Introduction to Biostatistics

Deepa Anwar

Biostatistics

- (a portmanteau word made from biology and statistics)
- The application of statistics to a wide range of topics in biology.

Biostatistics

It is the science which deals with development and application of the most appropriate methods for the:

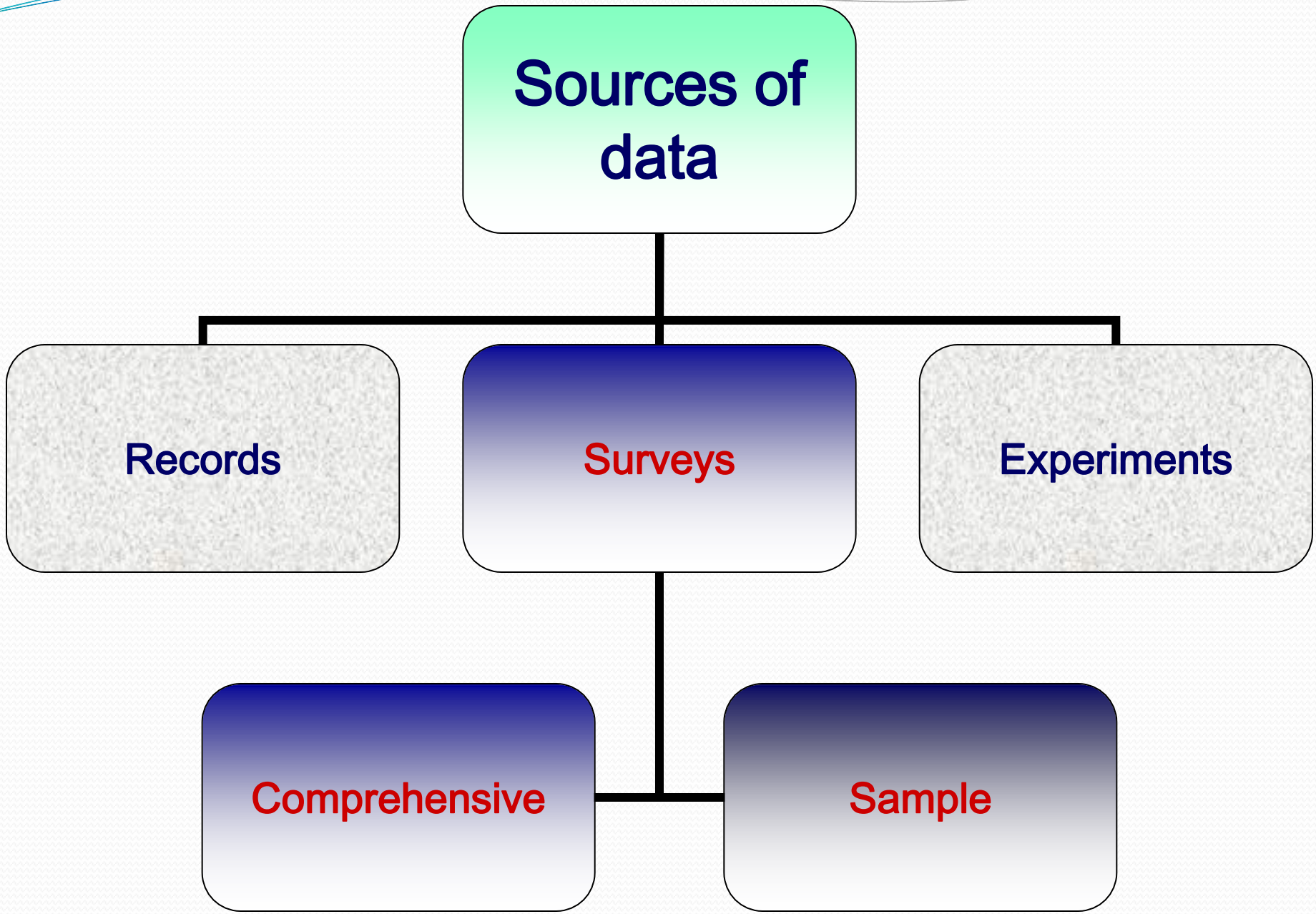
- Collection of data.
- Presentation of the collected data.
- Analysis and interpretation of the results.
- Making decisions on the basis of such analysis

Other definitions for “Statistics”

