variance component** (s_a^2), present in s_b^2, and also its proportion, $p(s_a^2)$, in the total variation due to both s_a^2 and s_w^2. Where each of the groups possesses the size n,

$$s_a^2 = \frac{s_b^2 - s_w^2}{n}; \quad p(s_a^2) = \frac{s_a^2}{s_a^2 + s_w^2}.$$

On the contrary, if the computed F is lower than the critical F_α for a chosen significance level (α) or for the 0.05 level, there is no significant added component, either treatment or random, in the between-groups variance. This leads to the inference that there is no significant difference between the group means $(P > \alpha)$.

Example 2.2.1.

Work out one-way anova with the following body weight data (g) of female crabs (group 1) and male crabs (group 2) to find whether or not there is a significant added variance component between the groups ($\alpha = 0.01$).

Group 1 (X_1) : 11.10, 14.20, 12.40, 15.20, 14.00, 14.25, 15.50, 13.50, 13.50, 14.40, 15.50, 14.45.

Group 2 (X_2) : 10.10, 11.40, 14.30, 13.50, 11.10, 13.25, 10.20, 12.15, 13.80, 10.75, 12.80, 10.65.

Critical F scores : $F_{0.01(1,23)} = 7.88$; $F_{0.01(1,22)} = 7.94$; $F_{0.01(2,23)} = 5.66$; $F_{0.01(2,22)} = 5.72$.

Solution :

The data are entered in Table 2.1 for further treatment and the X_1^2 and X_2^2 scores computed are also entered therein.

Table 2.1 : Table for computing anova with crab bodyweight data. ($n_1 = n_2 = 12$.)

	X_1	X_2	X_1^2	X_2^2
	11.10	10.10	123.21	102.01
	14.20	11.40	201.64	129.96
	12.40	14.30	153.76	204.49
	15.20	13.50	231.04	182.25
	14.00	11.10	196.00	123.21
	14.25	13.25	203.06	175.56
	15.50	10.20	240.25	104.04
	13.50	12.15	182.25	147.62
	13.50	13.80	182.25	190.44
	14.40	10.75	207.36	115.56
	15.50	12.80	240.25	163.84
	14.45	10.65	208.80	113.42
Σ	168.00	144.00	2369.87	1752.40

$$SS_t = \sum X_1^2 + \sum X_2^2 - \frac{\left(\sum X_1 + \sum X_2\right)^2}{n_1 + n_2} = 2369.87 + 1752.40 - \frac{(168.00 + 144.00)^2}{12 + 12} = 66.27.$$

$$df_t = (n_1 + n_2) - 1 = 12 + 12 - 1 = 23.$$

$$SS_b = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} - \frac{\left(\sum X_1 + \sum X_2\right)^2}{n_1 + n_2} = \frac{(168)^2}{12} + \frac{(144)^2}{12} - \frac{(168 + 144)^2}{12 + 12} = 24.00.$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 66.27 - 24.00 = 42.27.$$

$$df_w = N - k = (n_1 + n_2) - k = 12 + 12 - 2 = 22.$$

The sums of squares are entered in Table 2.2 and used in computing the variance and the F ratio.

Table 2.2: Anova table for crab body weight data.

Variations	Sums of squares	df	Variances	F
Between groups	24.00	1	24.00	12.50
Within groups	42:27	22	1.92	
Total	66.27	23	-	-

$$s_b^2 = \frac{SS_b}{df_b} = \frac{24.00}{1} = 24.00. \quad s_w^2 = \frac{SS_w}{df_w} = \frac{42.27}{22} = 1.92.$$

$$s_t^2 = \frac{SS_t}{df_t} = \frac{66.27}{23} = 2.88.$$

$$F = \frac{s_b^2}{s_w^2} = \frac{24.00}{1.92} = 12.50. \; df \; of \; F = (df_b, df_w) = 1, 22.$$

Critical F score : $F_{0.01(1,22)} = 7.94$.

Because the computed F is higher than the critical $F_{0.01}$, there is a significant added variance component (s_a^2) in the between-groups variance, leading to a significant difference between the group means ($P < 0.01$). Because $n_1 = n_2 = n = 12$,

$$s_a^2 = \frac{s_b^2 - s_w^2}{n} = \frac{24.00 - 1.92}{12} = 1.84.$$

$$p(s_a^2) = \frac{s_a^2}{s_a^2 + s_w^2} = \frac{1.84}{1.84 + 1.92} = 0.49.$$

Example 2.2.2

Work out one-way model I anova with the following tracheal ventilation scores (ml per minute) of two groups of animals from a water beetle species, exposed to two respective "fixed" levels of a pesticide, to find if the different levels of pesticide treatment have caused a significant difference between the mean tracheal ventilations of the two groups ($\alpha = 0.05$).

