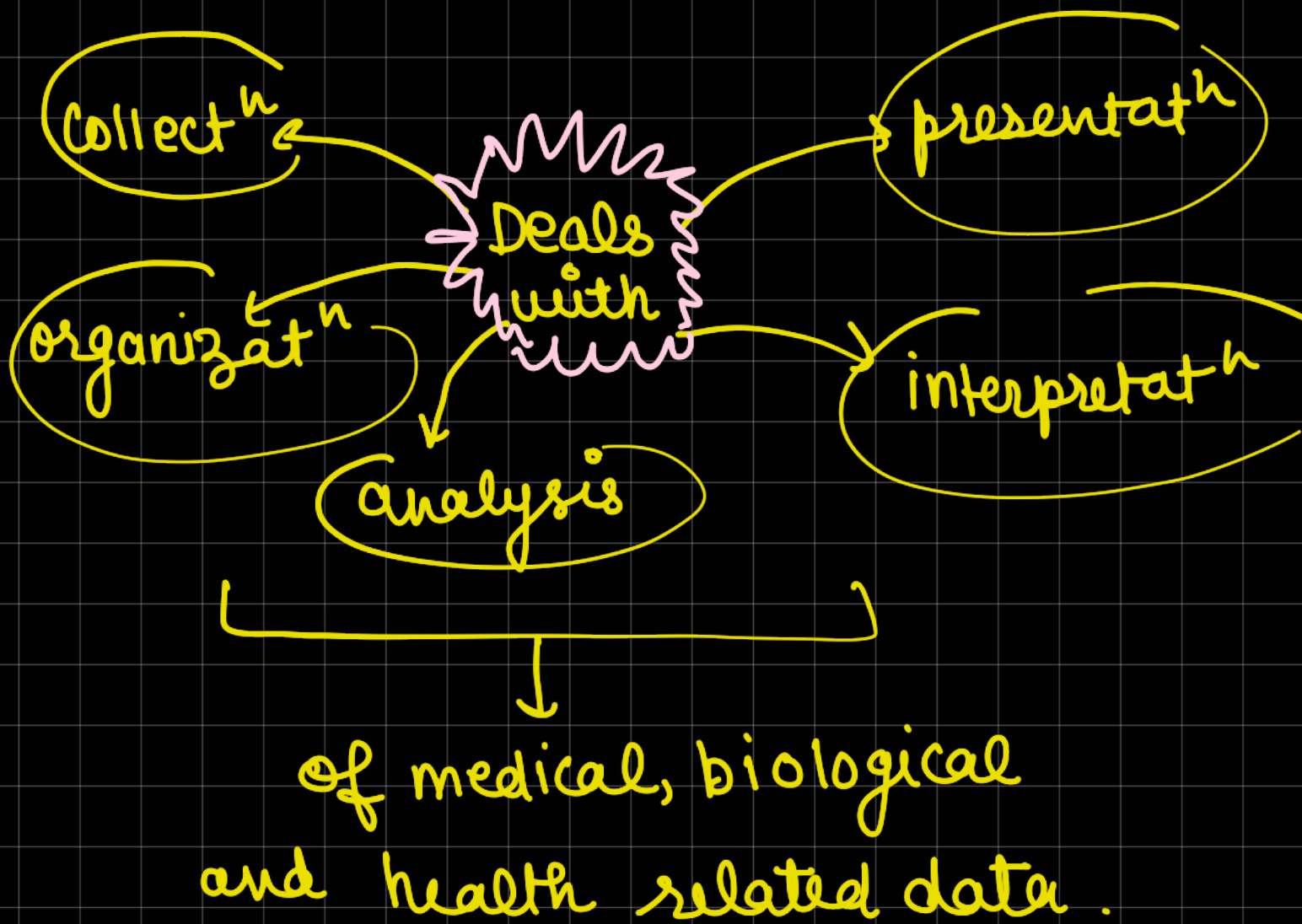


Biostatistics

- Branch of statistics



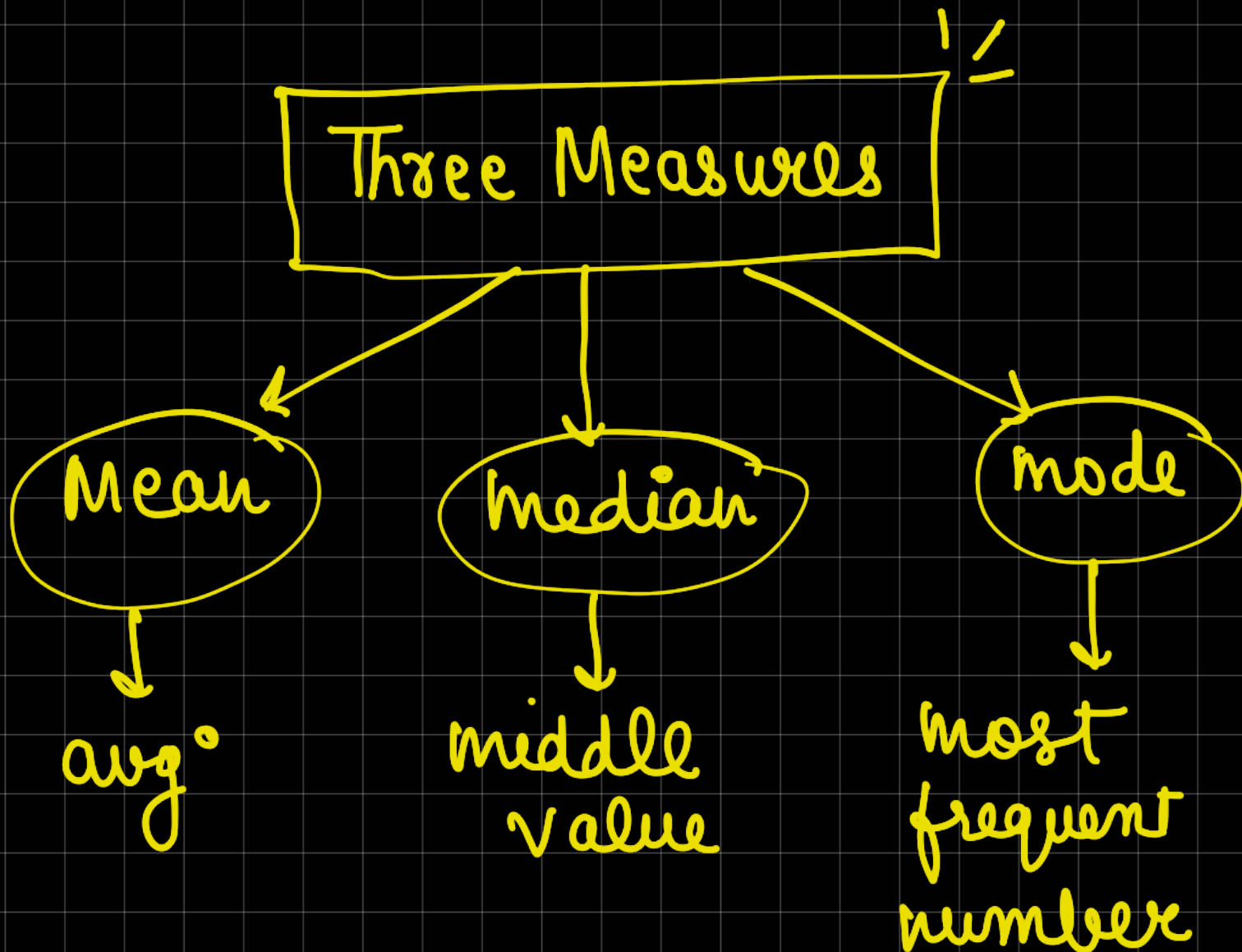
eg; comparing pain scores
 Before
 after
 treatment.

→ Importance ?

- helps in research
- helps interpret clinical studies
- assist in evidence-based physiotherapy.
- Helps compare treatment effectiveness
- Supports decision making using data.

