Date:- / /
Page:-

UNIT-3

Statistical Technique I :-

overview of measures of central tendency, Moments, Skewness, Kurtosis, Curve fitting, Method of least squares, fitting of straight lines, fitting of second degree parabola, exponential curves, correlation and Rank correlation, Regression Analysis : Regression lines of y on x and x on y.

Statistics :- Statistics is the Science which deals with methods of collecting, classifying, presenting, comparing and Interpreting numerical data collected to throw light on any sphere of enquiry.

Variable (variate) :-

A quantity which can vary from one individual to another is called a variable / variate.
eg. heights, weights, age, Rainfall records of cities.

Quantities which can take any numerical value within a certain range are called continuous variables e.g. as the child grows, his / her height takes all possible values from 50 cm to 100 cm.

Quantities which are incapable of taking all possible values are called discrete or discontinuous variables.

Date:- 11
Page:-

(#) Measure of Central Tendency :→
or Averages.

A figure which is used to represent a whole (data) Series should neither have the lowest value nor the highest in the Series, but a value somewhere between these two limits, possibly in the centre, where most of the items of the Series cluster.
such figures are called 'measure of central tendency.' (or averages).

"Averages are Statistical Constants which enable us to comprehend in a single effort the significance of the whole."

There are five types of averages

Average

Mathematical Average
1. Arithmetic mean / mean
2. Geometric mean
3. Harmonic Mean

Positional Average
1. Median
2. Mode.

Arithmetic Mean can be found for
Individual series (i.e where frequency is not given)
For discrete Series
For Continuous Series.

Date:- / /
Page:-

Arithmatic Mean :→

I) For Individual Series / ungrouped data.

if $x_1, x_2, \ldots, x_n$ are n variables Then

1) Direct Method :-

$$\text{Mean } (\bar{x}) = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\Sigma x}{n}$$

2) Short cut method :- This is shift of origin shifting the origin to an arbitrary point a, the formula $\bar{x} = \frac{\Sigma x}{a}$ becomes

$$\bar{x} - a = \frac{1}{n}\Sigma(x-a)$$

$$\bar{x} = a + \frac{1}{n}\Sigma(x-a)$$

$$\bar{x} = a + \frac{1}{n}\Sigma d_x \quad , \text{ where } d_x = x - a$$

a is called assumed mean : ↓ deviation of x from a

and n is the no. of observations.

3) step deviation :-

$$\bar{x} = a + \frac{c}{n}\Sigma d_x' \quad , \; d_x' = \frac{x-a}{c}$$

$$= a + \frac{\Sigma d_x'}{n} \times c$$

Where c is common value.

Date:- / /
Page:-

Ex:- Daily income of 10 families is given as follows:

S.No	1	2	3	4	5	6	7	8	9	10
Daily Income (₹)	100	120	80	85	95	130	200	250	225	275

find Average daily Income.

Ans:-

S.No	Daily income (in Rs) (x)	$d_x = x - a$	$d_x' = \frac{d_x}{c}$
1	100	20	2
2	120	40	4
3	80	0	0
4	85	5	0.5
5	95	15	1.5
6	130	50	5
7	200	120	12
8	250	170	17
9	225	145	14.5
10	275	195	19.5
	$\Sigma x = 1560$	$\Sigma d_x = 760$	$\Sigma d_x' = 76$

By using Direct Method -

Here $n = 10$

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

$$= \frac{1560}{10}$$

$$= 156$$

Hence average daily income is Rs. 156

By using short cut method -

Let assumed mean $a = 80$

$$\bar{x} = a + \frac{\Sigma d_x}{n} = 80 + \frac{760}{10}$$

$$= 80 + 76 = 156$$

Date:- / /
Page:-

step deviation —

let $c = 10$

Then

Mean $\bar{x} = a + \frac{\Sigma d_x'}{n} \times c$

$= 80 + \frac{76}{10} \times 10$

$= 80 + 76$

$= 156$

Asithmetic Mean for Discrete Series :—

1.) Direct Method :— If the frequency distribution is

$x : x_1, x_2, \ldots, x_n$

$f : f_1, f_2, \ldots, f_n$ Then

Mean $\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n}$

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

N = Sum of frequencies $= \Sigma f$

2.) Short cut Method :—

$$\boxed{\bar{x} = a + \frac{1}{N}\Sigma f d_x}$$

$d_x = x - a$

a = assumed mean

$N = \Sigma f$

3) step deviation Method :—

$$\boxed{\bar{x} = a + \frac{\Sigma f d_x'}{N} \times c}$$

$d_x' = \frac{x-a}{c}$

c = step deviation

Date:- / /
Page:-

Ex:- Calculate the mean of the following frequency distribution of marks in a test in Mathematics:

marks	10	20	30	40	50	60	70	80
No. of Students	3	6	10	12	9	6	2	2

Sol^n:-

Marks x	No. of students f	fx	$d_x = x-a$	fd_x	$d'_x = \frac{x-a}{h}$	fd'_x
10	3	30	-30	-90	-3	-9
20	6	120	-20	-120	-2	-12
30	10	300	-10	-100	-1	-10
40	12	480	0	0	0	0
50	9	450	10	90	1	9
60	6	360	20	120	2	12
70	2	140	30	60	3	6
80	2	160	40	80	4	8
	$\Sigma f = 50$	$\Sigma fx = 2040$		$\Sigma fd_x = 40$		$\Sigma fd'_x = 4$

By Direct Method:-

A.M. $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{2040}{50} = \frac{204}{5} = 40.8$

By Shortcut Method :- $a = 40$

$$\bar{x} = a + \frac{\Sigma f d_x}{\Sigma f} = 40 + \frac{40}{50}$$

$$= 40 + .8 = 40.8$$

Date:- / /
Page:-

By Step deviation Method :-

deviation $h = 10$

$$\bar{x} = a + \frac{\Sigma f d_x'}{\Sigma f} \times h$$

$$= 40 + \frac{4 \times 10}{50} = 40 + 0.8$$

$$= 40.8.$$

Arithmetic Mean for Continuous :-

1.) Direct Method :- $\boxed{\bar{x} = \frac{\Sigma fm}{N} = \frac{\Sigma fm}{\Sigma f}}$

2.) Short cut Method :- $\boxed{\bar{x} = a + \frac{\Sigma f d_x}{N}}$

$d_x = x - a$

3.) Step deviation Method :-

$$\boxed{\bar{x} = a + \frac{\Sigma f d_x'}{N} \times h}$$

$d_x' = \frac{x - a}{h}$

Where a is assumed Mean
m is the mid pt of various class
h is step deviation.

Date:- / /
Page:-

Ex:- Find the Arithmetic Mean for the follow data.