Group 1 (X_1) : 80.3, 84.2, 76.0, 80.5, 77.2, 85.4, 83.6, 84.2, 78.5, 76.1, 74.0

Group 2 (X_2) : 70.8, 80.0, 72.2, 70.0, 73.4, 74.0, 73.5, 70.2, 74.1, 73.4, 71.4

Critical F scores : $F_{0.05(1,18)} = 4.41$; $F_{0.05(1,19)} = 4.38$; $F_{0.05(1,20)} = 4.35$; $F_{0.05(2,19)} = 3.52$; $F_{0.05(2,18)} = 3.55$; $F_{0.05(1,22)} = 4.30$; $F_{0.05(1,23)} = 4.28$.

Solution :

The data are entered in Table 2.3. The squared value of each score is computed and entered in that table.

Table 2.3: Table for computing anova with beetle tracheal ventilation data ($n_1 = n_2 = 11$).

	X_1	X_2	X_1^2	X_2^2
	80.3	70.8	6448.09	5012.64
	84.2	80.0	7089.64	6400.00
	76.0	72.2	5776.00	5212.84
	80.5	70.0	6480.25	4900.00
	77.2	73.4	5959.84	5387.56
	85.4	74.0	7293.16	5476.00
	83.6	73.5	6988.96	5402.25
	84.2	70.2	7089.64	4928.04
	78.5	74.1	6162.25	5490.81
	76.1	73.4	5791.21	5387.56
	74.0	71.4	5476.00	5097.96
Σ	880.0	803.0	70555.04	58695.66

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{880.0}{11} = 80.0 \text{ ml.} \quad \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{803.0}{11} = 73.0 \text{ ml.}.$$

$$SS_t = \sum X_1^2 + \sum X_2^2 - \frac{\left(\sum X_1 + \sum X_2\right)^2}{n_1 + n_2} = 70555.04 + 58695.66 - \frac{(880.0 + 803.0)^2}{11+11} = 501.20.$$

$$df_t = N - 1 = (n_1 + n_2) - 1 = 11 + 11 - 1 = 21.$$

$$SS_b = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} - \frac{\left(\sum X_1 + \sum X_2\right)^2}{n_1 + n_2} = \frac{(880)^2}{11} + \frac{(803)^2}{11} - \frac{(880+803)^2}{11+11} = 269.50.$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 501.20 - 269.50 = 231.70.$$

$$df_w = N - k = (n_1 + n_2) - k = 11 + 11 - 2 = 20.$$

The sums of squares are entered in Table 2.4 and used in computing the variances and F ratio.

Table 2.4 Anova table for tracheal ventilation data

Variations	Sums of squares	*df*	Variances	*F*
Between groups	269.50	1	269.50	23.25
Within groups	231.70	20	11.59	
Total	501.20	21	-	-

$$s_b^2 = \frac{SS_b}{df_b} = \frac{269.50}{1} = 269.50. \quad s_w^2 = \frac{SS_w}{df_w} = \frac{231.70}{20} = 11.59.$$

$$F = \frac{s_b^2}{s_w^2} = \frac{269.50}{11.59} = 23.25. \quad df \text{ for } F: (df_b, df_w) = (1, 20.).$$

Critical $F_{0.05(1,20)} = 4.35$.

As the computed F is found to be higher than the critical $F_{0.05}$, the computed F is considered significant ($P < 0.05$). So, there is a significant added treatment component in the between-groups variance. Because there are only two groups here, the significant F ratio directly leads to the inference that there is a significant difference between the mean tracheal ventilations of the two groups ($P < 0.05$).

2.3 MULTIPLE COMPARISON TESTS :

A significant F ratio in a one-way model I anova involving more than two groups gives no indication about which of the group means are significantly different from each other — it may be that all the group means differ significantly from each other, or it may also be that some of them differ significantly while others do not. To find this out, a *before-design multiple comparison test* (e.g., multiple comparison t test and Scheffe's F test) is done with chosen pairs of group means, decided upon before even doing the experiment. But an *after-design multiple comparison test* (e.g., Gabriel's SS-STP) is undertaken where the group means cannot be chosen before obtaining and inspecting the experimental data.