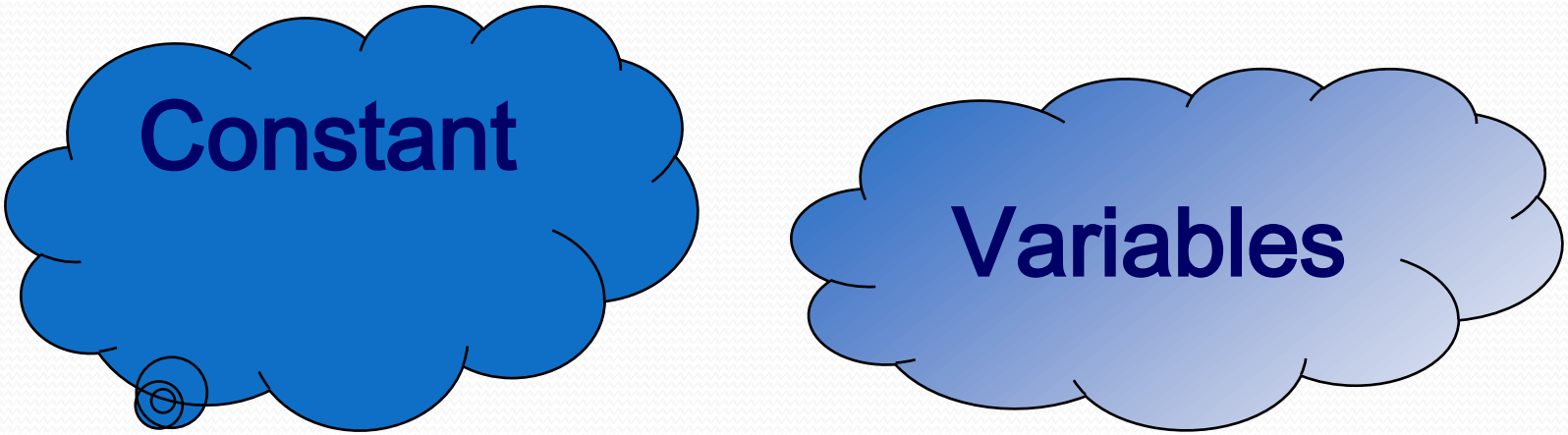
- Frequently used in referral to recorded data
- Denotes characteristics calculated for a set of data : sample mean

Role of statisticians

- To guide the design of an experiment or survey prior to data collection
- To analyze data using proper statistical procedures and techniques
- To present and interpret the results to researchers and other decision makers


Sources of data
Records
Surveys
Experiments
Comprehensive
Sample

Types of data



Types of variables

Quantitative variables

- Quantitative continuous
- Quantitative descrete

Qualitative variables

- Qualitative nominal
- Qualitative ordinal

Methods of presentation of data

1. Numerical presentation
2. Graphical presentation
3. Mathematical presentation

1- Numerical presentation

Tabular presentation (simple – complex)

Simple frequency distribution Table (S.F.D.T.)

Title

Name of variable (Units of variable)	Frequency	%
- - Categories -		
Total		

Table (I): Distribution of 50 patients at the surgical department of Alexandria hospital in May 2008 according to their ABO blood groups

Blood group	Frequency	%
A	12	24
B	18	36
AB	5	10
O	15	30
Total	50	100

Table (II): Distribution of 50 patients at the surgical department of Alexandria hospital in May 2008 according to their age

Age (years)	Frequency	%
20-<30	12	24
30-	18	36
40-	5	10
50+	15	30
Total	50	100

Complex frequency distribution Table

Table (III): Distribution of 20 lung cancer patients at the chest department of Alexandria hospital and 40 controls in May 2008 according to smoking

Smoking	Lung cancer				Total	
	Cases		Control			
	No.	%	No.	%	No.	%
Smoker	15	75%	8	20%	23	38.33
Non smoker	5	25%	32	80%	37	61.67
Total	20	100	40	100	60	100

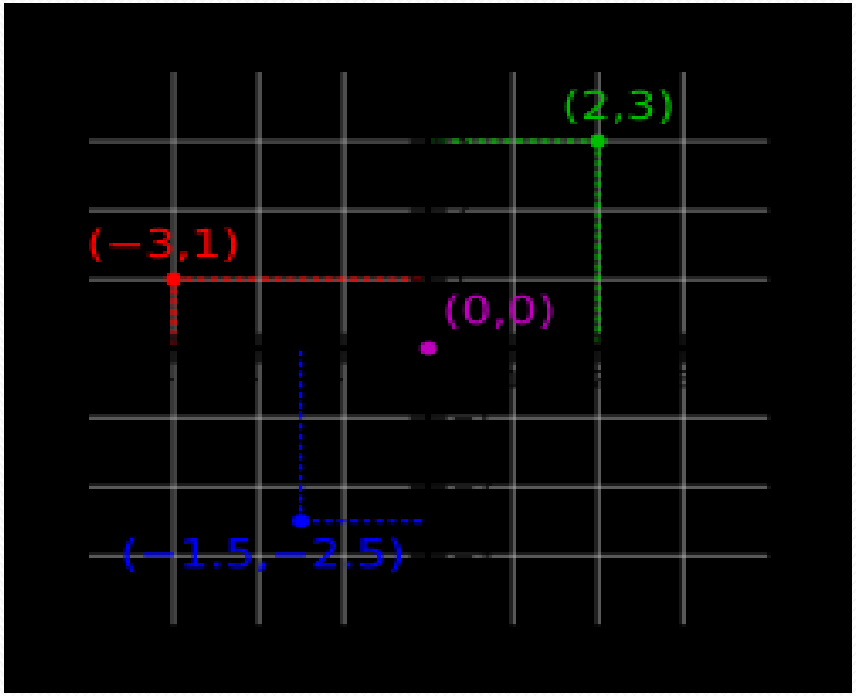
Complex frequency distribution Table

Table (IV): Distribution of 60 patients at the chest department of Alexandria hospital in May 2008 according to smoking & lung cancer

