BIO-STATISTICS

STATISTICS

Is the science of Collecting, Presenting, Summarizing, Analyzing & Interpreting sets of data.

BIO-STATISTICS

Is the application of statistics to living organisms and human beings.

BIO-STATISTICS

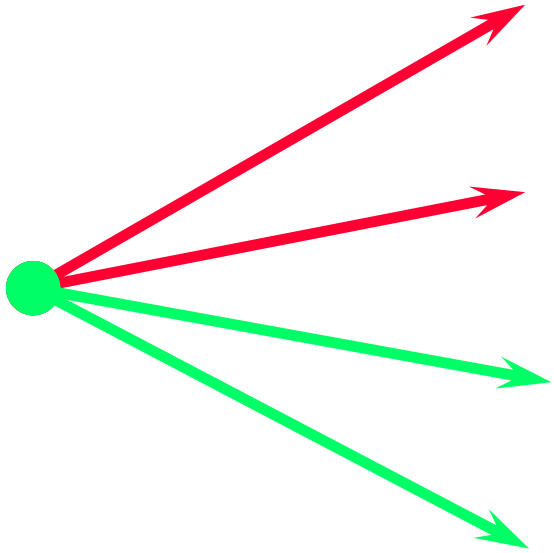
Is the science of Collecting, Presenting, Summarizing, Analyzing & Interpreting sets of data related to living organisms and human beings.

BASIC PRINCIPLES OF BIOSTATISTICS

1. Collection of Data
2. Presentation of Data
3. Summarization of Data
4. Analysis of Data
5. Interpretation of Data

TYPES OF DATA

DATA



1.Quantitative

2.Qualitative

3.Primary

4.Secondary

METHODS OF DATA COLLECTION

COLLECTION OF PRIMARY DATA

- Observation Method
- Interview Method
- Through Questionnaires
- Through Schedules
- Clinical Examination
- Laboratory investigations
  (microbiological, biochemical, histopathological, radiographic)

COLLECTION OF SECONDARY DATA

- Published data
- Unpublished data

Rules for secondary data

1. The data should be accurate

2. The data should be sufficient for analysis

3. It should be fulfills the aims and objectives


SAMPLING

- SAMPLE

A group of sampling units that forms part of a population generally selected so as to be representative of the population whose variables are under study.

- SAMPLING

It is the process of obtaining information about an entire population by examining only a part of it.

NECCESSARIES:

- Sampling can save time & money
- Physical impossibility to check all items in a population
- Sampling remains the only way when population contains infinitely many members

SAMPLING TECHNIQES

There are two types of sampling techniques.

1) Probability sampling

2) Non probability sampling

THE PROBABILITY SAMPLING METHODS ARE

1) Simple Random Sampling

2) Systematic Random Sampling

3) Stratified Random Sampling

4) Cluster Sampling

5) Multistage Sampling

6) Multiphase Sampling

NON PROBABILITY SAMPLING METHODS ARE :

1) quota sampling

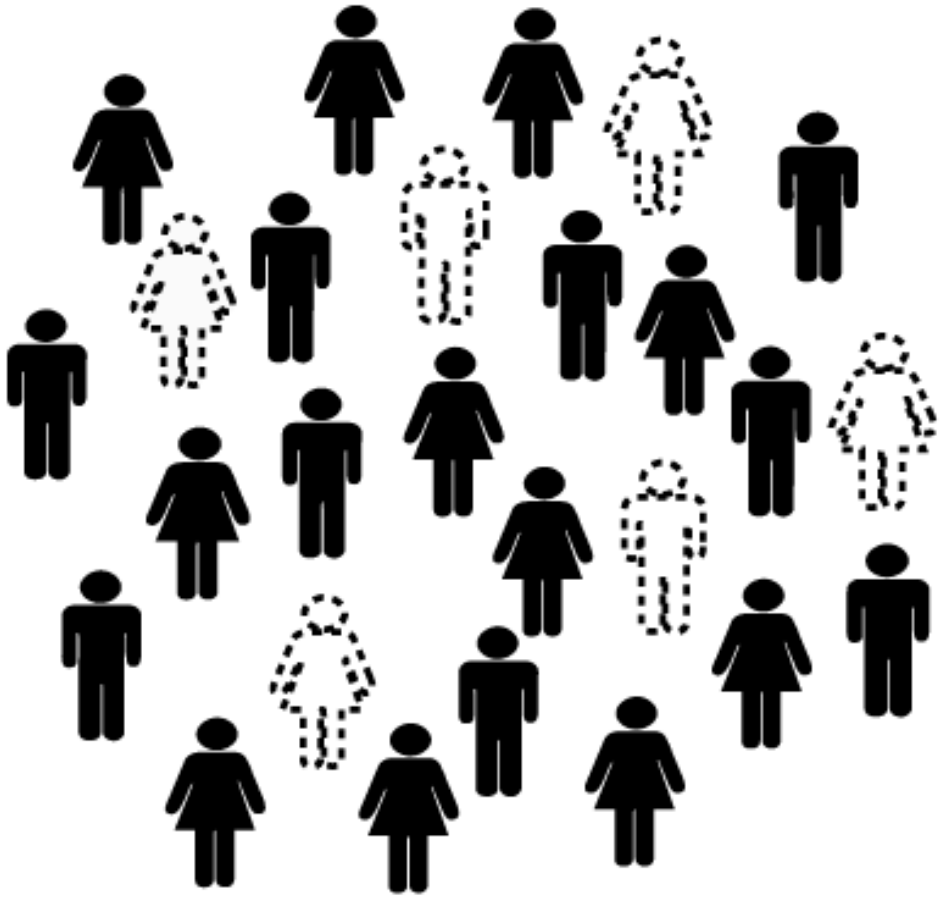
2) judgment sampling

3) convenience sampling

4) Snow ball sampling

Population

Sample



1. SIMPLE RANDOM SAMPLING

- Applicable when the population is small, homogeneous and readily available
- Principle is that every member or every unit of the population has an equal chance of being selected.
- Two Methods

  1) Lottery Method

  2) Table of Random Number Method

Simple Random Sampling

2, 6, 7, 12, 18

Each member of the population is listed in fashion (e.g., numerically) and then a sample is drawn by randomly selecting members of the population

II. SYSTEMATIC RANDOM SAMPLING

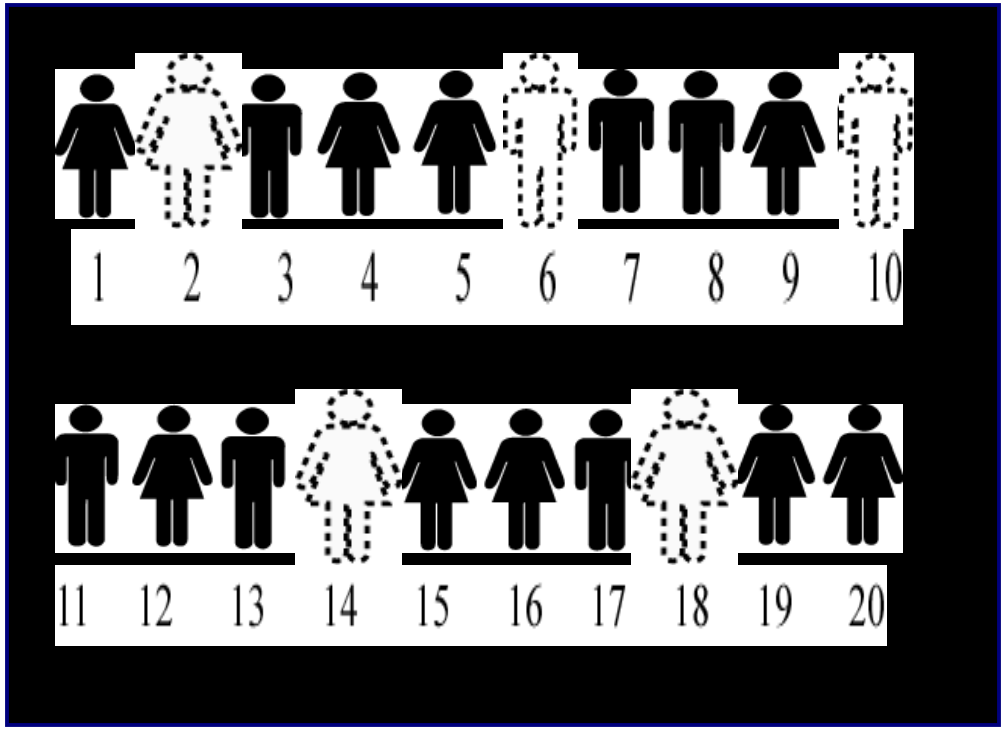
- Used in those cases where a complete list of population from which sample is to be drawn is available.

- More often applied to field studies where population is large, scattered & homogeneous.

- Choose every K^{th} house where 'K' refers to the SAMPLING INTERVAL

$$K = \frac{\text{Total population}}{\text{Sample size desired}}$$

Systematic/Sequential Random Sampling



Desired Sample Size: 5
Population Size: 20
Increment: 20/5 = 4

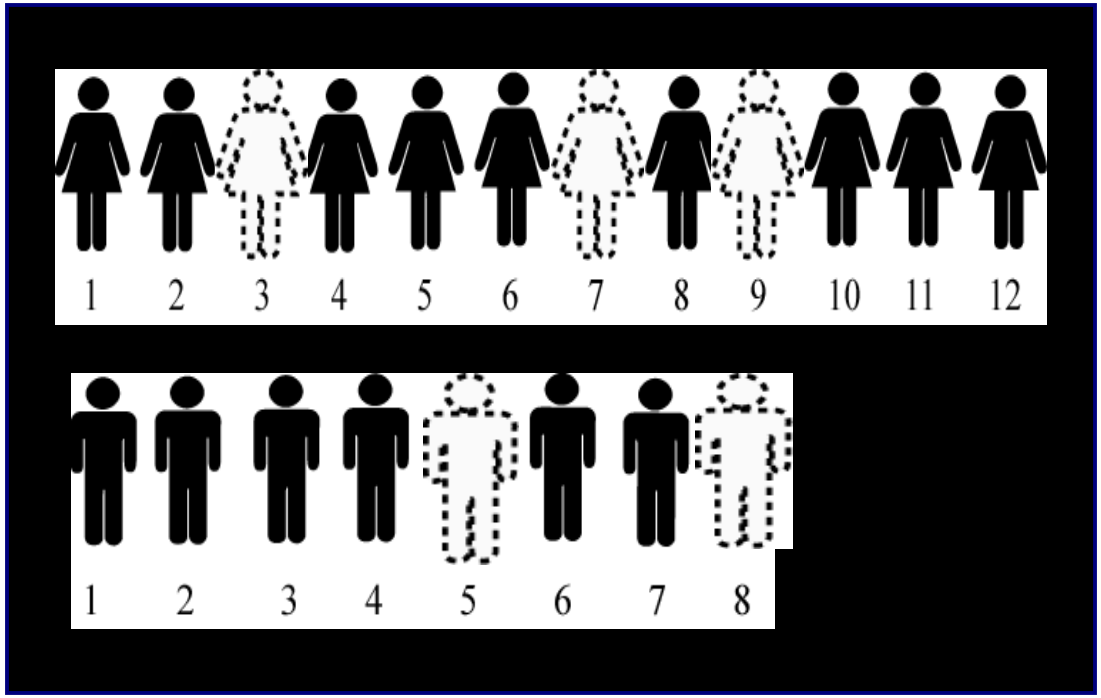
Random Start: 2

A random start in the sequence is selected, and sample is selected by selecting cases sequentially in the list to produce the desired sample size