Multiple comparisons by t test

The difference between the means of any chosen pair of groups, e.g., $\overline{X}_1 - \overline{X}_3$, is transformed into Student's t score using the respective group sizes and the within-groups variance $\left(s_w^2\right)$ worked out in the preliminary F test of anova (see Section 2.2).

$$t = \frac{\text{Difference between the means}}{SE \text{ of the difference between the means}}; \text{ so, for } \overline{X}_1 - \overline{X}_3,$$

$$t = \frac{\overline{X}_1 - \overline{X}_3}{s_{\overline{X}_1 - \overline{X}_3}} = \frac{\overline{X}_1 - \overline{X}_3}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_3}}}.$$

$$df = df_w = N - k.$$

The computed t, having the same df as that of the s_w^2 used in working it out, is compared with the critical t_α scores with that df. The difference between the means is considered significant only if the computed t either equals or exceeds the critical t score for a significance level (α) not higher than 0.05 ($P \leq \alpha$). This procedure is repeated for the difference between means of each chosen pair of groups.

Multiple comparisons by Scheffe's F test

This test is more powerful than the Student's t test. For each chosen pair of groups, the difference between their means (e.g., $\overline{X}_1 - \overline{X}_3$) is transformed into F score which actually amounts

to the squared t score between two group means.

$$F = \left[\frac{\text{Difference between the means}}{SE \text{ of the difference between the means}}\right]^2; \text{ so, for } \bar{X}_1 - \bar{X}_3,$$

$$F = \left(\frac{\bar{X}_1 - \bar{X}_3}{s_{\bar{X}_1 - \bar{X}_3}}\right)^2 = \frac{\left(\bar{X}_1 - \bar{X}_3\right)^2}{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_3}}.$$

The computed F score is then compared with the critical F' scores for 0.05 and 0.01 levels of significance, computing them from the respective critical F scores, having the degrees of freedom (df_b and df_w) of the two variances used in working out the preliminary F test in anova (see Section 2.2).

$F'_\alpha = df_b \times F_{\alpha(df_b, df_w)}$; thus,

$F'_{0.05} = (k-1)F_{0.05(k-1, N-k)}$ and $F'_{0.01} = (k-1)F_{0.01(k-1, N-k)}$.

The difference between the given group means is considered significant, only if the computed F score either equals or exceeds the critical F' score for a significance level (α) not higher than 0.05.

This procedure is repeated for the difference between the means of each chosen pair of groups.

Gabriel's SS-STP

In this *Sum of Squares Simultaneous Test Procedure*, a critical sum of squares (SS_α) is first worked out, using the number (k) of groups in the experiment, s_w^2 and the critical $F_{\alpha(df_b, df_w)}$ score for a chosen α. The SS_α is then compared with the SS_b, worked out between those groups (e.g., groups 1 and 3) that have too close means, using their scores (X_1 and X_3) and group sizes (n_1 and n_3, here).

$$SS_\alpha = (k-1)s_w^2 F_{\alpha(df_b, df_w)};$$

$$SS_b = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_3\right)^2}{n_3} - \frac{\left(\sum X_1 + \sum X_3\right)^2}{n_1 + n_3}.$$

Only if SS_b is either higher than or equal to SS_α, the given group means (here, $\bar{X}_1$ and $\bar{X}_3$) differ from each other significantly.

Example 2.3.1

Work out one-way model I anova with the following winglength scores (mm) of three groups of cockroaches, treated respectively with three chosen levels of a pesticide, to find if there is a significant added treatment component in their between-groups variance. Are there significant differences between the means of groups 1 and 2, and between those of groups 1 and 3? ($\alpha = 0.01$).

Group 1 (X_1) : 41, 46, 40, 37, 42, 40, 39, 37, 40, 38.

Group 2 (X_2) : 36, 33, 34, 38, 37, 32, 35, 34, 36, 37, 33.

Group 3 (X_3) : 25, 27, 28, 26, 27, 30, 24, 29, 27.

Critical F scores : $F_{0.01(1,29)}$=7.60; $F_{0.01(1,28)}$=7.64; $F_{0.01(2,27)}$=5.49; $F_{0.01(2,29)}$=5.42; $F_{0.01(2,28)}$=5.45.

Critical t scores : $t_{0.01(29)}$=2.756; $t_{0.01(28)}$=2.763; $t_{0.01(27)}$=2.771; $t_{0.01(30)}$=2.750.

Solution :

The data are entered in Table 2.5. The squared scores are computed, entered in the table and used in further computations.

Table 2.5 : Table for computing anova with cockroach winglength data (n_1 = 10, n_2=11, n_3=9).