Marks	No. of students
0-10	5
10-20	10
20-30	40
30-40	20
40-50	25

Sol^n:-

marks	mid-values m	No of students f	fxm	d	fd'
0-10	5	5	25	-20	-100
10-20	15	10	150	-10	-100
20-30	25	40	1000	0	0
30-40	35	20	700	10	200
40-50	45	25	1125	20	500
		N = Σf = 100	Σfm = 3000		Σfd = 500

By D.M

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

$$= \frac{3000}{100}$$

$$= 30$$

Short cut Method :- $a = 25$

$$\bar{x} = a + \frac{\Sigma fd}{N}$$

$$= 25 + \frac{500}{100} = \underline{30}$$

step

d'	fd'
-2	-70
-1	-10
0	0
1	20
2	50
	Σfd' = 50

(#) Mat

1. The
the

Proof:-

Then

Date:- / /
Page:-

Step deviation Method :-

here $h = 10$

Then

$$\bar{x} = a + \frac{\Sigma fd'}{N} \times h$$

$$= 25 + \frac{50}{100} \times 10$$

$$= 25 + 5$$

$$= 30.$$

d'	fd'
-2	-40
-1	-10
0	0
1	20
2	50
	$\Sigma fd' = 50$

(#) Mathematical Properties of Arithmetic Mean :-

1. The Total of the deviations of the items from the mean is equal to zero.

Proof :-

Let $x_1, x_2, \ldots, x_n \} \rightarrow x_i$, $i = 1, 2, \ldots, n$

Then

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$$

$$= \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} 1 \qquad \because \sum_{i=1}^{n} 1 = n$$

$$= \sum_{i=1}^{n} x_i - \bar{x}(n)$$

$$= \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - n \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= 0$$

Date:- / /
Page:-

2. Combined Mean of two groups :-

Let $\bar{X}_1$ = mean of first series
$\bar{X}_2$ = Mean of second series
N_1 = No. of observation in first group
N_2 = " " in second group
Then combined mean of two group

$$\bar{X}_{1,2} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

Ex:- A cooperative bank has two branches employing 50 & 70 workers respectively The Average salaries paid by two respective branches are Rs. 360 and Rs 390 per month Calculate the mean of the salaries of all the employes.

Soln:-

Mean of the salaries of all the employes

$$\bar{X}_{1,2} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$= \frac{50 \times 360 + 70 \times 390}{50 + 70}$$

$$= \frac{45300}{120}$$

$$= 377.5$$

Date:- / /
Page:-

Que:1 The Mean of 200 items was 50. Later on it was discovered that two items were misread as 92 and 8 instead of 192 and 88. Find correct mean.

Sol:-

Here incorrect value of $\bar{x} = 50$, $n = 200$

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

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

Using incorrect value of $\bar{x}$

$\Sigma x = 200 \times 50 = 10,000$

$\therefore$ Corrected Value of $\Sigma x = 10000 - (92+8) + (192+88)$

$= 10000 - 100 + 280$

$= 10180$

$$\text{Corrected Mean} = \frac{\text{Corrected } \Sigma x}{n}$$

$$= \frac{10180}{200}$$

$$= 50.9$$

Weighted Arithmetic Mean :-

If the variate-values are not of equal importance, we may attach to them 'weight' $w_1, w_2, \ldots, w_n$ as measures of their importance.

The weighted mean $\bar{x}_w$ is defined as

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\Sigma wx}{\Sigma w}$$

Date:- / /
Page:-

Ex:- calculate weighted mean by weighting each price by the quantity consumed.

Soln:-

Articles of food	Quantity consumed in kg (W)	Price in Rs (per kg) (X)
Flour	11.50	5.6
Ghee	5.60	58.4
Sugar	0.28	8.2
Potato	0.16	2.5
Oil	0.35	20.0
	$\Sigma W = 17.89$	

Soln:-

WX
66.700
327.04
2.296
0.400
7.000
$\Sigma WX = 403.436$

$$\bar{X}_w = \frac{\Sigma WX}{\Sigma W}$$

$$= \frac{403.436}{17.89}$$

$$= 22.55$$

weighted Mean price.

Date:- / /
Page:-

Ques 2 Compute the arithmetic mean for the following data:

class Height (in cm)	219	216	213	210	207	204	201	198	195
No. of persons	2	4	6	10	11	7	5	4	1

Sol^{n}:-

Height in (cm) x	No. of Persons f	fx
219	2	438
216	4	864
213	6	1278
210	10	2100
207	11	2277
204	7	1428
201	5	1005
198	4	792
195	1	195
	$\Sigma f = 40$	$\Sigma fx = 10,377$

Arithmetic Mean By Direct Method

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

$$= \frac{10,377}{40}$$

$$= 259.425$$

Date:- / /
Page:-

Geometric Mean :—

(a) Geometric mean for Individual series →

Geometric mean (G.M) of n individual observations $x_1, x_2, \ldots, x_n$ $(x_i \neq 0)$ is the n^{th} root of their product.

Thus $$\boxed{G = (x_1 \cdot x_2 \cdots x_n)^{1/n}}$$

after taking log on both sides, we get

$$G = \text{antilog}\left[\frac{1}{n}\sum_{i=1}^{n} \log x_i\right]$$

(b) Geometric mean for discrete series →

If $x_1, x_2, \ldots, x_n$ occur $f_1, f_2, \ldots, f_n$ times resp. and $N = \sum_{i=1}^{n} f_i$, then G.M is given by

$$G.M,\ G = (x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n})^{1/N}$$

taking logarithm on both sides

$$\log G = \log (x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n})^{1/N}$$

$$= \frac{1}{N}[f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n]$$

$$\boxed{G = \text{antilog}\left[\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i\right]}$$

Date:- / /
Page:-

(c) G.M for continuous Series →

The Geometric mean for continuous series can be obtained by finding out the mid-value of the interval and by using the concept of G.M for discrete series.

Ex:-1 Calculate the Geometric mean for the following data:

x	12	13	14	15	16	17
f	5	4	4	3	2	2

Sol:-

x	f	log x	f x log x
12	5	1.0792	5.3960
13	4	1.1139	4.4556
14	4	1.1461	4.5844
15	3	1.1761	3.5283
16	2	1.2041	2.4092
17	2	1.2304	1.2304
	$\Sigma f = 20$		$\Sigma f \log x = 21.6029$

we have

$$G.M = \text{Antilog}\left[\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i\right] \qquad N = \Sigma f_i$$

$$= \text{Antilog}\left[\frac{21.6029}{19}\right]$$

$$= \text{Antilog}(1.137)$$

$$= 10^{1.1373} = 13.71$$

Date:- / /
Page:-

Ex:-2 find the Geometric mean for the following data:

marks	0-10	10-20	20-30	30-40	40-50
no. of students	4	8	10	6	7

Sol^{n}:-

marks	No. of students	mid values	$\log x$	$f \log x$
	f	x		
0-10	4	5	0.6990	2.7960
10-20	8	15	1.1761	9.4088
20-30	10	25	1.3979	13.9790
30-40	6	35	1.5441	9.2646
40-50	7	45	1.6532	11.5724
	$\Sigma f = N = 35$			$\Sigma f \log x = 47.0208$

So $G.M = \text{Antilog}\left[\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i\right]$

$= \text{Antilog}\left[\frac{47.0208}{35}\right]$

$= \text{Antilog}\,[1.3435]$

$= 10^{1.3435} = 22.06$

(#)

Harmonic Mean:- →

~~For~~ Individual Series -

Harmonic mean of a no. of observations is the reciprocal of the Arithmetic mean of the reciprocals of the given values. Thus. H.M of n observations $x_1, x_2, \ldots x_n$ is

Date:- / /
Page:-

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

for discrete Series.

if $x_1, x_2, \ldots, x_n$ $(\neq 0)$ have the frequencies $f_1, f_2, \ldots f_n$ resp. then Harmonic Mean

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}} = \frac{N}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_n}{x_n}}, \quad N = \sum_{i=1}^{n} f_i$$

for continuous Series (class-Intervals)

x is taken to be the mid-value of class-Int.

Ex:- An Aeroplane flies along the four sides of a square at speeds of 100, 200, 300 and 400 Km/hr resp., what is the average speed of the aeroplane in its flight around the square?

