

Biostatistics

It is a method of collecting, compilation, analysis, interpretation of data.

Data - These are the discrete observations of events which carry a little mean when considered alone. (Numerical countings.)

Information - when data are reduced, summarized and adjusted to the variable factor such as (age, sex) are collectively known as information.

Intelligence - Integration of various data on the basis of social and political values.

(Data → Information → Intelligence)

Population - No. of individuals eligible for participating in any experiment. are collectively known as population.
e.g - Hypertensive patients.

Sample - Actual no. of individuals drawn from the population who are eligible for the experiment. Every individual is known as element.

Collection :-

there are 2 ways for it — i) Sources
ii) Collection

Sources — i) Census
ii) Records & documents
iii) Research paper & Publication

Method — i) Interview
ii) Surveys.
iii) Questionnaire

Compilation -

(A) i) Data summarisation ii) Data presentation

Quantitative (Measurable)	Qualitative or Categorical (Countable)
— B.P	— Gender wise
— Height	— Religion wise
— Wt., Hg.	— ABO grouping
	Anaemia $\begin{cases} \text{Yes} \\ \text{No} \end{cases}$

(B)

Primary

Secondary

Data collected by 1st hand

→ Data collected by already published, research, documents.

(C)

Continuous - (fractional Value) — Hb
Discrete (Non-frac.) — (No. of children in family)

Measurement :- Scales of ^ (Types of data).

Name - Nominal - only name of categories involve. e.g. - (religion, gender, colour, ^{name})
Order - Ordinal - Order of classification involve. (1, 2, 3rd division)
Interval - (Temp °C)
Ratio - (Wt in gm.)
Metric Scale. (used for quantitative data).

Interval - Interval keeps on changing.
No fixed zero. (0°C)

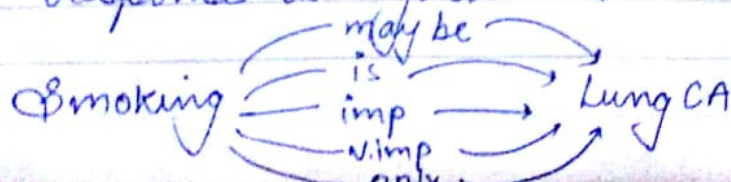
Ratio - fixed interval
fixed zero. (Temp. in Kelvin)

Special -

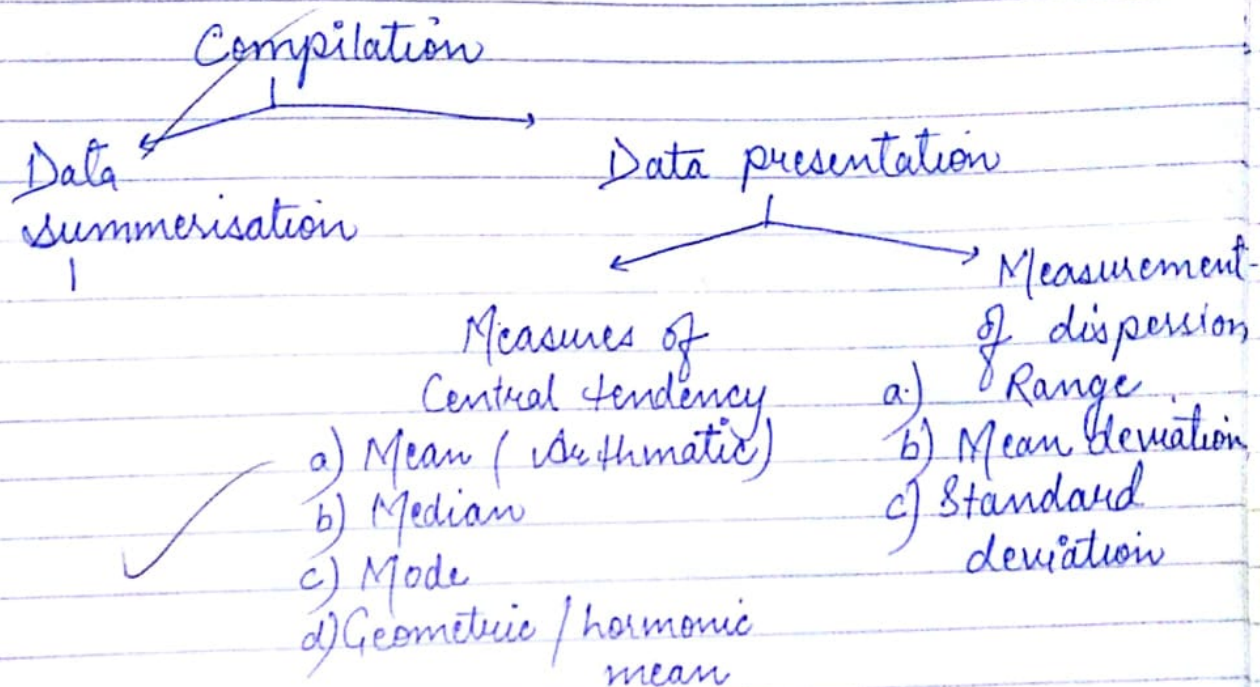
i) Likert scale - Strongly when the responses are graded in agreed or disagreed.
(strongly disagree, Not agreed, No comment, Agreed, Strongly agreed).