	X_1	X_2	X_3	X_1^2	X_2^2	X_3^2
	41	36	25	1681	1296	625
	46	33	27	2116	1089	729
	40	34	28	1600	1156	784
	37	38	26	1369	1444	676
	42	37	27	1764	1369	729
	40	32	30	1600	1024	900
	39	35	24	1521	1225	576
	37	34	29	1369	1156	841
	40	36	27	1600	1296	729
	38	37		1444	1369	
		33			1089	
Σ	400	385	243	16064	13513	6589

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{400}{10} = 40.0 \text{ mm}, \ \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{385}{11} = 35.0 \text{ mm}, \ \bar{X}_3 = \frac{\sum X_3}{n_3} = \frac{243}{9} = 27.0 \text{ mm}.$$

$$SS_t = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 - \frac{\left(\sum X_1 + \sum X_2 + \sum X_3\right)^2}{n_1 + n_2 + n_3} = 16064 + 13513 + 6589 - \frac{(400 + 385 + 243)^2}{10 + 11 + 9}$$

$$= 939.87.$$

$$df_t = N - 1 = (n_1 + n_2 + n_3) - 1 = (10 + 11 + 9) - 1 = 29.$$

$$SS_b = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \frac{\left(\sum X_3\right)^2}{n_3} - \frac{\left(\sum X_1 + \sum X_2 + \sum X_3\right)^2}{n_1 + n_2 + n_3}$$

$$= \frac{(400)^2}{10} + \frac{(385)^2}{11} + \frac{(243)^2}{9} - \frac{(400 + 385 + 243)^2}{10 + 11 + 9}$$

$$= 809.87.$$

$$df_b = k - 1 = 3 - 1 = 2.$$

$$SS_w = SS_t - SS_b = 939.87 - 809.87 = 130.00.$$

$$df_w = N - k = (n_1 + n_2 + n_3) - k = (10 + 11 + 9) - 3 = 27.$$

The computed sums of squares are entered in Table 2.6 and used in working out the variances and the F ratio.

Table 2.6 : Anova table for cockroach winglength data.

Variations	Sums of squares	*df*	Variances	*F*
Between groups	809.87	2	404.94	84.19
Within groups	130.00	27	4.81	
Total	939.87	29	-	-

$$s_b^2 = \frac{SS_b}{df_b} = \frac{809.87}{2} = 404.94, \quad s_w^2 = \frac{SS_w}{df_w} = \frac{130.00}{27} = 4.81.$$

$$F = \frac{s_b^2}{s_w^2} = \frac{404.94}{4.81} = 84.19, \quad df \text{ for } F = (df_b, df_w) = (2, 27).$$

Critical $F_{0.01(2,27)}$=5.49.

Because the computed F is higher than the critical $F_{0.01}$, there is a significant added treatment component in the between-groups variance ($P < 0.01$).

To test for the significances of $(\bar{X}_1 - \bar{X}_2)$ and $(\bar{X}_1 - \bar{X}_3)$, either multiple comparison t test or Scheffe's F test can be done.

(a) Multiple comparison t test :

(i) For $\bar{X}_1 - \bar{X}_2$:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}}} = \frac{40.0 - 35.0}{\sqrt{\frac{4.81}{10} + \frac{4.81}{11}}} = 5.218;\ df = df_w = 27.$$

Critical $t_{0.01(27)} = 2.771$.

As the computed t is higher than the critical $t_{0.01}$, $P < 0.01$. So, there is a significant difference between the means of groups 1 and 2.

(ii) For $\bar{X}_1 - \bar{X}_3$:

$$t = \frac{\bar{X}_1 - \bar{X}_3}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_3}}} = \frac{40.0 - 27.0}{\sqrt{\frac{4.81}{10} + \frac{4.81}{9}}} = 12.901;\ df = df_w = 27.$$

Critical $t_{0.01(27)} = 2.771$.

As the computed t is higher than the critical $t_{0.01}$, there is a significant difference between the means of groups 1 and 3 ($P < 0.01$).

(b) Schefe's F test :

(i) For $\bar{X}_1 - \bar{X}_2$:

$$F = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}} = \frac{(40.0 - 35.0)^2}{\frac{4.81}{10} + \frac{4.81}{11}} = 27.23.$$

Critical $F'_\alpha = (k-1)F_{\alpha(df_b, df_w)}$; thus

$$F'_{0.01} = (3-1) \times F_{0.01(2,27)} = (3-1) \times 5.49 = 10.98.$$

As the computed F is higher than the critical $F'_{0.01}$, there is a significant difference between the means of groups 1 and 2 ($P < 0.01$).

