18/10/21

UNIT-3.

Central tendency:- (Measures).

1. Arithmetic mean (value at the middle).
2. Median (middle value).
3. Mode (
4. Geometric mean
5. Harmonic mean

Averages and are statistical constants which enable us to comprehend in a single effort the ~~simp~~ significance of the entire data (or) whole. (This gives us an idea abt the concentration of values in the central part of the data)

Requisites for an ideal measure:-

(i) It has to be define rigidly. (To evaluate it need a crt formula to evaluate the mean)

(ii) It should be readily comprehensible and easy to calculate.

(iii) It is based on all the observations.

Classification of data:

Examples.

(i) Individual observations → 30, 35, 37, 30, 43, 46.

(ii) Discrete observations →

x :	30	35	37	43	46
f :	2	1	1	1	1

(iii) Continuous observations → 30-35 - 3 (including 35)

35-40 - 1.

40-45 - 1.

45-50 - 1.

6

(or)

30-35 - 2

35-40 - 2

40-45 - 1

45-50 - 1

6

Definition of Average:: (X̄)

(i) Individual observations: $\bar{x}$ (Arithmetic $\bar{x}$ - mean)

$$\bar{X} = \frac{\sum_i x_i}{n}$$ {Add all the values) divided by no. of observation that would be our average)

(ii) Discrete Observations::

$$\bar{X} = \frac{\sum_i f x_i}{\sum_i f}$$ → frequency. (Multiplication of frequent divided by the frequencies)

(iii) continous observations:: (frequency distribution) — we can use these formula.

$$\bar{X} = A + \frac{\sum_i fd}{\sum_i f} \times i$$

A ↓ Assumed mean (middle value)

i → length of the class interval.

$d = \frac{m - A}{i}$ ← m → mid value of class interval.

i → length of the CI

19/10/21

Properties of Arthimetic Mean (AM):

① Algebric sum of the deviations of a set of values from their AM is zero '0'.

T.P: $\sum_i f_i (x_i - \bar{x}) = 0$ {Discrete case}

$(x_i - \bar{x})$ → deviations from AM.

[If x_i and $\bar{x}$ is provided we can find the AM].

(If the f_i is '1' then it is individual observations)

Proof: LHS:

Consider $\sum_i f_i (x_i - \bar{x})$

$= \sum_i f_i x_i - \sum_i f_i \bar{x}$.

$\bar{X} = \frac{\sum_i f_i x_i}{\sum_i f_i}$

$= \bar{x} \sum_i f_i - \bar{x} \sum_i f_i = 0$.

② The sum of the squares of the deviations of the set of values is the minimum when taken about mean.

Proof:-

$$X : x_1 \quad x_2 \quad \dots \quad x_n$$
$$f : f_1 \quad f_2 \quad \dots \quad f_n$$

$$\boxed{Z = \sum_{i=1}^{n} f_i (x - A)^2}$$

I.P:-

The minimum value of Z is obtained when $A = \bar{x}$.

Dif. w.r.t. A;

$$\frac{dz}{dA} = \sum_{i=1}^{n} 2 f_i (x - A)(-1).$$

$$\frac{d^2 z}{dA^2} = \sum_{i=1}^{n} 2 f_i (-1)(-1).$$

$$\Rightarrow \sum 2 f_i > 0$$

$\Rightarrow$ Minimum value is obtained when

$$-2 \sum f_i (x_i - A) = 0 \; //$$

(-2 cannot be zero)

$$\Rightarrow \sum f_i x_i - A \sum f_i = 0$$

$$\boxed{A = \frac{\sum f_i x_i}{\sum f_i} = \bar{x}}$$

Dispersion

S.D
M.D
all these are about mean and will always be minimum.

③ Combined AM

If $\bar{x}_i$ $(i = 1, 2, \dots k)$ are the means of k component series of size n_i $(i = 1, 2, \dots k)$ respectively, then the mean $\bar{x}$ of the

composite series obtained by combining the component series is given by;

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \cdots + n_k \bar{x}_k}{n_1 + n_2 + \cdots n_k}$$ } combined mean

$\frac{\bar{x}_1}{n_1}$ comp: 1 $\frac{\bar{x}_2}{n_2}$ Comp·2 $\frac{\bar{x}_k}{n_k}$ Comp K.

Proof:

$$\bar{x}_1 = \frac{(x_{11} + x_{12} + \cdots x_{1n_1})}{n_1}$$

$$\bar{x}_k = \frac{(x_{k1} + x_{k2} + \cdots x_{kn_k})}{n_k}$$

$$\bar{x} = \frac{(x_{11} + x_{12} + x_{13} + \cdots + x_{1n_1}) + \cdots + (x_{k1} + x_{k} \ldots + x_{kn_k})}{n_1 + n_2 + \cdots + n_k}$$

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \cdots + n_k \bar{x}_k}{(n_1 + n_2 + \cdots + n_k)}$$ no. of observa

1. Find the mean of 32, 45, 28, 13, 0, 16, 32, ~~20~~ 27.

Soln:

$$\frac{\text{sum of observations}}{\text{No. of observations}}$$ Formula.

$$\bar{X} = \frac{\sum_1^n x_i}{n}$$

$$\bar{X} = \frac{32 + 45 + 28 + 13 + 0 + 16 + 32 + 27}{8}$$

$= \frac{193}{8}$

$\bar{x} = 24.125.$

---

2. The algebric sum of the deviations of 20 observations measure from 30 is 2. Find the mean?

Soln: $\sum_{i=1}^{20} (x_i - A) = 2$

$\sum_{i=1}^{20} (x_i - 30) = 2.$

$\sum_{i=1}^{20} x_i - 20 \times 30 = 2$

$\sum_{i=1}^{20} x_i = 602.$

$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{602}{20}$ → $\boxed{\bar{x} = \frac{\sum x_i}{n}}$

$= 30.1 //$

---

3. Calculate the mean for the following.

x:	0	10	20	30	40	50	60
f:	8	10	11	16	20	25	15.

Soln:

x	f	fx
0	8	0
10	10	100
20	11	220
30	16	480
40	20	800
50	25	1250

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

Sum of the frequencies = 105.

$$\bar{x} = \frac{3750}{105}$$

$\bar{x} = 37\ 35.71$ (should be 2 or 3 decimal places)

4. calculate the mean for the following frequency distribution:

CI -	0-8	8-16	16-24	24-32	32-40	40-48
f -	4	12	20	28	36	44.

Soln:

$\rightarrow$ Assumed mean (middle value)

$$\bar{x} = A + \frac{\Sigma' fd}{\Sigma' f} \times i \rightarrow \text{length of the CI.}$$

$\rightarrow$ mid value of CI

where $d = \frac{m - A}{i}$

mid pt = $\frac{\text{CI} \rightarrow \text{Add CI}}{2}$ $\frac{}{2}$

CI	midpt.	f	$d = \frac{m-A}{i}$	fd	fm
0-8	4	4	$d = \frac{4-28}{8} = -3$	4x-3 = -12	16
8-16	12	12	-2	12x-2 = -24	144
16-24	20	20	-1	-20	400
24-32	28 (A)	28	0	0	784
32-40	36	36	1	36	1296
40-48	44	44	2	88	1936

$\Sigma' f = 144$

$\Sigma' fd = 68$

$\Sigma' fm = 4576$

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

$$= 28 + \frac{68}{144} \times 8$$

$$= \frac{4032 + 68 \times 8}{144} = \frac{4576}{144}$$

$$= 31.777../31.778$$

(or)

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

$$= \frac{4576}{144} = 31.77../31.778$$

---

5. Find the mean of the following:

CI	0-7	8-14	15-22	23-30	31-38	39-46
f:	8	7	16	24	15	7

Soln:

CI	f	midpt.	d = m - A	fd
0-7.5	8	3.5	-23	-184
7.5-14.5	7	11	-15.5	-108.5
14.5-22.5	16	18.5	-8	-128
22.5-30.5	24	26.5	0	0
30.5-38.5	15	34.5	+8	120
38.5-46.5	7	42.5	+16	112
	77			-188.5

$\Sigma f = 77$

$\Sigma fd = -188.5$

$$\bar{x} = A + \frac{\Sigma fd}{\Sigma f}$$

$$\bar{x} = 26.5 + \left(\frac{-188.5}{77}\right)$$

$$= 24.05//$$

20/10/21

Recall:

* AM:- Its a measure to represent the entire data.

* Individual observation : $\bar{x} = \frac{\sum x_i}{n}$ → no. of observations

* Discreate observation : $\bar{x} = \frac{\sum fx}{\sum f}$ → frequency

* continous data : $\bar{x} = A + \frac{\sum fd}{\sum f} \times i$ → Assumed mean

$d = \frac{m - A}{i}$ → mid-pt of the CI

i → length of the CI.

* combined AM:-

$\frac{n_1}{\bar{x}_1} \cdot \frac{n_2}{\bar{x}_2} \cdots \frac{x_k}{\overline{x_{nk}}}$

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

(1.) The average salary of male employees in a form was Rs. 5200 and that of females was Rs. 4200. The mean salary of all employees was Rs. 5000. Find the percentage of male and female employees.

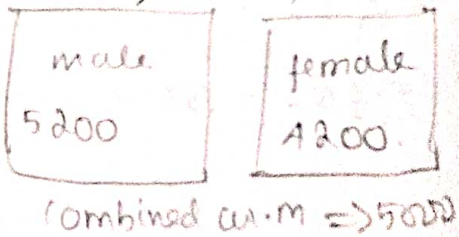
combined mean.

% of m & f employees.

Soln:

$\bar{x_1} = 5200$

Average salary of male employee



$\bar{x}_2 = 4200$ (Average salary for female employee)

$\bar{x} = 5000$.

n_1 → is no. of male employees

n_2 → no. of female employees.

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$ → Formula

$$5000 = \frac{n_1(5200) + n_2(4200)}{n_1 + n_2}$$

% of mal
fem

$$5000n_1 + 5000n_2 = 5200n_1 + 4200n_2.$$

$$5000n_2 - 4200n_2 = 5200n_1 - 5000n_1$$

$$800n_2 = 200n_1$$

$$\frac{n_1}{n_2} = \frac{800}{200} = \frac{4}{1}$$

4/50

% of male employees $= \frac{4}{5} \times 100 = 80$

% of female employees $= \frac{1}{5} \times 100 = 20$.