Smoking	Lung cancer				Total	
	positive		negative			
	No.	%	No.	%	No.	%
Smoker	15	65.2	8	34.8	23	100
Non smoker	5	13.5	32	86.5	37	100
Total	20	33.3	40	66.7	60	100

2- Graphical presentation

1. *Graphs drawn using Cartesian coordinates*

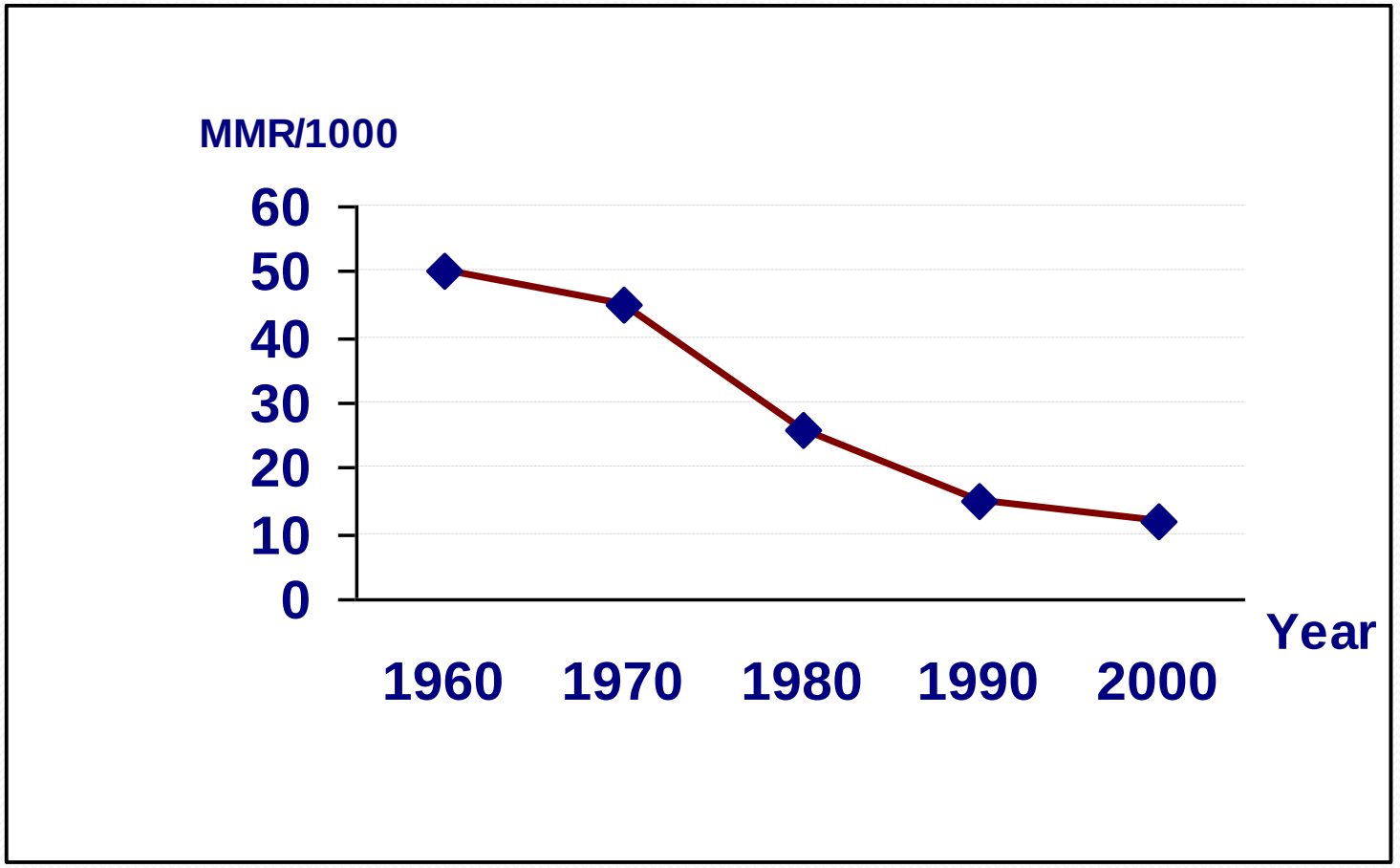


- Line graph
- Frequency polygon
- Frequency curve
- Histogram
- Bar graph
- Scatter plot