III. STRATIFIED RANDOM SAMPLING

- Used when the population is not homogeneous.
- The population under study is first divided into homogeneous groups or classes called ‘strata’ & sample is drawn from each stratum at random in proportion to its size.
- Gives representation to all ‘*strata’* of the society or population.

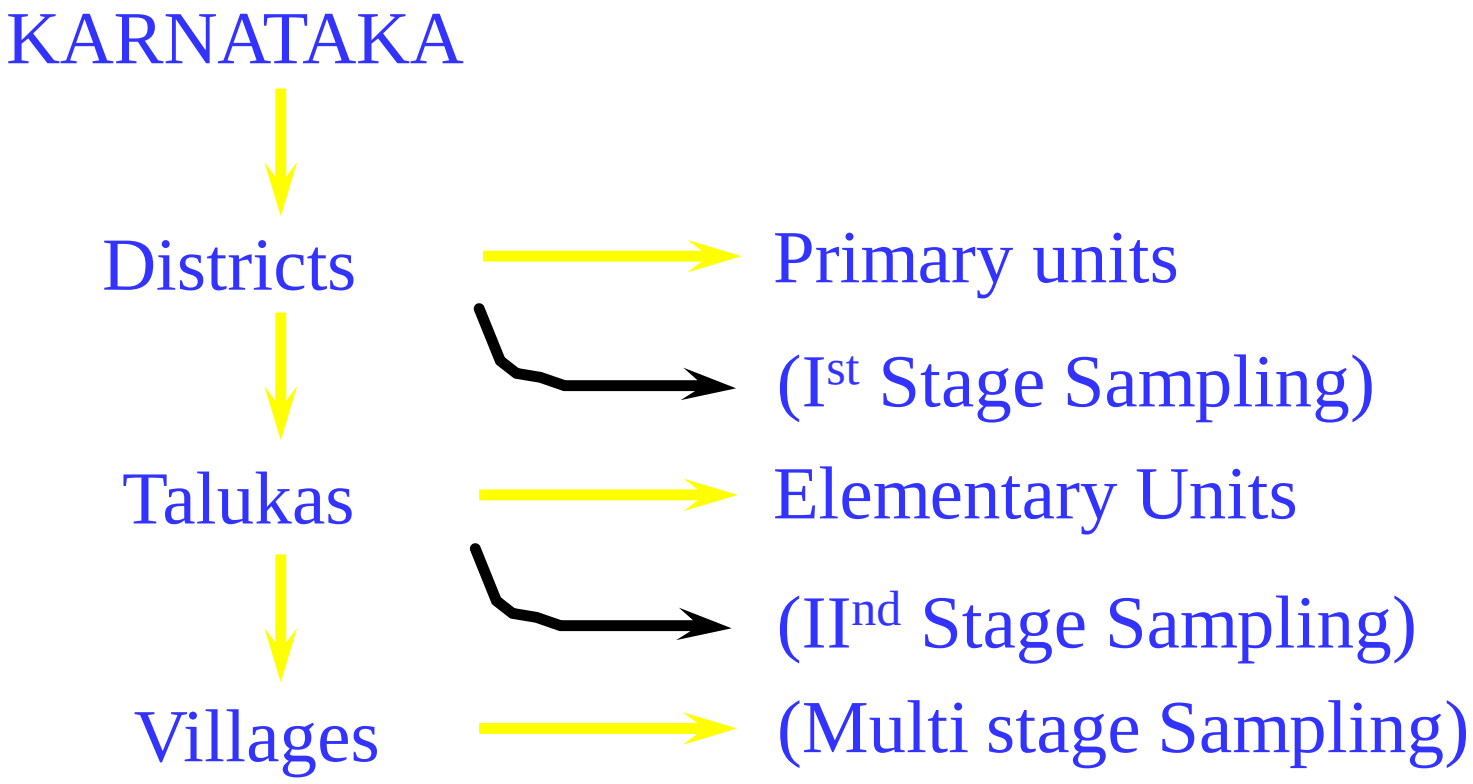
Stratified Random Sampling



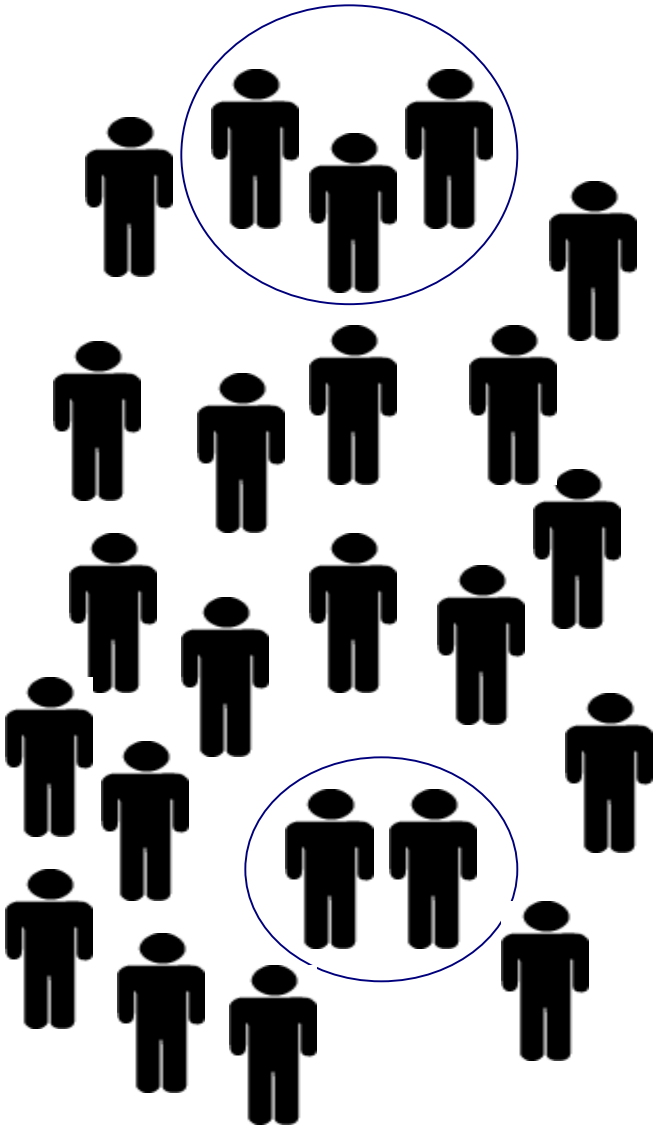
Subgroups (“strata) created that separate members of the population on some important attribute (e.g., sex, race). A random sample from each stratum is then drawn.

IV. CLUSTER RANDOM SAMPLING

Population is heterogeneous, Vast and Scattered over a wide area

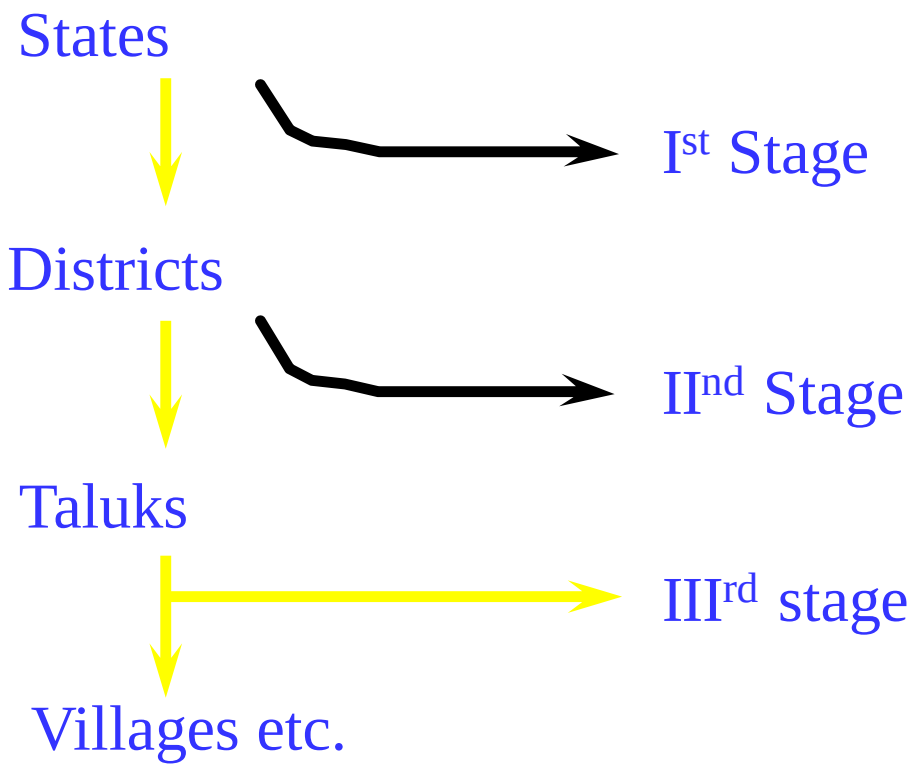


Cluster Sampling



V. MULTISTAGE SAMPLING

- Here, the sampling procedure are carried out in several stages using simple random sampling techniques.