Measures of Central Tendency

- numerical values
- represent the center or typical of data set.



MEAN

↳ • arithmetic mean

• average of all observations

formula (Individual / Ungp. data)

★
$$\text{Mean} = \frac{\sum X}{N}$$
 ★

where , $\sum X$ = sum of all observations

N = Total number of observations.

(2023)

Example,

Pain scores : 8, 6, 7, 9, 5

$$\text{Mean} = \frac{8 + 6 + 7 + 9 + 5}{5}$$

$$\text{Mean} = \frac{35}{5}$$

$$\text{Mean} = 7$$

*) *)

→ Mean for individual
or ungrouped data

→ mean for Grouped data

formula:

$$\bar{X} = \frac{\sum fX}{\sum f}$$

where,

f = frequency

X = midpoint of class interval.

Example; Suppose the marks of 30 students are grouped as follows:

given given

Lower limit + Upper limit
2

Marks (class interval)	frequency (f)	Midpoint (x)	f × x
0 - 10 LL UL	2	5	= 10
10 - 20 LL UL	5	15	= 75
20 - 30	8	25	= 200
30 - 40	10	35	= 350
40 - 50	5	45	= 225
Total	30		860

$$\sum f = 30$$

$$\sum fx = 860$$

$$\bar{x} = \frac{\sum fx}{\sum f}$$

put values

$$\bar{x} = \frac{860}{30}$$

$$\bar{x} = 28.67$$

Steps:

- ① find Midpoint (x)
- ② calculate $f \times x$
- ③ find $\sum f$
 $\searrow$
 $\sum fx$
- ④ Apply formula.

Advantages

- uses every observatⁿ
- easy to calculate
- most commonly used.

Disadvantages

- affected by extreme values
↓
(outliers)
- can't be used for categorical
data.

MEDIAN

↳ • middle value.

- after arranging data in ascending or descending order

Example;

3, 7, 9, 10, 20

↓

median

★ ★ Odd Number (ungp. data)

Eg; 2, 4, 6, 8, 10

↓
median.

Eg; 8, 9, 20, 40, 6

median = ?

^(unq. data) Even Number of Observatⁿ

Median = Avg. of 2 middle observations

Example;

2, 4, 6, 8

2 middle observatⁿ

$$\frac{4+6}{2} = \frac{10}{2} = 5$$

↓
median

mostly asked

Median (grouped data)

formula:

$$\text{Median} = L + \frac{\left(\frac{N}{2} - CF\right) \times h}{f}$$

where;

L = Lower boundary of median class

N = Total frequency

CF = Cumulative frequency before median class.

f = frequency of median class

h = Class width

Steps

- ① find Total frequency (N)
- ② Calculate $\frac{N}{2}$. ($\frac{20}{2} = \underline{\underline{10}}$)
- ③ find Median class.

↳ where does 10th observatⁿ lie?

Read the CF column:

- CF = 3 → observatⁿ 1-3 = 0-10
- CF = 8 → observatⁿ 4-8 = 10-20
- CF = 15 → observatⁿ 9-15 = 20-30

The 10th observation falls b/w
9 and 15, so it must be in

20-30 class. → median class

* Now identify the values:

$L = 20$ (lower limit of median class)

$CF = 8$ (cumulative freq. before median class)

$f = 7$

$h = 10 = ?$

$20-30$

gap = ? 10

question

Total = N

Marks	frequency (f)	Cf
0-10	3	3
10-20	5	8
20-30 m.c	7 f	15
30-40	4	19
40-50	1	20

$$\text{Median} = L + \frac{\frac{N}{2} - Cf}{f} \times h$$

$$\text{Median} = L + \frac{\left(\frac{N}{2} - cf\right)}{f} \times h$$

$$= 20 + \frac{(10-8) \times 10}{7}$$

$$= 20 + \frac{20}{7}$$

$$= 20 + 2.86$$

$$= \underline{22.86}$$

Advantages

↳ not affected by extreme values.