---

80 % are males and 20 % are females.

Disadvantages of AM:

1. It is affected by the extreme values.
2. It can't be determined by inspection nor it can be located graphically.
3. AM can't be calculated if any observation is missing.
4. In extremely asymmetrical distribution AM is not a suitable measure of location.

Median:

↳ The middle value of the data. (No. of observations)

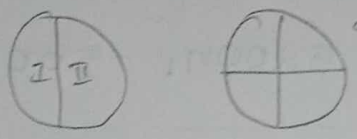
Individual observations: size of $\left(\frac{N+1}{2}\right)^{th}$ item.

Discrete observations: size of $\left(\frac{N+1}{2}\right)^{th}$ item. ($N = \Sigma f$ = sum of all frequencies; Σf → summation)

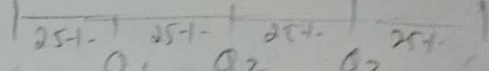
Continous observations: Median $= L + \frac{\frac{N}{2} - cf}{f} \times i$

- L → lower limit of the median class
- f → frequency of median class
- cf → cumulative frequency preceding the median class
- i → length of the class

Median is nothing but cutting the two parts equally.



quartiles

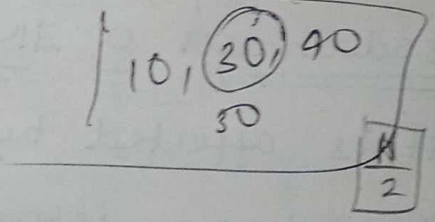


$$Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times i \quad (1^{st} \text{ quartile})$$

$\frac{3N}{4}$

$$Q_3 = L + \frac{\frac{3N}{4} - cf}{f} \times i \quad (3^{rd} \text{ quartiles})$$

What reason for indi. & discrete observations we have written $\left(\frac{N+1}{2}\right)^{th}$ and for continous observation we have written $\left(\frac{N}{2}\right)$ →?



$\Rightarrow \frac{20+30}{2}$

$\Rightarrow \frac{50}{2}$

$\Rightarrow 25$

Eg:

(i) Find the median of 20?

There is no median.

(ii) Find the median of 10, 20, 30, 40.

In this median is 25.

Advantage:

* It is unaffected by the extreme values. (outliers)

1. Find the median for the data: (Individual observations)

40, 25, 60, 35, 44, 50

Soln:

* Arrange in ascending / decending order

25, 35, 40, 44, 50, 60.

No. of observation (N) = 6.

Median :- size of $\left(\frac{N+1}{2}\right)^{th}$ item. = size of 3.5^{th} item

= Avg. of 3^{rd} & 4^{th} position

$= \frac{40+44}{2}$

$= 42.$

The median is 42 //

---

2. Find the median of 30, 60, 20, 15, 70.

Soln:

* Arrange in ascending order

15, 20, 30, 60, 70.

N = 5.

Median := size of $\left(\frac{N+1}{2}\right)^{th}$ item = size of 3^{rd} item

= Avg

Median = 30 //

---

3. Obtain median for the following (Discrete observation)

x :	10	20	30	40	50	60	70	80
f :	3	15	20	6	10	8	7	11.

Soln:

* Form the table as follows;

x	f	cf
10	3	3
20	15	18
30	20	38
40	6	44
50	10	54
60	8	62
70	7	69
80	11	80
	80	

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

$= \text{size}\left(\frac{80+1}{2}\right)^{th}$ item

$N = \sum f = 80.$

$= \text{size}(40.5)^{th}$ item.

In cf column we should see the no's that is $(\geq)$ to the size of th item

i.e., (40.5)

In this soln it is 44

40 (40.5) 41

$\geq \quad \geq \quad \geq$

Median = 40//

Suppose if we want to calculate

$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)^{th}$ item $\qquad Q_1 = \frac{N+1}{4}$

$= \text{size of}\left(\frac{8}{4}\right)^{th}$ item $\qquad = \frac{82}{4}$

$= 20.25$//

$Q_1 = 30$, Median = 40.

4. Find the median. (continous observation)

CI:	10 – 20	20-30	30-40	40-50	50-60
f:	20	90	150	100	70.

Soln:

Form the table:-

P. T. O.

CI	f	cf
10-20	20	20
20-30	90	110 cf
30-40	150 f	260 → median class
40-50	100	360
50-60	70	430
	430	

$\frac{N}{2} = \frac{430}{2} = 215.$

$\frac{430}{2}$ gets $\geq$

Median $= L + \frac{\frac{N}{2} - cf}{f} \times i$

$= 30 + \frac{215 - 110}{150} \times 10$

$= 30 + \frac{105}{15} \Rightarrow 37$

5- Given median is 46. Determine the missing frequencies

CI:	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f:	12	30	? x	65	? y	25	18.

· sum of all frequencies is 229.

Soln:

Form a table.

frequency column la edha edutthukoromo annuku mela ulla cf tha cf value

CI	f	cf
10-20	12	12
20-30	30	42
30-40	x	42+x cf
40-50	65 f	107+x
50-60	y	107+x+y
60-70	25	132+x+y
70-80	18	150+x+y
	229	

Given:

Median = 46.

Median class - 40-50.

$150 + x + y = 229$

$x + y = 229 - 150$

$\geq$

$x + y = 79$.

$L = 40, \frac{N}{2} = \frac{229}{2} = 114.5$

$cf = 42 + x ; f = 65$

$i = 10$

$\frac{229}{2}$.

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

$$46 = 40 + \frac{114.5 - 42 - x}{65} \times 10.$$

$$46 = 40 + \cdot$$

$$6 = \frac{76.5 - x}{65} \times 10$$

$$6 \times 65 = 76.5 - x \times 10.$$

$$\frac{6 \times 65}{10} = 76.5 - x$$

$$x = 76.5 - \frac{390}{10}$$

$$x = 76.5 - 39$$

$$\boxed{x = 37.5}$$

$$x + y = 79.$$

$$37.5 + y = 79$$

$$y = 79 - 37.5$$

$$\boxed{y = 41.5}$$

21/10/21

Recall:

Median: → Middle value of the data.

Formulas:

Individual observations - size of $\left(\frac{N+1}{2}\right)^{th}$ item (N → No. of observations)

Discrete observations - size of $\left(\frac{N+1}{2}\right)^{th}$ item, $N = \sum f$

continous observations - $L + \frac{\frac{N}{2} - cf}{f} \times i$

(L → lower limit of median class; cf → $\sum f$ → cumulative fre. preceding the median class; i - length of CI; f → frequency of median class)

Mode:

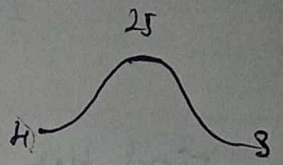
Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster density.

Eg:1-

Consider the following frequency distribution.

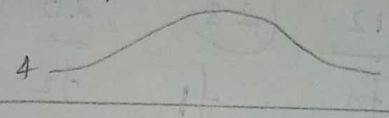
x:	1	2	3	4	5	6	7	8
f:	4	9	16	25	22	15	7	3

→ mode



Find Mode = ?

Mode = 4 //



(How to find mode?: Which is maximum frequency that should be considered as the mode

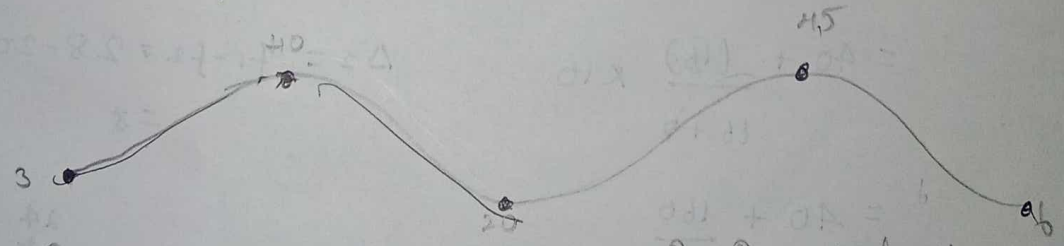
Eg:2

Consider the following frequency distribution:

x:	1	2	3	4	5	6	7	8	9	10	11	12
f:	3	8	15	23	35	40	32	28	20	45	14	6

Find mode = ?

Mode = 10 // (Highest frequency).



(When we have like this data it is called as irregular data)

Relation blw mean, median, mode.

Mode = 3median − 2 Mean → Formula.

Continous observation:-

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

L → lower limit of modal class (CI which have highest frequency)

i → length of the CI.

f_1 → frequency of modal class (highest frequency)

f_0 → frequency preciding the modal class.

f_2 → frequency succeding the modal class.

Where = $\Delta_1 = f_1 - f_0$

$\Delta_2 = f_1 - f_2$

---

1. Find the mode for the following:-

CI ::	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f :	5	8	7	12	28	20	10	10
				f_0	f_1	f_2		

Soln:-

Highest frequency $f_1 = 28$, $f_0 = 12$, $f_2 = 20$.

Modal class → 40-50 :- $L = 40$, $i = 10$.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i.$$

$$= 40 + \frac{(16)}{16 + 8} \times 10$$

$$= 40 + \frac{160}{24}$$

$$= \frac{960 + 160}{24}$$

$$= \frac{1120}{24}$$

$$= 46.1666\ldots \text{ (approx)}$$

$\Delta_1 = f_1 - f_0 = 28 - 12 = 16$

$\Delta_2 = f_1 - f_2 = 28 - 20 = 8$

24 | 1120 (46

24

960

1160

1120

Page Checking

x	f	C_1	column 2 C_2	column 3 C_3	C_4	C_5	C_6
1	3	→ 11			26		
2	8			→ 23		46	
3	15	→ 38					78
4	23			→ 58			
5	35	→ (75)			(98)		
6	10			(72)		(107)	
7	32	→ 60					(100)
8	28			→ 48	80	93	
9	20	→ 65					
10	45			→ 59			79
11	14	→ 20			65		
12	6						

This is called grouping method

Find the maximum value in column 2 i.e; 75// → adding 1st & 2nd frequencies

Find the maximum value in colom 3 i.e; 72// → 2nd & 3rd.

In column 4 -> we should add 1st '3' frequencies.