Soln:-

When equal distances are covered with unequal speeds, the Harmonic Mean is the Proper average.

∴ Average speed = Harmonic Mean

$$= \frac{4}{\frac{1}{100}+\frac{1}{200}+\frac{1}{300}+\frac{1}{400}}$$

$$= 192 \text{ Km/hr.}$$

Date:- / /
Page:-

Ex:-2 Find out the Harmonic mean of the following data:

Marks (out of 150)	No. of students
10	2
20	3
40	6
60	5
120	4

Sol^n:-

Marks x	No. of students f	$\frac{1}{x}$	$f \times \frac{1}{x}$
10	2	0.100	0.200
20	3	0.050	0.150
40	6	0.025	0.150
60	5	0.017	0.085
120	4	0.008	0.032
	$\Sigma f = N = 20$		$\Sigma \frac{f}{x} = 0.617$

Harmonic Mean

$$= \frac{N}{\Sigma \frac{f}{x}}$$

$$= \frac{20}{0.617} = 32.4$$

Date:- / /
Page:-

Median :→

Median is the central value of the variables when the values are arranged in ascending or descending order of magnitude.

1. for ungrouped / Individual frequency Distribution→

If the n values of the variate are arrange in ascending or descending order of magnitude.

a.) if n is odd Then

$$\text{median} = \left(\frac{n+1}{2}\right)^{th} \text{ value} = \text{middle value}$$

if n is even Then

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

2. For discrete frequency distribution →

Median is obtained by cumulative frequencies

Find $\frac{N+1}{2}$, where $N = \Sigma f_i$

Then $\text{Cf}_{\text{just}} \geq \frac{N+1}{2}$ The corresponding value of x is median.

3. For a Grouped frequency distribution →

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Median class is the class corresponding to cumulative frequency $\geq \frac{N+1}{2}$ or $\frac{N}{2}$

Date:- / /
Page:-

where, l = lower limit of median class
h = width of Median class
f = frequency of Median class
$N = \Sigma f$
C = C.f. of the class preceding the median class.

Ex:-1 Find the median of 6, 8, 9, 10, 11, 12, 13.

Sol:-

Arranging given values in ascending order

6, 8, 9, 10, 11, 12, 13

here $n = 7$ (odd)

So

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

$$= \left(\frac{7+1}{2}\right)^{th} \text{value} = 4^{th} \text{value}$$

$$= 10 \quad \text{Ans.}$$

Ex:-2. Obtain the median for the following frequency distribution -

x :	1	2	3	4	5	6	7	8	9
f :	8	10	11	16	20	25	15	9	6

Sol:-

The given frequency distribution is discrete Then cumulative frequency distribution table-

Date:- / /
Page:-

x	f	c.f
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	$N = \Sigma f = 120$	

Here $N = 120$

$\therefore \frac{N+1}{2} = \frac{120+1}{2}$

$= 60.5$

c.f just $\geq \frac{N+1}{2} = 60.5$ is 65

and the value of x corresponding to c.f 65 is 5

Hence Median is 5.

Ex:3. find the median for the following data:

Marks	No. of students	Marks	No. of students
Below 10	15	Below 50	94
Below 20	35	Below 60	127
Below 30	60	Below 70	198
Below 40	84	Below 80	249

Sol:- The cumulative freq. table with class Intervals

Marks	No. of students (f)	c.f
0-10	15	15
10-20	20	35
20-30	25	60
30-40	24	84
40-50	10	94
50-60	33	127
60-70	71	198
70-80	51	249

$\Sigma f = N = 249$

$\frac{N}{2} = 124.5$

$\therefore$ median class is 50-60 $\quad l = 50$

$h = 10, \quad f = 33, \quad c = 94$

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 50 + \frac{10}{33}(124.5 - 94)$$

$$= 59.24 \text{ marks.}$$

(#) **Mode :** $\rightarrow$

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

OR Mode is that value of the variate for which frequency is maximum.

a) for discrete frequency distribution mode is the value of x corresponding to maximum frequency.

But in case of the following

i) maximum freq. is repeated

ii) maximum freq. occurs in the very beginning or at the end of the distribution

iii) if there is irregularities in the distribution, the value of the mode is determined by method of grouping.

Date:- / /
Page:-

b) for continuous frequency distribution

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

where l is the lower limit
h is the width
f_m is the frequency of modal class
f_1, f_2 are the frequencies of the classes preceding and succeeding the model class respectively.

for a Symmetric Distribution, mean, median, mode coincide.

if Method of grouping fails then

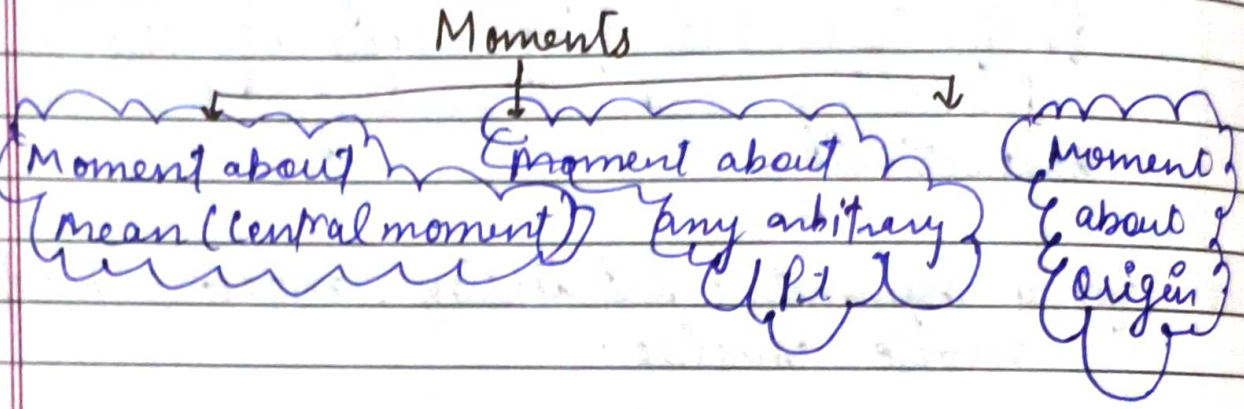
$$\boxed{\text{mode} = 3\,\text{Median} - 2\,\text{mean}}$$

called empirical mode.

Date:- / /
Page:-

Imp Moments : →

Moments are statistical tools, used in statistical investigations.
Moments of a distribution are the arithmetic means of various powers of the deviations of items from some given number.



1. Moments about Mean (Central Moments)

(a) for an Individual Series →
If $x_1, x_2, \dots, x_n$ are the values of the variable under consideration,
The r^{th} moment μ_r about mean $\bar{x}$ is

$$\mu_r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^r}{n} \quad ; \quad r = 0, 1, 2, \dots$$

(b) for a frequency distribution →
if the frequency distribution is

x :	x_1,	x_2,	$\dots$	x_n
f :	f_1,	f_2,	$\dots$	f_n

Then the r^{th} moment μ_r about the mean $\bar{x}$ is defined as

Date:- / /
Page:-

$$\mu_r = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^r}{N} \quad ; r = 0, 1, 2, 3 \ldots$$

$$N = \sum_{i=1}^{n} f_i$$

for $r = 0$,

$$\mu_0 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^0 \qquad \because (x_i - \bar{x})^0 = 1$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i = \frac{N}{N}$$

$$\mu_0 = 1$$

∴ for any distribution, $\boxed{\mu_0 = 1}$

for $r = 1$,

$$\mu_1 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i x_i - \frac{1}{N} \sum_{i=1}^{n} f_i \bar{x}$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i x_i - \bar{x} \frac{1}{N} \cdot \sum_{i=1}^{n} f_i \qquad N = \sum_{i=1}^{n} f_i$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i x_i - \bar{x} \cdot 1$$

$$= \bar{x} - \bar{x} = 0$$

∴ for any distribution $\boxed{\mu_1 = 0}$

for $r = 2$

$$\mu_2 = \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \text{variance} = (S.D)^2$$

∴ for any distribution, μ_2 = variance of the distribution.