Disadvantages

↳ doesn't use all observedⁿ

MODE

↪ value that occurs most frequently.

example; (individual or ungrouped data)

① 2, 5, 5, 4, 4, 6, 6, 6, 7

Mode = 6.

② 3, 4, 3, 6, 7, 7, 8, 3

Mode = 3.

Advantages

- very easy to understand.
- useful for categorical data
- not affected by extreme values.

Disadvantages

- Sometimes there is no mode.
- Sometimes more than one mode exists.

Difference

Mean



- Avg.
- use all observatⁿ
- affected by outliers
- Used for continuous data

Median



- Middle value
- use middle observatⁿ
- less affected
- Used for Skewed Data

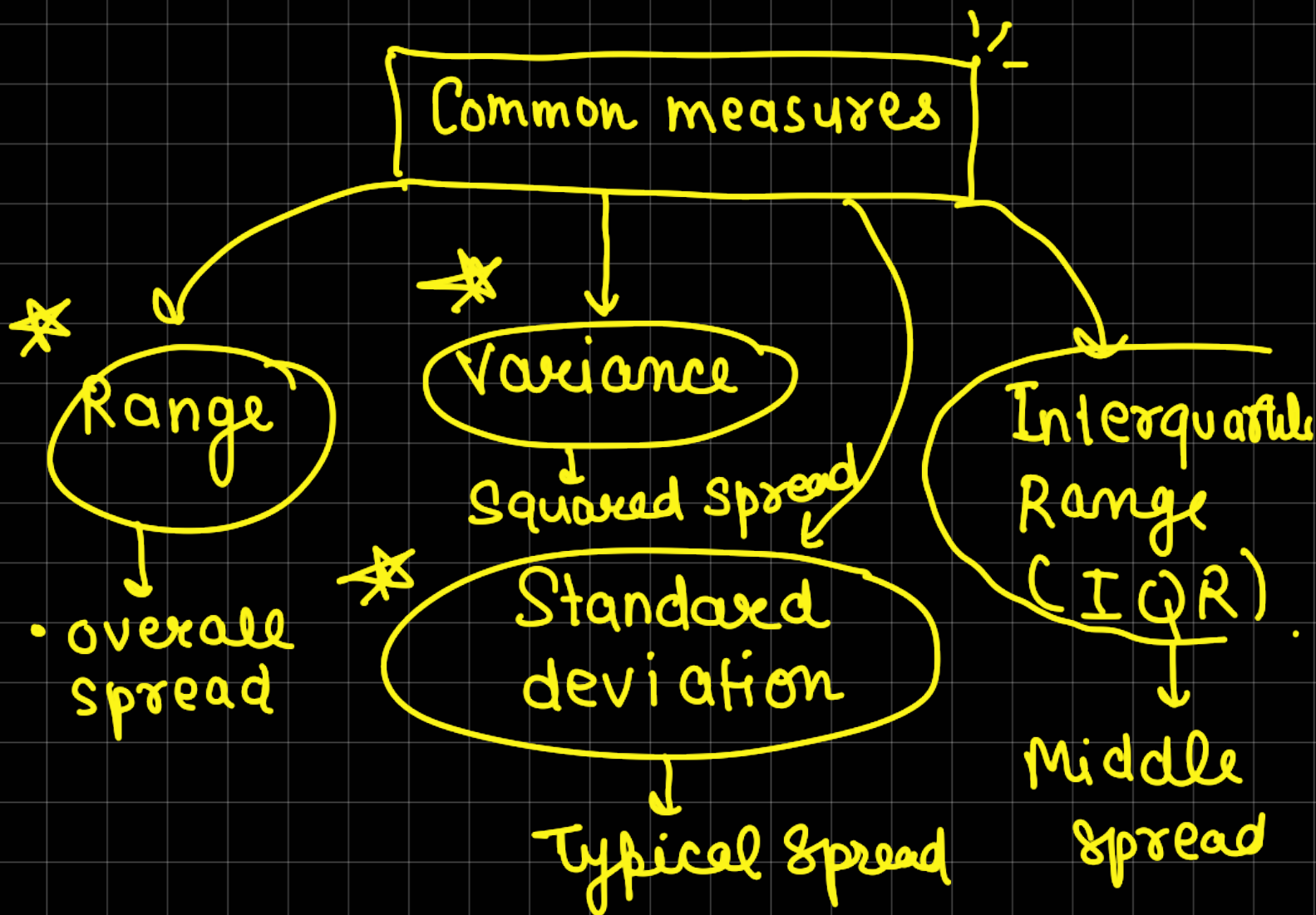
Mode



- Most freq. value
- Uses frequency
- Not affected
- Used for categorical data

Measure of Dispersion

↳ Tells how far the data is spread from the center.



range —

→ Difference between the largest and the smallest value.

eg; class A : 10,10,10,10,10

Class B : 5, 8, 10, 12, 15

Formula

Range = Highest value - Lowest value

Example: $5, 7, 9, 11, 15$.
 (5) is a LV
 (15) is a HV

$$\text{Range} = \text{HV} - \text{LV}$$

$$= 15 - 5$$

$$= 10$$

Advantages

- Very easy to calculate.
- Quick measure of variability.

Disadvantages

- uses only two observations.
- affected by extreme values.

Standard deviatⁿ

- ↳ Square root of variance.
- tells us how much the data usually deviate from the mean.

Small SD



Data are close
to the mean

Large SD



Data are
widely scattered

eg; Two physiotherapy groups have
Same avg. muscle strength.

group A: $SD = 2\text{kg}$ → Most patients
have similar
strength.

group B: $SD = 12\text{kg}$ → Patients vary
greatly in
strength.

Advantages

↳ uses all observations.

- easy to compare different data
set.

Disadvantages

- affected by extreme values.
- slightly more difficult to calculate than range.

Variance

How far the observations are spread from the mean.

Small variance

↓
values are close to the mean

eg; group A : 20, 21, 20, 19, 20

Mean = 20

Large variance

↓
values are spread far from the mean

The values are very close to mean, so variance is low.

group B : 5, 15, 20, 25, 35.

Mean = 20

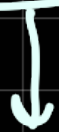
The values are much more spread out so the variance is high

Advantages



- uses every observation
- More accurate than range.