↳ Find the maximum value in C_4 i.e; 98//

In C_5 leave the 1st frequency and from 2nd add three frequencies ie; 8+15+23

↳ Find the maximum value in C_5 i.e; 107

(x) x=5, 6, 7. has got a role to play calculation of the mode.

In C_6 omit the 1st-2' frequencies and add sum of '3' frequency

↳ Find the maximum value in C_5 ie; 100.

C_1 → frequency

C_2 → sum of two frequency (consecutive)

C_3 → omit the 1st frequency and sum the two (consecutive freq.)

C_4 → the sum of the 1st '3' frequency

C5 -> omit the 1st frequency and sum the '3' consecutive

C6 -> omit the 1st '2' frequency and sum the '3' " " "

{Identify the maximum in each column}

Max in C1 -> 45 ie; (10). (x)

Max in C2 -> 75 ie; 5, 6.

Max in C3 -> 72 ie; 6, (7)

Max in C4 -> 98 ie; (4), (5), (6)

Max in C5 -> 107 ie; 5, 6, 7

Max in C6 -> 100 ie; 6, 7, (8).

(When we have irregular data like this we want to do.)

x	f
4	1
5	3
6	(5)
7	3
8	1
10	1

x: 4 -> (1) that like we want to calculate.

-> Higest frequency -> 5.

So, the mode is 6 //

25/10/21

Recall: (last week)

central tendency.

concentrated in middle part of the data.

1) Mean, 2) Median, 3) Mode, 4) Geometric mean

5) Harmonic mean.

classifications of data:

1) Individual observation, 2) Discrete observation

3) Continous observation.

* Mean: $\bar{x} = \frac{\Sigma_i x}{n}$ → Individual observations

$\bar{x} = \frac{\Sigma_i fx}{\Sigma_i f}$ → Discrete observations.

$\bar{x} = A + \frac{\Sigma_i fd}{\Sigma_i f} \times i$ → Continous observation

A → Assumed mean

where $d = \frac{m - A}{i}$ (m → mid pt, i → length of CI)

* Median: $\bar{x}_z$ size of $\left(\frac{N+1}{2}\right)^{th}$ item → Indi. Obs

($N = \Sigma_i f$)

size of $\left(\frac{N+1}{2}\right)^{th}$ item → Discrete-Obs

$L + \frac{\frac{N}{2} - cf}{f} \times i$ → continous Obser

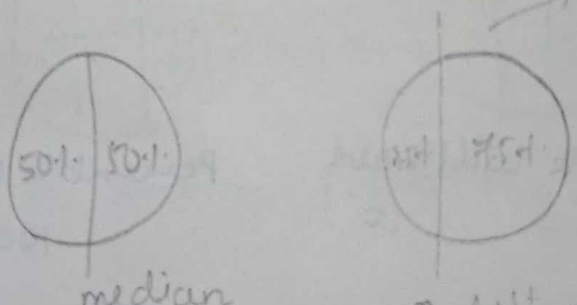
L → lower limit of the median cls

cf → cummulative freq. preceding the median class

f → frequency of the median class

i → length of the CI.

Putting the values:-



Q_1 = size of $\left(\frac{N+1}{4}\right)^{th}$ item.

$Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times i$

* Median: $\bar{X}_7$ size of $\left(\frac{N+1}{2}\right)^{th}$ item → Indi. Obs

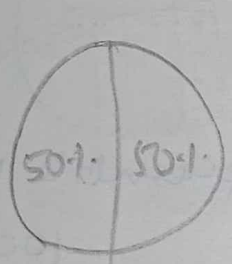
$\lceil \rightarrow N = \Sigma f$

size of $\left(\frac{N+1}{2}\right)^{th}$ item → Discrete-obs

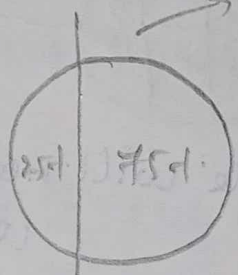
$$L + \frac{\frac{N}{2} - cf}{f} \times i$$

- L → lower limit of the median cls
- cf → cummulative freq. preceding the median class
- i → continous obser ↳ width of the CI.
- f → frequency of the median class

Partining the values:-



median

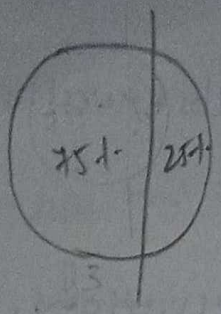


Q_1 (1st quartile)

$\rightarrow \left(\frac{N+1}{4}\right)^{th}$ item.

Q_1 = size of $\left(\frac{N+1}{4}\right)^{th}$ item.

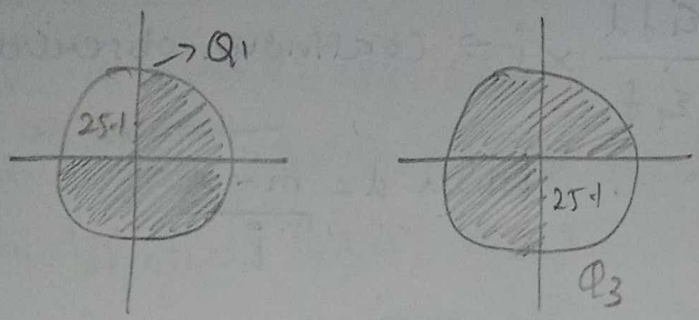
$$Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times i$$



Q_3 = size of $\left(\frac{N+1}{4}\right)^{th}$ item.

$$Q_3 = L + \frac{\frac{3N}{4} - cf}{f} \times i$$

L ↓ lower limit of the 3rd quartile

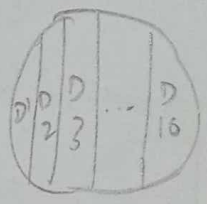


Percentile, Decile

$$\text{Decile} = L + \frac{\frac{N}{10} - cf}{f} \times i.$$

$$K^{th} \text{ decile} = L + \frac{\frac{KN}{10} - cf}{f} \times i.$$

$30\% \Rightarrow K = 3$



Decile will have 10

Percentile will have 100

If we want 60^{th} percentile $= L + \frac{\frac{60N}{100} - cf}{f} \times i$

↓

6^{th} decile.

If we want 50^{th} percentile $= L + \frac{\frac{50N}{100} - cf}{f} \times i.$

& ↳ 5^{th} decile.

median

(half half so it is median)

[5th decile and 50th percentile are same as median]

Mode :- Most replaced value

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

↓ modal class

where, $\Delta_1 = f_1 - f_0$ → frequency preceding the modal class

$\Delta_2 = f_1 - f_2$ → frequency of the modal class

→ frequency succeeding the modal class.

Mode = 3 Median − 2 Mean.

↳ Relationship b/w mean, median, mode.

Geometric mean:

Geometric mean of a set of n observations is the n^{th} root of their product.

Individual observations:

$x_1, x_2, \ldots, x_n$ are the given values

$$G.M = G = (x_1 x_2 \ldots x_n)^{1/N}.$$

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \ldots + \log x_n)$$

$$G = \text{Antilog}\left\{\frac{1}{n}\sum_{i=1}^{n} \log x_i\right\},$$

Formula for frequency (Discrete data):

$$G = \text{Antilog}\left\{\frac{1}{n}\sum_{i=1}^{n} f_i \log x_i\right\} \quad \& \quad N = \sum_{i=1}^{n} f_i.$$

$$GM = \left(x_1^{f_1} \cdot x_2^{f_2} \ldots x_n^{f_n}\right)^{1/N}. \quad \rightarrow \text{where } N = \Sigma f$$

Continous data:

GM = Antilog{

x :	5	15	25	35
f :	3	4	7	6

This is the discrete data

convert into continous data.

Sol:

CI :-	0-10	10-20	20-30	30-40
f :-	3	4	7	6

If this is given we want to discrete data then ~~the~~ we should find mid-pt

Disadvantages of Geometric mean (GM):

(1) If one value is (-)ve then GM becomes difficult.

(2) If any one value is zero, then GM = 0.

Relationship blw AM & GM.

AM ≥ GM.

~~Proof~~.

a, b are the two values.

$$AM = \frac{a+b}{2}.$$

$$GM = \sqrt{ab}$$

To prove:

AM ≥ GM. Equivalently, to prove AM − GM ≥ 0.

Consider $A.M - G.M = \frac{a+b}{2} - \sqrt{ab}$

$$= \frac{a+b-2\sqrt{ab}}{2}$$

$$= \frac{(\sqrt{a}-\sqrt{b})^2}{2} \geq 0.$$

$\therefore AM \geq G.M.$

Hence proved //

---

1. Find the GM of $-3, -5$.

$GM = \sqrt{-3 \times -5}$

$GM = -\sqrt{15}$.

$\sqrt{(-1)^2} = 1.$

$i = -\sqrt{1}$

$\sqrt{-3}$

$-1 = i^2 = \sqrt{i^4} = \sqrt{i^2 \cdot i^2}$

$= \sqrt{(-1)(-1)}$

$\sqrt{(a)(-b)} = \pm\sqrt{ab}$

$= \sqrt{1}$

$= 1$ //

$\therefore -1 = 1$

Geometric mean of combined grp:

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2 + \cdots + n_k \log G_k}{n_1 + n_2 + \cdots + n_k}$$

$G_1 = (x_{11}\, x_{12} \cdots x_{1n_1})^{1/n_1}$

$\log G_1 = \frac{1}{n_1}\left\{\sum^{n_1} \log x_{1i}\right\}$

$$n_1 \log G_1 = \sum_{i=1}^{n_1} \log x_{1i}$$

$$G_k = (x_{k1}\, x_{k2} \cdots x_{kn_k})^{1/n_k}$$

$$n_k \log G_k = \sum_{i=1}^{n_k} \log x_{ki}.$$

Combined grp

$$G = (x_{11}\, x_{12} \cdots x_{1n_1}\, x_{21}\, x_{22} \cdots x_{2n_2} \cdots x_{k1}\, x_{k2} \cdots x_{kn_k})^{\frac{1}{n_1+n_2+\cdots+n_k}}$$

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2 + \cdots + n_k \log G_k}{n_1 + n_2 + \cdots + n_k}$$

26/10/21: Recall:-

Geometric mean:

$$G_M = (x_1 x_2 \cdots x_n)^{1/n}$$

For calculation purpose:

(Individual observation)

$$G = \text{Antilog}\left\{\frac{1}{n}\sum_i \log x_i\right\}$$

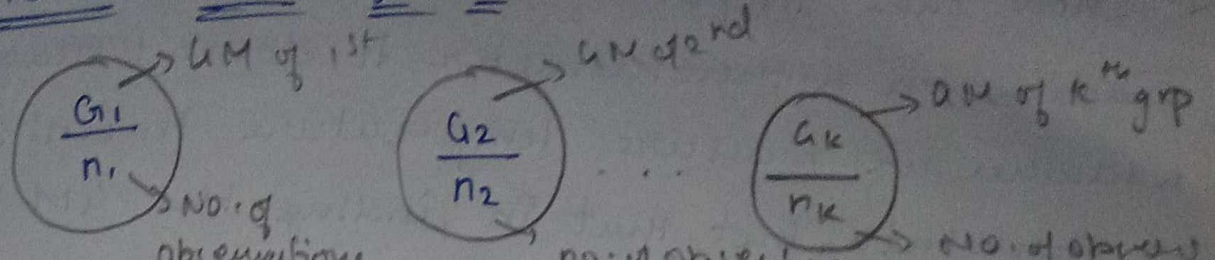
(Discrete distributions):

$$G = \text{Antilog}\left\{\frac{1}{N}\sum_i f_i \log x_i\right\}, \quad N = \sum_i f_i.$$

for both same

(continous distribution)

combined data for GM:



$$G = \text{Antilog}\left\{\frac{\sum_{i=1}^{k} n_i \log a_i}{\sum_{i=1}^{k} n_i}\right\}$$

→ Formula.

---

1. the GM of 10 observations on a certain variable was calculated as 16.2. It was later discovered that one of the variable/observation was wrongly recorded as 12.9 instead of 21.9. calculate the correct GM?

soln:

Let x_1 be 12.9. → was wrongly entered

x_1' be 21.9. , $n = 10$.

$G' = (x_1 x_2 \cdots x_n)^{1/n} = 16.2$.

$= (12.9 \times x_2 \times x_3 \cdots x_{10})^{1/10} = 16.2$

$= (x_2 x_3 \cdots x_{10})^{1/10} = \frac{16.2}{(12.9)^{1/10}}$.

corrected GM $= (x_1' x_2 x_3 \cdots x_{10})^{1/10} = (x_1')^{1/10} (x_2 x_3 \cdots x_{10})^{1/10}$

$= (21.9)^{1/10} \times \frac{16.2}{(12.9)^{1/10}}$.

(or)

$\left(16.2 \left(\frac{\times 21.9}{12.9}\right)\right)^{1/10}$.

2. Show that in finding the AM of a set of readings on a thermometer it doesn't matter whether we measure temperature in centigrade or Fahrenheit, But that in finding the GM it does matter which scale we use.

Ans: $\left\{\frac{F-32}{180} = \frac{C}{100}\right\}$

Soln:

$$F - 32 = \frac{\cancel{180}^{9}}{\cancel{100}_{5}} \times C$$

$$\boxed{F = 32 + \frac{9}{5} C.}$$

Suppose $C_1, C_2, \cdots C_n$ are in centigrade

$$AM \Rightarrow \bar{C} = \frac{1}{n}\{C_1 + C_2 + \cdots + C_n\}$$

$$GM \Rightarrow G_C = (C_1 C_2 \cdots C_n)^{1/n}.$$

The observations corresponding to Fahrenheit would be;

$$32 + \frac{9}{5} C_1 \to 32 + \frac{9}{5} C_2, \cdots, 32 + \frac{9}{5} C_n$$

$$\bar{C_f} = \frac{1}{n}\left\{\left(32 + \frac{9}{5} C_1\right) + \left(32 + \frac{9}{5} C_2\right) + \cdots + \left(32 + \frac{9}{5} C_n\right)\right\}$$

$$= 32n + \frac{9}{5}\left\{\frac{C_1 + C_2 + \cdots + C_n}{n}\right\} = 32 + \frac{9}{5}\bar{C}.$$

G.M

$$G_{C_f} = \frac{1}{n}\left\{\left(32 + \frac{9}{5} C_1\right)\left(32 + \frac{9}{5} C_2\right) \cdots \left(32 + \frac{9}{5} C_n\right)\right\}$$

$$\neq \frac{9}{5}(C_1 C_2 \cdots C_n)^{1/n} + 32$$

3. In a frequency table the upperbound of each class interval has a constant ratio to the lower bound. show that GM is given by $\log G$

$$\log G = x_0 + \frac{c}{N} \sum_i f_i (i-1)$$

where; x_0 - logarithm of midvalue of 1st CI.

c - " " " ~~constant ratio~~.

ratio of upperbound and lowerbound

Soln:

Let the i^{th} CI is denoted as;

$I_i - I_{i+1}$ & the frequency is f_i

$i = 1, 2, \ldots, n$

$I_1 - I_2$ f_1

$I_2 - I_3$ f_2

$I_3 - I_4$ f_3

Given: $\frac{I_2}{I_1} = \frac{I_3}{I_2} = \cdots = \frac{I_i}{I_{i-1}} \cdots = \lambda$

$I_i = \lambda I_{i-1}$

$= \lambda(\lambda I_{i-2}) = \cdots = \lambda^{i-1} I_1$.

Let x_i be the mid-pt of i^{th} class

$$x_1 = \frac{1}{2}(I_1 + I_2) = \frac{1}{2}(I_1 + \lambda I_1)$$

$\{I_i = \lambda^{i-1} I_1\}$.

$I_2 = \lambda I_1$

$$= \frac{I_1}{2}(1+\lambda)$$

$$x_i = \frac{1}{2}(I_i + I_{i+1}) = \frac{1}{2}\{\lambda^{i-1} I_1 + \lambda^i I_1\}$$

$$= \frac{1}{2} \lambda^{i-1} I_1 (1+\lambda).$$

$$x_1 = \left(\frac{1+\lambda}{2}\right) I_1 \quad \& \quad x_i = \frac{1}{2} \lambda^{i-1} I_1 (1+\lambda)$$

$$x = \lambda^{i-1} x_1$$

Let G be the GM;

$$\log G = \frac{\sum f_i \log x_i}{\sum f_i N}$$

$$= \frac{1}{N} \sum f_i \log (\lambda^{i-1} x_1)$$

$$\log G = \frac{1}{N} \sum_i f_i \log (x_1 \lambda^{i-1})$$

$$= \frac{1}{N} \{ \sum_i f_i [\log x_1 + \log \lambda^{i-1}) \}$$

$$= \frac{1}{N} \sum_i f_i \log x_1 + \frac{1}{N} \sum_i f_i \log \lambda^{i-1}.$$

$$= \frac{1}{N} \sum_i f_i \log x_1 + \frac{1}{N} \sum_i f_i (i-1) \log \lambda.$$

$$= \frac{\log x_1}{N} \sum_i f_i + \frac{\log \lambda}{N} \sum_i f_i (i-1).$$

one and one same

$$= x_0 + C \sum_i f_i (i-1) //$$

Harmonic Mean:- (HM)

HM of a no. of observations, none are zero ie; (nothing should be zero'0' in the given data) is the reciprocal of A.M of the reciprocal of the given values.

ie; $x_1, x_2, \ldots, x_n$ is the given data then the harmonic mean is;

$$HM = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$ (Individual observation)

$$HM = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}, \quad N = \sum_{i}^{1} f_i$$

(Both for discrete & continous)

In case in continous observation take x_i as the mid pt, and in discrete case as such as ie; f_i/x_i.

---

1. A cyclist speads from his house to his college at a speed of 10 km/Hr and back from college to his house at 15 km/Hr. Find the average speed.

Soln:

clg → H
H → clg
(2x)

To find the average speed use HM

Let 'x' be the distance b/w house and clg

(i) House to College → $\frac{x}{10}$ Km/Hr (distance travelled)

(ii) College to house → $\frac{x}{15}$ Hr (distance travelled).

the total distance 2x is covered in $\frac{x}{10} + \frac{x}{15}$ hrs.

Average speed = $\frac{\text{Total distance}}{\text{Total time taken}}$ = $\frac{2x}{x\left(\frac{1}{10}+\frac{1}{15}\right)}$ = $\frac{2}{\left(\frac{1}{10}+\frac{1}{15}\right)}$ = $\frac{2 \times 10 \times 15}{25}$ = 12 hrs//

27/10/21

Dispersion:

Dispersion is scatteredness.

Why to study this?

* To have an idea about the homogeneity and heterogeneity of the gn. data / ~~obs.~~ distribution.

Defn:

The degree to which numerical data tend to spread about an average value is called dispersion / variation of ~~data~~ data.

Measures of dispersion which are widely used:

- Range → co-efficient of range
- Quartile deviation → co-efficient of quartile dev.
- Mean deviation → co-efficient of mean dev.
- Standard deviation → co-efficient of variation.
  - ↳ moments.
    - Skewness → Size of the data
    - Kurtosis → Flatness of the data. (height of the gn-data)

All these are independent of units

suppose if we wish to compare dispersion of 2 samples.

we use co-efficient of dispersion.

(It is independent of units) Dispersion.

Formulas:

* Range = $L - S$

L → largest value
S → smallest value

Co-efficient of range = $\frac{L-S}{L+S}$

(The same formula for all 3 observations)

ie; Individual, Discrete, Continous

* Quartile deviation = $\frac{Q_3 - Q_1}{2}$ → 3rd quartile, → 1st quartile.

Co-eff. of QD = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Mean deviation:

MD from α = $\frac{1}{N} \sum_i f_i |x_i - \alpha|$

α => mean → mean deviation abt mean

α => median → mean deviation abt median

α => mode → mean deviation abt mode

* Co-eff. of MD = $\frac{MD}{mean}$

* Co-eff of median = $\frac{MD}{median}$

* Co-eff of mode = $\frac{MD}{mode}$

Standard deviation: (σ sigma)

$$\sigma = \sqrt{\frac{1}{N} \sum_i f_i (x_i - \bar{x})^2}$$ Formula.

$$\sum f_i = N \text{ (Total)}.$$

* called as Root mean square value.

* σ^2 = variance of the distribution.