S.D = Standard deviation

Date:- / /
Page:-

Similarly,

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^3$$

&

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^4 \quad \text{& soon}$$

Ex:- Find the first four moments for the following individual series:

x	3	6	8	10	18

Sol:-

S.No.	x	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
1	3	-6	36	-216	1296
2	6	-3	9	-27	81
3	8	-1	1	-1	1
4	10	1	1	1	1
5	18	9	81	729	6561
$n=5$	$\Sigma x = 45$	$\Sigma(x-\bar{x})=0$	$\Sigma(x-\bar{x})^2 = 128$	$\Sigma(x-\bar{x})^3 = 486$	$\Sigma(x-\bar{x})^4 = 7940$

$$\text{mean } \bar{x} = \frac{\Sigma x}{n} = \frac{45}{5} = 9$$

first four moments

$$\mu_1 = \frac{\Sigma(x-\bar{x})}{n} = 0$$

$$\mu_2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{128}{5} = 25.6$$

$$\mu_3 = \frac{\Sigma(x-\bar{x})^3}{n} = \frac{486}{5} = 97.2$$

$$\mu_4 = \frac{\Sigma(x-\bar{x})^4}{n} = \frac{7940}{5} = 1588.$$

Date:- / /
Page:-

Ex-2. Calculate $\mu_1, \mu_2, \mu_3, \mu_4$ for the following frequency distribution

marks	5-15	15-25	25-35	35-45	45-55	55-65
No. of students	10	20	25	20	15	10

Sol:-

Marks	No. of students f	mid pt (x)	fx	$x-\bar{x}$	$f(x-\bar{x})$
5-15	10	10	100	-24	-240
15-25	20	20	400	-14	-280
25-35	25	30	750	-4	-100
35-45	20	40	800	6	120
45-55	15	50	750	16	240
55-65	10	60	600	26	260
	$N = \Sigma f = 100$		$\Sigma fx = 3400$		$\Sigma f(x-\bar{x}) = 0$

$$\text{mean } \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{3400}{100} = 34.$$

$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
5760	-138240	3317760
3920	-54880	768320
400	-1600	6400
720	4320	25920
3840	61440	983040
6760	175760	4569760
$\Sigma f(x-\bar{x})^2 = 21400$	$\Sigma f(x-\bar{x})^3 = 46800$	$\Sigma f(x-\bar{x})^4 = 9671200$

$$\mu_1 = \frac{\Sigma f(x-\bar{x})}{N} = \frac{0}{100} = 0$$

$$\mu_2 = \frac{\Sigma f(x-\bar{x})^2}{N} = \frac{21400}{100} = 214$$

Date:- / /
Page:-

$$\mu_3 = \frac{\sum f(x-\bar{x})^3}{N} = \frac{46800}{100} = 468$$

$$\mu_4 = \frac{\sum f(x-\bar{x})^4}{N} = \frac{967200}{100} = 9672$$

2. Moments about an arbitrary number :→ (Raw Moments)

(a) For ungrouped data or Individual Series →

If $x_1, x_2, \ldots, x_n$ are the values of the variable x, then moments about any point A is denoted by μ_r' and is defined as

$$\boxed{\mu_r' = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r}, \quad r = 0, 1, 2, 3, \ldots$$

For grouped data or for frequency distribution (Discrete or continuous) :—

If $x_1, x_2, \ldots, x_n$ are the values of a variable x with the corresponding frequencies $f_1, f_2, \ldots, f_n$ respectively, then the r^{th} moment about any point A is denoted by μ_r' and is defined as

$$\boxed{\mu_r' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^r} \quad r = 0, 1, 2, 3$$

$$N = \sum_{i=1}^{n} f_i$$

Date:- / /
Page:-

for $r=0$, $\mu_0' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^0 = 1$ $\boxed{\mu_0' = 1}$

for $r=1$,

$$\mu_1' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i - \frac{1}{N}\sum_{i=1}^{n} f_i A$$

$$= \bar{x} - \frac{A}{N}\sum_{i=1}^{n} f_i$$

$$= \bar{x} - \frac{A}{N} \times N$$

$$\boxed{\mu_1' = \bar{x} - A}$$

for $r=2$,

$$\mu_2' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^2$$

for $r=3$,

$$\mu_3' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^3 \quad \text{and so on.}$$

3. Moments about the origin :→

If $x_1, x_2, \ldots, x_n$ be the values of a variable x with corresponding frequencies $f_1, f_2, \ldots, f_n$ respectively then r^{th} moment about the origin ν_r is defined as

$$\boxed{\nu_r = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^r}, \quad r = 0, 1, 2, \ldots \qquad N = \sum_{i=1}^{n} f_i$$

For $r=0$, $\nu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^0 = \frac{N}{N} = 1 \Rightarrow \boxed{\nu_0 = 1}$

For $r=1$, $\nu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i = \bar{x}$ $\boxed{\nu_1 = \bar{x}}$

(#) **Relation between μ_r and μ_r' :→**

We know that

$$\mu_r = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^r}{N} \quad \text{— (1)}$$

and

$$\mu_r' = \frac{\sum_{i=1}^{n} f_i (x_i - A)^r}{N} \quad \text{— (2)}$$

from (1),

$$\mu_r = \frac{1}{N} \sum_{i=1}^{n} f_i [(x_i - A) - (\bar{x} - A)]^r$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i [(x_i - A) - \mu_1']^r \qquad \because \mu_1' = \bar{x} - A$$

$$= \frac{1}{N} \sum_{i=1}^{n} f_i [{}^rC_0 (x_i - A)^r (\mu_1')^0 - {}^rC_1 (x_i - A)^{r-1} (\mu_1') + {}^rC_2 (x_i - A)^{r-2} (\mu_1')^2 - \dots + (-1)^r (\mu_1')^r]$$

Binomial Theorem.

$$= \frac{1}{N} \sum_{i=1}^{n} f_i (x_i - A)^r - \frac{{}^rC_1}{N} \sum_{i=1}^{n} f_i (x_i - A)^{r-1} (\mu_1') + \frac{{}^nC_2}{N} \sum_{i=1}^{n} f_i (x_i - A)^{r-2} (\mu_1')^2 + \dots + (-1)^r \frac{1}{N} \sum_{i=1}^{n} f_i (\mu_1')^r$$

$$= \mu_r' - {}^rC_1 \mu_{r-1}' (\mu_1')^1 + {}^rC_2 \mu_{r-2}' (\mu_1')^2 - {}^rC_3 \mu_{r-3}' (\mu_1')^3 + \dots (-1)^r (\mu_1')^r$$

we know that $\boxed{\mu_1 = 0}$

Putting $r = 2, 3, 4$, we get

$$\mu_2 = \mu_2' - 2\mu_1'^2 + \mu_0' \cdot \mu_1'^2$$

$$= \mu_2' - 2\mu_1'^2 + \mu_1'^2 \qquad \mu_0' = 1$$

$$\boxed{\mu_2 = \mu_2' - \mu_1'^2}$$

Date:- / /
Page:-

$r=3,$

$$\mu_3 = \mu_3' - {}^3C_1 \mu_2' \mu_1'^2 + {}^3C_2 \mu_1' \mu_1'^2 - {}^3C_3 \mu_0' \mu_1'^3$$