VI. MULTIPHASE SAMPLING

- Part of the information is collected from the whole sample & part from the sub sample.

- Number on the sub-samples decrease in 2nd, 3rd & 4th phases & will become smaller & smaller.

ADVANTAGES

1) Less costly & more purposeful

2) Less laborious

NON PROBABILITY SAMPLING METHODS

I. Quota sampling

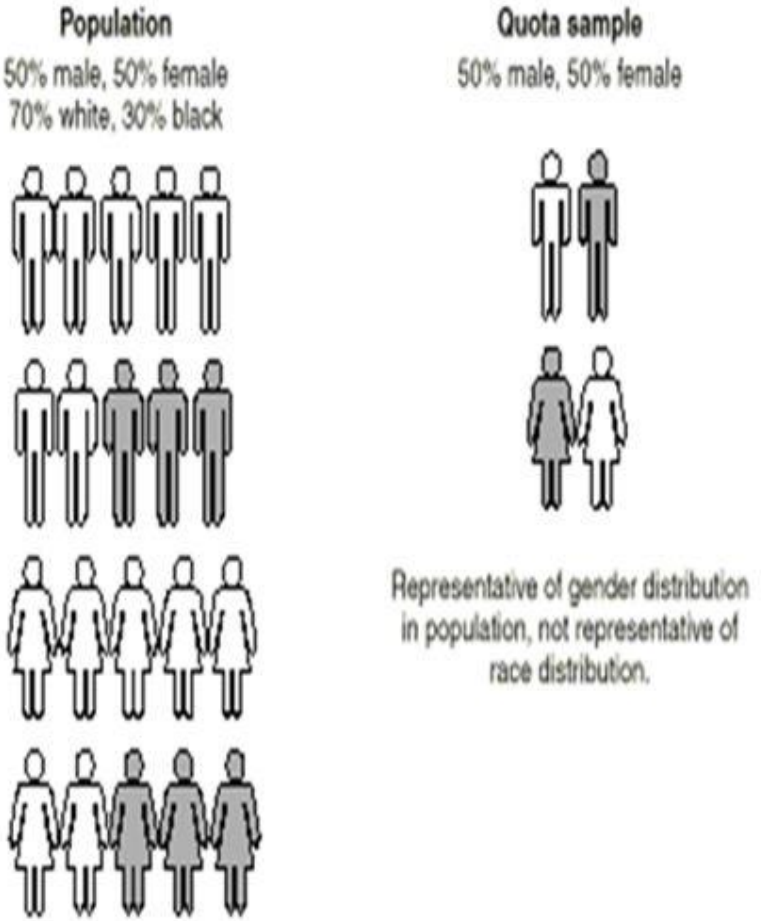
II. Sample of convenience

III. Judgment sampling

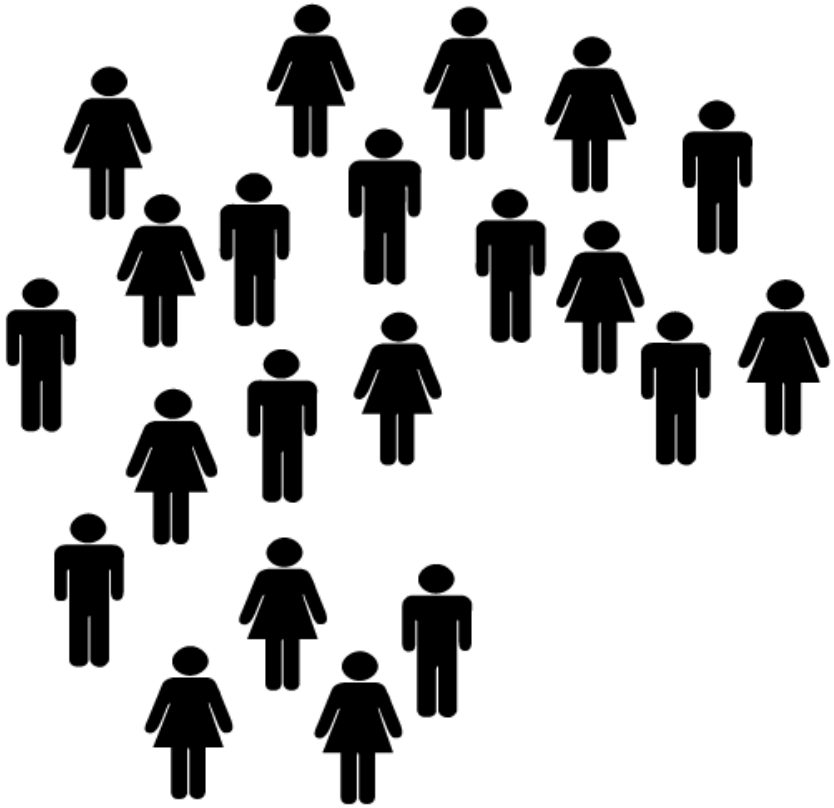
IV. Snow ball sampling

Quota Sampling

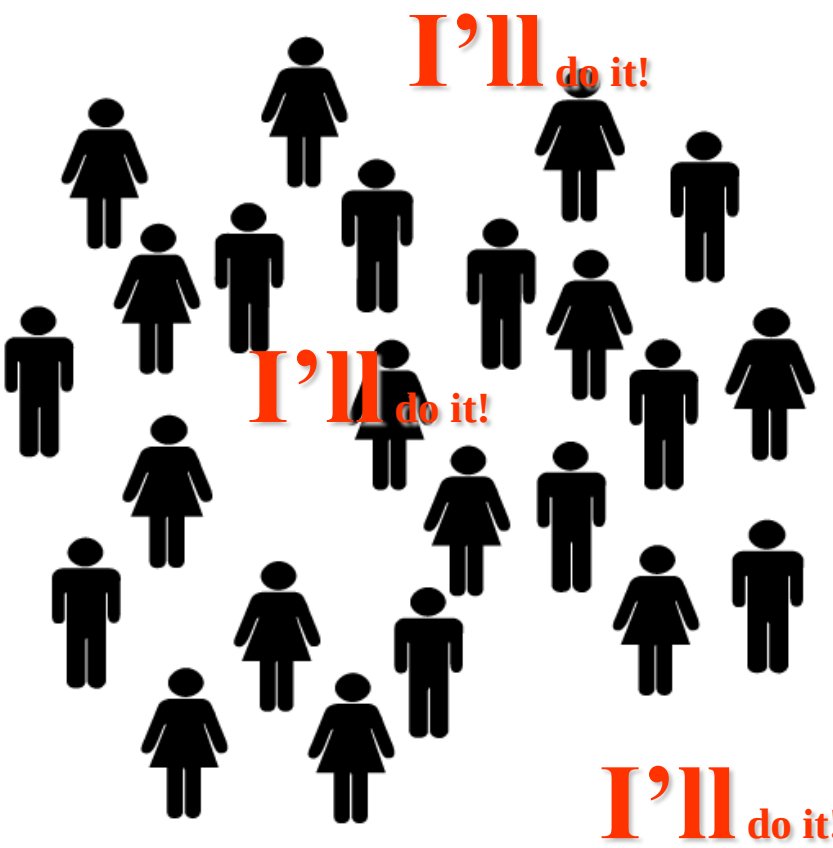
Example: Quotas have been set for gender only. Under the circumstances, it's no surprise that the sample is representative of the population only in terms of gender, not in terms of race. Interviewers are only human;.



Convenience Sampling

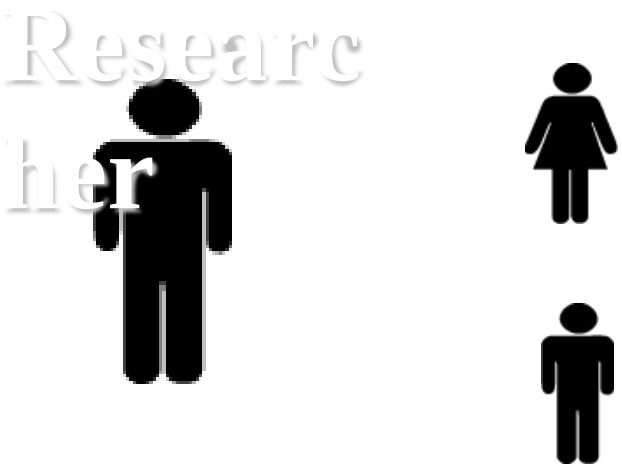


Volunteer Sampling



Researcher uses only people who volunteer to participate in the research

Network/Snowball Sampling



Researcher selects a few participants, who then suggests others who may be willing to participate

METHODS OF PRESENTATION OF DATA

1. Tables

2. Diagrams

3. Graphs

TABULATION

Types of Tables

1. General Purpose Table

2. Analytical Table (Summery Table)

3. Simple Table

4. Complex Table

DIAGRAMMATIC REPRESENTATION

(Qualitative Data)

