BIOSTATSTICS AND RESEARCH METHODOLOGY

UNIT- Ist

Syllabus :-

- Introduction : Statistics, Biostatistics, frequency distribution
- Measures of central tendency : Mean, Median, Mode - Pharmaceutical examples
- Measures of dispersion - Dispersion, Range, standard deviation, pharmaceutical problems
- Correlations : Definition, karl pearson's coefficient of correlation, Multiple correlation - pharmaceutical examples ..

⊙ Introduction to Biostatics

Statistics →

It is defined as "it is the collection, presentation, analysis and interpretation of numerical data.

(meant, explanation, conclusions ...)

- Statistics
  - derived from latin word 'status'
    - वो कोन है,
    - उसके बारे में information
      - data

(eg) Weight of a person, height of a person etc ..

Now, statistics are formed by using Raw data

- the data which is collected in its original form.
- experiments, survey & records

* According to Crexton and Cowden.

Collection of data → Presentation of data → Analysis of data → Interpretation of data

- Components of statistics

* It can be of two types :—

① Descriptive statistics → the collection of data is described in a summary.
- mean, median, mode etc..

② Inferential statistics → Drawing conclusion from descriptive statistics.
- It includes hypothesis testing & estimation (analysis)...

* BIOSTATISTICS →

Bio + statistics
Medical field

- When statistics or tools of statistics are applied in the field of biological and medical science, than it is known as ~~biostatistics~~ biostatistics.
- It includes Human biology, medicine, public health etc...

(eg.) No. of patients attending a hospital,

- It includes :—

* Medical statistics - It deals with application of statistics to the study of disease, efficacy of vaccine etc..

Health statistics - It deals with application of statistical methods to information of public health importance.

Applications and Uses of Bio-statistics

- find out the normal values (eg. pulse rate, blood pressure etc..)
- to define, what is normal or healthy in a population.
- to find out efficacy/potency of a new drugs by comparing with standard drugs.
- test usefullness of vaccines
- Others :- leading cause of death, important causes of sickness, Rise and fall of particular disease.. etc..

FREQUENCY DISTRIBUTIONS

- frequency → It is the number of occurance of the any value in a data.

OR

It is the number of times that any particular value come in a data.

(eg). Akash take a table thrice in a day –
– So, Akash take a tablet (3) time in a day
– (3) is the frequency of that data..

- frequency distributions → It is a tabular or graphical representation of data that displayed the number of observation (frequency) within a given intervals.

OR

In simple words,

It is the distribution of frequencies (no. of observations) with respect to their given interval or particular categories.

(eg) Akash, Jass, Pradeep drink a tea respectively (A), (J), (P), (A), (J), (A), (P), (A), (J), (J), (A)

So, in this Raw data, Akash drink a tea for (5) times, Jass for (4) times, Pradeep for (2) times
– So (5), (4) & (2) are the frequencies

⊙ But in raw data, it is difficult to find so we can create a short tables in which we show that data – It is known as frequency distr–

S.NO	Content	frequency
01.	Akash	(5)
02.	Jass	(4)
03.	Pradeep	(2)

these are frequencies

✻ So, overall organisation of frequencies in a simple form is known as frequency Distributions.

- Types of frequency distributions →

i) Discrete f.D.

ii) Continuous f.D. → Exclusive
Inclusive

iii) Commulative frequency distributions

i) Discrete frequency distributions →

In this, class interval are not given..

- In this, categorical data are presented in the form of ungrouped...

eg.

S.No.	Variables	frequency
01.	Akash	5
02.	Jass	4
03.	Pradeep	2

Variable	N (frequency)
• Male	19
• female	22

→ frequencies

ii) Continuous frequency distributions →

In this, class intervals are used as a variables.

- In this, variables will be continuous, such as age, salary etc. are examined.
- In this, the data may be divided into number of groups / classes called class intervals.
- Also called as Grouped frequency distributions

eg. Result out, student got a marks..

* It is of two types :-

i) Exclusive series → In this, the class interval does not includes upper class limit.

ii) Inclusive series → In this, the class interval that includes the upper class limit.

Exclusive →

S.NO	Marks	No. of students
01.	0-20	06
02.	20-40	11
03.	40-60	32

Inclusive →

S.NO	Marks	No. of student
01.	1-20	06
02.	21-40	11
03.	41-60	32

iii) Cumulative frequency Distributions

It is a type of distribution in which the frequencies of variables are summed, as one moves from the top of the table to the bottom.

- Thus the bottom category would have a cumulative frequency equivalent to the total size

Eg.

S.NO.	Variables (Name)	frequency(N)	Cumulative frequency	Cumulative percentage
01.	Akash	7	7	21.8%
02.	Jass	6	13	40.6%
03.	Pradeep	5	18	56.2%
04.	Neha	4	22	68.75
05.	Sizan	2	24	75%
06.	froz	4	28	87.5%
07.	Nikhil	2	30	93.75%
08.	Ganesh	2	32	100%

total no. of chai

⑥ Graphical Presentation of frequency Distributions

It is the graphical representation of data that displayed the no. of observation (frequency) within a given intervals.

There are follows:—

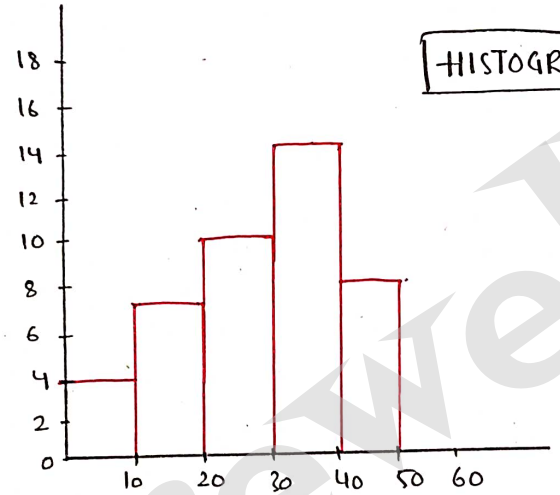
i) Histogram
ii) Polygon
iii) Line graph
iv) Bar graph
v) Cumulative frequency curve or Ogive

i) Histogram →

It is a graph used to represent the frequency distributions of numerical data that are organised into intervals (continuous)