$$= \mu_3' - 3\mu_2'\mu_1' + 3\mu_1'^3 - 1.1\mu_1'^3 \quad \because \mu_0' = 1$$

$$\boxed{\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3}$$

$r=4,$

$$\mu_4 = \mu_4' - {}^4C_1 \mu_3' \mu_1' + {}^4C_2 \mu_2' \mu_1'^2 - {}^4C_3 \mu_1' \mu_1'^3 + {}^4C_4 \mu_0' \mu_1'^4$$

$$\boxed{\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4}$$

Relation between ν_r and μ_r :→

we know that

$$\nu_r = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^r \quad , r = 0, 1, 2, \ldots$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x} + \bar{x})^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i \left[{}^rC_0 (x_i - \bar{x})^r \bar{x}^0 + {}^rC_1 (x_i - \bar{x})^{r-1} \bar{x} + {}^rC_2 (x_i - \bar{x})^{r-2} \bar{x}^2 + \cdots \bar{x}^r \right]$$

$$\nu_r = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^r + \frac{{}^rC_1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^{r-1} \bar{x} + \frac{{}^rC_2}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^{r-2} \bar{x}^2 + \cdots + \frac{1}{N}\sum_{i=1}^{n} f_i \bar{x}^r$$

$$= \mu_r + {}^rC_1 \mu_{r-1} \bar{x} + {}^rC_2 \mu_{r-2} \bar{x}^2 + {}^rC_3 \mu_{r-3} \bar{x}^3 + \cdots + \bar{x}^r$$

Putting $r = 1, 2, 3, 4, \ldots$

$$\nu_1 = \mu_1 + \mu_0 \bar{x} = \bar{x} \qquad [\mu_1 = 0, \mu_0 = 1]$$

$$\boxed{\nu_1 = \bar{x}}$$

Date:- / /
Page:-

$r=2$,

$$\nu_2 = \mu_2 + {}^2C_1 \mu_1 \bar{x} + {}^2C_2 \mu_0 \bar{x}^2$$

$$= \mu_2 + 2 \cdot 0 \cdot \bar{x} + 1 \cdot 1 \cdot \bar{x}^2$$

$$\boxed{\nu_2 = \mu_2 + \bar{x}^2}$$

$r=3$,

$$\nu_3 = \mu_3 + {}^3C_1 \mu_2 \bar{x} + {}^3C_2 \mu_1 \bar{x}^2 + {}^3C_3 \mu_0 \bar{x}^3$$

$$\boxed{\nu_3 = \mu_3 + 3\mu_2 \bar{x} + \bar{x}^3} \qquad [\because \mu_1 = 0, \ \mu_0 = 1]$$

$r=4$,

$$\nu_4 = \mu_4 + {}^4C_1 \mu_3 \bar{x} + {}^4C_2 \mu_2 \bar{x}^2 + {}^4C_3 \mu_1 \bar{x}^3 + {}^4C_4 \mu_0 \bar{x}^4$$

$$\nu_4 = \mu_4 + 4\mu_3 \bar{x} + 6\mu_2 \bar{x}^2 + \bar{x}^4$$

$$\boxed{\nu_4 = \mu_4 + 4\mu_3 \bar{x} + 6\mu_2 \bar{x}^2 + \bar{x}^4}$$

(#) **Karl Pearson's β and γ coefficients:-**

These coefficients are based upon the first four moments of a frequency distribution about its mean:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\gamma_1 = +\sqrt{\beta_1}, \quad \gamma_2 = \beta_2 - 3$$

The Practical use of these coefficients is to measure the skewness & kurtosis of a frequency distribution

Date:- / /
Page:-

Ex:1. The first four moments of a distribution about the value 4 of a variable are -1.5, 17, -30 & 108. Find the moments about mean, about origin. Also find the moments about the point $x = 2$. Also β_1 & β_2

Sol:-

Here $A = 4$

$\mu_1' = -1.5$

$\mu_2' = 17$

$\mu_3' = -30$

$\mu_4' = 108$

Moments about mean →

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = 17 - (-1.5)^2 = 14.75$

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

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

$= 39.75$

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

$= 142.3125$

Also $\mu_1' = \bar{x} - A$

So $\bar{x} = \mu_1' + A$

$= -1.5 + 4$

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

6.25
14.75
.00

Moment about origin →

$\nu_1 = \bar{x} = 2.5$

$\nu_2 = \mu_2 + \bar{x}^2 = 14.75 + 2.5^2 = 21$

$\nu_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3$

$= 166$

$\nu_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4 = 1132$

Date:- / /
Page:-

Now Karl Pearson's coef.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.875)^2}{(14.75)^3} = 0.492377$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{(141.3125)}{(14.75)^2} = 0.654122$$

$$= 0.654122.$$

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{0.492377}$$

$$\gamma_2 = \beta_2 - 3 = 0.654122 - 3$$

$$=$$

Moments about point $x = 2$

$$\mu_1' = \bar{x} - A = 2.5 - 2 = 0.5$$

$\because \mu_2 = \mu_2' - \mu_1'^2$

$$\mu_2' = \mu_2 + \mu_1'^2 = 14.75 + 0.5^2$$

$$= 15$$

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

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

$$= 39.75 + 3 \times 15 \times 0.5 - 2 \times 0.5^3$$

$$= 62$$

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

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

$$= 244$$

Date: / /
Page:

Ex:-2 The first three moments of a distribution about the value 2 of the variable are 1, 16, -40 show that the mean is 3, variance is 15, & $\mu_3 = -86$.

Sol^n:

$A = 2$

$\mu_1' = 1, \quad \mu_2' = 16, \quad \mu_3' = -40$

We have,

$\mu_1' = \bar{x} - A$

$\bar{x} = \mu_1' + A = 1 + 2 = 3$

$\because \mu_2 = \mu_2' - \mu_1'^2 = 16 - 1^2 = 15$

$\therefore$ variance $\mu_2 = 15$

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

$= -40 - 3 \times 16 \times 1 + 2 \times 1^3$

$= -86.$

Ex:-3. The first four moments of a distribution about $x = 2$ are 1, 2.5, 5.5 and 16. Calculate the first four moments about the mean and about origin.

Sol^n:- We have $A = 2, \quad \mu_1' = 1, \quad \mu_2' = 2.5, \quad \mu_3' = 5.5$

$\mu_4' = 16$

Moments about mean.

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = (2.5)^2 - 1^2 = 2.5 - 1 = 1.5$

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

$= 5.5 - 3 \times 2.5 \times 1 + 2 \times 1^3$

$= 5.5 - 7.5 + 2 = 0$

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

$= 16 - 4 \cdot (5.5)(1) + 6 \times 2.5 \times 1^2 - 3 \times 1^4$

$= 6$

Date:- / /
Page:-

Moments about origin →

$$\nu_1 = \bar{x} - A$$
$$\Rightarrow \bar{x} = \nu_1$$

$\because \mu_1' = \bar{x} - A$

$\Rightarrow \bar{x} = \mu_1' + A = 1 + 2 = 3$

So $\nu_1 = \bar{x} = 3$

$\nu_2 = \mu_2 + \bar{x}^2$

$= 1.5 + 3^2$

$= 1.5 + 9$

$= 10.5$

$\nu_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3$

$= 0 + 3 \times 1.5 \times 3 + 3^3$

$= 40.5$

$\nu_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4$

$= 6 + 4(0)(3) + 6(1.5)(3)^2 + (3)^4$

$= 168.$

Ex:-4. For a distribution, the mean is 10, variance is 16, $\gamma_1 = 1$ & β_2 is 4. Find the first four moments about the origin.