1.One Dimensional Diagrams

i. Simple Bar Diagram

ii. Multiple Bar Diagram

iii.Component Bar Diagram

iv. Percentage Bar Diagram

2.Two Dimensional Diagram

i. Rectangles

ii. Pie-Diagram

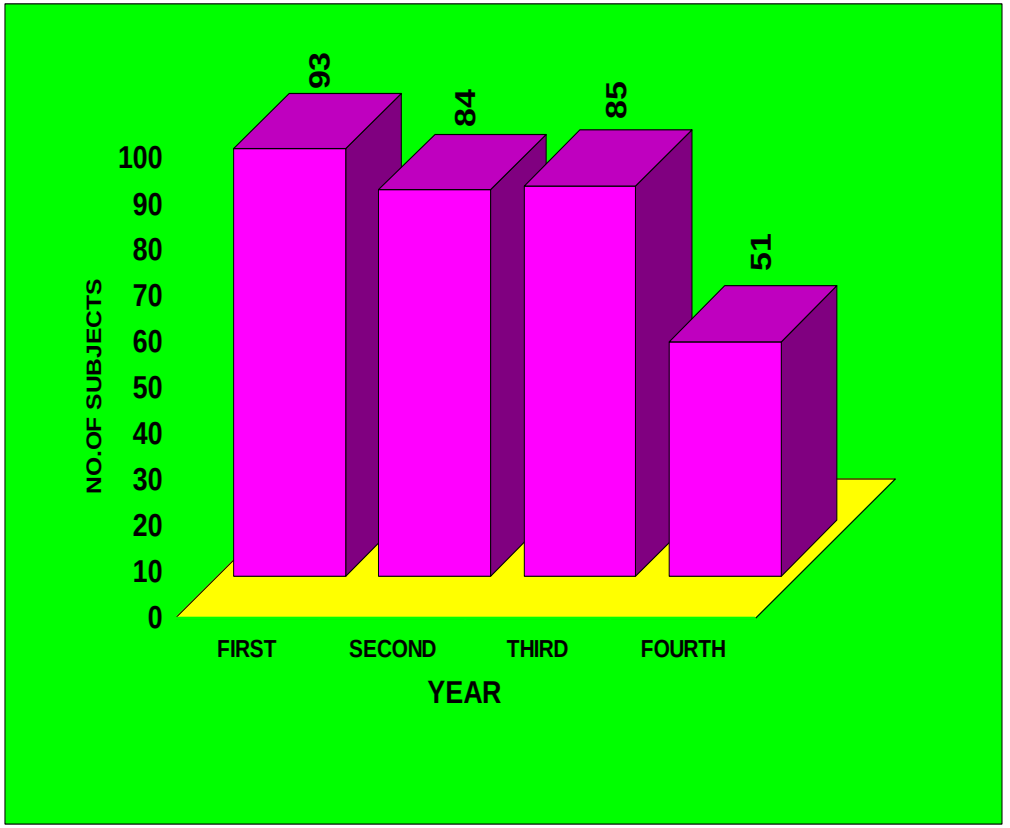
3. Pictograms

SIMPLE TABLE

DISTRIBUTION OF STUDY BDS SUBJECTS, YEAR WISE

B. D. S. YEAR	TOTAL NO. OF SUBJECTS
FIRST	93
SECOND	84
THIRD	85
FOURTH	51
TOTAL	313

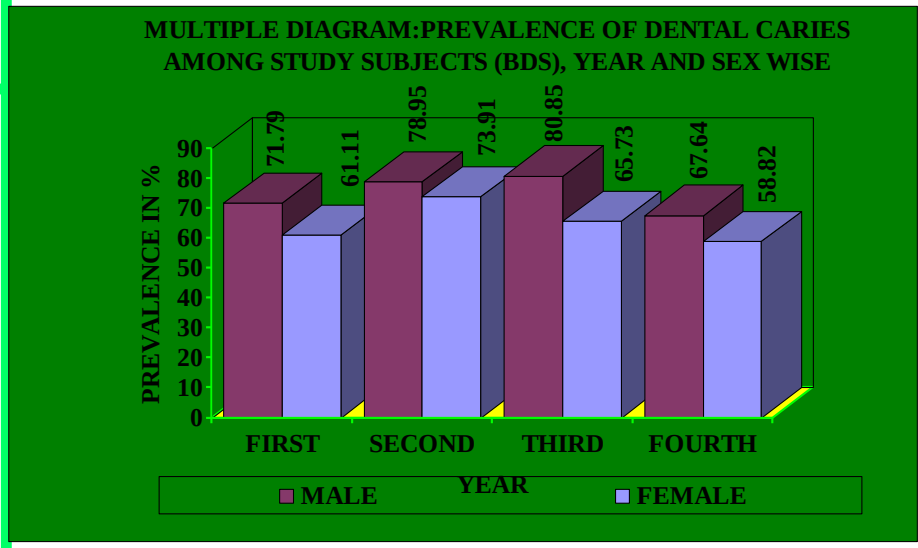
SIMPLE BAR DIAGRAM



ANALYTICAL TABLE

Distribution of prevalence of dental caries among study subjects,year and sex wise

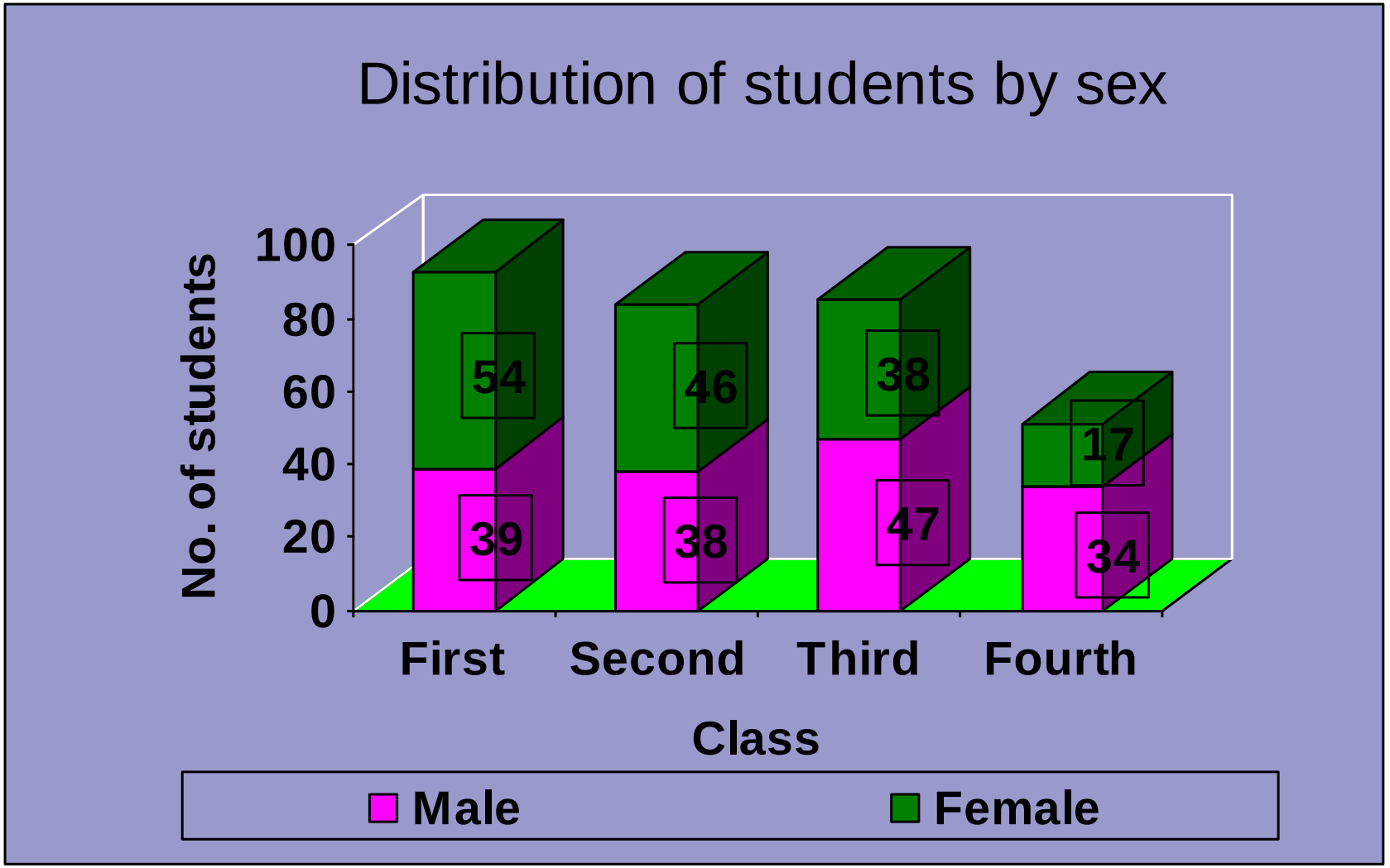
YEAR	NO.OF STUDENTS			
	F	M	F	M
FIRST	39(28)	54(33)	71.79	61.11
SECOND	38(30)	46(34)	78.95	73.91
THIRD	47(38)	38(25)	80.85	65.78
FOURTH	34(23)	17(10)	67.64	58.82



COMPONENT BAR DIAGRAM

The bars are constructed on the basis of total and The total divided into its components.