(ii) For $\bar{X}_1 - \bar{X}_3$:

$$F = \frac{(\bar{X}_1 - \bar{X}_3)^2}{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_3}} = \frac{(40.0 - 27.0)^2}{\frac{4.81}{10} + \frac{4.81}{9}} = 166.43$$

$$F'_{0.01} = (k-1)F_{0.01(df_b, df_w)} = (3-1) \times 5.49 = 10.98.$$

As the computed F is higher than the critical $F'_{0.01}$, there is a significant difference be een the means of groups 1 and 3 ($P < 0.01$).

2.4 KRUSKAL-WALLIS NONPARAMETRIC ANOVA :

This test is used if the assumptions for the parametric method of one-way anova are found not to be justifiable.

2.4.1 Kruskal-Wallis statistic H :

Kruskal-Wallis test is a rank-dependent and nonparametric one-way anova. Unlike the parametric one-way anova, it may be used for both continuous and discontinuous, normally or non-normally distributed, and ordinal variables. Its *assumptions* are: the scores of the dependent variable in the experiment can be given ranks, and each score or rank of the variable occurs randomly in the sample in accordance to the laws of probability.

To compute the Kruskal-Wallis statistic H, ranks are given in an ascending order to the scores or the already existing ranks of all the k number of groups, in a composite manner taking all groups together. An average rank is given to each of a tied set of identical scores even if they belong to separate groups – the average rank is the arithmetic mean of the expected ranks of the tied scores if they were consecutive scores. The rank sums (R_1, R_2, etc.) are worked out for the respective groups of sizes n_1, n_2, etc., by totalling their ranks separately, and are used in computing the mean ranks $(\bar{R}_1, \bar{R}_2$, etc.) of the respective groups. The mean rank $(\bar{R})$ of all the groups is also computed from the rank sums of different groups. These are next used in working out H.

$$N = n_1 + n_2 + n_3 + \ldots + n_k.$$

$$\overline{R}_1 = \frac{R_1}{n_1};\ \overline{R}_2 = \frac{R_2}{n_2};\ \overline{R}_3 = \frac{R_3}{n_3};\ldots\ \overline{R}_k = \frac{R_k}{n_k}.$$

$$\overline{R} = \frac{R_1 + R_2 + \ldots + R_k}{N}$$

$$H = \frac{12}{N(N+1)}\left[n_1\left(\overline{R}_1 - \overline{R}\right)^2 + n_2\left(\overline{R}_2 - \overline{R}\right)^2 + \ldots + n_k\left(\overline{R}_k - \overline{R}\right)^2\right]$$

$df = k - 1$.

The computed H is compared with the critical chi square for the chosen level of significance (α). There is a significant difference between the groups if the H score is either higher than or equal to the critical $\chi^2_{\alpha(df)}$.

If there are more than two groups in an experiment ($k > 2$), a significant H should be followed by multiple comparison U test between the groups of each chosen pair to find if they differ significantly.

2.4.2 Multiple comparison Mann-Whitney *U* test :

This is a rank-dependent nonparametric alternative to Student's t test. To test the difference between the groups of each chosen pair (say, groups 1 and 2 of sizes n_1 and n_2), the scores of those two groups (X_1 and X_2) are given ranks in a composite manner and an ascending order, and the separately totalled rank sums (R_1 and R_2) of the two groups are used in computing the Mann-Whitney U scores. Any of the two rank sums (either R_1 or R_2) may be used in computing U.

Either, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$,

or, $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$.

Any of the two U scores may be transformed into z score, using the U_e expected from the H_0 and the SE of U (s_U). The computed z score is then used in working out the probability (P) of the H_0 being correct.

$$U_e = \frac{n_1 n_2}{2};\quad s_u = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}};$$

Either, $z = \frac{U_1 - U_e}{s_U}$, or, $z = \frac{U_2 - U_e}{s_U}$.

$P = 2$ [0.5000 – (Area of unit normal curve from μ to the computed z)].

There is a significant difference between the means of the given pair of groups, only if P is either equal to or lower than the chosen level of significance (α).

Example 2.4.1

(a) Work out the Kruskal-Wallis anova to find whether or not there is a significant difference between the heart rates (/ minute) of the following three samples ($\alpha = 0.01$).

(b) If yes, use the Mann-Whitney U test to find if there is a significant difference between groups 2 and 3 ($\alpha = 0.05$).

Group 1 (X_1) : 70, 72, 75, 79, 70, 77, 80, 68, 79.

Group 2 (X_2) : 82, 88, 90, 80, 87, 85, 81, 88, 89, 92.

Group 3 (X_3) : 92, 96, 95, 80, 88, 98, 102, 99, 100.