ii) VAS scale - Visual analogous scale - $\rightarrow \begin{matrix} 0-10 \\ 0-100 \end{matrix}$
scale which measures intensity of pain used by surgeons & anaesthetics.

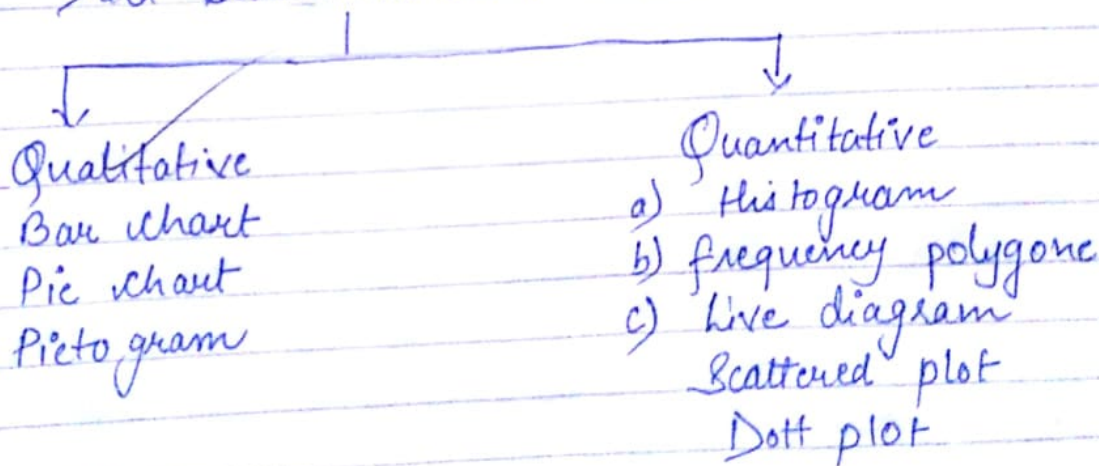
iii) Guttman's Score - when importance of data or response is graded.



iv) Thurstone score - which measure attitude of the patient.



✓ Data summarisation Presentation.



✓ Data summarisation

Qualitative

frequency %

Quantitative

Mode of central tendency

Mode of dispersion

Mean
Median
Mode

Range
Mean deviation
SD.

Data Presentation - Qualitative

① Bar charts



Magnitude $\times$ length of Bar

② Pie charts -



(Maternal Mortality)

Magnitude $\times$ Surface area

H	-	Haemorrhage
A	-	Anaemia
S	-	Sepsis
O	-	Obstructive labour
T	-	Toxaemia
O	-	Others.

3) Pictogram - Pictorial representation of the numerical data.

also k/a "Man in the street" graph.