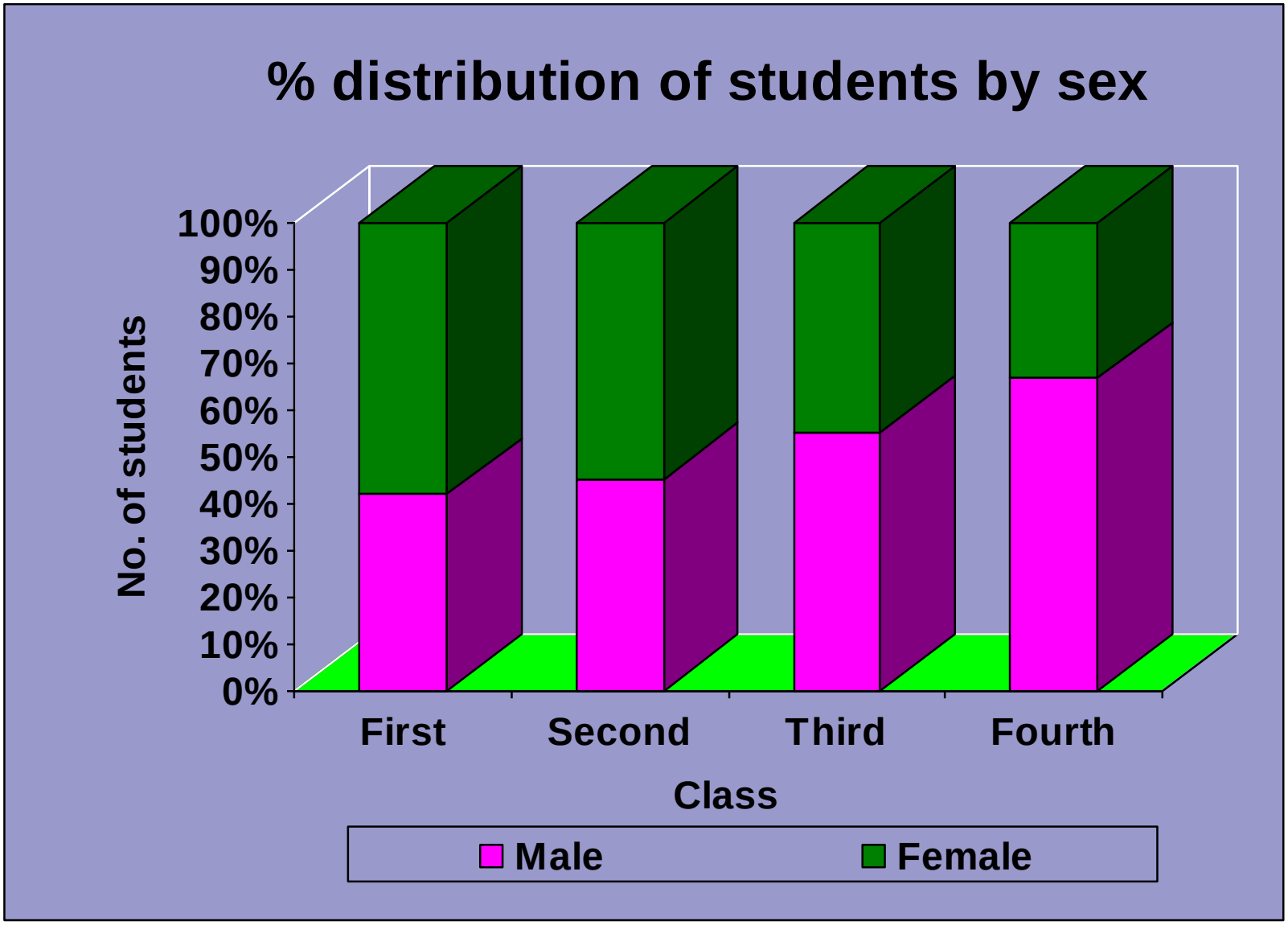
COMPONENT BAR DIAGRAM



PERCENTAGE BAR DIAGRAM

The previous bars are based on the absolute values, some times those not gives any information. The absolute are converted into percentage, the presentation is known as percentage bar diagram. In this bar, the height and width are equal to 100.

PERCENTAGE BAR DIAGRAM



TWO DIMENSIONAL DIAGRAMS

Two dimensional diagrams are also known as area diagrams.

i.e. Area = Breadth x Length

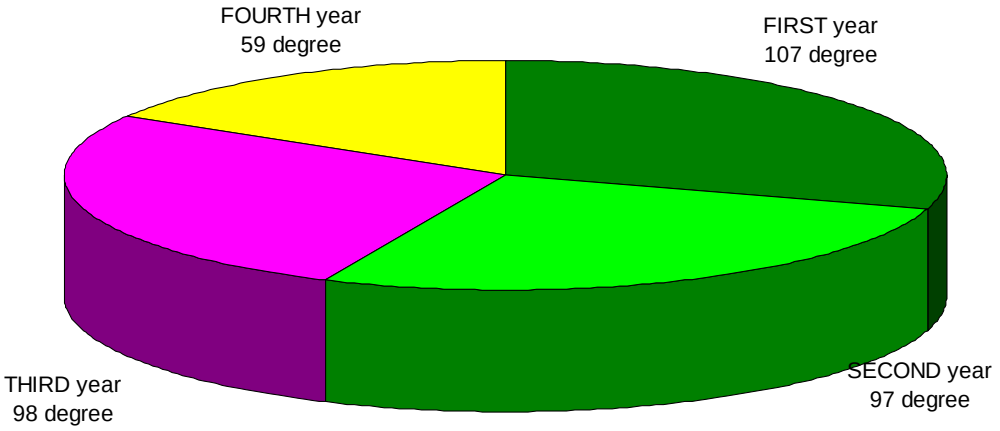
Rectangles

Pie-Diagram

PIE-DIAGRAM

DISTRIBUTION OF STUDY POPULATION ,YEAR - WISE

Class	NO.	Angles
FIRST year	93	107
SECOND year	84	97
THIRD year	85	98
FOURTH year	51	59
Total	313	

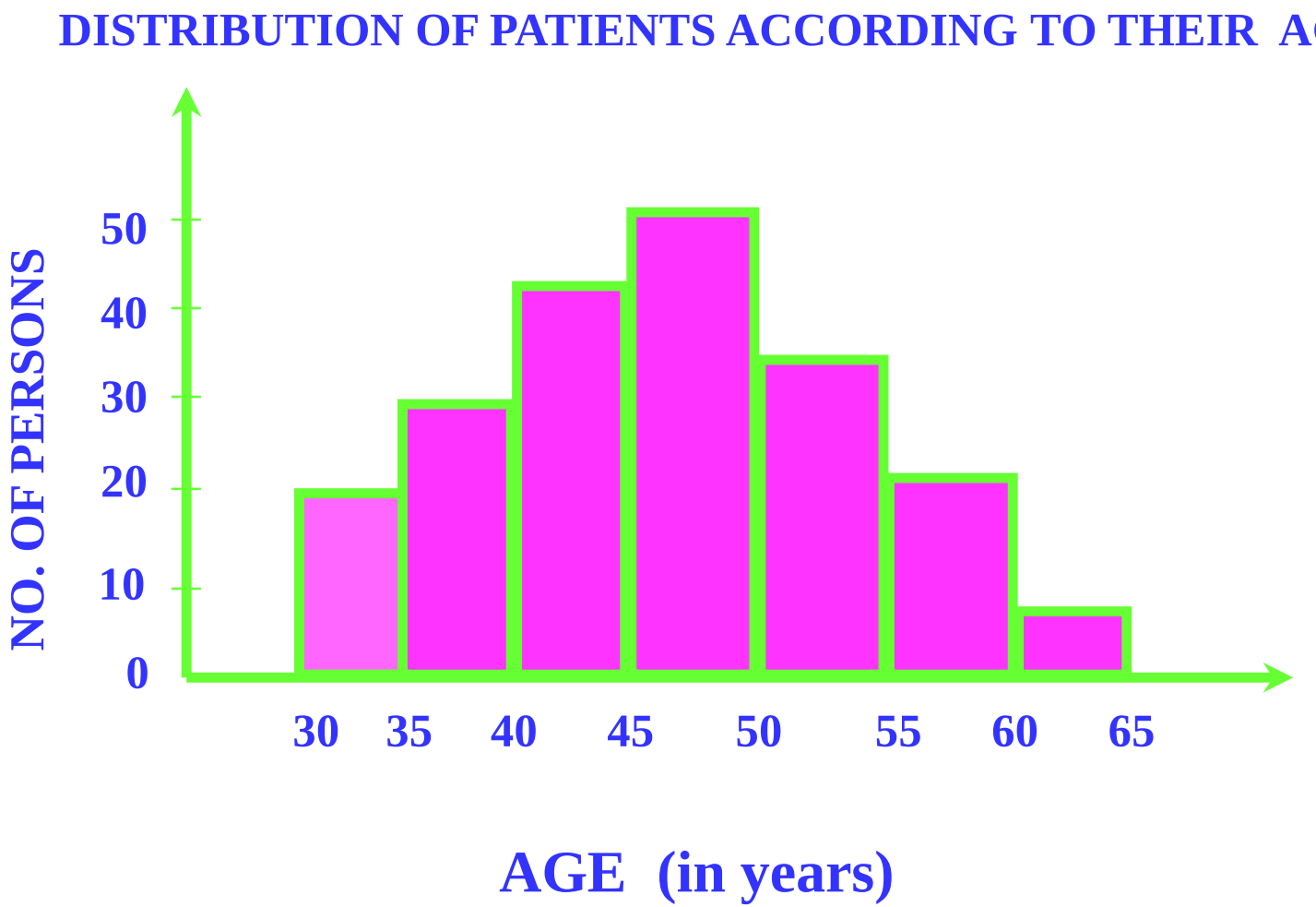


GRAPHICAL REPRESENTATION
(QUANTITATIVE DATA)

Frequency distribution table can be presented by any one of the following graphs:

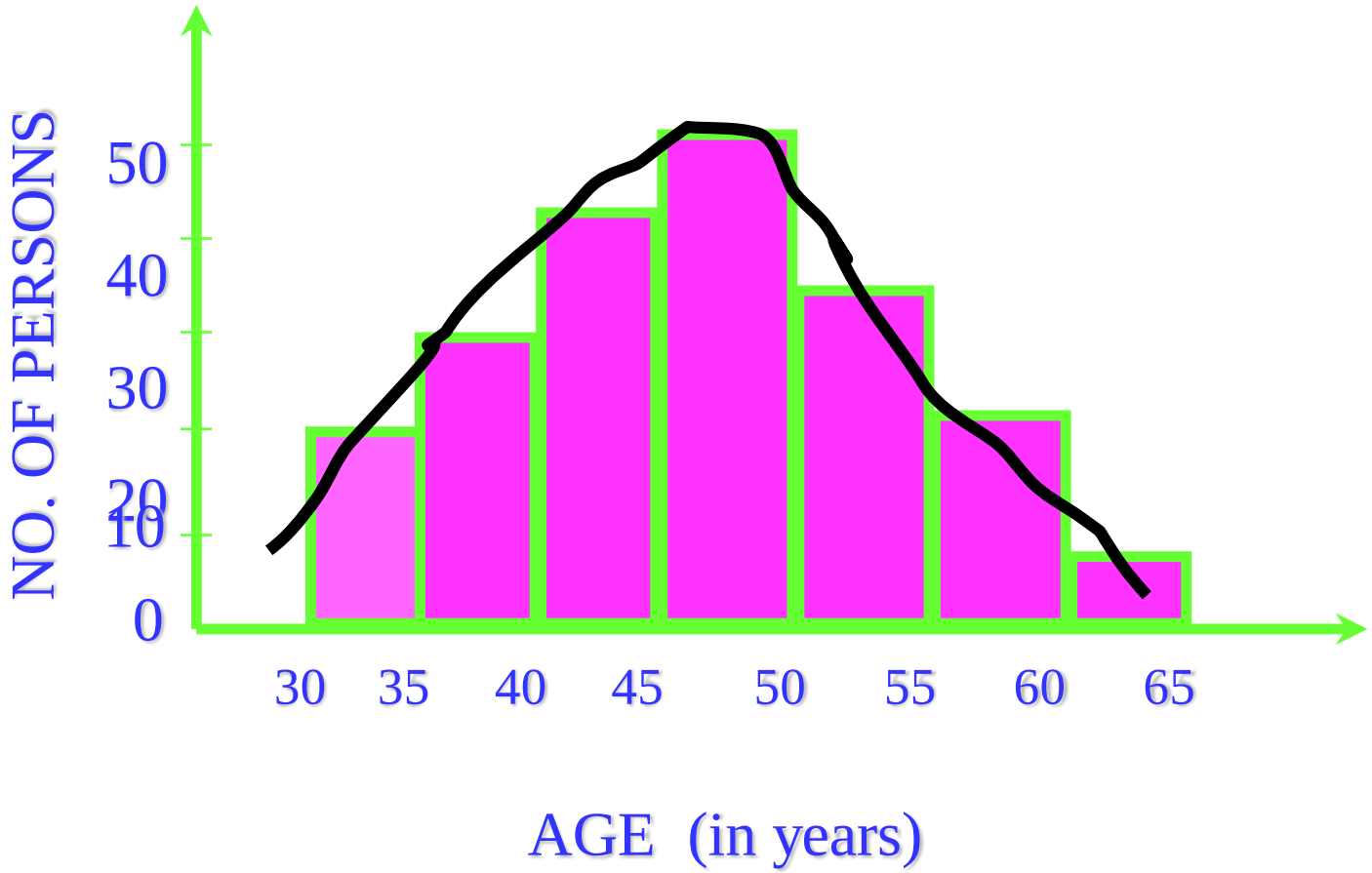
1. Histogram
2. Frequency Polygon
3. Frequency curve
4. Cumulative (or Ogive) curves