Critical χ^2 scores : $\chi^2_{0.01(27)} = 46.96$; $\chi^2_{0.01(2)} = 9.21$; $\chi^2_{0.01(26)} = 45.64$.

z score : 2.51, area : 0.4940; z score : 2.52, area : 0.4941;

z score : 2.53, area 0.4943; z score : 2.54, area : 0.4945;

z score : 2.55, area : 0.4946.

Solution :

(a) Composite ranks are given in an ascending order to the scores of all three groups taken together (Table 2.7) and the rank sums of the respective groups are worked out separately.

Table 2.7 Ranking of group scores for computing H.

X_1		X_2		X_3	
Scores	Ranks	Scores	Ranks	Scores	Ranks
70	2.5	82	13	92	21.5
72	4	88	17	96	24
75	5	90	20	95	23
79	7.5	80	10	80	10
70	2.5	87	15	88	17

Contd...Table 2.7

X_1		X_2		X_3	
Scores	Ranks	Scores	Ranks	Scores	Ranks
77	6	85	14	98	25
80	10	81	12	102	28
68	1	88	17	99	26
79	7.5	89	19	100	27
		92	21.5		
Σ	46.0 (R_1)		158.5 (R_2)		201.5(R_3)

$n_1 = 9;\ n_2 = 10;\ n_3 = 9;\ N = n_1 + n_2 + n_3 = 9+10+9 = 28.$

$$\overline{R}_1 = \frac{R_1}{n_1} = \frac{46.0}{9} = 5.11;\ \overline{R}_2 = \frac{R_2}{n_2} = \frac{158.5}{10} = 15.85;\ \overline{R}_3 = \frac{R_3}{n_3} = \frac{201.5}{9} = 22.39$$

$$\overline{R} = \frac{R_1 + R_2 + R_3}{N} = \frac{46.0+158.5+201.5}{28} = 14.50.$$

$$H = \frac{12}{N(N+1)}\left[n_1(\overline{R}_1 - \overline{R})^2 + n_2(\overline{R}_2 - \overline{R})^2 + n_3(\overline{R}_3 - \overline{R})^2\right]$$

$$= \frac{12}{28(28+1)}\left[9(5.11-14.50)^2 + 10(15.85-14.50)^2 + 9(22.39-14.50)^2\right] = 20.28$$

$df = k-1 = 3-1 = 2.$

Critical $\chi^2_{0.01(2)} = 9.21.$

As the computed H is higher than the critical $\chi^2_{0.01}$, $P < 0.01$. Hence, there are significant differences between the heart rates of the three groups $(P < 0.01)$.

(b) To find if there is a significant difference between groups 2 and 3, multiple comparison *Mann-Whitney U test* is undertaken between those two groups. For this, composite ranks are given in an ascending order to the heart rate scores of groups 2 and 3 (Table 2.8) and the rank sums of the respective groups are worked out.

Any of the two rank sums (e.g., R_2) is then used to compute the Mann-Whitney U score which is next transformed into the z score.

Table 2.8 : Composite ranking of scores of groups 2 and 3 for Mann-Whitney U test.

X_2		X_3	
Scores	Ranks	Scores	Ranks
82	4	92	12.5
88	8	96	15
90	11	95	14
80	1.5	80	1.5
87	6	88	8
85	5	98	16
81	3	102	19
88	8	99	17
89	10	100	18
92	12.5		
Σ	69.0 (R_2)		121.0 (R_3)

$$U_2 = n_2 n_3 + \frac{n_2(n_2+1)}{2} - R_2$$

$$= 10 \times 9 + \frac{10(10+1)}{2} - 69.0$$

$$= 76.0.$$

$$U_e = \frac{n_2 n_3}{2} = \frac{10 \times 9}{2} = 45.0.$$

$$s_U = \sqrt{\frac{n_2 n_3 (n_2 + n_3 + 1)}{12}} = \sqrt{\frac{10 \times 9(10+9+1)}{12}} = 12.25.$$

$$z = \frac{U_2 - U_e}{s_U} = \frac{76.0 - 45.0}{12.25} = 2.53$$

$$P = 2[0.5000 - (\text{Area of unit normal curve from } \mu \text{ to the computed } z)]$$

$$= 2[0.5000 - (\text{Area from } \mu \text{ to } 2.53)]$$

$$= 2\ [0.5000 - 0.4943]$$

$$= 0.01.$$

As the computed P is lower than the α of 0.05, P is considered too low. So, the H_0 is rej. cted and it is inferred that there is a significant difference between the groups 2 and 3 ($P<0.05$).