Equavalent formula for SD: (Alternate formula)

$$\sigma = \sqrt{\frac{\sum f d^2}{\sum f} - \left(\frac{\sum f d}{\sum f}\right)^2} \times i$$

where $d = \frac{m - A}{i}$

SD considered to be one bcz it is used to compare

Co-eff of variation:-

$$CV(x) = \frac{\sigma}{\bar{x}} \times 100$$

σ - σ^2

1. For a group of 200 students the mean and standard deviation (SD) of scores are found to be 40 and 15 respectively. Later it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the correct mean and SD?

soln:

Gr: $n = 200$, $\sigma = 15$

$\bar{x} = 40$

To calculate corrected mean & SD.

$$\bar{x} = \frac{\sum x}{n} = \frac{\bar{x} \cdot n}{40 \cdot 40}$$

34 → 43 (correct value)

53 → misread 35

34 batha 43

53 " " 35

$\sum x = n \cdot \bar{x} = 200(40) = 8000$.

corrected $\sum x = \sum x$ - 34 - 53 + 43 + 35

$= 8000 - 9$.

$= 7991$

Corrected mean $= \frac{7791}{200}$ $\quad \frac{}{n}$

$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ (Individual observation)

Corrected $\sum x^2 =$ $\quad$ 40

$15^2 = \frac{\sum x^2}{200} - (40^2)$ $\quad \bar{x} = \frac{\sum f}{n}$

$225 = \frac{\sum x^2}{200} - (1600)$

$\frac{\sum x^2}{200} = 1825$

$\sum x^2 = 1825 \times 200$

$\sum x^2 = 365000$

Corrected $\sum x^2 = \sum x^2 - 34^2 - 53^2 - 43^2 + 35^2$

$= 365000 - 1156 - 2809 + 1849 + 1225$

$= 364109$

Corrected $\sigma^2 = \frac{364109}{200} - \left(\frac{7991}{200}\right)^2$

$= 224.1429$

$\sigma = \sqrt{224.1429}$

$\sigma = 14.97.$

---

2. The 1st of the two samples has 100 items with mean 15 and SD 3. If the whole group

has 250 items with mean 15.6 and SD $\sqrt{13.44}$. Find the SD of the 2nd grp.

Soln:

{ Combined SD is gn by

$$\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + \cdots + n_k(\sigma_k^2 + d_k^2)}{n_1 + n_2 + \cdots + n_k}$$

where $d_i = \bar{x}_i - \bar{x}$ $(i = 1, 2, \ldots, k)$

$\bar{x}$ = combined mean }

Soln:

$\bar{x}_1 = 15$, $n_1 = 100$, $n_2 = 150$

$\bar{x} = 15.6$ (combined mean)

$\sigma_1 = 3$, $\sigma^2 = 13.44$ (combined variance)

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \Rightarrow 15.6 = \frac{100(15) + 150(\bar{x}_2)}{250}$$

$$\bar{x}_2 = \frac{15.6 \times 250 - 1500}{150}$$

$$\bar{x}_2 = \frac{2400}{150} = 16$$

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$$

$d_1 = \bar{x}_1 - \bar{x}$
$= 15 - 15.6$
$= -0.6$

$d_2 = \bar{x}_2 - \bar{x}$
$= 16 - 15.6$
$= 0.4$

$$13.44 = \frac{100(9 + 0.36) + 150(\sigma_2^2 + 0.16)}{250}$$

$$1344 \times 250 = 936 + 150(\sigma_2^2 + 0.16)$$

$3360 - 936 = 150\,(\sigma_2^2 + 0.16)$

$2424 = 150\,(\sigma_2^2 + 0.16)$

$\sigma_2^2 = 16$ //

$\sigma_2 = 4$ //

$\sigma_2^2 = 16$

$\sigma_2\sqrt{10} = 4$

28/10/21

Recall:

* Measures of dispersion:

=> Range ⟶ L-S

=> Quartile deviation ⟶ $\frac{Q_3 - Q_1}{2}$

=> Mean deviation (MD) ⟶ $\Sigma f|m - \bar{x}| \,/\, \Sigma f$ (divide by)

=> Standard deviation (SD) ⟶ $\sigma = \sqrt{\frac{1}{n}\Sigma x^2 - \left(\frac{\Sigma x}{n}\right)^2}$

=> moments. ⟶

σ^2 = variance (when we square the SD).

Co-eff. of dispersion:

=> co-eff of range ⟶ $\frac{L-S}{L+S}$

=> co-eff of Q.D ⟶ $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

=> Co-eff of MD ⟶ $\frac{MD}{mean}$

=> Co-eff of SD variation ⟶ $\frac{\sigma}{\bar{x}} \times 100$.

(Independent of units)

Combined SD:

$$\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2) + \cdots + n_k(\sigma_k^2 + d_k^2)}{n_1 + n_2 + \cdots + n_k}$$

where $d_i = \bar{x}_i - \bar{x}$

↳ combined mean

$$\frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$$

1. calculate QD and co-eff. of QD for the foll.

QD

CI:	20-30	30-40	40-50	50-60	60-70	70-80	80-90
f:	3	61	132	153	140	51	2

Soln:

$\frac{Q_3 - Q_1}{2}$ →

CI	f	cf
20-30	3	3
30-40 Q_1	61 f	64 cf
40-50	132	196
50-60	153	349 cf
60-70 Q_3	140 f	489
70-80	51	540
80-90	2	542
	542	

$QD = \frac{Q_3 - Q_1}{2}$

$Q_3 = L + \frac{\frac{3N}{4} - cf}{f} \times i$

$\frac{3N}{4} = \frac{3(542)}{4}$

$= 406.5$

Σ the comming no ie; 489

$Q_3 = 60 + \frac{406.5 - 349}{140} \times 10.$

$Q_3 = 64.107$

$Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times i$

$\frac{N}{4} = \frac{542}{4} = 135.5$

$Q_1 = 40 + \frac{135.5 - 64}{132} \times 10$

$40 + 0.5416 \times 10$

$Q_1 = 45.416$

$Q_1 = 45.416$, $Q_3 = 64.107$

$$QD = \frac{Q_3 - Q_1}{2}$$

$$= \frac{64.107 - 45.416}{2}$$

$$= 9.345$$

$$\text{co-eff of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{9.345}{109.523}$$

$$= 0.085 \quad // \quad 0.17068$$

---

2. Find the SD for the following.

Marks:	0-4	4-8	8-12	12-16	16-20	20-24.
No. of st. students	10	12	18	4	5	3.

Soln:-

Whenever we can take the assumed mean but for calculation it is better to take big no.

CI	f	m	$d = \frac{m - A}{4}$	fd	fd^2
0-4	10	2	-3	-30	90
4-8	12	6	-2	-24	48
8-12	18	10	-1	-18	18
12-16	4	14 (A)	0	0	0
16-20	5	18	1	5	5
20-24	3	22	2	6	12
	55			-61	173

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

$$= \sqrt{\frac{173}{55} - \left(\frac{-61}{55}\right)^2} \times 4$$

$= \sqrt{3.145 - 123} \times 4.$

$= 1.3838 \times 4$

$= 5.535$

$$\sigma = \sqrt{\frac{\Sigma fdm^2}{\Sigma f} - \left(\frac{\Sigma fm}{\Sigma f}\right)^2}$$ (or this formula)

3. Find the mean deviation about mean and co-eff of M.D for the foll..

CI:	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
f:	35	45	70	105	90	74	51	30

Soln:

CI	f	m	$d = \frac{m-A}{5}$	fd
20-25	35	22.5	-3	-105
25-30	45	27.5	-2	-90
30-35	70	32.5	-1	-70
35-40	105	37.5	0	0
40-45	90	42.5	1	90
45-50	74	47.5	2	148
50-55	51	52.5	3	158
55-60	30	57.5	4	120
	500			246

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

$= 37.5 + \left(\frac{246}{500}\right) \times 5$ (2.46)

$= 37.5 + 2.46$

$\bar{x} = 39.96$

continue the table

$\lvert m - 39.96 \rvert$	$f\lvert m - 39.96 \rvert$
17.46	611.1
12.46	560.7
7.46	522.2
2.46	258.3
2.54	228.6
7.54	557.96
12.54	639.54
17.54	526.2

$$MD = \frac{\Sigma f \lvert m - 39.96 \rvert}{\Sigma f}$$

$= \frac{3904.6}{500}$

$MD = 7.809$

Co-eff. MD = $\frac{7.809}{39.96}$ $\frac{MD}{mean}$.

$= 0.1954$.

$MD = \frac{\sum f|m - \bar{x}|}{\sum f}$

---

4. An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the foll.

	firm A	firm B
No. of workers n	500 → n_1	600 → n_2
Avg. daily wage $\bar{x}$	₹ 186 → $\bar{x}_1$	₹ 175 → $\bar{x}_2$
variance of wage σ^2	81 → σ_1^2	100 → σ_2^2

$\sigma_1^2 = 59$ (combined mean & SD. we want to find in this)

(i) Which firm A or B has large wage bill?

(ii) Which firm A or B is there greater variability in individual wages? → comb. variance

(iii) calculate the avg. (combined mean) daily wage and variance of the distribution of the wages of all the workers in firm A and firm B together?

Soln:

(ii) $CV(A) = \frac{\sigma_1}{\bar{x}_1} \times 100$

$CV(x) = \frac{\sigma}{\bar{x}} \times 100$.

$= \frac{9}{186} \times 100 = \frac{900}{186}$

$= 4.838$.

$CV(B) = \frac{\sigma_2}{\bar{x}_2} \times 100$

$CV(x) = \frac{\sigma}{\bar{x}} \times 100$

$\sqrt{100} = 10$.

$= \frac{10}{175} \times 100 = \frac{1000}{175} = 5.714$.

$CV(B) > CV(A)$.

$\Rightarrow$ B has got greater variability when compared with A.

(ii) $n_1 = 500$, $n_2 = 600$

$\bar{x}_1 = 186$, $\bar{x}_2 = 175$

Total wages in Firm A $= n_1 \bar{x}_1$

$= 500 \times 186$

$= 93000$

186
53
93000

Total wages in Firm B $= n_2 \bar{x}_2$

$= 600 \times 175$

$= 105,000$

More wage is paid by firm B.