HISTOGRAM

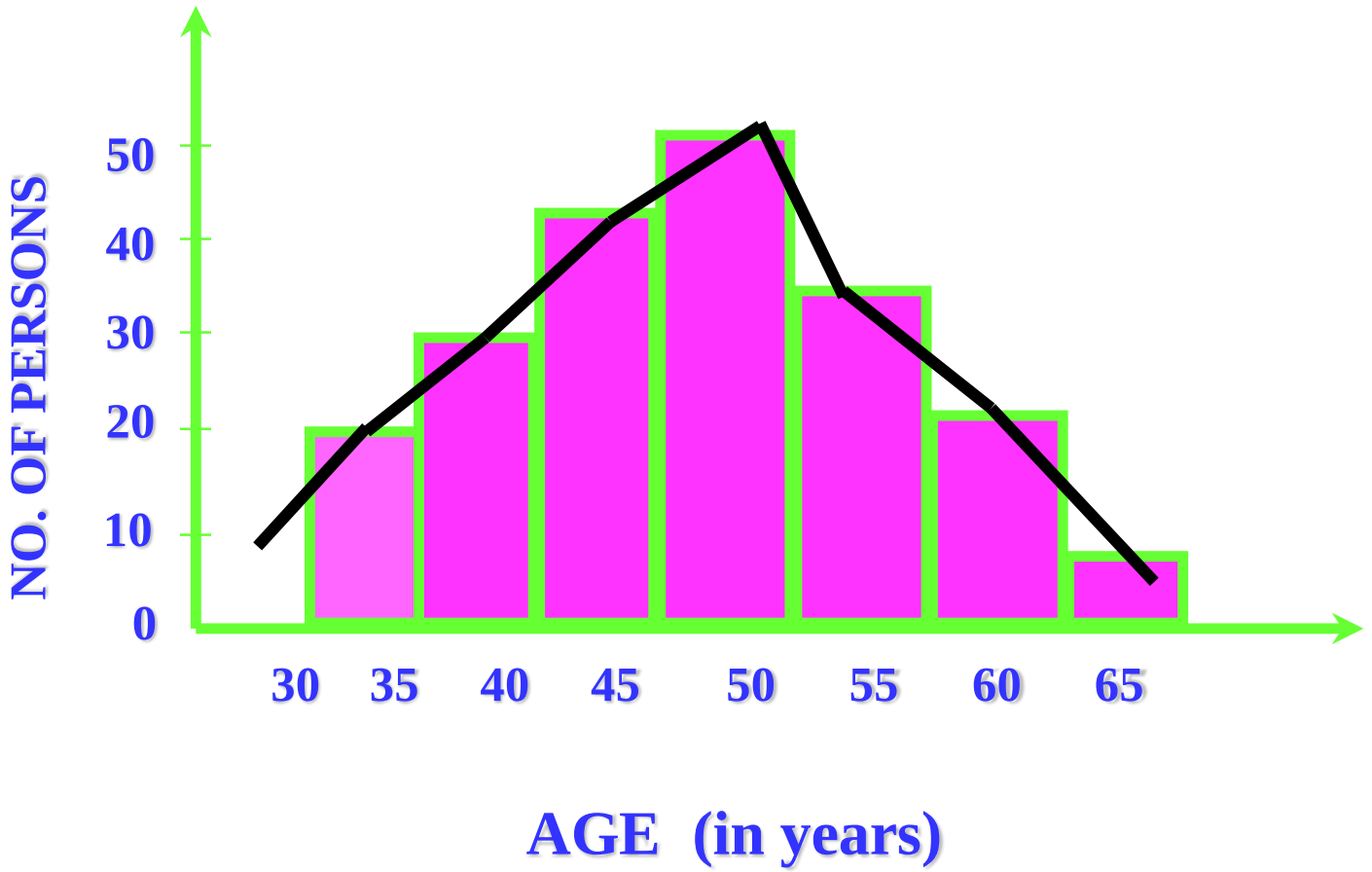

DISTRIBUTION OF PATIENTS ACCORDING TO THEIR AGE
NO. OF PERSONS
50
40
30
20
10
0
30
35
40
45
50
55
60
65
AGE (in years)

FREQUENCY CURVE

DISTRIBUTION OF PATIENTS ACCORDING TO THEIR AGE



FREQUENCY POLYGON



METHODS OF SUMMARIZATION

(a) Measures of Central Tendency

1. Mean
2. Median
3. Mode

(b) Measures of Dispersion

1. Range
2. Mean Deviation
3. Standard Deviation

MEAN

Mean is obtained by summing up all the observations and dividing the total by the number of observations. It is given by

X1 ,X2 ,X3,……..,Xn ,n observations

$$\text{MEAN } (\bar{x}) = \frac{X1+X2+X3+\ldots\ldots\ldots+Xn}{n}$$

$$= \Sigma X/n$$

Where, x= variable, n= sample size

Ex:The following data gives the plaque scores in 5 students. Calculate Mean plaque score.

1.1267, 0.9834, 1.5634, 1.0267, 2.3245

Solution:

Mean = [1.1267+0.9834+1.0267+2.3245] / 5

= 7.0247 / 5

= 1.4049

(Mean plaque score of 5 students)

MEDIAN

Median is the *middle value* (frequencies) after arranging them *either* in the ascending or in the descending order.

Median = Size of the $[(n+1)/2]^{th}$ item

If n is odd number,

Median divides the observations EXACTLY into half (Middle Term).

If n is even number,

Median is the MEAN of the middle two terms

Ex: The following data gives the plaque cores in 5 students. Calculate Median plaque score.

1.1267, 0.9834, 1.5634, 1.0267, 2.3245

Ascending order:

0.9834, 1.0267, 1.1267 1.5634, 2.3245

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

= Size of the [(5+1)/2]th item

= Size of the 3th item

= 1.1267

MODE

Mode is the one which is the most repeated in the particular series of observations.

OR

It is the value of the variable which occurs most frequently in a series of observation.

Ex: Calculate Mode from the following Plaque scores

1.25, 3.10, 0.95, 0.75, 1.81, 1.81, 2.72, 2.50

1.81 repeated 2 times ,

Therefore, the frequency of 1.81 is 2

The value of mode is 1.81

MEASURES OF DISPERSION

1. Range

2. Mean Deviation

3. Standard Deviation

Measures of variability of observations help to find how individual observations are dispersed around the mean of a series

RANGE

It is the difference between the highest and lowest values in the series.

RANGE = H - L

= Maximum value - Minimum value

- It is simplest measure of dispersion
- It is no based upon all the observations
- It is based on extreme values
- it is affected by sampling fluctuations

Ex: Find range of incubation period of measles in 9 patients

14, 13, 11, 15, 10, 7, 9, 12, 10 (in days)

Here,

Maximum Value = 15

Minimum Value = 7

Range = 15 - 7 = 8

MEAN ABSOLUTE DEVIATION

OR MEAN DEVIATION

It is the arithmetic mean of the deviations of the values from a measure of central tendency without taking plus and minus signs into consideration.

If X is the variable, n the number of observations and $\bar{x}$ is mean and it is given by

$$M.D. = \sum |x - \bar{x}| / n$$

Where | | indicates ignoring the negative signs

Steps involved in calculation of mean deviation:

- Calculate mean.
- Find deviations, i.e.$(x-\bar{x})$
- Find absolute deviation of $(x-\bar{x})$ i.e. $I x-\bar{x} I$
- find the sum of absolute deviations,

$$\Sigma I x-\bar{x} I$$

- Divide sum by n , i.e. $\Sigma I x-\bar{x} I / n$

Ex. The following data gives the respiration rate per minute.

Find the Mean Deviation.

23, 22, 24, 16, 17, 18, 19, 21, 20

Respiration Rate per min (x)	$(x - \bar{x})$	$\lvert x - \bar{x} \rvert$
23	3	3
22	2	2
24	4	4
16	-4	4
17	-3	3
18	-2	2
19	-1	1
21	1	1
20	0	0
Total ($\bar{x} = 20$)	0	22

Mean Deviation

$= \Sigma \lvert x - \bar{x} \rvert / n$

$= 22/9$

$= 2.4444$

Standard Deviation is the square root of the variance.

OR

It is defined as "square-root of the summation of squares of all the deviations being measured from the mean of the observations".

S.D. = $\sqrt{\sum(\bar{x} - x)^2 / n-1}$ for small samples

= $\sqrt{\sum (x - \bar{x})^2 / n}$ for large samples

$\bar{x}$ = mean of the observations

STEPS INVOLVED IN CALCULATING STANDARD DEVIATION

1. Calculate Mean
2. Find Deviations, i.e.$(x - \bar{x})$
3. Find Deviations square, i.e. $(x - \bar{x})^2$
4. Find sum of the squares, $\sum (x - \bar{x})^2$
5. Divide sum by n-1 or n

   n-1 for small samples

   n for large samples
6. Take square root in step no.5

Ex. The following data gives the respiration rate per minute. Find the Standard Deviation.