2. *Pie chart*

3. *Statistical maps*

Line Graph



Year	MMR
1960	50
1970	45
1980	26
1990	15
2000	12

Figure (1): Maternal mortality rate of (country), 1960-2000

Frequency polygon

Age (years)	Sex		Mid-point of interval
	Males	Females	
20 -	3 (12%)	2 (10%)	(20+30) / 2 = 25
30 -	9 (36%)	6 (30%)	(30+40) / 2 = 35
40-	7 (8%)	5 (25%)	(40+50) / 2 = 45
50 -	4 (16%)	3 (15%)	(50+60) / 2 = 55
60 - 70	2 (8%)	4 (20%)	(60+70) / 2 = 65
Total	25(100%)	20(100%)	

Age	Sex		M-P
	M	F	
20-	(12%)	(10%)	25
30-	(36%)	(30%)	35
40-	(8%)	(25%)	45
50-	(16%)	(15%)	55
60-70	(8%)	(20%)	65

Frequency polygon

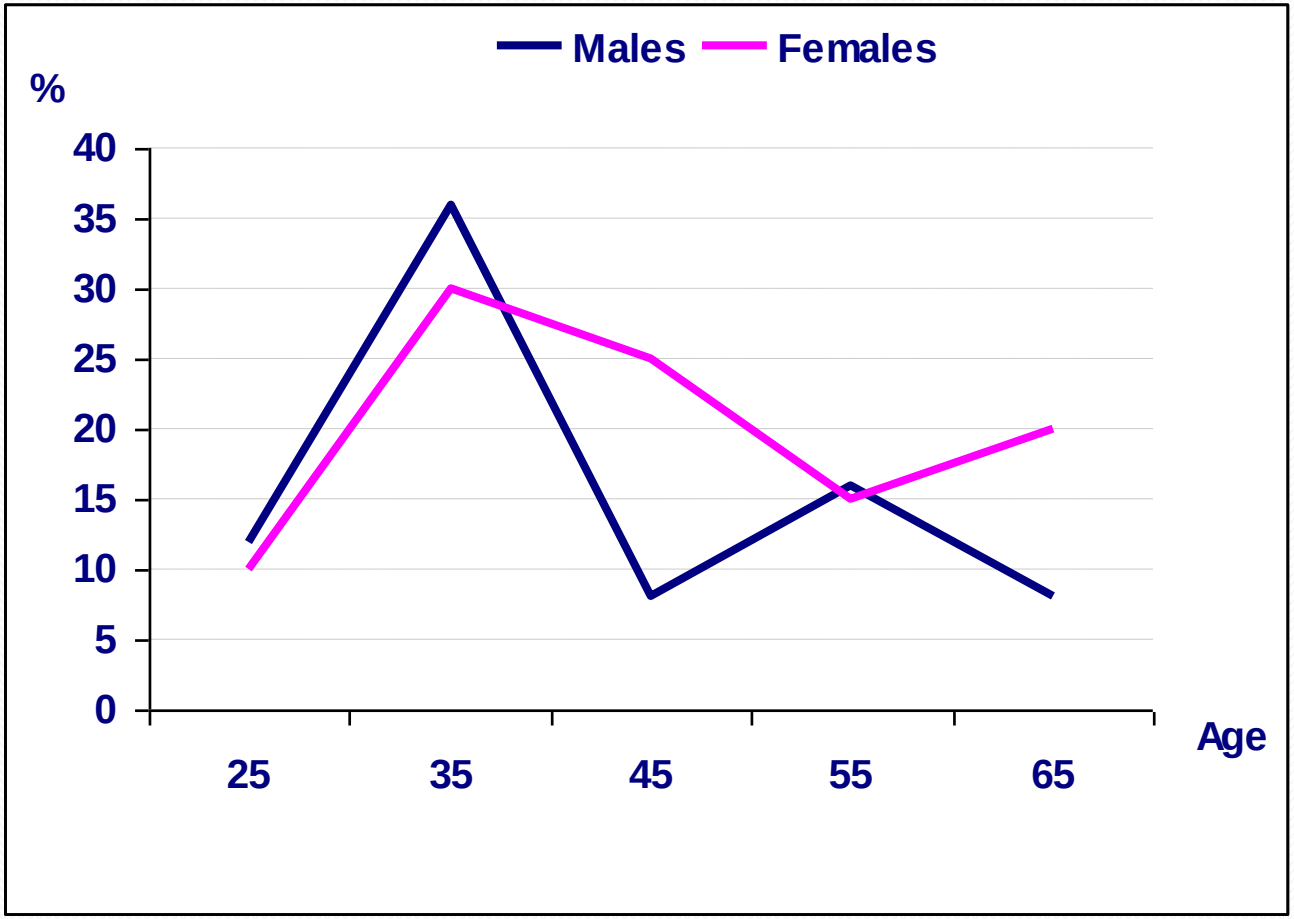
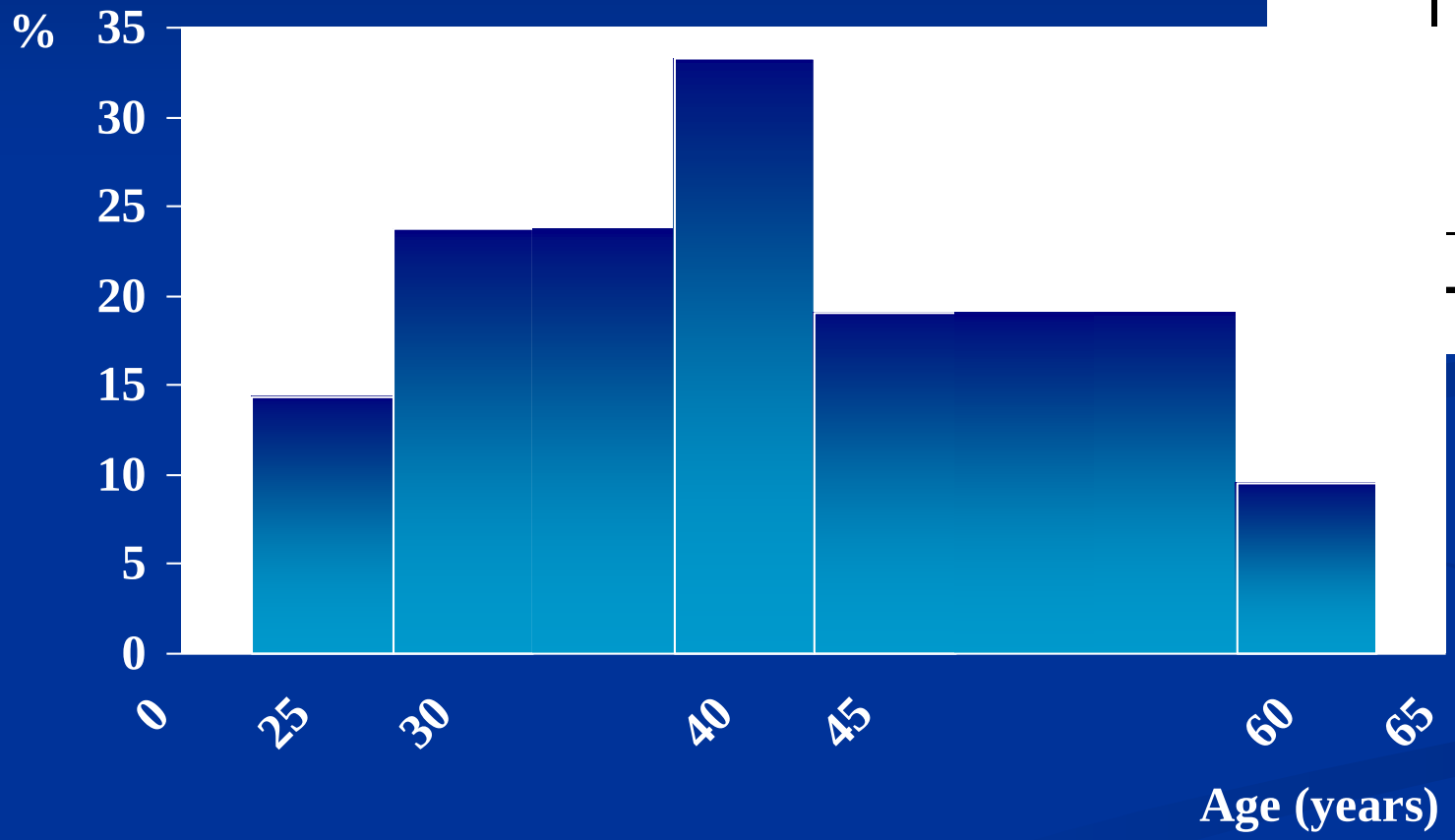