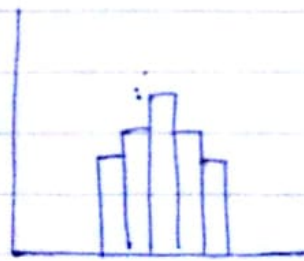


Developed	500
Developing	3500
undeveloped	10,000

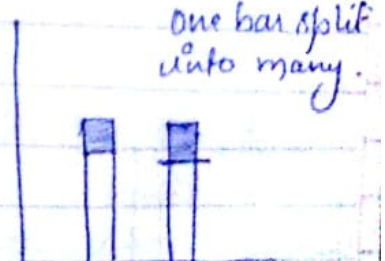
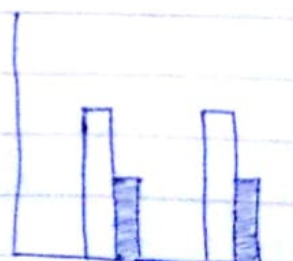
Quantitative data -

1) Histogram (X-axis = Quantitative continuous data are presented)

Simple Multiple Compound.

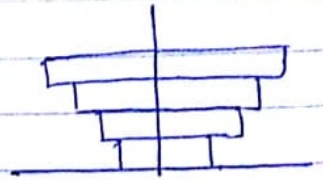


Simple histogram

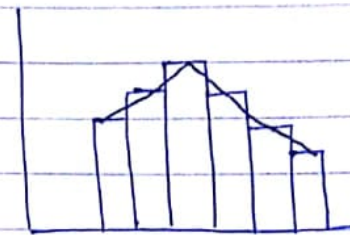


one bar split into many.

Age pyramid of developing country
double simple histogram.

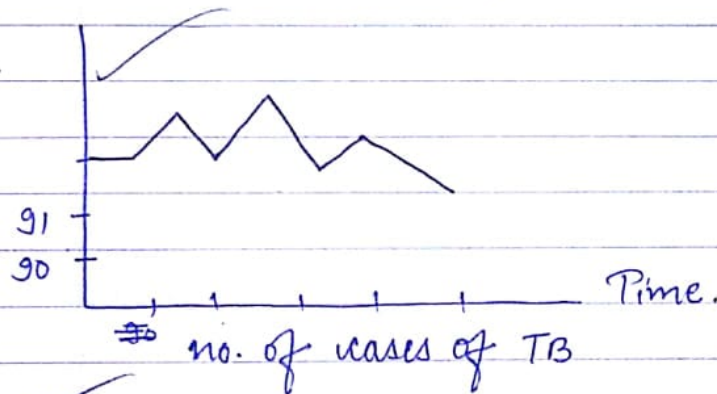


2) frequency polygon.



3) Line diagram -

With respect to time distribution of any event or disease. (Time trend).



4) Scattered plot
Dotted plot



⊕ Correlation ↑ Cholesterol intake B.P ↑

⊖ ↑ SES IMR ↓.

No. Alcohol/smoking ≠ PG Rank

It is a correlation btw two variables. one may
↑ or ↓ or does not affect the other.

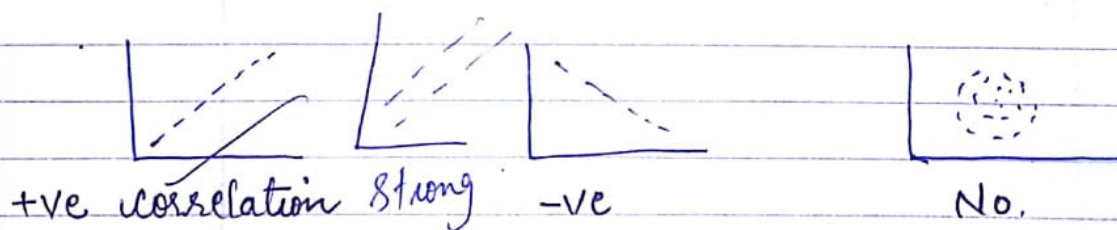
'r' = correlation coefficient.
(-1 0 +1)

↑ SES ↓ IMR r = $\begin{cases} \rightarrow +1 \\ \rightarrow 0 \\ \rightarrow -0.8 (\checkmark) \\ \rightarrow -1 \end{cases}$

(IMR not only depend on SES).

'r²' = coefficient of Determination.
(0 - 1) (r² = (0.8)² = 0.64 = 64%)

In IMR, 64% Variability due to SES.
— Always represent in %.