(eg.) ① Histogram for equal class interval

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	4	7	10	15	8



② for unequal

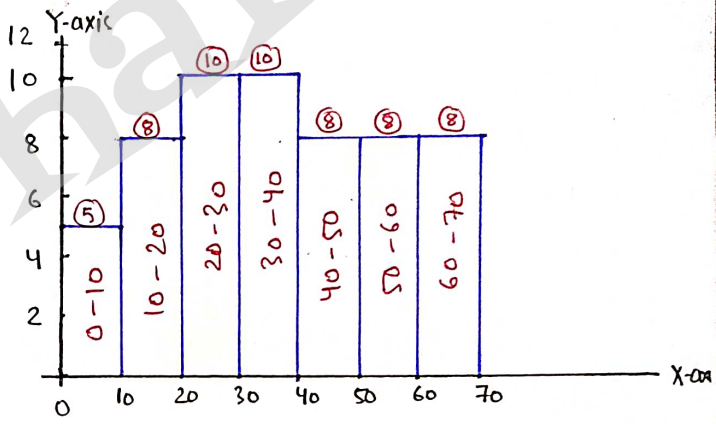
Class interval	0-10	10-20	20-40	40-70
frequency	5	8	20	24

Rearrange it →

20/2 = 10

24/2 = 8

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f(n)	5	8	10	10	8	8	8



③ for inclusive data

It includes that class interval which has upper class limit.

Marks	0-10	10-20	21-30	31-40	41-50
No. of students	4	7	10	15	8

Solution ↴

10 + 11 = 21/2 = 10.5

Marks	0-10.5	10.5-20.5	20.5-30.5	30.5-40.5	40.5-50
No. of students	4	7	10	15	8

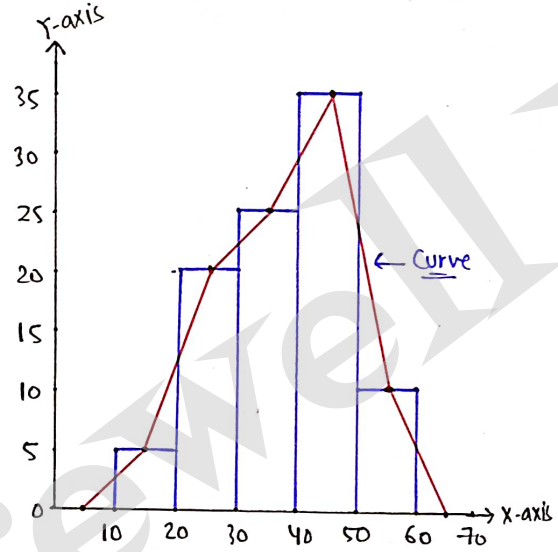
ii) Polygon →

It is a another method of representing frequency distribution graphically.

In this, curve is obtained by joining the mid-points of the tops of the rectangular

eg.

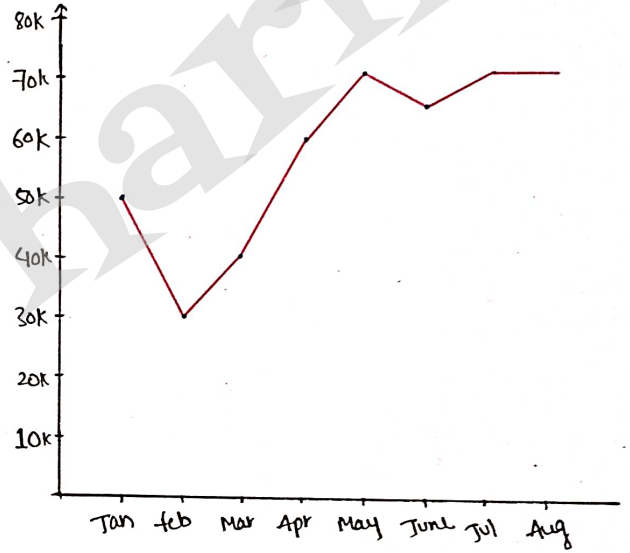
Age in year	No. of patients
10-20	5
20-30	20
30-40	25
40-50	35
50-60	10



iii) Line graph →

Line graph or linear graph is used to display the continuous data and useful for predicting future events over time.

Month	Jan	feb	March	April	May	June
No. of PCM Sold	50K	30K	40K	60K	70K	65K

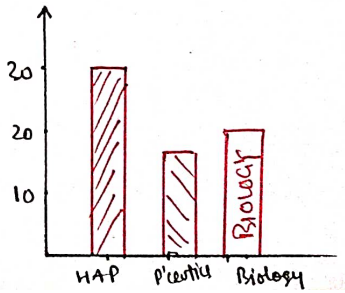


iv) Bar graph →

Used to display the category of data and it compares the data using solid bars.

eg.

Subject	No. of students
HAP	30
Pharmaceutics	15
Biology	20



v) Cumulative frequency curve / Ogive

those graphs which are used to represent the cumulative frequency data.

- for this, firstly ordinary frequency distribution table is converted into cumulative frequency distri..

- Two types :—

① less than Ogive → the frequencies are added cumulatively in an increasing order..

② more than Ogive → the frequencies of diff. classes are estimated in a decreasing order..

eg.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Students	4	7	9	10	15	22	14	6

Rearrange it, in the form of less than Ogive → choose upper limit

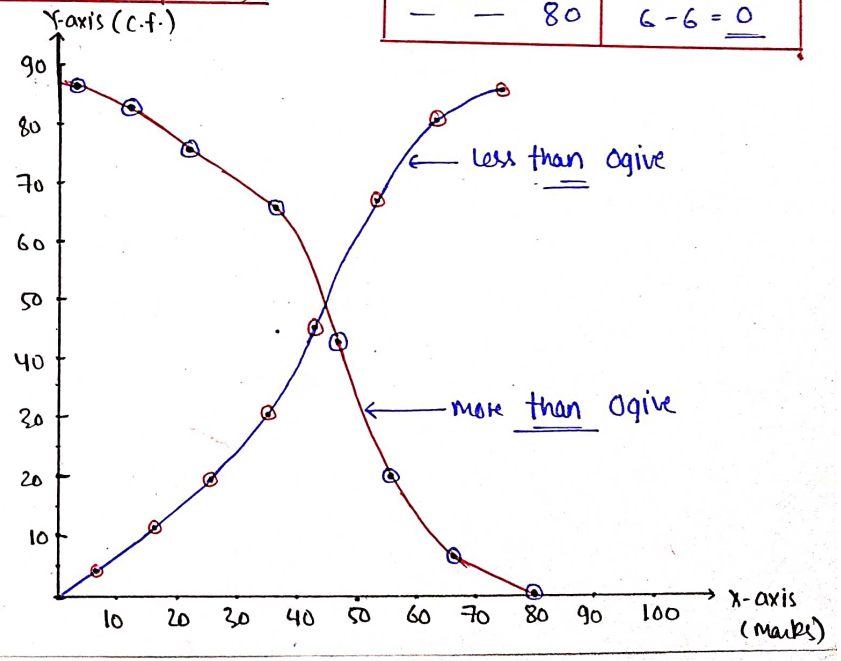
and more than Ogive → choose lower limit