23, 22, 24, 16, 17, 18, 19, 21, 20

Respiration Rate per min (x)	$(x - \bar{x})$	$(x - \bar{x})^2$
23	3	9
22	2	4
24	4	16
16	-4	16
17	-3	9
18	-2	4
19	-1	1
21	1	1
20	0	0
Total ($\bar{x} = 20$)	0	60

Standard Deviation

$= \Sigma (x - \bar{x})^2 / n-1$

$= 60/9-1$

$= 7.5$

$= 2.7386$

NORMAL DISTRIBUTION AND NORMAL CURVE

After collection of large samples, prepare a frequency Distribution with small class intervals, see the following points:

1) Some observations are above the mean and others are below the mean.

2) If they are arranged in order, deviating towards the extremes from the mean, on plus or minus side, maximum number of frequencies will be seen in the middle around the mean and fewer at the extremes, decreasing smoothly on both sides.

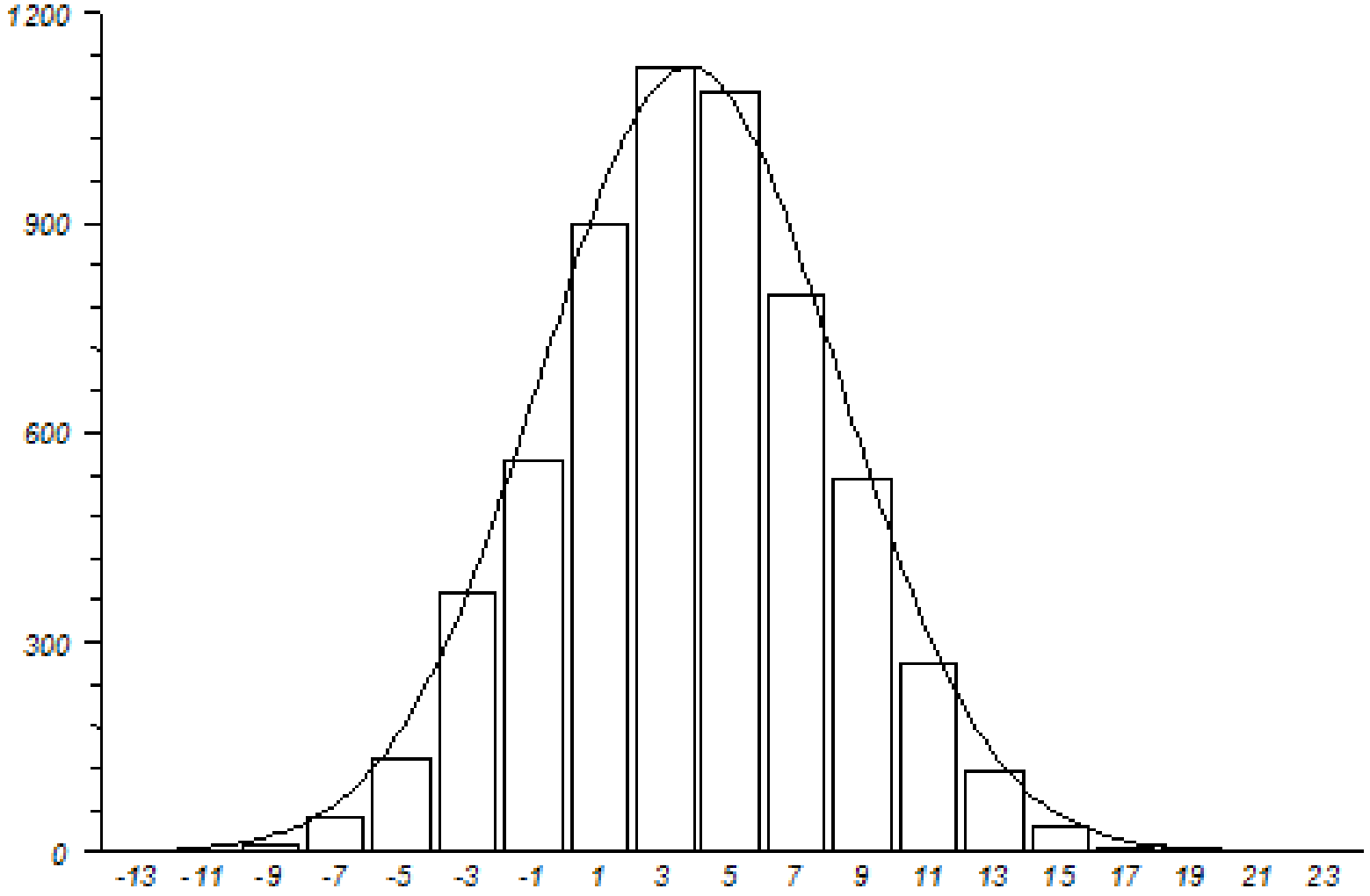
3) Normally almost half the observations lie above and half lie below the mean.

4) All the observations are symmetrically distributed on each side of the mean.

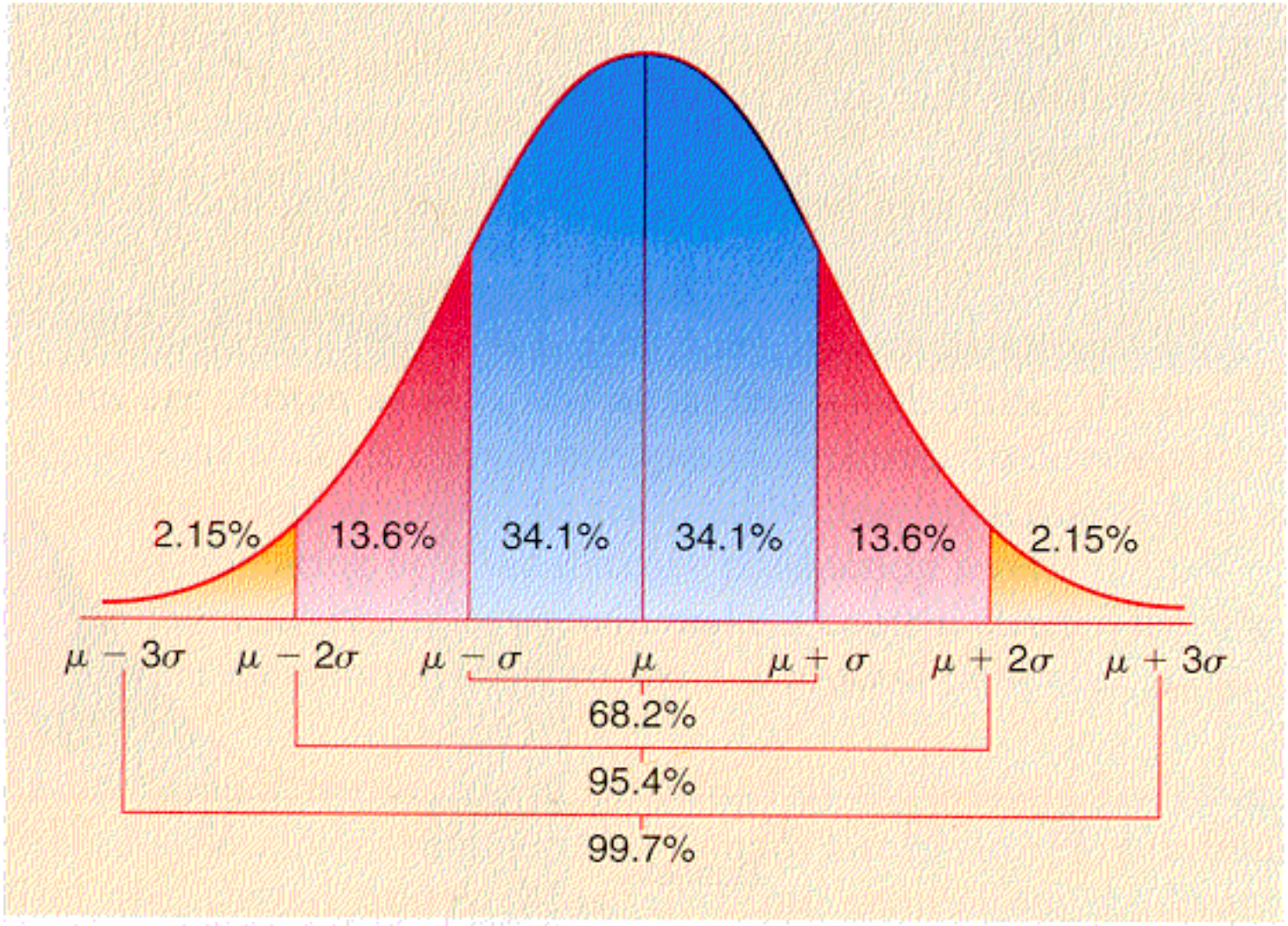
A distribution of this nature is called as NORMAL DISTRIBUTION OR *GAUSSIAN DISTRIBUTION*.

Histogram of the frequency distribution gives a frequency curve which is symmetrical in nature called as the *NORMAL CURVE* OR *GAUSSIAN CURVE*.

NORMAL CURVE



NORMAL CURVE



CHARACTERISTICS OF A NORMAL CURVE

1. It is bell shaped curve and symmetrical about the mean which at the max. ordinate
2. The first and third quartiles are equidistant from median.
3. Mean, Median and Mode are coincides.
4. The curve approaches horizontal axis but never reaches horizontal axis.
5. Area Property.
    i. Mean ± SD includes 68.26% area
    ii. Mean ± 2SD includes 95.44% area
    iii. Mean ± 3SD includes 99.74% area

BASIC CONCEPTS OF STATISTICAL INFERENCES

HYPOTHESIS

Any assumption regarding the parameter or the nature of the probability behavior.

NULL-HYPOTHESIS

It is a hypothesis which reflects no change or no difference, usually denoted by H_0.

ALTERNATIVE HYPOTHESIS

Any alternative assumption to null hypothesis. OR

It is a statement which reflects the situation anticipated to be correct if the null hypothesis is wrong, usually denoted by H_1.

LEVEL OF SIGNIFICANCE

The maximum probability of rejecting a correct null hypothesis (H_0).

TESTS OF SIGNIFICANCE

Are the mathematical methods by which probability (P) or relative frequency of an observed difference occurring by chance is found.