(iii) combined mean $= \dfrac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$

$= \dfrac{93000 + 105000}{500 + 600}$

105000
93000
198000

$= \dfrac{198000}{1100} = 180$

Combined SD $= \dfrac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}$

$= \dfrac{500(186) + 600(175)}{500 + 600}$

$= \dfrac{500(81 + 36) + 600(125)}{1100}$

$= \dfrac{500(117) + 600(125)}{1100} = \dfrac{133500}{1100}$

$= 121.36$

01/11/21.

Recall:

Measures of dispersion:

(1) Range = L - S

(2) Quartile deviation = $\frac{Q_3 - Q_1}{2}$

(3) Mean deviation abt mean $\Big\} = \frac{\Sigma f |x - \bar{x}|}{\Sigma f}$

(4) Standard deviation $\Rightarrow \sigma = \sqrt{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2} \times i$

$\left[\text{where } d = \frac{m - A}{i}\right]$

Variance $= (\sigma^2)$.

Co-eff of dispersion:

(1) co-eff of range $= \frac{L - S}{L + S}$

(2) co-eff of mean deviation $= \frac{\text{M.D of Mean}}{\text{Mean}}$

(3) co-eff of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

(4) co-eff of variation $= \frac{\sigma}{\bar{x}} \times 100$

(variability of the data

Moments:

Quantification of the SD.

The r^{th} moment of a variable x about any pt. $x = A$, denoted by μ_r' is given by

(raw moment) $\mu_r' = \frac{1}{\sum f_i} \sum f_i (x_i - A)^r$

$\mu_1' =$

The r^{th} moment of a variable x about the mean $\bar{x}$, denoted by μ_r is defined as $\mu_r = \frac{1}{\sum f_i}$

$$\mu_r = \frac{1}{\sum f_i} \sum f_i (x_i - \bar{x})^r.$$

Note:-

$$\mu_r = \frac{\sum f_i (x - \bar{x})^r}{\sum f_i}$$

(i) when $r = 0$;

$$\mu_0 = \frac{\sum f_i}{\sum f_i} = 1.$$

(ii) $r = 1$;

$$\mu_1 = \frac{\sum f_i (x - \bar{x})}{\sum f} = \frac{\sum f_i x_i}{\sum f_i} - \bar{x} \frac{\sum f_i}{\sum f_i}$$

$$= \bar{x} - \bar{x} \cdot 1 = 0.$$

(iii) $r = 2$;

$$\mu_2 = \frac{\sum f_i (x - \bar{x})^2}{\sum f_i} = \sigma^2.$$

Relationship b/w M_r and M_r':

Let $d_i = x_i - A$ & $\bar{x} = A + \frac{\sum f_i d_i}{\sum f_i}$

$\bar{x} - A = M_1'$

$= A + M_1'$

$\Rightarrow \bar{x} - A = M_1'$ (1)

$N = \sum f_i$

$\bar{x} - A = M_1'$

$\bar{x} - A = M_1'$

$$M_r' = \frac{1}{N} \sum f_i (x - \bar{x})^r \cdot$$

$$= \frac{1}{N} \sum f_i \{x_i - A + A - \bar{x}\}^r$$

$$= \frac{1}{N} \sum f_i \{d_i - M_1'\}^r$$

$$M_r = \frac{1}{N} \sum_i f_i (d_i - M_1')^r$$

$$= \frac{1}{N} \sum_i f_i \{d_i^r - r_{C_1} d_i^{r-1} + r_{C_2} d_i^{r-2} (M_1')^2 + \ldots + (-1)^r (M_1')^r\}$$

$$= \frac{1}{N} \sum f_i d_i^r - r_{C_1} M_1' \frac{\sum f_i d_i^{r-1}}{N} + r_{C_2} (M_1')^2 \frac{\sum f_i d_i^{r-2}}{N} + \ldots + (-1)^r (M_1')^r$$

$$M_r = M_r' - r_{C_1} M_{r-1}' M_1' + r_{C_2} M_{r-2}' (M_1')^2 + \ldots + (-1)^r (M_1')$$

$$M_r = M_r' - r_{C_1} M_{r-1}' M_1' + r_{C_2} M_{r-2}' (M_1')^2 + \ldots + (-1)^r (M_1')^r.$$

when $r = 2$;

$$M_2 = M_2' - 2 C_1 M_1' M_2' + (-1)^2 (M_1')^2 = M_2' - (M_1')^2$$

(variance)

when $r=3$;

$$\mu_3 = \mu_3' - 3C_1\mu_2'\mu_1' + 3C_2\mu_1'(\mu_1')^2 - {}_3C_3(\mu_1')^3$$

$$\mu_3 \Rightarrow \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3.$$

when $r=4$;

$$\mu_4 = \mu_4' - 4C_1\mu_3'\mu_1' + 4C_2\mu_2'(\mu_1')^2 - ({}_4C_3)\mu_1'(\mu_1')^3 + (-1)^4(\mu_1')^4$$

$$= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1' - 3(\mu_1')^4.$$

$$\mu_1 = \frac{\sum f(x-\bar{x})}{\sum f} = \frac{\sum fx}{\sum f} - \bar{x} - \bar{x} = 0$$

$$\boxed{\mu_1 = \mu_1' - \mu_1' = 0} \rightarrow (1)$$

Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$ {Nature of the data}.

Kurtosis $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$ {Flatness of the curve}

For a symmetrical distribution:

$$\boxed{\beta_2 = 3}$$

---

① Raw moments

② central moments

③ Relationship blw raw and central moments

④ Particular cases in central moments.

μ_2 = variance $= \mu_2' - (\mu_1')^2$

μ_3 & μ_4.

⑤ $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

① The first four moments about the value $x=4$ are $-1.5, 17, -30, 108$. Find skewness & kurtosis.

Soln:- $\mu_1' \quad \mu_2' \quad \mu_3' \quad \mu_4'$

Gn: $\mu_1' = -1.5, \mu_2' = 17, \mu_3' = -30, \mu_4' = 108$.

To find: $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ & $\beta_2 = \frac{\mu_4}{\mu_2^2}$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

$\beta = \frac{\mu_4}{\mu_2^2}$

$\mu_2 = \mu_2' - (\mu_1')^2$

$= 17 - (-1.5)^2$

$= 17 - 2.25$

$= 14.75$.

$\mu_3 = \mu_3' - 3\mu_2'(\mu_1')^2 + 2(\mu_1')^3$

$= -30 - 3(17)(-1.5)^2 + 2(-1.5)^3$.

$= -30 + 51(1.5) - 2(1.5)^3$

$= -30 + 76.5 - 6.75$

$= 39.75$

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{39.75^2}{(14.75)^3}$ Skewness.

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$.

$= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4$

$= 108 - 120(1.5) + 6(17)(2.25) - 3(2.25)(2.25)$

$= 142.3125$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2}$. Kurtosis

$= 0.6541$

Note:

$\bar{x} = A + M_1' = 4 + (-1.5) = 2.5$

$\boxed{\bar{x} = 2.5}$

How will you find moments abt $x=0$?

$\bar{x} = A + M_1'$

Here assume $A=0$,

$\Rightarrow \bar{x} = M_1' = 2.5$.

$\bar{x} = A + M_1'$

$\therefore M_1' = \bar{x}$

$M_2 = M_2' - (M_1')^2$

$\Rightarrow M_2' = M_2 + (M_1')^2 = 14.75 + (2.5)^2$

$= 14.75 + 6.25$

$= 21$ //

2. Find the 1st 4 moments for the following.

x:	0	1	2	3	4	5	6	7	8
f:	1	8	28	56	70	56	28	8	1

Soln:

x	f	$d = x - A$	fd	fd^3	fd^2	fd^4
0	1	-4	-4	-64	16	256
1	8	-3	-24	-216	72	648
2	28	-2	-56	-224	112	448
3	56	-1	-56	-56	56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	224	112	448
7	8	3	24	216	72	648
8	1	4	4	64	16	256

$\sum fd = 0$.

$\sum fd^2 = 512$

$\sum fd^3 = 0$.

$\sum fd^4 = 2816$

$$\bar{x} = A + \frac{\sum fd}{\sum f} = 4 + \frac{0}{\sum f} = 4.$$

$M_1 = 0$

$$M_2 = \frac{\sum fd^2}{\sum f} = \frac{512}{256} = 2$$

$$M_3 = \frac{\sum fd^3}{\sum f} = \frac{0}{256} = 0$$

$$M_4 = \frac{\sum fd^4}{\sum f} = \frac{2816}{256} = 11$$

2/11/21

correlation.

Amount of relationship b/w two variables.

linear correlation:

A relationship b/w two variables

x	↑	↓	↑	↓
y	↑	↓	↓	↑

Multiple correlation & partial correlation

$r_{12.3}$ -> partial

$R_{12.3}$ -> Multiple

Karl Pearson's co-efficient of correlation r or ρ (rho).

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2}\sqrt{N\sum y^2 - (\sum y)^2}}$$

Individual (correlation formula)

$r = \frac{\text{covariance}(x, y)}{\sigma_x \sigma_y}$ i.e., $r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$

SD of x & y

If x and y are independent then it is zero.

When; $r = 0$ x and y are independent.

Properties. (xi)

Properties of correlation co.eff.

1. r lies blw -1 & 1 ie; $-1 \le r \le 1$
2. correlation is unaffected by the change of scale & and origin. (Reduce big to smaller no's)

* Individual observations:

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

* Discute formulas:

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

$x = \Sigma f$, $dx = \frac{x-A}{h}$, $dy = \frac{y-B}{K}$

individual

1. Calculate correlation co-eff.

x: 65 66 67 67 68 69 70 72.

y: 67 68 65 68 72 72 69 71.

Soln:-

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

(or)

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

$\rightarrow dx = \frac{x-A}{h}$

$\rightarrow dy = \frac{y-B}{K}$

x	y	$dx = x-68$	$dy = y-69$	dx^2	dy^2	$dxdy$
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
		0	0	36	14	24

dx = Total Add all Total no

$N = 8$

$\Sigma dx = 0 = \Sigma dy$

$\Sigma dx^2 = 36$

$\Sigma dy^2 = 44$

$\Sigma dxdy = 24$

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

$$= \frac{8(24) - 0 \times 0}{\sqrt{8(36) - 0^2}\sqrt{8(44) - 0^2}}$$

$$= \frac{192}{8\sqrt{8 \times 352}} = \frac{192}{8\sqrt{2816}}^{32}$$

$$= \frac{32}{\sqrt{2816}} = 0.603 //.$$

2. The following tables gives the number of third per lakh of population in different age groups, Find out the correlation blw age and blindness.