less than Ogive

Marks	c.f.
less than 10	4
" " 20	4+7 = 11
" " 30	11+9 = 20
" " 40	20+10 = 30
" " 50	30+15 = 45
" " 60	45+22 = 67
" " 70	67+14 = 81
" " 80	81+6 = 87

more than Ogive

Marks	c.f.
more than 0	87
more than 10	87 - 4 = 83
— — 20	83 - 7 = 76
— — 30	76 - 9 = 67
— — 40	67 - 10 = 57
— — 50	57 - 15 = 42
— — 60	42 - 22 = 20
— — 70	20 - 14 = 6
— — 80	6 - 6 = 0



MEASURES OF CENTRAL TENDENCY

- Central tendency refer to the average value of any data.

- Central Tendency
  - Centre की ओर → toward centre (middle)

- It is a measurement in which we calculate a single number (known as average) that represent the whole data.

eg. (A) → 3, 4, 4, 5, 4, 3, 4, 5, 3, 5etc.. (Roti)

average → $\frac{40}{10} = 4$ average

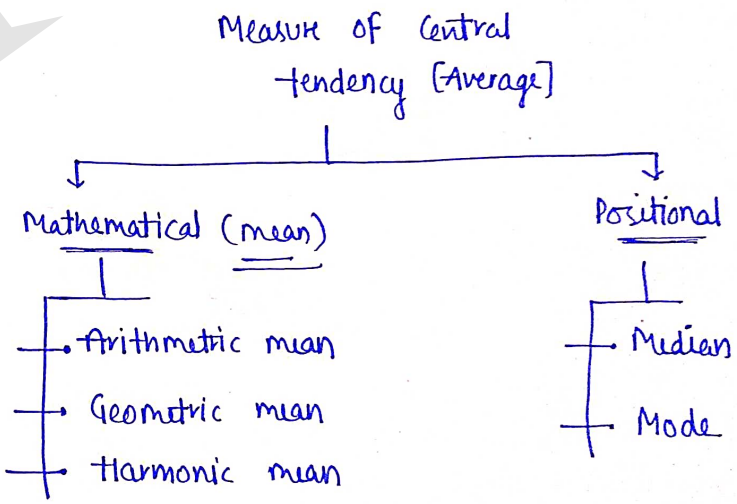
(A) daily eat 4 chappati --

So, Now, (4) represents the diet of (A)..

- Also known as measure of central value, average of first order.

- Objectives
  - to find single value that represents the whole data.
  - to help for comparision
  - to help in decision making

- Types of Measure of Central tendency



- Mathematical →

① Arithmetic mean →

It is defined as the sum of all observation divided by the total number of observation.

- It is denoted by $\bar{x}$ and also called as average value

eg. Akash → 3, 2, 4, 4, 3, 2, 3, 2, 3, 4 etc..

$$\bar{x} = \frac{3+2+4+4+3+2+3+2+3+4}{10} = \frac{30}{10} \quad (3)$$

$$\bar{x} = 3$$

- Two methods are used for calculating arithmetic mean
  - Direct method for simple data
  - Short-cut method for large data.
    - i) for individual series
    - ii) for descrete series
    - iii) for continuous series

i) for individual series →

for example, if $x_1, x_2, x_3, x_4 \ldots x_n$ be the value of n observation then arithmetic mean is given as

$$\bar{x} = \frac{\text{Sum of the value}}{\text{total no. of the value}}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$\boxed{\bar{x} = \frac{\Sigma x}{n}}$$

eg. Calculate the arithmetic mean of marks obtained in Biostatics by 10 students of Carewell pharma.

⇒

Students	A	B	C	D	E	F	G	H	I	J
Marks (x)	06	10	15	06	18	17	12	14	08	14

there are two methods

i) Direct method ii) Short-cut method

i) Direct method →

Σ = sum

$$\bar{x} = \frac{\Sigma x}{n}$$

$$\bar{x} = \frac{6+10+15+6+18+17+12+14+8+14}{10}$$

$$= \frac{120}{10} = 12$$

ii) By short-cut method → (indirect method)

$$\bar{x} = a + \frac{\Sigma d}{n}$$

where, a = assumed mean

d = deviation $\bar{x}$ = Mean

$d = x - a$

Student	A	B	C	D	E	F	G	H	I	J
Marks(x)	6	10	15	6	18	17	(12)	14	8	14
d = x-a	-6	-2	3	-6	6	5	0	2	-4	2

- Assumed mean → take any middle value than find (d) for deviation.. (eg.) A = 12

$a = 12$, $n = 10$

$$\Sigma d = \frac{3+6+5+0+2+2-6-2-6-4}{10}$$

$$= \frac{18-18}{10} = \frac{0}{10}$$

$$\bar{x} = 12 + \frac{0}{10}$$

$$\bar{x} = 12$$

eg.2 the birth weight of 6 babies are 2.0, 2.4, 2.6, 3.1, 3.4 & 2.5 kg. find the mean birth weight.

Sol → weight (x) = 2.0, 2.4, 2.6, 3.1, 3.4, 2.5

~~d (x-a) = -1.1, -0.7, -0.5, 0, 0.3, 0.6~~

~~take $a = 3.1$, $\Sigma d = 3.2$, $n = 6$~~

~~$\bar{x} = a + \frac{\Sigma d}{n}$~~ | ~~$\bar{x} = 3.1 + \frac{3.2}{6}$~~

~~$\bar{x} = 3.1 + 0.53$~~ ⇒

2.667

ii) for Discrete series / ungrouped series

If for a discrete series with value $x_1, x_2 \ldots x_n$ and frequencies $f_1, f_2 \ldots f_n$

then..

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

where $\bar{x}$ = mean

Σfx = sum of all f.x

Σf = sum of all frequencies.

eg.1 Calculate the arithmetic mean from the following table that shows marks secured in p'ceutics by student of Carewell pharma.