2.5 SUMMARY :

Anova is a powerful method for finding whether or not the independent variable has produced a significant change in the dependent variable in an experiment. According to the number of one, two or still higher number of independent variables in an experiment, one-way, two-way or a still higher way of anova has to be used. Models I, II or III of anova has to be used according as the independent variable is "fixed", random or a combination of both.

One-way anova partitions the total variance into the between-groups and within-groups variances, and works out the F ratio between them. Only when F exceeds or equals the critical F for the chosen α, there is a significant additional component in the between-groups variance, which is absent in the within-groups variance and so, there is a significant difference between the group means owing to the exposure to the independent variable.

A significant F ratio in an anova of the data from more than two groups should be followed by either a before-design multiple comparison test such as Scheffe's F test or an after-design multiple comparison test such as Gabriel's SS-STP.

Where assumptions for the parametric one-way anova are found not justifiable, Kruskal-Wallis nonparametric anova may be undertaken, followed by the nonparametric multiple comparison Mann-Whitney U test.

2.6 GLOSSARY :

Anova : Analysis of variances of the scores of the dependent variable of an experiment for testing of hypothesis.

Between-groups variance : Variance of scores belonging to different groups exposed to different levels of the independent variable in an experiment.

F ratio : Ratio of the variance under investigation such as the between-groups variance, and an error variance like the within-groups variance.

Gabriel's SS-STP : An after-design multiple comparison test, known as the sum of squares - simultaneous test procedure.

Homoscedasticity : Homogeneity of variances.

Kruskal-Wallis test : A nonparametric rank-dependent one-way anova.

Mann-Whitney *U* test : A nonparametric alternative to *t* test, also used as a nonparametric, rank-dependent, multiple comparison test.

Model I anova : Anova of the data of an experiment with "fixed" experimental treatment(s) as independent variable(s).

Model II anova : Anova of the data of an experiment with only random and uncontrolled independent variable(s).

Model III anova : Anova of an experiment where independent variables of both "fixed" and random natures have been applied.

Multiple comparison test : Tests which should follow the finding of a significant *F* ratio, to investigate whether or not the means of the chosen pairs of groups of an experiment differ significantly.

One-way anova : Anova of an experiment in which a single independent variable has been used.

Scheffe's *F* test : A powerful before-design multiple comparison test.

Two-way anova : Anova of an experiment in which two independent variables have been used.

Within-groups variance : Variance of scores within the groups as a measure of their random variations giving rise to error terms.

2.7 QUESTIONS :

1. (a) State the assumptions to be satisfied for applying anova to the experimental data.

   (b) Describe how you partition the total variance into the between-groups and within-groups variances in a one-way anova and use them in computing the variance ratio, mentioning the formulae to be used.

   (c) Work out a one-way anova to find whether or not the mean winglength (mm) of male cockroaches (group 1) differ from that of females (group 2) in the following data ($\alpha = 0.05$).

   Group 1 : 40, 42, 40, 39, 37, 42, 38, 39, 36, 34.

   Group 2 : 35, 34, 33, 36, 30, 30, 37, 32, 33, 31, 35, 31.

Critical F scores: $F_{0.05(1,21)} = 4.32$; $F_{0.05(2,21)} = 3.47$; $F_{0.05(1,20)} = 4.35$; $F_{0.05(2,20)} = 3.49$.

2. (a) Explain the different models and ways of anova.

   (b) Write what you understand by greater and lesser mean squares and variance ratio, giving their importance in one-way anova.

   (c) Use the following hemoglobin scores (g per dl) of three groups of subjects, treated with three different controlled levels of a folate-antagonist, to compute the F ratio, test its significance, and write the appropriate inference ($\alpha = 0.05$). Also apply Scheffe's F test to find whether or not the means of groups 1 and 3 differ significantly ($\alpha = 0.05$).

   Group 1 : 14.2, 15.0, 13.8, 15.0, 14.4, 14.2, 13.6, 14.8, 14.0, 14.2.

   Group 2 : 13.8, 14.6, 14.0, 14.8, 14.0, 13.6, 13.8, 13.2, 13.6, 14.0.

   Group 3 : 10.8, 12.2, 11.6, 12.0, 10.6, 9.8, 11.0, 10.0, 11.2, 10.6.

   Critical F scores: $F_{0.05(2,29)} = 3.33$; $F_{0.05(1,29)} = 4.18$; $F_{0.05(2,28)} = 3.34$; $F_{0.05(1,28)} = 4.20$; $F_{0.05(2,27)} = 3.35$; $F_{0.05(1,27)} = 4.21$.