Sol:- Here $\bar{x} = 10$, $\mu_2 = 16$, $\gamma_1 = 1$, $\beta_2 = 4$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 4$$

$\Rightarrow \mu_4 = 4 \times 16^2 = 4 \times 256 = 1024$

also $\gamma_1 = 1 \Rightarrow +\sqrt{\beta_1} = 1 \Rightarrow \beta_1 = 1$

$\Rightarrow \frac{\mu_3^2}{\mu_2^3} = 1$

$\Rightarrow \mu_3^2 = 1 \times 16^3 \Rightarrow \mu_3^2 = (64)^2$

Date:- / /
Page:-

$\mu_3 = 64$

Moments about the origin →

$\nu_1 = \bar{x} = 10$, $\nu_2 = 116$, $\nu_3 = 1544$, $\nu_4 = 22184$

(#) **Skewness :** →

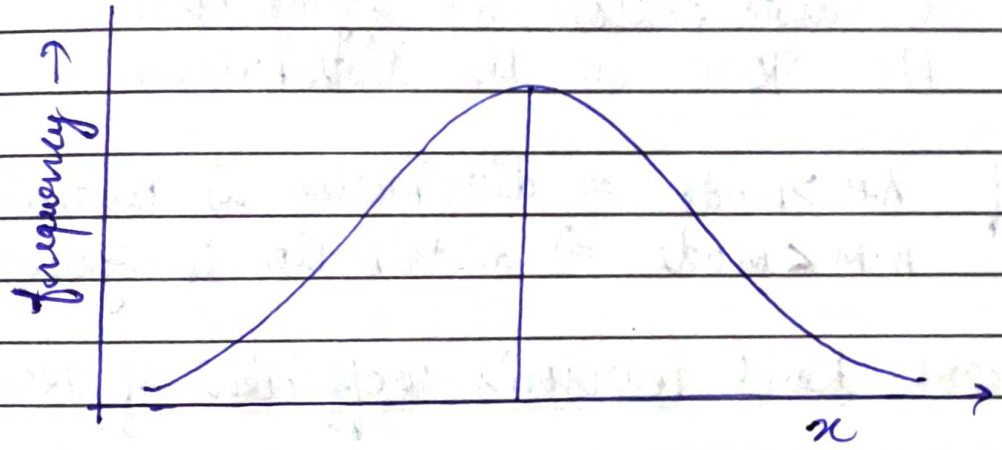
Symmetric Distribution

For a symmetrical distribution, the frequencies are symmetrically distributed about the mean i.e variates equidistant from the mean have equal frequencies.

In symmetric distribution,

Mean = Mode = Median.

and median lies half-way between the two quartiles.



$M = M_o = M_d$

Skewness :- Lack of symmetry in a frequency distribution is called skewness

Skewness indicates whether the curve is turned more to one side than to other side. i.e whether the curve has a longer tail on one side.

Date:- / /
Page:-

$\mu_3 = 64$

Moments about the origin →

$\nu_1 = \bar{x} = 10$, $\nu_2 = 116$, $\nu_3 = 1544$, $\nu_4 = 22184$

(#) **Skewness :** →

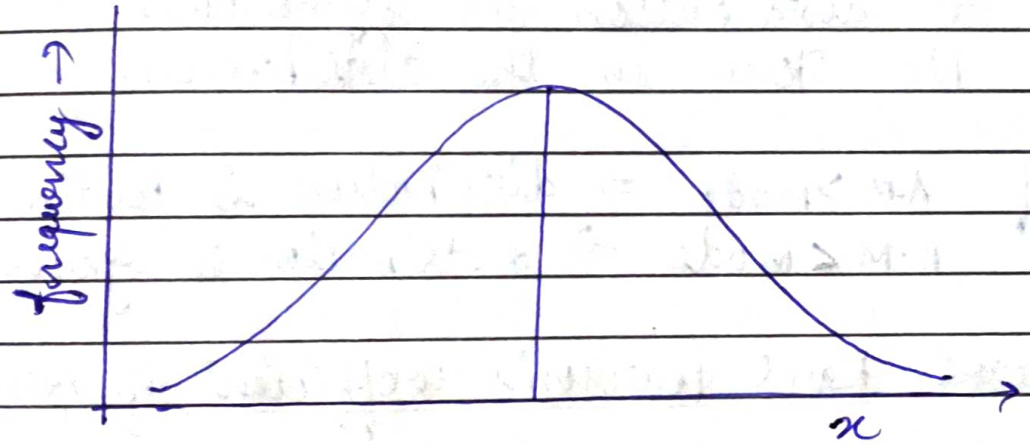
Symmetric Distribution

For a symmetrical distribution, the frequencies are symmetrically distributed about the mean i.e variates equidistant from the mean have equal frequencies.

In symmetric distribution,

Mean = Mode = Median.

and median lies half-way between the two quartiles.



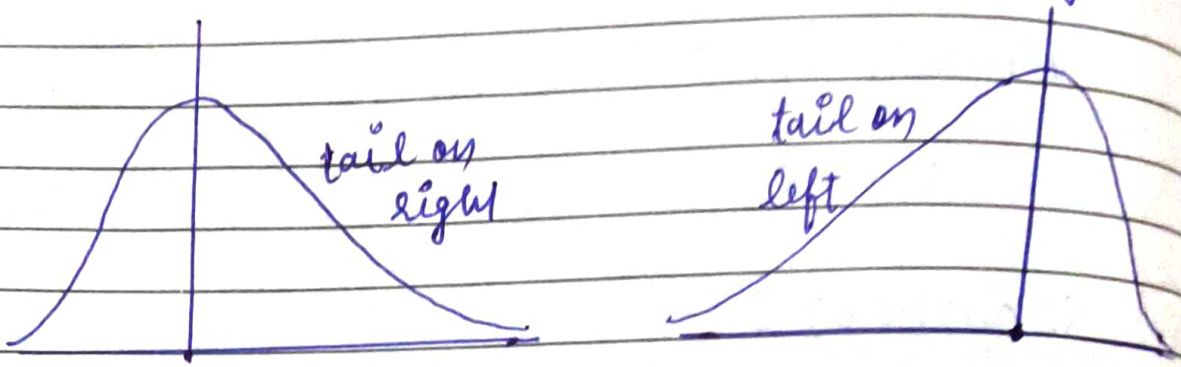
$M = M_o = M_d$

Skewness :- Lack of symmetry in a frequency distribution is called skewness

Skewness indicates whether the curve is turned more to one side than to other side. i.e whether the curve has a longer tail on one side.

Date:- / /
Page:-

Skewness can be positive as well as negative



Positively skewed distribution

Negatively skewed distribution

* Skewness is positive if the longer tail of the distribution lies towards the right and negative if it lies towards the left.

* if A.M = Mode = Median
⇒ distribution is symmetric
⇒ No skew in the distribution.

* if A.M > Mode ⇒ distribution is positively skewed
A.M < Mode ⇒ distribution is negatively skewed

~~Koard~~ Karl Pearson's coefficient of skewness:

measure of skewness are called coef. of skewness

Karl Pearson's coefficient of skewness

$$SK_p = \frac{A.M - Mode}{S.D}$$

Date:- / /
Page:-

As we know that,

Mode = 3 Median − 2 A.M.