Marks (x)	40	48	52	58	64	69	74	78
No. of student (f)	5	2	7	8	5	3	2	1

i) Direct method →

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

Marks (x)	No. of students (f)	f.x
40	5	200
48	2	96
52	7	364
58	8	464
64	5	320
69	3	207
74	2	148
78	1	78
	$\Sigma f = 33$	$\Sigma fx = 1877$

$$\bar{x} = \frac{1877}{33} = 56.87$$

ii) By short cut method

$$\bar{x} = a + \frac{\Sigma fd}{(\Sigma f)}$$

Marks (x)	No. of Student (f)	d = x − A (deviation)	fd
40	5	−24	−120
48	2	−16	−32
52	7	−12	−84
58	8	−6	−48
(64) ← A	5	0	0
69	3	5	15
74	2	10	20
78	1	14	14
	Σf → 33		Σfd → −235

$$\bar{x} = 64 + \frac{(-235)}{33}$$

$$= 64 - 7.121$$

$$= \boxed{56.87}$$

iii) for Continuous series / Grouped f.D.

for continuous series, we take the midpoints of each class.

- If $m_1, m_2, m_3 \ldots m_n$ are the midpoints of the class with frequencies $f_1, f_2, f_3 \ldots f_n$ then

$$\bar{x} = \frac{f_1m_1 + f_2m_2 + f_3f_m - f_nm_n}{f_1 + f_2 + f_3 + \ldots f_n}$$

$$\boxed{\bar{x} = \frac{\Sigma fm}{\Sigma f}}$$

(eg.) Calculate the arithmatic mean of the following distribution of patient of covid-19 as per their weights.

⇒

weight (x)	30-40	40-50	50-60	60-70	70-80
No. of patients (f)	5	8	7	12	14

It can be solved by two methods :-

i) Direct method ii) Step deviation methods

i) Direct methods →

firstly midpoint of 'x' is calculated

↓

then midpoint (m) is multiplied to (f)

$\frac{L.L. + U.L}{2}$

↓

Calculate (fm) then divide by (n)

weight (x)	30-40	40-50	50-60	60-70	70-80	80-90
No. of patients (f)	5	8	7	12	14	17
Mid point (m)	35	45	55	65	75	85
f(m)	175	360	385	780	1050	1445

$\Sigma fm = 4195$

$\Sigma f = 63$

$\bar{x} = \frac{\Sigma fm}{\Sigma f}$

$= \frac{4195}{63} \Rightarrow 66.58$

So, mean weight of patients is 66.58

ii) Short cut method & step deviation

firstly findout the midvalue (m)

↓

then calculate deviation (d) by choosing assumed mean (a) ⇒ $d = m - a$

↓

then multiply frequency (f) to the deviation (d) and calculate fd

then, $\bar{x} = a + \frac{\Sigma fd}{\Sigma f}$ — for short cut

for step deviations → for this, calculate

step deviation (d') = d/i

where, d = deviation

i = common factor

then multifiy frequency (f) to the de

step deviation (d')

$\bar{x} = a + \frac{\Sigma fd'}{\Sigma f} \times i$

Weight (x)	No. of patients (f)	Mid points (M)	d = m − a (a = 55)	d' = d/i (i = 10)	Direct fM	Short-cut fd	Step-deviation fd'
30 - 40	5	35	35 − 55 = −20	−20/10 = −2	175	−100	−10
40 - 50	8	45	45 − 55 = −10	−10/10 = −1	360	−80	−8
50 - 60	7	55 → A	55 − 55 = 0	0	385	0	0
60 - 70	12	65	65 − 55 = 10	10/10 = 1	780	120	12
70 - 80	14	75	75 − 55 = 20	20/10 = 2	1050	280	28
80 - 90	17	85	85 − 55 = 30	30/10 = 3	1445	510	51
	63 (Σf)				4195 (Σfm)	730 (Σfd)	73 ($\Sigma fd'$)

Short cut →

$$\bar{x} = 55 + \frac{730}{63}$$

$$= 55 + 11.58 \Rightarrow \boxed{66.58}$$

Step-deviation →

$$\bar{x} = a + \frac{\Sigma fd'}{\Sigma f} \times i$$

$$= 55 + \frac{73}{63} \times 10 = 55 + 11.58 \Rightarrow \boxed{66.58}$$

(eg.) Calculate the combined arithmatic mean of medical and paramedical staff working in Goverment hospitals of pune district in last five year.

$\Rightarrow$ $n_1 = 50$ $\quad n_2 = 70$

$\bar{x}_1 = 145$ $\quad \bar{x}_2 = 148$

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

$$= \frac{50 \times 145 + 70 \times 148}{50 + 70}$$

$$= \frac{7250 + 10360}{120} \Rightarrow \frac{17610}{120} \Rightarrow 146.75$$

(eg.2) In one of pharmaceutical company, 120 pharmacists are working, out of that 50 girls and 70 boys with mean height of 145 and 148 cm, respectively. find out the mean for the boys and girl together (combined) --

- Merits →
  - simple to understand, easy to calculate
  - helps in comparision & decision making
- Demerits →
  - Not for qualitative analysis
  - not for extreme value

0, 250, 500, 0

- **Geometric mean (GM)**

It is defined as the n^{th} root of the product of n value.

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

$$\log GM = \log\left(\sqrt[n]{x_1 \cdot x_2 \cdots x_n}\right)$$

$$\log GM = \frac{1}{n}\left[\log x_1 + \log x_2 + \cdots + \log x_n\right]$$

$$\log GM = \frac{1}{N}\sum \log x$$

$$\boxed{G.M. = \text{Antilog}\left[\frac{1}{N}\sum \log x\right]}$$

(eg.) Daily income of workers are given below. Calculate Geometric Mean (G.M.).

700, 900, 800, 850, 750

→

Income (x)	700	900	800	850	750
$\log(x)$	2.845	2.954	2.903	2.929	2.875

So, $\sum \log x = 14.506$ & $N = 5$

$$GM = \text{Antilog}\left[\frac{1}{5} \times 14.506\right]$$

$$= \text{Antilog } 2.901$$

$$= \boxed{796.74}$$

- **Harmonic mean (HM)**

It is defined as the reciprocal of the arithmetic mean of the reciprocal of given items

$$\boxed{H.M. = \frac{n}{\sum \frac{1}{x}}}$$

(eg.) weight of 5 tablets..

x	1.20	1.18	1.23	1.17	1.19	
$1/x$	0.83	0.84	0.81	0.85	0.84	4.17

$$H.M = \frac{5}{4.17}$$

$$= \boxed{1.199}$$

H. mean

MEDIAN

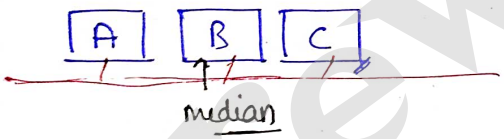
It is a positional average.