Age:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of blidness.	55	67	100	111	150	200	300	500.

Soln:

Age	m	fy	$dx = \frac{m-45}{10}$	fdx^2	$fdy = fy - 150$	dy^2	dxdy
0-10	5	55	-4	16	-95	9025	380
10-20	15	67	-3	9	-83	6889	249
20-30	25	100	-2	4	-50	2500	100
30-40	35	111	-1	1	-49	2401	49
40-50	45	150	0	0	0	0	0
50-60	55	200	1	1	50	2500	50
60-70	65	300	2	4	150	22500	300
70-80	75	500	3	9	350	122500	1050
			-4	44	273	168315	2178

$N = 8, \sum dx = -4, \sum dx^2 = 44, \sum dy = 273, \sum dy^2 = 168315$

$\sum dxdy = 2178$

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

$$= \frac{8(2178) - (-4)(273)}{\sqrt{8(44) - (-4)^2}\sqrt{8(168315) - (273)^2}}$$

$$= \frac{18516}{\sqrt{336}\sqrt{1271991}}$$

$$= \frac{18516}{18.333 \times 1127.82}$$

$$= 0.8956$$

09/11/21

The following table gives the marks obtained by 100 students. Find the co-eff. of correlation.

Age / Marital	18	19	20	21	Total.
10-20	4	2	2	-	8
20-30	5	4	6	4	19
30-40	6	8	10	11	35
40-50	4	4	6	8	22
50-60	2	2	4	4	10
60-70	-	2	3	1	6
Total	19	22	31	28	100.

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2}\sqrt{N\sum y^2 - (\sum y)^2}}$$

(cov(x, y) ↑)

↓ SD(x) ↓ S.D(y)

Properties:

(i) $-1 \le r \le 1$

(ii) Co-eff of correlation is unaffected change of scale and origin.

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

$$r = \frac{N\sum f\,dx\,dy - \sum f\,dx \sum f\,dy}{\sqrt{N\sum f\,dx^2 - (\sum f\,dx)^2}\sqrt{N\sum f\,dy^2 - (\sum f\,dy)^2}}$$

where, No. of observation
$N = \sum f$
$dx = \frac{x - A}{h}$
$dy = \frac{y - B}{k}$

where;
$N = \sum f$
$dx = \frac{y - A}{h}$
$dy = \frac{y - B}{k}$

X / dx / y / dy	18	19	20	21	Total
	-2	-1	0	1	
10-20	4	2	2	0	8
20-30	5	4	6	4	19
30-40	6	8	10	11	35
40-50	4	4	6	8	22
50-60	0	2	4	4	10
60-70	0	2	3	1	6
f	19	22	31	28	120

$dx = x - 20$

$dy = \frac{y - 45}{10}$

x / y		18	19	20	21	f	fdy	fdy^2	fdxdy
	dx	-2	-1	0	1	-			
	dy	4	2	2	0	8	-24		
10-20	-3	4_{24}	2_6	2_0	0_0	8	-24	72	30
20-30	-2	5_{20}	4_8	6_0	4_{-8}	19	-38	76	20
30-40	-1	6_{12}	8_8	10_0	11_{-11}	35	-35	35	9
40-50	0	4_0	4_6	6_0	8_0	22	0	0	0
50-60	1	0_0	2_{-2}	4_0	4_4	10	10	10	2
60-70	2	0_0	2_{-4}	3_0	1_2	6	12	24	-2
f	-	19	22	31	28	100	*	*	59
	fdx	-38	-22	0	28				
	fdx^2	76	72	0	28				
	fdxdy	56	16	0	-13	59			

$dx = x - 20$

$dy = \frac{y - 45}{10}$

$\Sigma f dx dy = 59$

$\Sigma f dx = -32$

$\Sigma f dy = -75$

$\Sigma f dx^2 = 126$

$\Sigma f dy^2 = 217$

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

$$= \frac{100(59) - (-32)(-75)}{\sqrt{100(126) - (-32)^2}\sqrt{100(217) - (-75)^2}}$$

$$= 0.2566 //$$

2. A computer while calculating co-eff of correlation blw x and y from 25 pairs of observations obtained the following.

$\Sigma xy = 508, n = 25; \Sigma x = 125, \Sigma x^2 = 650, \Sigma y = 100, \Sigma y^2 = 460$

* It was later discovered that at the time of checking he had copied two pairs wrongly and whereas correct values are

x	y
6	14
9	6

x	8	6
y	6	8

obtain correct correlation co-eff

x	y
8	6
6	8

Soln:

$$r = \frac{N\Sigma' yx - \Sigma' x \Sigma' y}{\sqrt{N\Sigma' x^2 - (\Sigma' x)^2}\sqrt{N\Sigma' y^2 - (\Sigma' y)^2}}$$

corrected $\Sigma' x = 125 - 6 - 9 + 8 + 6 = 124$

corrected $\Sigma' y = 100 - 14 - 6 + 6 + 8 = 94$.

corrected $\Sigma' x^2 = 650 - 36 - 81 + 64 + 36 = 633$.

corrected $\Sigma' y^2 = 460 - 196 - 36 + 36 + 64 = 328$

corrected $\Sigma' xy = 508 - 6 \times 14 - 9 \times 6 + 6 \times 8 + 8 \times 6$

$= 466$.

$$\text{corrected } r = \frac{25(466) - (124)(94)}{\sqrt{25(633) - (124)^2}\sqrt{25(328) - (94)^2}}$$

(-)ve

Rank correlation:- (Spearman's correlation (co-eff)).

Data is qn. in according to two different characteristics if we require co-eff of correlation, we use rank correlation.

X: → Data arranged according to heigh of the student

Y: → Data arranged according to marks.

We 1st assign ranks and find the correlation

$$\rho = 1 - \left[\frac{6\sum D^2}{N(N^2-1)}\right]$$

(Rho)

Rank x, Rank y

$D = R_x - R_y$

N → No. of observations

For repeated ranks;

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m^3-m) + \dots\right\}}{N(N^2-1)}$$

m = no. of times a particular rank gets repeated

⇒ If rank '3' is repeated then take 'm' as '3'

⇒ If rank '5' is repeated then take 'm' as '5'

Example: Find the rank correlation co-eff of the following

X:	65	67	63	62	65
Y:	19	18	23	18	17

Soln:

X	Y	R_x	R_y	$D = R_x - R_y$	D^2
65	19	2.5	2	-0.5	0.25
67	18	1	3.5	-2.5	6.25
63	23	4	1	3	9.00
62	18	5	3.5	1.5	2.25
65	17	2.5	5	-2.5	6.25
					24.00

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right\}}{N(N^2-1)}$$

$$= 1 - \frac{6\left\{24 + \frac{1}{12}(8-2) + \frac{1}{12}(2-2)\right\}}{5(25-1)}$$

$$= 1 - \frac{6\left\{24 + \frac{1}{2} + \frac{1}{2}\right\}5}{5\times 24}$$

$= 1 - \frac{30}{24} \frac{5}{4} = -\frac{1}{4}$

$= -0.25$

10/11/21

Rank correlation:

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \cdots\right\}}{N(N^2-1)}$$

$D = R_x - R_y$

$m \rightarrow$ no. of times a particular rank gets repeated

Quality of two observations:-

$x \rightarrow$ Height of the students

$y \rightarrow$ Marks scored by the students.

---

10 Ten competitions in a beauty contest were ranked by 3 judges:

R_x:	4	2	8	6	1	5	3	9	10	7
R_y:	2	5	9	3	6	7	1	10	8	4
R_z:	4	3	6	9	2	8	7	5	1	10.

Which pair of judges has the nearest approach to common taste?

Soln:

R_x	R_y	R_z	$D_1 = R_x - R_y$	D_1^2	$D_2 = R_x - R_z$	D_2^2	$D_3 = R_y - R_z$	D_3^2
4	2	4	+2	4	0	0	-2	4
2	5	3	-3	9	-1	1	2	4
8	9	6	-1	1	2	4	3	9
6	3	9	3	9	-3	9	-6	36
1	6	2	-5	25	-1	1	4	16
5	7	8	-2	4	-3	9	-1	1
3	1	7	2	4	-4	16	-6	36
9	10	5	-1	1	4	16	5	25
10	8	1	2	4	9	81	7	49
7	4	10	-3	9	-3	9	-6	36

$\sum D_1^2 = 70$

$\sum D_2^2 = 146$

$\sum D_3^2 = 216$

$$P_{xy} = 1 - \frac{6\sum D_1^2}{N(N^2-1)}$$

$$= 1 - \frac{6(70)}{10(99)} = 0.542$$

$$P_{xz} = 1 - \frac{6\sum D_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6(146)}{10(99)} = 0.115$$

$$P_{yz} = 1 - \frac{6\sum D_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6(216)}{10(99)} = -0.309.$$

$P_{xy} = 0.542$ → Judges x & y have the common taste

$P_{yz} = -0.309$

$P_{xz} = 0.115$

---

2. Find the rank correlation co-eff, given

(x) Marks in math: 48 57 82 65 39 78 26 47 50.85 48 57

(y) Marks in stats: 56 43 80 45 79 50 46 72 65 71 32 59.

R_x	R_y	$D = R_x - R_y$	D^2
8.5	7	1.5	2.25
5.5	11	5.5	30.25
2	1	1	1.00
4	10	-6	36
11	2	9	81
3	8	-5	25
12	9	3	9
10	3	7	49
7	5	2	4
1	4	-3	9
9.5	12	-3.5	12.25
	6	0.5	0.25

$\sum D^2 = 259$

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m^3-m) + \dots\right\}}{N(N^2-1)}$$

$$= 1 - \frac{6\left[259 + 1/12(2^3-2) + \frac{1}{2}(2^3-2)\right]}{130(12(143))}$$

$$= 1 - \frac{6(259+1)}{12(143)}$$

$= 1 - \frac{130}{143}$

$= \frac{13}{143}$

Regression lines ::

Approx value for the gn. variable.