∴ Coeff. of Skewness = $\frac{\text{A.M} - \text{Mode}}{\text{S.D}}$

$= \frac{\text{A.M} - (3\,\text{Median} - 2\,\text{A.M})}{\text{S.D}}$

$= \frac{3\,\text{A.M} - 3\,\text{Median}}{\text{S.D}}$

∴ Karl Pearson's Coefficient of Skewness $= \frac{3(\text{A.M} - \text{Median})}{\text{S.D}}$

Note:- i) The value of SK_p lies between −1 and 1.

ii) for symmetric distribution, $SK_p = 0$

iii) if $SK_p > 0$, then distribution is Positively Skewed.

iv) if $SK_p < 0$, then distribution is negatively Skewed.

Ex:-1. Karl Pearson's coefficient of skewness of a distribution is 0.32. Its Standard deviation is 6.5, mean is 29.6 find the mode of the distribution.

Sol:- we have $SK_p = 0.32$, $S.D = 6.5$

mean $\bar{x} = 29.6$

$SK_p = \frac{\text{A.M} - \text{Mode}}{\text{S.D}}$

$0.32 = \frac{29.6 - \text{Mode}}{6.5}$

$29.6 - \text{Mode} = 0.32 \times 6.5$

Mode = 27.52

Date:- / /
Page:-

Ex:-2. The sum of 20 observations is 300 and sum of their squares is 5000. The median is 15. find the Karl Pearson's coefficient of skewness.

Sol":- we have,

$n = 20$, $\Sigma x = 300$, x is a variable

$\Sigma x^2 = 5000$

Median $= 15$

Mean $\bar{x} = \frac{\Sigma x}{n} = \frac{300}{20} = 15$

Variance $= \mu_2 = \nu_2 - \bar{x}^2$

$= \frac{\Sigma x^2}{n} - \bar{x}^2$

$= \frac{5000}{20} - 15^2$

$= 250 - 225$

$= 25$

So S.D $= \sqrt{\mu_2}$

$= \sqrt{25}$

$= 5$

Now Karl Pearson's coeff. of skewness

$SK_p = \frac{3(A.M - \text{median})}{S.D}$

$= \frac{3(15-15)}{5}$

$= \frac{3 \times 0}{5}$

$= 0$

$\Rightarrow$ distribution is symmetric.

Date:- / /
Page:-

Ex:-3 find the coefficient of skewness by Karl Pearson's method for the following data:

Value	6	12	18	24	30	36	42
Frequency	4	7	9	18	15	10	3

Sol^{n}:- Calculation of $\bar{x}$, S.D

value x	f	fx	fx^2
6	4	24	144
12	7	84	1008
18	9	162	2916
24	18	432	10368
30	15	450	13500
36	10	360	12960
42	3	126	5292
	$N = 66$	$\Sigma fx = 1638$	

A.M $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{\Sigma fx}{N} = \frac{1638}{66}$

$= 24.82$

Standard deviation, S.D $= \sqrt{\mu_2}$

$= \sqrt{\nu_2 - \bar{x}^2}$ $\because \nu_2 = \mu_2 + \bar{x}$

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

$= \sqrt{699.82 - 24.82^2}$

$= \sqrt{83.7876}$

$= 9.15$

Date:- / /
Page:-

Mode = value of x corresponding to maximum freq. 18
= 24

So coefficient of skewness

$$SK_p = \frac{A.M - Mode}{S.D}$$
$$= \frac{24.82 - 24}{9.15}$$
$$= 0.0896$$

(#) **Method of Moments :→**

In this Method, Second and third central moments of the distribution are used. This measure of skewness is called the Moment coefficient of skewness, denoted by SK_M or γ_1.

Moment coefficient of skewness $= \frac{\mu_3}{\sqrt{\mu_2^3}}$

$$SK_M = \gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}}$$

So $$\boxed{SK_M = \frac{\mu_3}{\sqrt{\mu_2^3}}}$$

Ex:1. The first three central moments of a distribut^n are 0, 15, -31. find the moment coef. of skewness.

Sol^n:- we have $\mu_1 = 0$, $\mu_2 = 15$, $\mu_3 = -31$

So moment coef of skewness $= \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-31}{\sqrt{15^3}} = \frac{-31}{58.09} = -0.53$

Date:- / /
Page:-

Ex:-2. The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Calculate the moment coef. of skewness.

Sol^{n}:-

We have,

$A = 5$, $\mu_1' = 2$, $\mu_2' = 20$, $\mu_3' = 40$, $\mu_4' = 50$

Now, $\mu_2 = \mu_2' - \mu_1'^2$

$= 20 - 2^2$

$= 16$

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

$= 40 + 3 \times 20 \times 2 + 2 \times 2^3$

$= 40 - 120 + 16$

$= 56 - 120$

$= -64$

Moment coefficient of skewness

$$SK_M = \frac{\mu_3}{\sqrt{\mu_2^3}}$$

$$= \frac{-64}{\sqrt{16^3}}$$

$$= \frac{-64}{64}$$

$$= -1$$

Date:- / /
Page:-

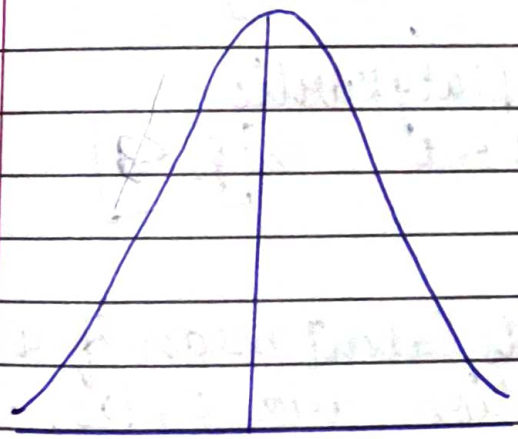
(#) **Kurtosis :→** A frequency curve may be symmetrical but it may not be equally flat topped with the normal curve. The bulginess or the Relative flatness of the curve of a freq. distribution is called Kurtosis.

Hence, the sharpness of the peak of a frequency-distribution curve is called Kurtosis.

(1) curves which are neither flat nor sharply peaked are called normal curves or mesokurtic curves.

(2) Curves which are flatter than the normal curve are called Platykurtic curves.

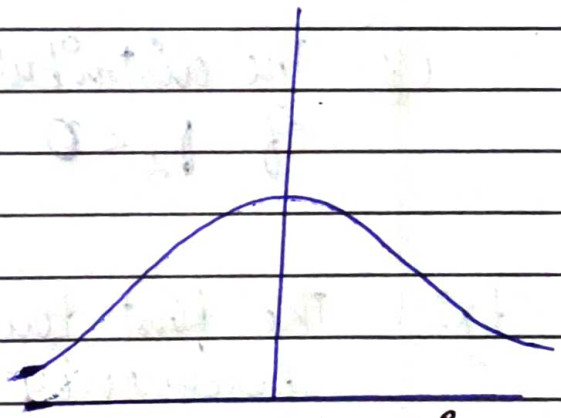
(3) Curves which are more sharply peaked than the normal curve are called leptokurtic curves.