Definition → It is the middle value of any data when the data is arranged in ascending or a descending order.

OR

Median is the middle value that divides the whole data into two equal parts.

eg.



4, 6, 2, 10, 8 → arrange it in ascending or descending order

↓

2, 4, (6), 8, 10 → (6) is the median

Median → for Individual series
→ for discrete series
→ for continuous series (grouped)

⊙ for individual series

they have only one data (observation)

- If the no. of observation are **odd**

then, $M =$ Size of $\left(\frac{N+1}{2}\right)^{th}$ item

- If the no. of observation are **even**

then, $M = \frac{\text{Size of } \left(\frac{N}{2}\right)^{th} + \left(\frac{N}{2}+1\right)^{th} \text{ term}}{2}$

where, (N) is the no. of observation

Steps → Allot a serial no. → arrange it in ascending/descending order → put the value in formula..

eg. Calculate the median of the following data gives the height of five student in cm.

168, 173, 153, 163, 158 = 5

⇒ Arrange it in ascending order..

① 153, ② 158, ③ 163, 168, 173

• total no. of observation are 5 = N

so, put the formula of odd

$M = \left(\frac{N+1}{2}\right)^{th}$ term

$M = \left(\frac{5+1}{2}\right)^{th} = \frac{6}{2}$ (3)

M = size of 3rd item

3rd item is 163

So, **Median height is 163**

eg.2 Calculate the median of following observation of Hb% of five female patients. 11, 9, 12, 7, 13

→ Solve yourself..

eg.3 find the median of following weights of capsule.

140, 138, 130, 150, 135, 145

→ Arrange in ascending order

130, 135, ③ 138, ④ 140, 145, 150

total no. of observation → 6 i.e. even

$\text{Median} = \frac{\text{size of } (N/2)^{th} \text{ \& } (N/2+1)^{th}}{2}$

$= \frac{\text{size of } (\overset{3}{6}/2)^{th} \text{ \& } (\overset{3}{6}/2+1)^{th}}{2}$

$= \frac{\text{size of } 3^{rd} + 4^{th} \text{ item}}{2}$

put the value of 3rd & 4th item

$= \frac{138+140}{2} \Rightarrow \frac{278}{2} \Rightarrow 139$ median

⊙ for discrete series →

they have two data (observation)..

Steps ↓

- Define the item by (x) and frequency by (f)
  ↓
- Calculate the commulative frequency ((cf))
  ↓

$$M = \text{size}\left(\frac{N+1}{2}\right)^{th} \text{ item}$$

where, M = Median, N = Cumulative frequency

(eg.) Calculate the median salary paid to the production supervisor working in parenteral department of p'ceutical industry as per..

Monthly salary (x)	25	30	26	27	28	31	29
No. of supervisor (f)	15	17	16	18	20	19	17

→ firstly convert into ascending order

Monthly salary (x)	25	26	27	28	29	30	31
No. of supervisor (f)	15	16	18	20	17	17	19
Cumulative frequency (c.f.)	15	31	49	69	86	103	122
	+16	+18	+20	+17	+17	+19	

Here, $N = \Sigma f = 122$.

$$\text{Median (M)} = \text{size}\left(\frac{N+1}{2}\right)^{th} \text{ item}$$

$$= \text{size}\left(\frac{122+1}{2}\right)^{th} \text{ value}$$

$$= \frac{122+1}{2} = \frac{123}{2} = 61.5^{th} \text{ value}$$

61.5 comes after the c.f. of 49..

So, the cumulative frequency is 69

∴ Median of salary paid is 28K

(eg.) find median of daily wages of employees in hospital :

Wages	145	170	180	190	200	210
No. of Employees	3	16	8	20	6	2

→ firstly arrange in ascending order

S.No	x	f	C.f.
01	145	3	3
02	170	16	19
03	180	8	27
04	190	20	47 (M)
05	200	6	53
06	210	2	55
		$\Sigma f = 55$ (N)	

$M = \left(\frac{N+1}{2}\right)^{th}$ observation

$= \frac{55+1}{2} \Rightarrow 56/2 \Rightarrow 28^{th}$ observation

So, 28 lies after 27 so the C.f. is 47

So, the median is 190

⊙ for continuous series which contain grouped data

Steps:- firstly define the item by (x), class interval → (i), frequency → (f)

↓

then find c.f. and put the value in formula ..

(M.C)

$$\text{Median} = L + \frac{\frac{N}{2} - c.f}{f} \times i$$

where, L = lower limit of class in which median lies

N = total no. of frequencies i.e. (Σf)

f = frequency of class in which median lies

C.f = cumulative frequency of before median class, i = width of C.T.

(eg.) Calculate the median from following data :-

x	60-70	70-80	80-90	90-100	100-120
f	8	10	12	16	14

→ firstly calculate cumulative frequency

x	60-70	70-80	80-90	90-100	100-120
f	8	10	12	16	14
c.f.	8	18	30	46	60

then find (N) = Σf i.e = 60

Now, find Median class = $N/2$

Median class = $60/2$ = 30

So, the median class is 30 → 80-90

So, Now, put the value of this median class to the formula —

$$M = L + \frac{N/2 - c.f}{f} \times i$$

$$= 80 + \frac{30 - 18}{12} \times 10$$

$$= 80 + \frac{12}{12} \times 10 \Rightarrow 80 + 10$$

Median ⇒ 90

eg. 2 Calculate the median forom following data

Wages	80-100	100-120	120-140	140-160	160-180
Workers NO.	8	12	16	8	6

⇒ Sol" — yourself

Hint, firstly find median class ($N/2$)

choose c.f. value of before median class

i = diff of class interval of median class..

MODE

It is a positional average.

Definition → It is that **value** in a series which contain the highest frequency.

OR

It is the **item** which occurs largest number of time in a frequencies distribution.

eg. (A) → 3, 3, 4, 4, 5, 3, 5, 4, 3, 4, 5, 4, 4 = 4 mode

3 → 4
4 → 6 — It has highest frequency..
5 → 3

Mode
- for individual
- for Discrete series
- for continuous (grouped)

① for individual → ungrouped & single data

eg. find the mode of following value of weight of 15 tablets in mg

→ 120, 121, 123, 122, 125, 124, 120, 121, 122, 123, 125, 124, 122, 120, 122

120 → 3
121 → 2
122 → 4
123 → 2
124 → 2
125 → 2

so, 122 have highest frequency,

so, **122 is mode**

② for discrete series

eg. the fine paid by student of one of pharmacy college given in the following distribution.