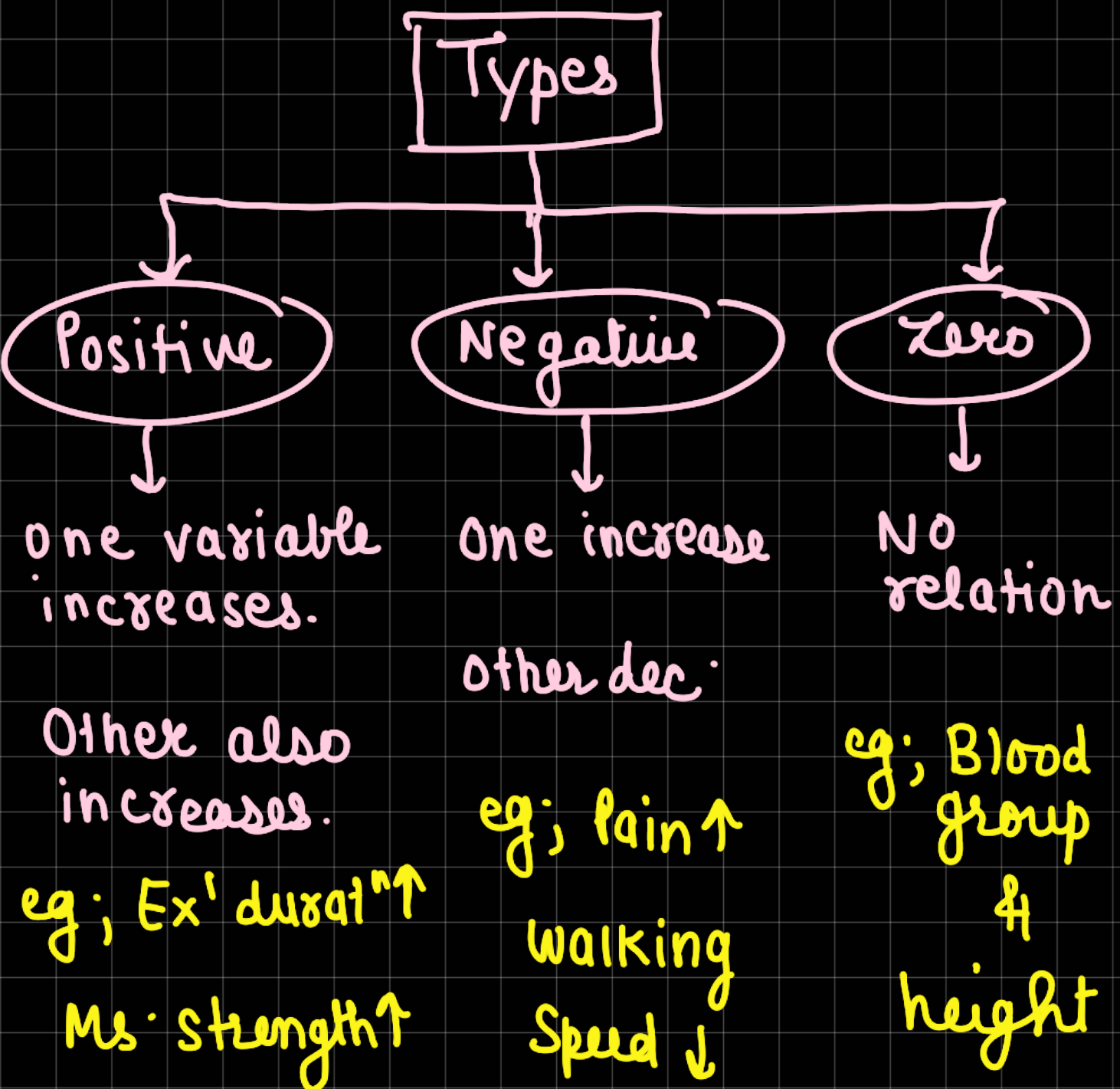
Disadvantages



- Difficult to interpret
- More calculation are required.

correlation

Statistical relationship between two variables.



Correlation Coefficient

(r)

Range: $-1 \leq r \leq +1$

Value	Interpretation
+1	Perfect +ve
+0.7 to 0.9	Strong +ve
+0.3 to +0.6	Moderate +ve
0	No correlat ⁿ
-0.3 to -0.6	Moderate -ve
-0.7 to -0.9	Strong -ve
-1	Perfect -ve

p-value

↓ tells us whether the observed result is likely due to chance.

interpretation

$p < 0.05 \rightarrow$ Statistically significant

$p > 0.05 \rightarrow$ Not statistically significant.

Common Stat test

① Independent t-test

compare means of two independent groups.