Measures of Central Pendency $\Rightarrow$

- a) Arithmetic mean = $\frac{\sum \text{Total no. of observation}}{\text{no. of observation}} = \frac{47}{10} = 4.7$
- b) Median = $\frac{4+5}{2} = 4.5$
- c) Mode = Most frequent Value = 2
- d) Geometric Mean = $\sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$
- e) Harmonic Mean = $\frac{n}{\sum \frac{1}{x_n}}$ = reciprocal of mean including reciprocal of observation

'14466237' "Bimodal series" = $3 \times \text{median} - 2 \times \text{mean}$

2 3 4 5 6.

Harmonic mean = $\frac{5}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}}$

Mode of dispersion -

↓ ↓ ↓
Range Mean deviation Standard deviation

1 2 3 4 5 6 7 8 9. (1000)
↓

Most affected Value = Mean extreme Value.
Least " = Mode. Skewed Value.
Most preferred Value = Median. Outlier Value.

Most affected Value = $\frac{\text{Range}}{\text{from, min} \rightarrow \text{max.}}$ (in mode of dispersion)

Mean deviation = $\frac{|x - \bar{x}|}{n}$ (modulus) $\left(\frac{\cancel{x} - \cancel{\bar{x}}}{\cancel{n}} \right)$

Standard deviation = $\sqrt{\frac{(x - \bar{x})^2}{n}}$

Sulbhan
7/9/18

15, 25, 25, 30, 33, 21, 25, 25, 18, 25, 40, 18

Range Meandeviation S.Deviation

15	Range - (15-40)		
18	a) Simplest measure (min → max.)		
18	difference b/w two extremes.		
21			
25			
25			
25			
25			
25			
25			
25			
30			
33			
40			
	Mean deviation -	$\frac{\sum x - \bar{x} }{n}$	$\bar{x} = \frac{300}{12} = 25$
		Standard deviation	
	$ 15 - 25 $	-10	$(x - \bar{x})^2$
	15	-10	100
	18	-7	49
	18	-7	49
	21	-4	16
	25	0	0
	25	0	0
	25	0	0
	25	0	0
	25	0	0
	30	5	25
	33	8	64
	40	15	225
			538
			Root Mean Square deviation
			$n > 30 \quad \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$
			$n < 30 \quad \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$
			$= \sqrt{\frac{538}{12}}$
			$= \sqrt{48}$
			$= 6.9$

Vital Statistics :-

1) Rate -



Multiples ✓
10,000
10,000

2) Ratio



(X)

3) Proportion



✓ (always 100%)

time specific +/- (X) instant.

Maternal Mortality Rate V.I = $\frac{CBR}{CDR}$

IMR

NMR

PNMR (postnatal)

CBR

CDR

AGR (Annual growth Rate)

SMR (specific MR)

IMR (Incidence)

e.g -

- 1) case fatality rate
- 2) proportional mortality rate
- 3) Prevalence.

↳ old + new cases

Maternal Mortality rate - $\frac{\text{total no. of maternal deaths (complicatⁿ + child birth + within 42 days)}}{\text{Total no. of live birth.}} \times 1000$

Developed = 10,0000

Developing = 1000

Infant Mortality death - $\frac{\text{Total no. of children at under 1 year age} \times 1000}{\text{Total no. of live birth}}$

7 days - early neonate

1 month - Neonate

1m - 1yr = post neonate

1 yr = Neonate

CBR = $\frac{\text{Total no. of live births} \times 1000}{\text{Mid year population}}$

$$IMR = \frac{16}{159} \times 1000$$

$$CDR = \frac{53}{5300} \times 1000$$

$$CBR = \frac{159}{5300} \times 1000$$

$$CFR = \frac{6}{300} \times 100$$

$$PMR \text{ (Hepatitis)} = \frac{6}{53} \times 100$$

$$AGR = \frac{CBR - CDR}{\frac{159}{5300} - \frac{53}{5300}}$$

$$SMR = \frac{6}{5300} \times 1000$$

$$V.I = \frac{CBR}{CDR}$$

$$NMR = \frac{10}{159} \times 1000$$

$$PNMR = \frac{6}{159} \times 1000$$

$$PMR \text{ (Inf)} = \frac{16}{53} \times 100$$

P

Dr. H. K. S.
21/7/18