→ Calculate mode

fine(Rs)	100	120	140 (mode)	160	180
No. of (f) students	15	18	25	20	17

- So, the max. frequency is (25), the mode is 140 //.

(3) for continuous series → firstly modal class

$$\text{Mode} = L_1 + \frac{f_1 - f_0}{(2f_1 - f_0 - f_2)} \times i$$

where,
- L_1 → lower limit of modal class
- f_1 → frequency of modal class
- f_0 → frequency of previous class
- f_2 → frequency of next class
- i → class interval of modal class

(eg.) Calculate mode for following data

Prize	2-6	6-10	10-14	14-18	18-22	22-26
frequency	1	9	21	47	52	36

→ firstly choose the modal class by checking the highest frequency i.e. (52) → 18-22

So, Highest frequency → 52
Modal class → 18-22

$L_1 \to 18$, $f_1 \to 52$, $f_0 \to 47$, $f_2 \to 36$
$i = 4$

$$\text{Mode} = 18 + \frac{52 - 47}{2 \times 52 - 47 - 36} \times 4$$

$$= 18 + \frac{5}{104 - 47 - 36} \times 4$$

$$= 18 + \frac{5}{21} \times 4 \Rightarrow 18 + \frac{20}{21}$$

$$= 18 + 0.9523$$

$$= 18.952$$

Mode = 18.95

- Relation b/w **Mean, Median & Mode**

  - A distribution in which mean, mode and median have same value the such distribution is known as symmetrical distribution.

- for others :-

  Mean − Mode = 3 (Mean − Median)

  Mode = 3 median − 2 mean

MEASURES OF DISPERSION - Chapter-8

- Dispersion → It is the measure of differences or variation between the observations.

- It is obtained by comparing the individual observation with its average value (or mean).

eg. (2) 4, 6, 8, 10

$\bar{x} \Rightarrow \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$

Variation → 2 ⟷ 4 ⟷ 6 ⟷ 8 ⟷ 10

(6) (6) (6) (6) (6)

(4) 2 0 2 4

- fluctuation - Calculate

- Dispersion means scatter, deviation, spread or fluctuation.

- Also known as measure of variation.

Uses of dispersion :—

- to determine reliability of an average to control the variability
- to compare two or more distributions regarding their variability.

Types :—

(1) Range & coefficient of Range
(2) Mean deviation & its coefficient
(3) Standard deviation
(4) Varience & coefficient of variance

(1) Range →

It is defined as it is the difference b/w the highest and lowest value in the observations.

Range (R) = H − L

where, H = Highest value
L = lowest value

coefficient of Range = $\frac{H-L}{H+L}$

(eg) Calculate the range and the coefficient of range for the following data regarding Hb% of 10 patients.

(L) 8.3, 9.6, 12.3, 10.2, 11.3, 9.6, 13.2 (H), 10.1 & 9.7

→ Highest value = 13.2, Lowest value = 8.3

$R = H - L$
$= 13.2 - 8.3$
$R = 4.9$

$C.\text{ of } R = \frac{H-L}{H+L}$
$= \frac{13.2-8.3}{13.2+8.3} = \frac{4.9}{21.5}$
$= 0.2279$

eg.2 for discrete

x (L → 10, H → 60)	10	20	30	40	50	60
frequency	8	16	15	14	12	10

take value from x

highest value → 60, lowest value → 10

$R = 60 - 10 = 50$

C.of R. ⇒ $\frac{60-10}{60+10}$

$50/70 = 0.714$

eg.3 for continuous

The following data show blood sugar level of one hundred final B. Pharm student. Calculate Range & coefficient of Range

→

Blood sugar	80-90	90-100	100-110	110-120	120-130
No. of students	8	12	13	17	30
Mid value (x)	85 (L)	95	105	115	125 (H)

Range = 125 − 85
= 40

Co. of Range = $\frac{125-85}{125+85}$
= $\frac{40}{210}$ ⇒ 0.190

Eg.4 — for yourself | Range & its coefficients

Data (x)	0-10	10-20	20-30	30-40	40-50
frequency (f)	1	5	10	13	9

② Standard Deviations →

It is widely used in the measure of dispersion.

- It is the square root of the arithmetic mean of squared deviation of items taken from the arithmetic mean.
- It is denoted by S.D. (σ) – Sigma

$$S.D./(\sigma) = \sqrt{\frac{\sum dx^2}{N}}$$

OR

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$$

- for discrete series →

$$\sigma = \sqrt{\frac{\sum f.(x-\bar{x})^2}{N}}$$

(N, $\sum f$)

- for continuous series →

Shortcut methods

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C$$

(N, $\sum f$)

where, $d = \frac{x - A}{C}$ — A: assumed mean; C: class interval

Steps →

- Calculate mean $\bar{x}$
- find the deviation of observation i.e. dx
- take the square of these deviation ie. dx^2
- take the summation of these squared deviation ie $\sum dx^2$
- Apply formula..

eg.1 find out standard deviation from the following data: 3, 7, 8, 9, 10

→ firstly prepare table $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$

x	3	7	8	9	10
$x-\bar{x}$ (dx)	-4.4	-0.4	0.6	1.6	2.6
$\lvert x-\bar{x}\rvert^2$ (dx^2)	19.36	0.16	0.36	2.56	6.76 ⇒ 29.2

$\bar{x}$ (mean) $= \frac{3+7+8+9+10}{5} = \frac{37}{5} = 7.4$

$\sigma = \sqrt{\frac{\Sigma dx^2}{N}} = \sqrt{\frac{29.2}{5}} \Rightarrow \sqrt{5.84} \Rightarrow 2.416$

eg.2 In a survey of 150 families in a village, the following distribution of ages of children was found:

Ages of children (x)	0-2	2-4	4-6	6-8	8-10
No. of families (f)	40	32	25	23	30
Mid-value(x)	1	3	5 (A)	7	9
$d = \frac{x-A}{C}$ (C=2)	-2	-1	0	1	2
d^2	4	1	0	1	4
fd	-80	-32	0	23	60
fd^2	160	32	0	23	120

→ firstly calculate mid value (x)
then calculate (d) by $d = \frac{x-A}{C}$
for this, choose Assumed mean (A) = 5
class interval = 2 → C

- then, calculate (d^2)
  then, calculate fd & fd^2 by multiplying with (f).
- then calculate Σfd & Σfd^2
  then put value in formula..