② Paired t-test

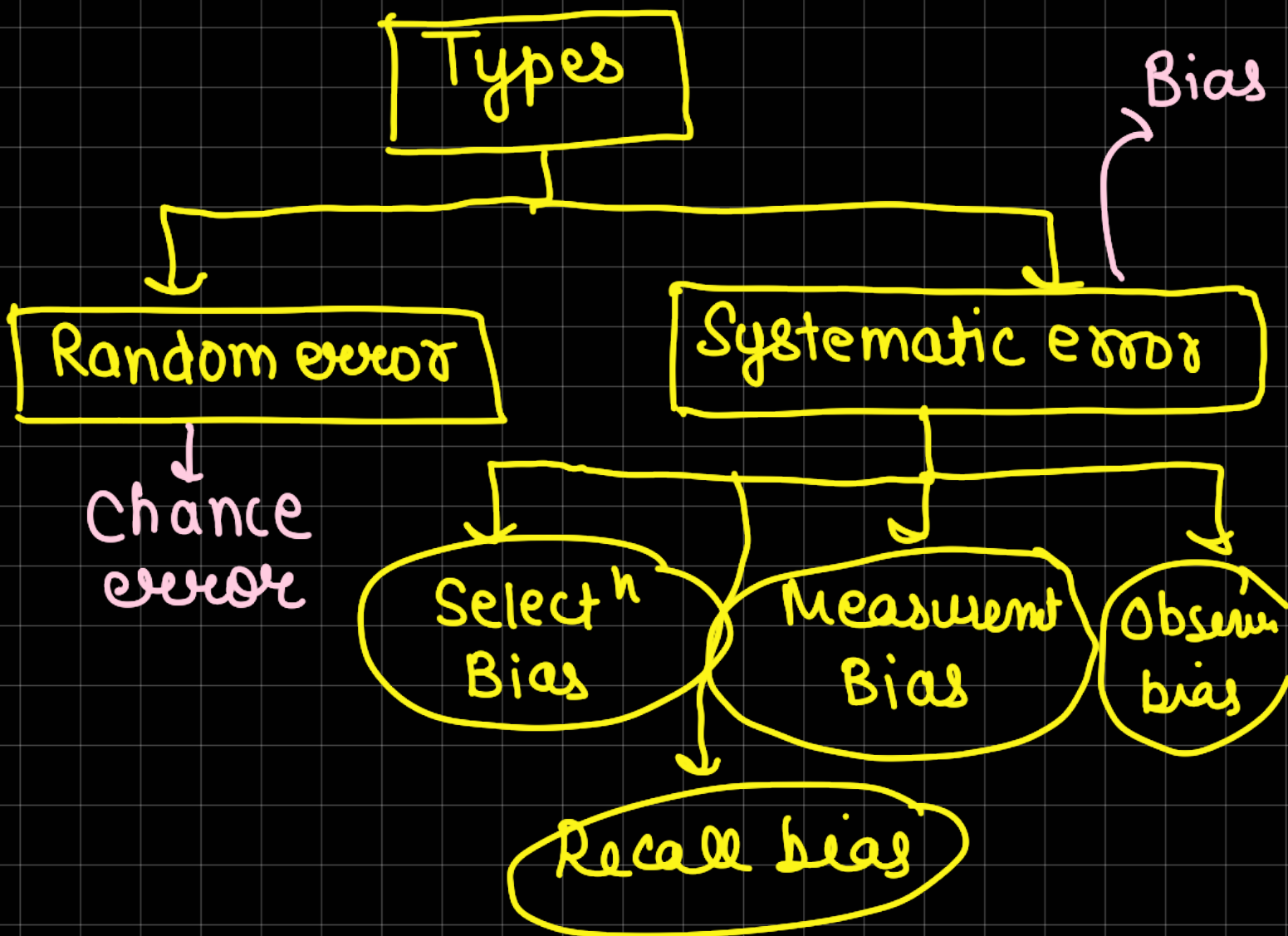
compare the means before and after treatment in the same group.

③ ANOVA

compare means of three or more groups.

ERROR

- difference b/w the observed (meas.) value and true value.
- in research errors can lead to incorrect conclusions.



* error in hypothesis testing

Type I Error $\rightarrow$ (α error)

$\hookrightarrow$ Reject null hypothesis
when its true $\left\{ \begin{array}{l} \text{false} \\ \text{+ve} \end{array} \right.$
(Reject — true)

example; A study concludes that
a new physiotherapy technique
reduces pain but in reality it
doesn't.

Type II error $\rightarrow$ (β error)

$\rightarrow$ failing to reject the null hypothesis when it is false.

(Accept $\rightarrow$ false)

example; a study concludes that a new balance training program has no benefit, but it actually improves balance.

$\left\{ \begin{array}{l} \text{false} \\ \text{-ve} \end{array} \right.$

Tips: Mean [★]
Mode
Median } ungrouped/
individual
data.

→ [★] Definitions

→ [★] Types of Research

MEQ - (4-5)
mostly one liners.
but practice numericals