3. (a) Write about the following and their applications in one-way model I anova : (i) Scheffe's F test and (ii) multiple comparison t test.

   (b) Give the assumptions of anova and state how you would compute the F ratio in a one-way anova.

   (c) How would you test the significance of the following in a one-way anova: (i) Fisher's variance ratio, (ii) Scheffe's F score and (iii) multiple comparison t scores.

4. (a) Give an account of the computation and significance test of one-way anova.

   (b) Describe how would you follow up a significant F ratio in Model I and Model II anova, mentioning all relevant formulae.

   (c) Apply one-way anova to find whether or not there is a significant added variance component in the between-groups variance of the body weight scores (g) of three groups of crabs of a particular species from three different habitats ($\alpha = 0.05$). Also work out the magnitude of the added variance component, provided the latter has been found significant.

Group 1 : 10.10, 14.00, 12.45, 15.00, 12.50, 13.25, 14.60, 13.82, 12.55, 14.00.

Group 2 : 9.20, 10.00, 10.20, 11.00, 11.00, 10.00, 9.80, 10.10, 11.20, 10.00.

Group 3 : 8.10, 9.00, 9.25, 10.00, 9.30, 8.50, 8.50, 7.80, 8.00, 9.50.

Critical F scores: $F_{0.05(2,27)} = 3.35$; $F_{0.05(2,28)} = 3.34$; $F_{0.05(1,27)} = 4.21$; $F_{0.05(1,28)} = 4.20$.

5. (a) What are the assumptions for Kruskal-Wallis one-way anova?

   (b) Work out Kruskal-Wallis test to find whether or not there is a significant difference between the group means of the data of Question 1(c) ($\alpha = 0.05$). Critical $\chi^2_{0.05(21)} = 32.67$; $\chi^2_{0.05(20)} = 31.41$; $\chi^2_{0.05(1)} = 3.84$.

BIBLIOGRAPHY FOR FURTHER READING :

1. Das, D. and Das, A., ***Statistics in Biology and Psychology***, 4th edition, Academic Publishers, Kolkata.
2. Sokal, R.R. and Rohlf, F.J., ***Introduction to Biostatistics***, 2nd edition, W.H. Freeman & Co., San Francisco.
3. Sokal, R.R. and Rohlf, F.J., ***Biometry***, W.H. Freeman & Co., San Francisco.
4. Bliss, C., ***Statistics in Biology***, Vols. I-II, McGraw-Hill, New York.
5. Ipsen, J. and Feigl, P., ***Bancroft's Introduction to Biostatistics***, Harper and Row, New York.
6. Armitage, P. and Berry, G., ***Statistical Methods in Medical Research***, Blackwell, Oxford.
7. Guenther, W.C., ***The Analysis of Variance***, Prentice-Hall, Englewood Cliffs, N.J.
8. Scheffe, H., ***The Analysis of Variance***, John Wiley, New York.
9. Edwards, A.L., ***Multiple Regression and the Analysis of Variance and Covariance***, Freeman, San Francisco.
10. Rao, C.R., ***Advanced Statistical Methods in Biometric Research***, John Wiley, New York.
11. Fisher, R.A., ***Statistical Methods for Research Workers***, Hafner, New York.
12. Altman, D.G., ***Practical Statistics for Medical Research***, Chapman and Hall.
13. Sprott, D.A., ***Statistical Inference in Science***, Springer-Verlag, New York.

Learner's Feed-back

After going through the Modules / Units please answer the following questionnaire.
Cut the portion and send the same to the Directorate.

To
The Director
Directorate of Distance Education.
Vidyasagar University
Midnapore - 721 102

1. The modules are : (give ✓ in appropriate box)
   ☐ easily understandable; ☐ very hard; ☐ partially understandable.

2. Write the number of the Modules/Units which are very difficult to understand :
   .....
   .....
   .....

3. Write the number of Modules / Units which according to you should be re-written :
   .....
   .....
   .....

4. Which portion / page is not understandable to you? (mention the page no. and portion)
   .....
   .....
   .....
   .....

5. Write a short comment about the study material as a learner.
   .....
   .....
   .....
   .....
   .....

.....
(Full Signature of the Learner)

Date : .....

Enrolment No .....

Phone / Mobile No. .....

Published by : Directorate of Distance Education,
Vidyasagar University, Midnapore 721 102, W.B.
Printed by : S P Communications Pvt. Ltd., 31B, Raja Dinendra Street, Kolkata - 700 009
E-mail : spcomkol@gmail.com