Figure (2): Distribution of 45 patients at (place) , in (time) by age and sex

Frequency curve

Histogram

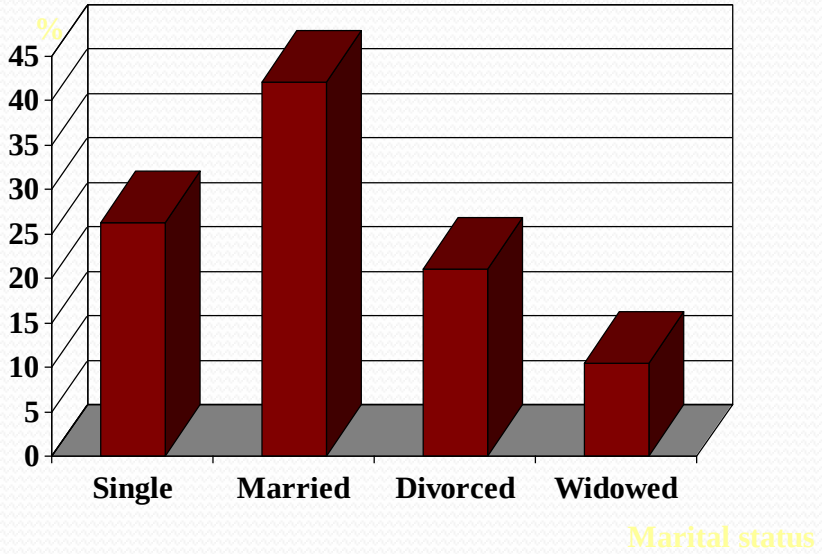


Distribution of a group of cholera patients by age

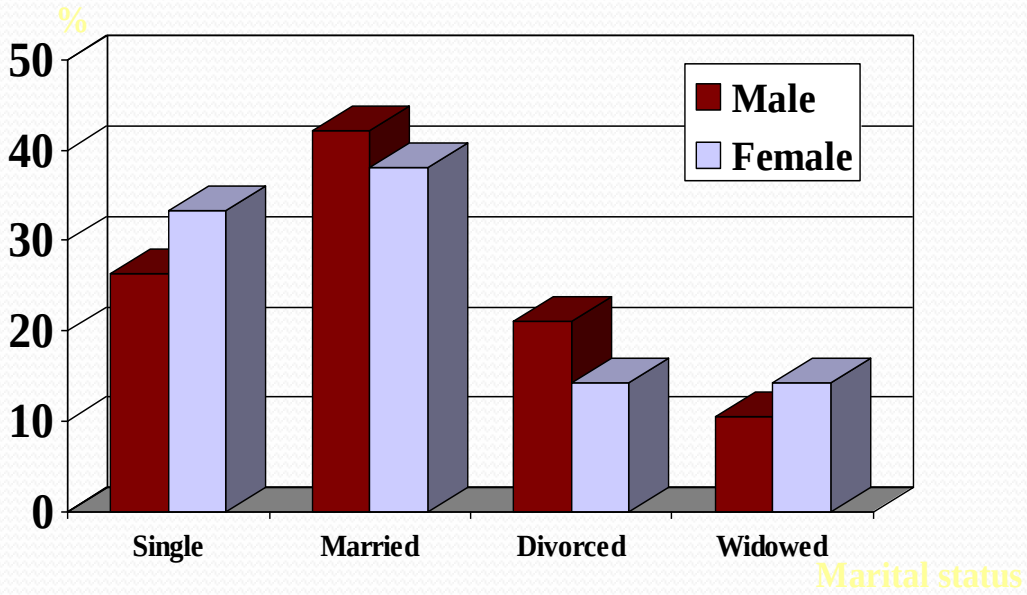
Age (years)	Frequency	%
25-	3	14.3
30-	5	23.8
40-	7	33.3
45-	4	19.0
60-65	2	9.5
Total	21	100