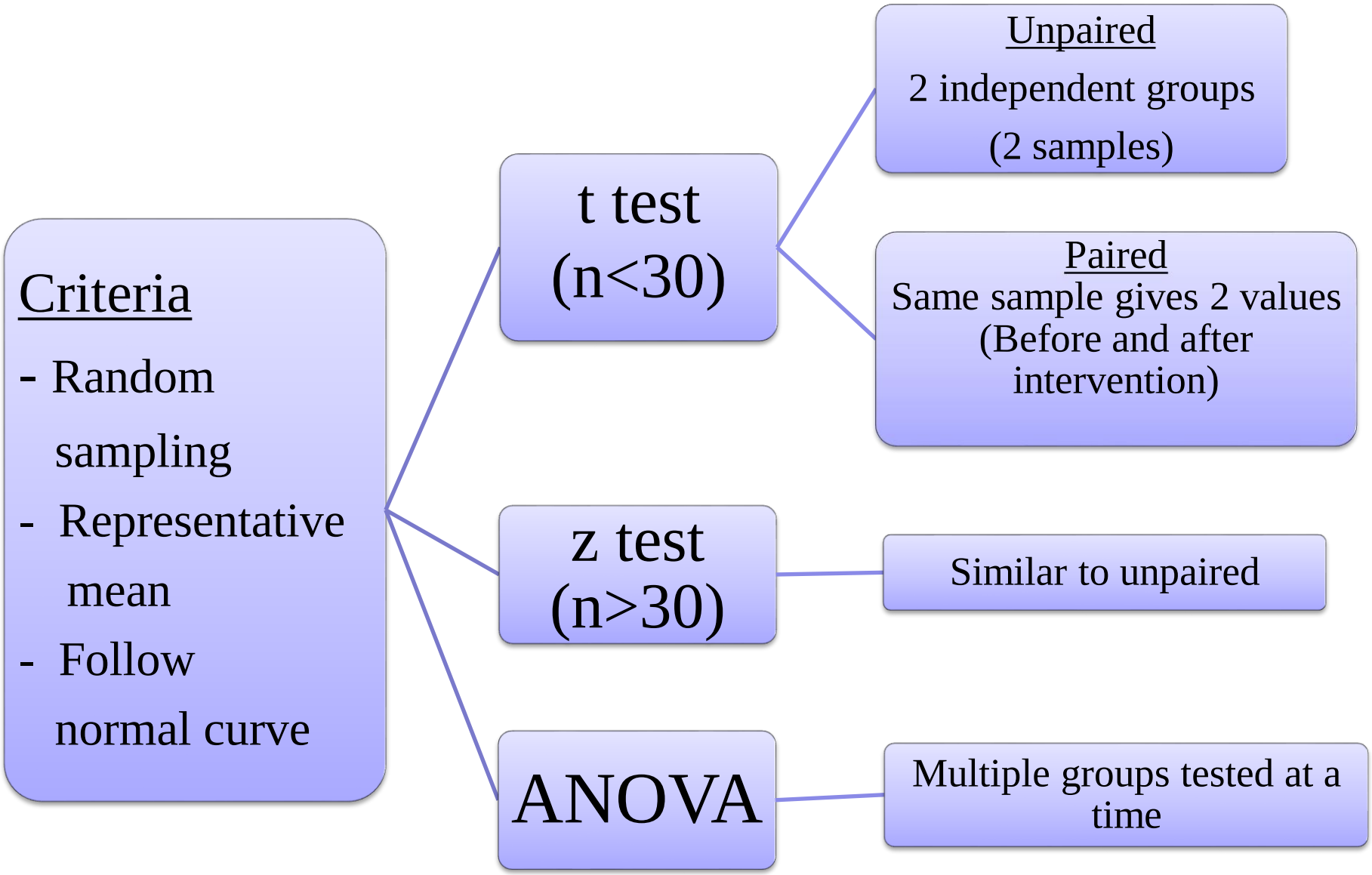
TYPES OF TESTS OF SIGNIFICANCE

Parametric tests

- Students t-test
  - Unpaired t-test for small samples
  - Paired t-test for small samples
- Z-test for large samples
- Analysis of variance

Non-parametric tests

- Chi-square test
- Mann Whitney U-Test
- Wilcoxon rank sum test
- Kruskal Wallis test


Criteria
- Random sampling
- Representative mean
- Follow normal curve
t test (n<30)
Unpaired
2 independent groups
(2 samples)
Paired
Same sample gives 2 values (Before and after intervention)
z test (n>30)
Similar to unpaired
ANOVA
Multiple groups tested at a time

STAGES IN PERFORMING A TEST OF SGNIFICANCE.

- Set up null hypothesis
- Set up alternative hypothesis
- Calculate test statistic i.e. t, Z, chi-square, etc.
- Calculate degrees of freedom
- Select table or critical value with DF
- Compare critical and calculated values
- Tab value < calculated value, Ho is rejected
- Tab value > calculated value, Ho is accepted
- Draw your conclusions

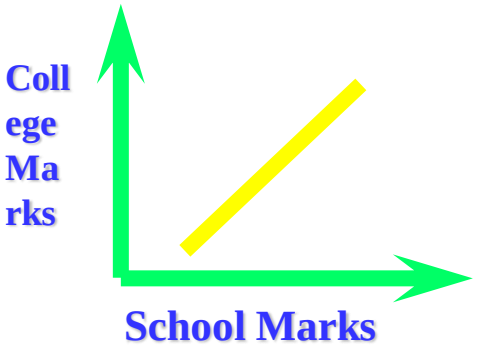
TYPES OF CORRELATION

- Positive correlation (r = +1)
- Negative correlation (r = -1)
- Zero Correlation (r = 0)
- Moderately +ve Correlation (0<r<1)
- Moderately - ve Correlation (-1<r<0)

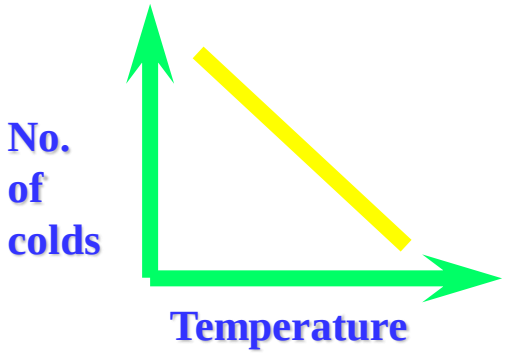
METHOD OF STUDYING CORRELATION

1. Scatter Diagram (Graphical Method):

Perfect +ve correlation (r=+1)

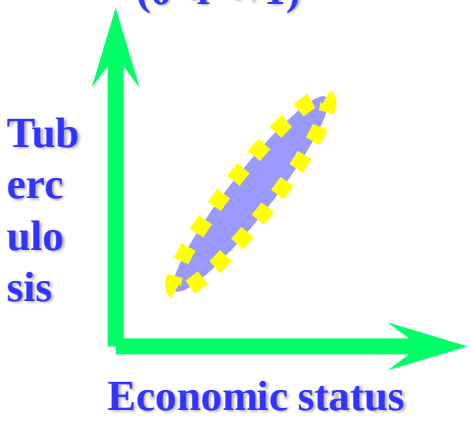


Perfect -ve correlation (r=-1)



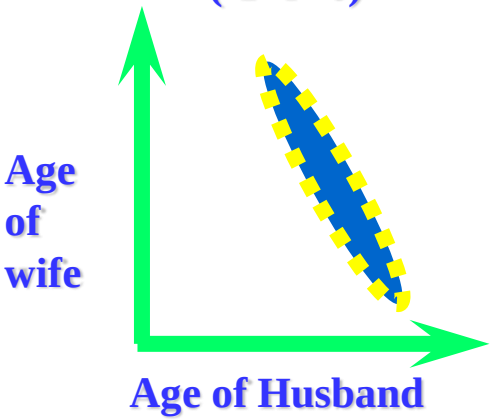
Moderately +ve Correlation

(0<r<+1)



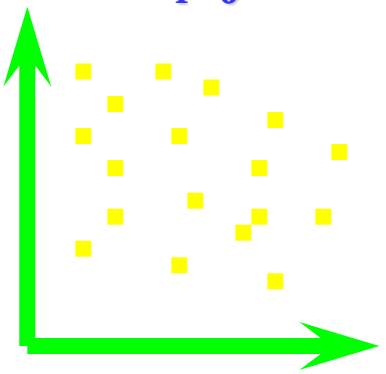
Moderately +ve Correlation

(-1<r<0)



Zero correlation

r=0



APPLICATIONS AND USES

In physiology and anatomy

- **To define the limits of normality for variables such as height, weight, blood pressure etc. in a population. (Variation more than natural limits may be pathological i.e abnormal due to play of certain external factors.)**
- **To find correlation between two variables like height and weight.**

In pharmacology

- **To find the action of drugs**
- **To compare the action of two drugs or two successive dosages of the same drug**
- **To find the relative potency of a new drug with respect to a standard drug**

In medicine

- **To identify signs and symptoms of disease.**
- **To find association between two attributes such as cancer and smoking.**
- **To compare the efficiency of a particular drug, operation or line of treatment.**
- **To test the efficacy of different treatments. eg. Surgical management vs medical management of angina patients**

In community medicine and public health

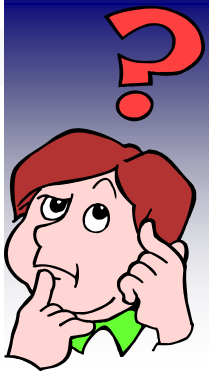
- To test usefulness of sera or vaccines in the field
- In epidemiologic studies the role of causative factors is statistically tested
- To compare the outcomes of different health care delivery systems

In research

- **It helps in compilation of data, drawing conclusions and making recommendations.**

For students of medicine/dentistry

- **By learning the methods in biostatistics a student learns to critically evaluate articles published in medical and dental journals or papers read in medical and dental conferences.**
- **To understand the basic methods of observation in clinical practice and research.**



Uses of statistics in dental science:

- **To find the statistical difference between means of two groups. Ex: Mean plaque scores of two groups.**

- **To assess the state of oral health in the community and to determine the availability and utilization of dental care facilities.**

- **To indicate the basic factors underlying the state of oral health by diagnosing the community and find solutions to such problems.**
- **To determine success or failure of specific oral health care programs or to evaluate the program action.**
- **To promote oral health legislation and in creating administrative standards for oral health care delivery.**

Misuse

"There are three kinds of lies: lies, damned lies, and statistics"

-Mark Twain

- **Inadvertent and intentional.**
- **By choosing (or rejecting, or modifying) a certain sample/study protocol results can be manipulated.**
- **Finding ways to interpret only the data that are favorable to the presenter.**