So, $\Sigma fd = -29$ $\quad \Sigma fd^2 = 335$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

($N = \Sigma f$)

$$= \sqrt{\frac{335}{150} - \left(\frac{-29}{150}\right)^2} \times C$$

$$= \sqrt{2.23 - 0.037} \times 2$$

$$= \sqrt{1.86} \times 2 \Rightarrow 1.363 \times 2$$

$$= 2.727 \text{ — approx.}$$

• By direct method → $\sigma = \sqrt{\frac{\Sigma f dx^2}{N}}$

$\bar{x} = 5$

x	1	3	5	7	9
$x - \bar{x}$	-4	-2		2	4
$(x-\bar{x})^2$	16	4	0	4	16 ⇒

$$\sigma = \sqrt{\frac{40}{5}} \Rightarrow \sqrt{8} \Rightarrow 2.828 \text{ — approx.}$$

Eg.3 the weight of nine ampicillin ampoules (in gm) are as follow. Calculate mean, standard deviation and coefficient of variation

2, 3, 2, 4, 3, 4, 5, 4, 2

$\rightarrow$ yourself

$$\text{coefficient of variation} = \frac{\text{S.D.}}{\text{mean}} \times 100$$

Eg.4 the following data reveal the pharmaceutical units having no. of ampoules filling machine in parenteral production facility of total 25 manufacturing plants present in Goa state

No. of Machines (x)	2	3	4 (A)	5	6	7
No. of manufacturing units (f)	2	4	5	6	4	4
d_x $x-A$, $A=4$	-2	-1	0	1	2	3
$f.dx$	-4	-4	0	6	8	12
d_x^2	4	1	0	1	4	9
fdx^2	8	4	0	6	16	36

• first take Assumed mean (A) = 4

find $d = x - A$ then multiply it with f

$\Sigma fdx = 18$, $\Sigma fdx^2 = 70$, $\Sigma f = 25 = N$

$$\sigma = \sqrt{\frac{\Sigma f d_x^2}{N} - \left(\frac{\Sigma f d_x}{N}\right)^2}$$

$$= \sqrt{\frac{70}{25} - \left(\frac{18}{25}\right)^2} = \sqrt{2.8 - 0.5184}$$

$$= \sqrt{2.281} \Rightarrow 1.510$$

- Mean $(\bar{x}) = A + \frac{\Sigma f d}{N} \Rightarrow 4 + \frac{18}{25} = 4.72$

⊙ By direct method

x	2	3	4	5	6	7	
f	2	4	5	6	4	4	$\Rightarrow 25$
fx	4	12	20	30	24	28	$\Rightarrow 118$
$(x-\bar{x})$	-2.72	-1.72	-0.72	0.28	1.28	2.28	
$(x-\bar{x})^2$	7.39	2.95	0.51	0.07	1.63	5.19	$\Rightarrow$
$f.(x-\bar{x})^2$	14.78	11.8	2.55	0.42	6.52	20.76	

$$\bar{x} = \frac{\Sigma fx}{\Sigma f(N)} = \frac{118}{25} = 4.72$$

$$\sigma = \sqrt{\frac{\Sigma f.(x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma f\, dx^2}{N}} \Rightarrow \sqrt{\frac{56.83}{25}}$$

$$= \sqrt{2.273} \Rightarrow 1.507 \Rightarrow 1.51$$

eg.5 Calculate the standard deviation of following data which show no. of peoples in the family suffering from viral infections.

No. of families	1	2	3	4	5
No. of patients	3	5	2	4	2

eg.6 The mean and standard deviation of two samples of size 80 and 40 are 15 and 10; and 8 and 5 respectively. Calculate the standard deviation of the combined sample

→ Here, $n_1 = 80$ and $n_2 = 40$

$\bar{x}_1 = 15$ and $\bar{x}_2 = 10$

$\sigma_1 = 8$ and $\sigma_2 = 5$

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{1200 + 400}{120} = \frac{1600}{120} = 13.33$$

$$\boxed{\bar{x}_{12} = 13.33}$$

x	15	10
$x - \bar{x}_{12}$ (d)	1.67	-3.33

$$\sigma_{12} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d^2}{n_1 + n_2}}$$

$$= \sqrt{\frac{80 \times 8^2 + 40 \times 5^2 + 80 \times (1.67)^2 + 40 \times (-3.33)^2}{80 + 40}}$$

$$= \sqrt{\frac{5120 + 1000 + 223.11 + 443.55}{120}}$$

$$= \sqrt{\frac{6786.66}{120}} = \sqrt{56.55} = \boxed{7.519}$$

- Standard deviation (σ) formula ..

— for Individual series → 1, 2, 3, 4, 5, 6

- Direct method → $\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}}$ | $\sigma = \sqrt{\frac{\Sigma dx^2}{N}}$

N → No. of observations

- Short cut method → $\sigma = \sqrt{\frac{\Sigma dx^2}{N} - \left(\frac{\Sigma dx}{N}\right)^2}$, where $dx = x - A$

N → No. of observation

— for discrete series → $x = 2, 4, 6$; $f = 1, 5, 8$; 19

- Direct method → $\sigma = \sqrt{\frac{\Sigma f \cdot (x-\bar{x})^2}{N}}$; $\Sigma f \cdot dx^2$

N → Σf

- Short cut method → $\sigma = \sqrt{\frac{\Sigma f dx^2}{N} - \left(\frac{\Sigma f dx}{N}\right)^2}$

where, $dx = x - A$

$N = \Sigma f$

— for continuous series → 0-10, 10-20 ; C

- Direct method → $\sigma = \sqrt{\frac{\Sigma f (x-\bar{x})^2}{N}}$

x → mid-value ; N → Σf

- Short cut method → $\sigma = \sqrt{\frac{\Sigma f dx^2}{N} - \left(\frac{\Sigma f dx}{N}\right)^2}$

where, $dx = m - A$ (m → mid-value)

$N = \Sigma f$

- Step-deviation method → $\sigma = \sqrt{\frac{\Sigma f dx^2}{N} - \left(\frac{\Sigma f dx}{N}\right)^2} \times C$

where, $dx = \frac{x - A}{C}$, C = class interval

- Range → $R = H - L$

- Coefficient of range → $\frac{H-L}{H+L}$

- Varience → σ^2

- Coefficient of varience → $\frac{S.D.}{Mean} \times 100$

CORRELATION

- It is the relation b/w two variables, in which change in the value of one variable changes the value of others.
- The value of correlation coefficient will vary from -1 to +1.

eg. x, y | $x = y + 1$ | if $y = 1$, then $x = 2$

1 = 0, | if $y = 2$, then $x = 3$

(A) → (J)

Types

- **Positive** → two variable moves in same directions.
  eg. Height & weight

x	1	2	3	4	5	6
y	10	20	30	40	50	60

- **Negative** → Value of both variable moves in opposite direction.

x	1	2	3	4	5	6	7
y	80	70	60	50	40	30	20

- **Zero** → Value of one variable doesnot affect the others..