Figure (2): Distribution of 100 cholera patients at (place) , in (time) by age

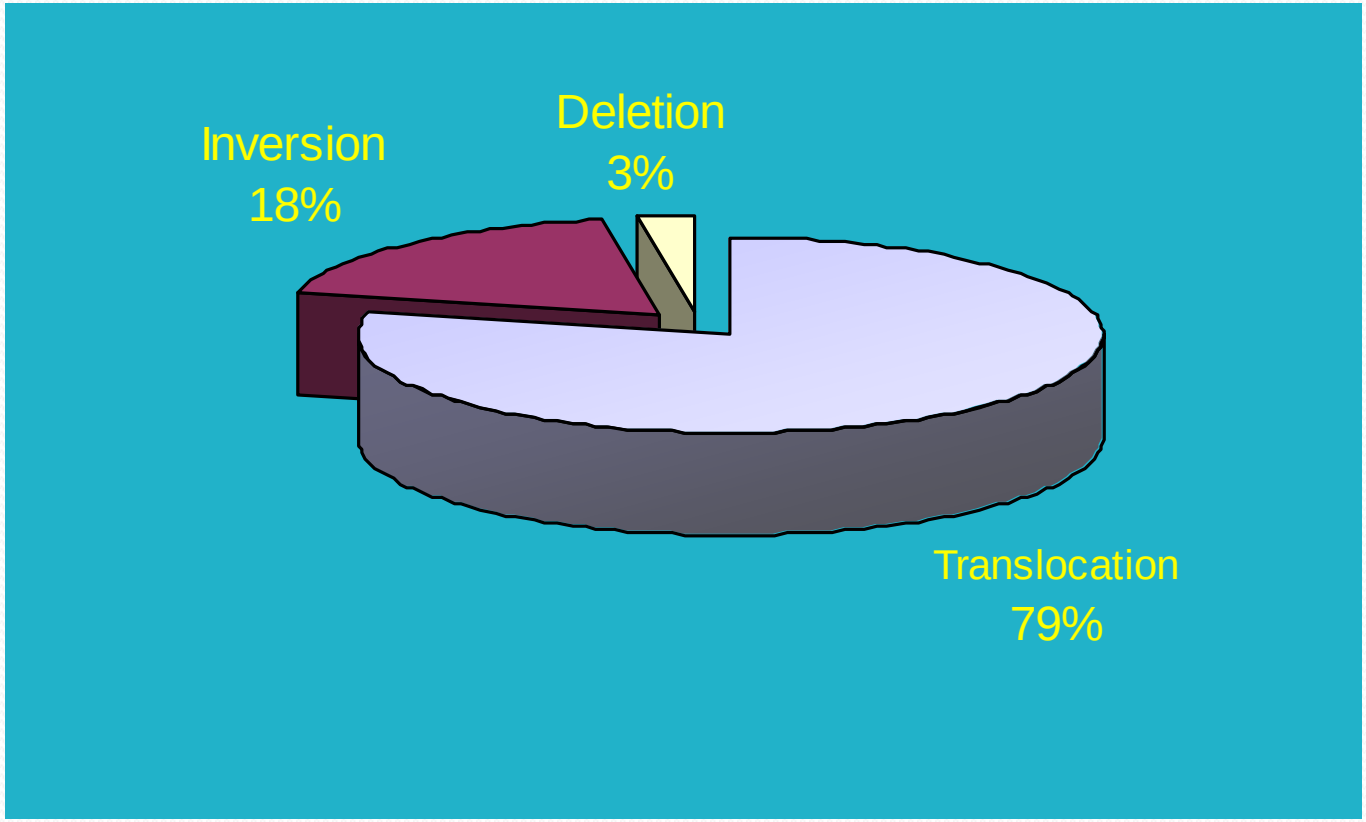
Bar chart


%
45
40
35
30
25
20
15
10
5
0
Single
Married
Divorced
Widowed
Marital status

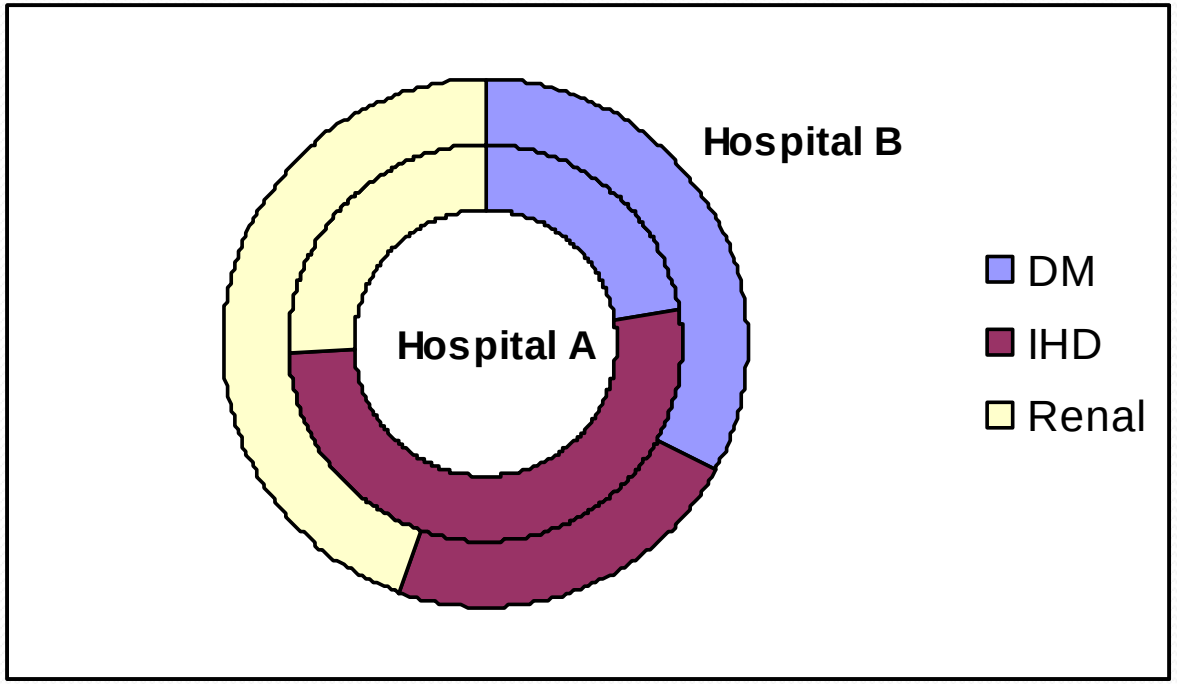
Bar chart


%
50
40
30
20
10
0
Male
Female
Single
Married
Divorced
Widowed
Marital status

Pie chart



Doughnut chart



3-Mathematical presentation

Summery statistics

- Measures of location

  1- Measures of central tendency

  2- Measures of non central locations

  (Quartiles, Percentiles)

- Measures of dispersion

Summery statistics

1- Measures of central tendency (averages)

- Midrange

$$\frac{\text{Smallest observation} + \text{Largest observation}}{2}$$

- Mode

the value which occurs with the greatest frequency i.e. the most common value

Summery statistics

1- Measures of central tendency (cont.)