* Regression line of x on y

(x depends on y)

$(x - \bar{x}) = b_{xy}(y - \bar{y})$

where; $b_{xy} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2}$

$\Rightarrow r \cdot \frac{\sigma_x}{\sigma_y}$ (standard deviation, SD)

b_{xy} also called as the regression of co-eff x on y.

* Regression line of y on x:

(y depends on x)

$(y - \bar{y}) = b_{yx}(x - \bar{x})$

where; $b_{yx} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2}$

$\Rightarrow r \cdot \frac{\sigma_y}{\sigma_x}$

x, y & z

↓ → $r_{12.3}$

Partial correlation (keep one variable constant)

Multiple correlation (effect of one variable over other variables)

It is denoted by $R_{1.23}$

Property :: (Regression co-eff)

(1) $r^2 = b_{xy} b_{yx}$

ie, when we multiply the regression co-eff then automatically square of the correlation co-eff

NOTE:

If both b_{xy} and b_{yx} are (-)ve the r is -ve.

Eg: $b_{xy} = -0.5$ & $b_{yx} = \frac{-1}{0.5}$

$$r^2 = (-0.5)\left(\frac{-1}{0.5}\right) = 1$$

$$\Rightarrow r = -1$$

Eg:

Calculate regression lines for the following.

x:	1	2	3	4	5
y:	10	12	7	3	8

Soln:-

	x	y	x^2	y^2	xy
	1	10	1	100	10
	2	12	4	144	24
	3	7	9	49	21
	4	3	16	9	12
	5	8	25	64	40
Total	15	40	55	366	107

$$b_{xy} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{5(107) - (15)(40)}{5(55) - (15)^2}$$

$$= -1.3$$

$$b_{yx} = \frac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{5(107) - (15)(40)}{5(366) - (40)^2}$$

$$= -0.283.$$

$b_{xy} = -1.3$

$b_{yx} = -0.283.$

$$\bar{x} = \frac{\Sigma x}{N} = \frac{15}{5} = 3; \quad \frac{\Sigma y}{N} = \bar{y} = \frac{40}{5} = 8.$$

⇒ x on y (x depends on y)

$x - \bar{x} = b_{xy}(y - \bar{y})$

$x - 3 = -0.283(y - 8)$

$x = 3 - 0.283y + 8(0.283)$

$x = 5.264 - 0.283y$

$\boxed{x = 5.264 - 0.283y}$

⇒ y on x (y depends on x)

$y - \bar{y} = b_{yx}(x - \bar{x})$

$y - 8 = -1.3(x - 3)$

$y = 8 - 1.3x + 3.9$

$\boxed{y = 11.9 - 1.3x}$

$x = 5.264 - 0.283y$

$y = 11.9 - 1.3x$

x on y ⇒ $x = 5.264 - 0.283y$.

y on x ⇒ $y = 11.9 - 1.3x$.

given x, to find y, use y on x
given y, to find x, use x on y.

Note.

Eg:

- find y, given $x = 6$

y on x is $y = 11.9 - 1.3x$

$\Rightarrow y(6) = 11.9 - 1.3(6)$

$= 11.9 - 7.8 = 4.1$.

$x = 6, y = 4.1$

- find x, given $y = 6$.

x on y is $x = 5.264 - 0.283y$

$\Rightarrow x(6) = 5.264 - 0.283(6)$

$= 3.566$.

y on x is $11.9 - 1.3x$.

$x = 1$ $\quad y = 11.9 - 1.3(1)$

$X=1$

$= 11.9 - 1.3$

$= 10.6$

$$\begin{array}{r} 11.9 \\ 1.3 \\ \hline 10.6 \end{array}$$

$\boxed{X=1,\ Y=10.6}$

11/11/21

1. For 2 variables x and y the equations of the regression lines are $9y - x - 288 = 0$ and $x - 4y + 38 = 0$. Find

(i) mean values of x & y.

(ii) Co-eff of ~~correlation~~ regression

(iii) The ratio of the SD. of y to that of x.

(iv) Most probable value of y when $x = 145$

(v) Most probable value of x when $y = 35$.

(i) $x - \bar{x} = b_{xy}(y - \bar{y})$

$y - \bar{y} = b_{yx}(x - \bar{x})$ } use this

Soln:

(1) $x - \bar{x} = b_{xy}(y - \bar{y})$

$y - \bar{y} = b_{yx}(x - \bar{x})$

$-x + 9y - 288 = 0$

$x - 4y + 38 = 0$

$5y - 250 = 0$

$5y = 250$

$y = \frac{250}{5}$

$\boxed{\bar{y} = 50}$

$-x + 9y - 288 = 0$

$-x + 9(50) - 288 = 0$

$-x + 450 - 288 = 0$

$x = -288 + 9y.$

$-x = -288 - 450$

$\boxed{\bar{x} = 172}$

(11) Let $9y - x - 288 = 0$ represent regression line of x on y.

$$x = 9y - 288$$

$$x = 9\left(y - \frac{288}{9}\right)$$

$$x = 9.$$

$$b_{xy} = 9.$$

The line $x - 4y + 38 = 0$ is ~~regression~~ regression line of y on x

$$4y = x + 38$$

$$y = \frac{1}{4}(x + 38)$$

$$y = \frac{1}{4}$$

$$b_{yx} = \frac{1}{4}$$

$$r^2 = b_{xy} b_{yx}$$

$$= 9 \times \frac{1}{4}$$

$$r = \frac{3}{2} > 1. \text{ (impossible)}$$

$\therefore$ $9y - x - 288 = 0$ is regression line of x on y

$x - 4y + 38 = 0$ is regression line of y on x

$x = 9y - 288 = 9\left(y - \frac{288}{9}\right)$ $b_{xy} = 9$

$+4y = x + 38 \Rightarrow y = \frac{1}{4}(x + 36)$ $b_{yx} = 1$

$9y = x + 288 \Rightarrow y = \frac{1}{9}(x + 288) \Rightarrow b_{yx} = \frac{1}{9}$

$\therefore x = 4y - 38 \Rightarrow x = 4\left(y - \frac{38}{4}\right) \Rightarrow b_{xy} = 4$

$r^2 = b_{xy} \times b_{yx} = \frac{4}{9} \Rightarrow r = 2/3$

(iii) $\frac{\sigma_y}{\sigma_x} = ?$

$b_{yx} = \frac{1}{9}$

$\frac{r\sigma_y}{\sigma_x} = \frac{1}{9}$

$\frac{\sigma_y}{\sigma_x} = \frac{1}{9} r = \frac{1}{9 \times \frac{2}{3}} = \frac{1}{6}$

$\frac{\sigma_y}{\sigma_x} = \frac{1}{6}$

(iv) Regression line of y on x is $9y - x - 297 = 0$

when $x = 145$

$9y - 145 - 297 = 0$

$9y = 297 + 145$

$y = \frac{433}{9}$

(v) Regression line of x on y is $4x - 4y + 38 = 0$

when $y = 35$

$x = 4y - 38$

$x = 4(35) - 38$

$\Rightarrow 140 - 38$

$= 102$

2. A panel of judges A and B graded seven detectors and independently awarded the marks. (Regression Y on X)

Judge A: 40 34 28 30 44 38 31 36

Judge B: 32 39 26 30 38 34 29 ?

Eighth debator was awarded 36 marks by judge A while judge B was not present. If judge B was also present, how many marks would be allot for the eighth debator.

Soln:-

Judge A = x, Judge B = Y.

To find regression line of y on x

then substitute x = 36 in the above to find y.

X	Y	$dx = x - 35$	$dy = y - 32$	$dx\,dy$	dx^2	dy^2
40	32	5	0	0	25	0
34	39	-1	7	-7	1	49
28	26	-7	-6	42	49	36
36	13	-5	-2	10	25	4
44	38	9	6	54	81	36
38	34	3	2	6	9	4
31	28	-4	-4	16	16	16
		0	3	121	206	145

$$Y = A + \frac{\xi dy}{n}$$

$$= 32 + \frac{3}{7}$$

$$= 32 + 0.4$$

$$= 32.4 //$$

$\bar{x} = 35$

$$byx = \frac{N\xi' dx dy - \xi' dx \cdot \xi' dy}{N\xi dx^2 - (\xi' dx)^2}$$

$$= \frac{7(121)}{7(206)} = 0.5873 //$$

Regression line of y on x is

$$y - \bar{y} = byx\,(x - \bar{x})$$

$$y - 32.4 = 0.5873x - 0.5873(35)$$

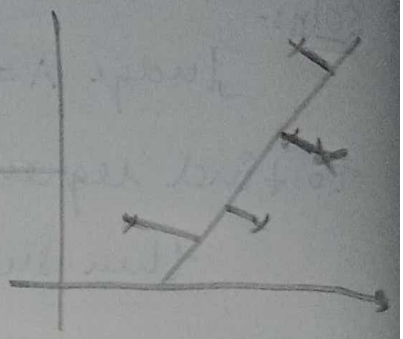
$$+32.4$$

$$y = 0.5873x + 32.4 - 0.5873(35)$$

Given $x = 36 \Rightarrow y = 0.5873(36) + 32.4 - 20.55$

$$= 53.54 - 20.55$$

$$= 32.99 //$$



Method of least squares:

★ Fit $y = a + bx$

Normal eqns $\begin{cases} \Sigma y = na + b\Sigma x \\ \Sigma xy = a\Sigma x + b\Sigma x^2 \end{cases}$ → (Fitting a line)

★ Fit $y = a + bx + cx^2$.

Normal eqn. $\begin{cases} \Sigma y = na + b\Sigma x + c\Sigma x^2 \\ \Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3. \\ \Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4. \end{cases}$ → (Fitting a parabola)

★ Fit $y = ae^{bx}$ (exponential curve).

$$y = ae^{bx}$$

$$\log_e y = \log_e (ae^{bx})$$

$$\log_e y = \log_e a + \log_e e^{bx}$$

$$\log_e y = A + bx \qquad (A = \log_e a)$$

$$\alpha = A + bx \qquad \alpha = \log_e y.$$

Normal eqn $\begin{cases} \Sigma \alpha = nA + b\Sigma x \\ \Sigma \alpha x = A\Sigma x + b\Sigma x^2 \end{cases}$