- Karl Pearson's coefficient of correlation
  ↓
  It is the value which measure the correlation of two variables..
- Coefficient of correlation always vary ~~with~~ b/w the two limits of +1 and -1
  - +1 → perfect positive correlation
  - -1 → perfect negative correlation
- Zero shows the absence of correlation.

Karl Pearson → The great Biologist and statistician, given the formula for calculation of coefficient of correlation.

- Also known as Pearsonian coefficient of correlation or Product moment correlation coefficient.
- It is denoted by (r) (ℓ) (γ) (R)

$$r = \frac{\sum xy}{n\sigma_1\sigma_2}$$

where, $x = x - \bar{x}$

$y = y - \bar{y}$

(σ) = standard deviation

i) Direct Methods → (Use

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}}$$

where, $x = x - \bar{x}$

$y = y - \bar{y}$

$\bar{x} = \frac{\Sigma x}{N}$ mean

ii) Short cut methods →

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx . \Sigma dy}{N}}{\sqrt{\Sigma (dx)^2 - \frac{[\Sigma (dx)]^2}{N}} \sqrt{\Sigma (dy)^2 - \frac{[\Sigma dy]^2}{N}}}$$

where, $dx = x - A$ (Assumed mean), $dy = y - B$

OR

$$r = \frac{n \Sigma xy - \Sigma x \Sigma y}{\sqrt{[n \Sigma x^2 - (\Sigma x)^2][n \Sigma y^2 - (\Sigma y)^2]}}$$

Eg. 1 find the coefficient of correlation b/w x and y for the following data.

(AKTU PREVIOUS)

x	2	4	4	7	5
y	8	8	5	6	2

→ ① Direct method →

firstly find mean for x & y

$$\bar{x} = \frac{2+4+4+7+5}{5} \Rightarrow \frac{22}{5} = 4.4 \quad (\bar{x})$$

$$\bar{y} = \frac{8+8+5+6+2}{5} \Rightarrow \frac{29}{5} = 5.8 \quad (\bar{y})$$

Draw the table →

x	y	$x - \bar{x}$	$y - \bar{y}$	① xy	② x^2	③ y^2
2	8	-2.4	2.2	-5.28	5.76	4.84
4	8	-0.4	2.2	-0.88	0.16	4.84
4	5	-0.4	-0.8	0.32	0.16	0.64
7	6	2.6	0.2	0.52	6.76	0.04
5	2	0.6	-3.8	-2.28	0.36	14.44
				-7.6	13.2	24.8

eq. 2 Calculate the karl pearson coefficient of correlation b/w the ages of plants and their height in a field given as follow :-

Age (x)	1	2	3	4	5	6
Height (y)	7	11	14	19	24	29

$\Rightarrow$

$$r = \frac{\Sigma dx dy - \frac{\Sigma dx . \Sigma dy}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

x	y	dx = x - A	dy = y - B	dx^2	dy^2	dx.dy
1	7	-2	-7	4	49	14
2	11	-1	-3	1	9	3
(3) A	(14) B	0	0	0	0	0
4	19	1	5	1	25	5
5	24	2	10	4	100	20
6	29	3	15	9	225	45

⊙ Multiple Correlation → (R)

In this, we study the relationship b/w three or more variable.

- Suppose, there are three variables x, y, z, in which one are dependent and others is are independent.
- So, it is relationship b/w a variable and a combined variable.

① If (z) is dependent variable [x & y] are independent. then,

$$R_{z,xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 r_{xz} r_{yz} r_{xy}}{1 - r_{xy}^2}}$$

② If (y) is dependent, then

$$R_{y,xz} = \sqrt{\frac{r_{zy}^2 + r_{xy}^2 - 2 r_{xy} r_{zy} r_{xz}}{1 - r_{xz}^2}}$$

③ If (x) is dependent, then

$$R_{x,yz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2 r_{xy} r_{yz} r_{zx}}{1 - r_{yz}^2}}$$

(eg.) $r_{12} = 0.86$, $r_{13} = 0.71$, $r_{23} = 0.66$

→ $R_{1,23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2}}$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}}$$

$$= \frac{-7.6}{\sqrt{13.2 \times 24.8}} \Rightarrow \frac{-7.6}{\sqrt{327.36}}$$

$$= \frac{-7.6}{18.09} \Rightarrow \boxed{-0.4201} \quad (-0.42)$$

(2) short cut method →

firstly draw the table –

x	y	$dx = x - A$	$dy = x - B$	dx.dy	dx^2	dy^2
2	8	-2	3	-6	4	9
4	8	0	3	0	0	9
(4) A	(5) B	0	0	0	0	0
7	6	3	1	3	9	1
5	2	1	-3	-3	1	9
		(2)	(4)	(-6)	(14)	(28)

$$r = \frac{\Sigma dxdy - \frac{\Sigma dx \Sigma dy}{N}}{\sqrt{\Sigma (dx)^2 - \frac{[\Sigma dx]^2}{N}} \sqrt{\Sigma dy^2 - \frac{[\Sigma dy]^2}{N}}}$$

$$r = \frac{-6 - \frac{(2 \times 4)}{5}}{\sqrt{14 - \frac{4}{5}} \sqrt{28 - \frac{16}{5}}}$$

$\Sigma dxdy = (-6)$
$\Sigma dx = 2$
$\Sigma dy = 4$
$\Sigma dx^2 = 14$
$\Sigma dy^2 = 28$

$$= \frac{-6 - 1.6}{\sqrt{14 - 0.8} \sqrt{28 - 3.2}}$$

$$= \frac{-7.6}{\sqrt{13.2} \sqrt{24.8}} \Rightarrow \frac{-7.6}{3.63 \times 4.979}$$

$$\Rightarrow \frac{-7.6}{18.07} \Rightarrow \boxed{-0.4205} \quad (-0.42)$$