- Median

the observation which lies in the middle of the ordered observation.

- Arithmetic mean (mean)

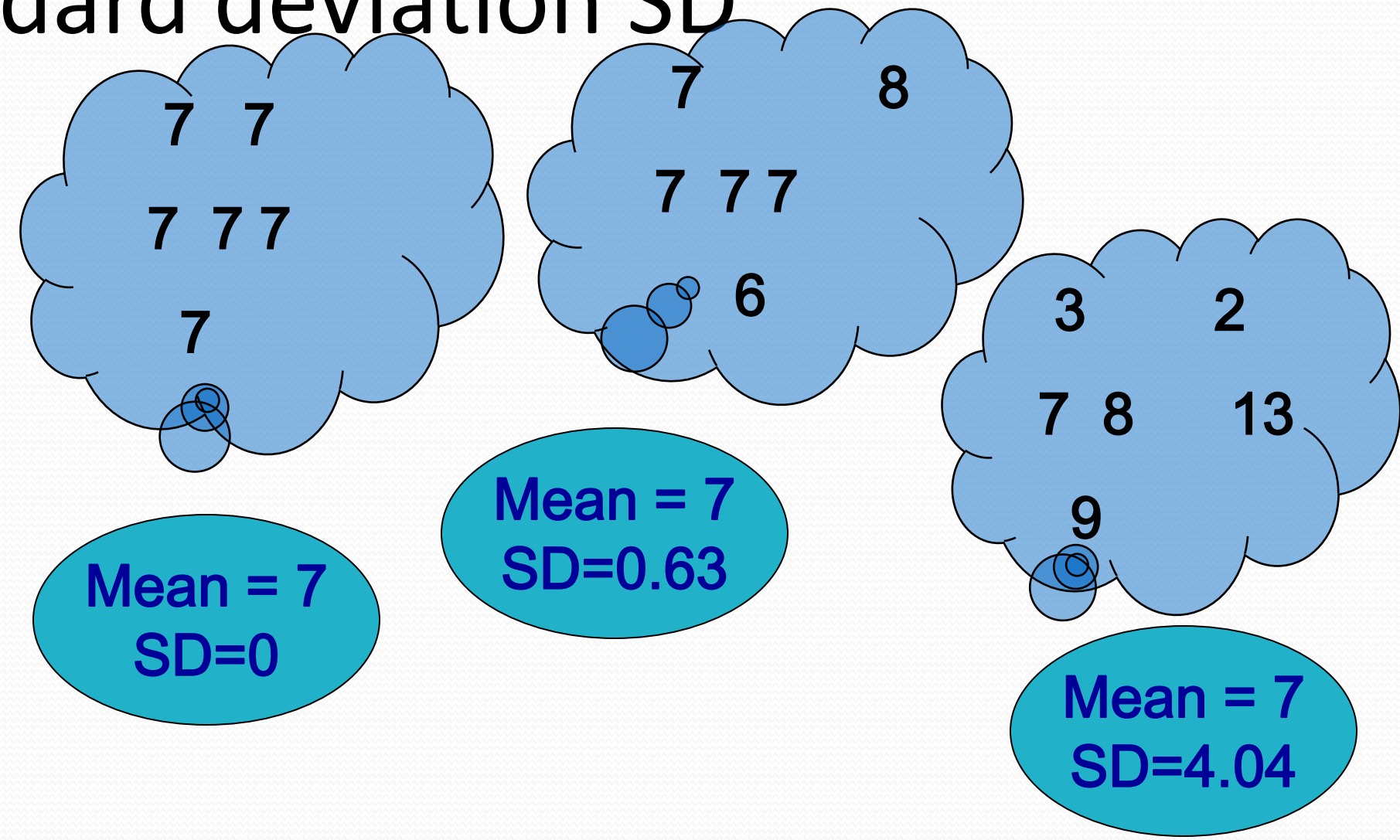
$$\frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Measures of dispersion

- Range
- Variance
- Standard deviation
- Semi-interquartile range
- Coefficient of variation

- “Standard error”

Standard deviation SD



Standard error of mean SE

A measure of variability among means of samples selected from certain population

$$SE\ (Mean) = \frac{S}{\sqrt{n}}$$