Leptokurtic

$\beta_2 > 3$

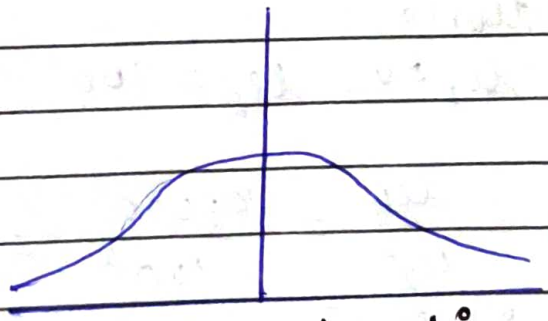
$\gamma_2 > 0$



Mesokurtic

$\beta_2 = 3$

$\gamma_2 = 0$



Platykurtic.

$\beta_2 < 3$

$\gamma_2 < 0$

Measure of Kurtosis :-

Measure of kurtosis is denoted by β_2 and is defined by

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$$

μ_2 and μ_4 are Second & fourth Central Moment

The kurtosis of a distribution is also measured by $\gamma_2 = \beta_2 - 3$.

Note :- (1.) The distribution is leptokurtic if $\gamma_2 > 0 \Rightarrow \beta_2 - 3 > 0 \Rightarrow \boxed{\beta_2 > 3}$

(2.) The distribution is Mesokurtic if $\gamma_2 = 0 \Rightarrow \beta_2 - 3 = 0 \Rightarrow \boxed{\beta_2 = 3}$

(3.) The distribution is platykurtic if $\gamma_2 < 0 \Rightarrow \beta_2 - 3 < 0 \Rightarrow \boxed{\beta_2 < 3}$

Ex :- The first four moments about mean of a frequency distribution are 0, 100, -7 and 35000. Discuss the kurtosis of the distribution.

Sol :- Given $\mu_1 = 0$, $\mu_2 = 100$, $\mu_3 = -7$, $\mu_4 = 35000$

Now $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{35000}{100^2} = 3.5 > 3$

$\Rightarrow$ The distribution is leptokurtic.

Date:- / /
Page:-

Ex:2. The first four moments of a distribution about $x=4$ are 1, 4, 10 and 45. obtain the various characteristics of the distribution on the basis of the given information. comment upon the nature of the distribution. (2016, 2018)

Sol:- We have $A=4$

$\mu_1' = 1, \mu_2' = 4, \mu_3' = 10, \mu_4' = 45$

Moments about mean (central moments)

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = 4 - 1^2 = 3$

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

$= 10 - 3\times4\times1 + 2\times1^3 = 0$

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

$= 45 - 4\times10\times1 + 6\times4\times1^2 - 3\times1^4$

$= 45 - 40 + 24 - 3$

$= 26$

Skewness:

Moment coefficient of skewness $= \frac{\mu_3}{\sqrt{\mu_2^3}}$

$= \frac{0}{\sqrt{3^3}} = 0$

∴ The distribution is symmetrical

Kurtosis:

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{26}{3^2} = \frac{26}{9} = 2.89 < 3$

∴ The distribution is platykurtic.

Date:- / /
Page:-

Ex:-3. The following table represents the height of a batch of 100 students. calculate Kurtosis & discuss skewness

Height (in cm)	59	61	63	65	67	69	71	73	75
No. of students	0	2	6	20	40	20	8	2	2

(2018)

Sol:- x f

Height (in cm)	No. of Students	$u = \frac{x-A}{2}$, $A=67$	fu	fu^2	fu^3	fu^4
59	0	-4	0	0	0	0
61	2	-3	-6	18	-54	162
63	6	-2	-12	24	-48	96
65	20	-1	-20	20	-20	20
67	40	0	0	0	0	0
69	20	1	20	20	20	20
71	8	2	16	32	64	128
73	2	3	6	18	54	162
75	2	4	8	32	128	512
$N = \Sigma f = 100$			$\Sigma fu = 12$	$\Sigma fu^2 = 164$	$\Sigma fu^3 = 144$	$\Sigma fu^4 = 1100$

Moments about the point $A = 67$ →

$$\mu_1' = \frac{\Sigma fu}{N} \times h$$

$$= \frac{12}{100} \times 2 = 0.24$$

$$\mu_2' = \frac{\Sigma fu^2}{N} \times h^2$$

$$= \frac{164}{100} \times 2^2 = 6.56$$

$$\mu_3' = \frac{\Sigma fu^3}{N} \times h^3 = \frac{144}{100} \times 2^3 = 11.52$$

Date:- / /
Page:-

$$\mu_4' = \frac{\Sigma f u^4}{N} \times h^4 = \frac{1100}{100} \times 2^4 = 176$$

Moments about mean →

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 6.56 - (0.24)^2 = 6.5024$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$= 11.52 - 3 \times 6.56 \times 0.24 + 2 \times 0.24^3$$
$$= 6.8244$$

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

$$= 167.19798$$

Skewness :-

Moment coef. of skewness $SK_M = \frac{\mu_3}{\sqrt{\mu_2^3}}$

$$= \frac{6.8244}{\sqrt{6.5024^3}} = 0.411570$$

⇒ ∴ The distribution is positively skewed

Kurtosis :-

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{167.19798}{(6.5024)^2}$$

$$= 3.9544 > 3$$

$$\Rightarrow \beta_2 > 3$$

Hence the distribution is leptokurtosis.

Ques The first four moments about the value x of a distribution are 0.294, 7.144, 42.409 and 454.98 Calculate the moments about mean. Also evaluate β_1, β_2 & comment upon the skewness & kurtosis of the distribution.

Date:- / /
Page:-

Curve fitting :→

Consider n-paired observations $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ of two variables x and y. In order to have an approximate idea about the relationship of these two variables, we plot these n paired points on a graph, we get a diagram showing the simultaneous variation in values of both the variables called Scatter or dot diagram.

Curve fitting means an exact relationship b/w two variables by Algebraic equations. In fact, this relationship is the eqn of the curve. Therefore, curve fitting means to form an eqn of the curve from the given data.

It is useful in the Study of Correlation and Regression.

for this we use Method of least squares.

Method of least squares :—

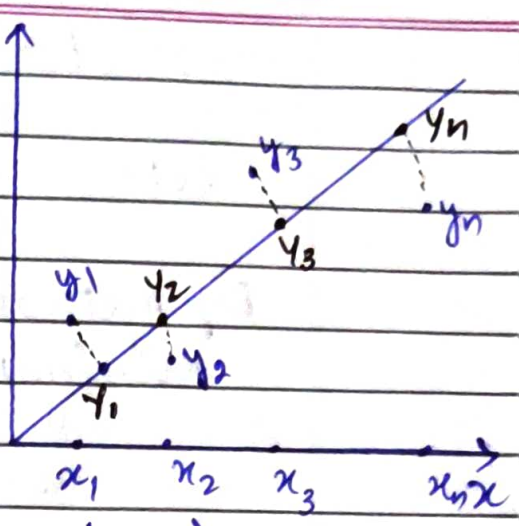
Method of ~~list~~ least squares provides a unique set of values to the constants and hence suggests a curve of best fit to the given data.

Suppose we have n-paired observations $(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)$ of x & y.

Date:- / /
Page:-

here y_i → observed value
or given value for x_i

Y_i → Expected value
$(Y_i = f(x_i)$
$i = 1, 2, \ldots n)$



Error or ~~Error~~ Residual

$E_i = y_i - Y_i$ (may be +ve / -ve)

$\Rightarrow E_i^2 = (y_i - Y_i)^2$ (for equal weightage)

Now Introducing E = Total Error

$$E = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

[if $E = 0 \Rightarrow E_i = 0 \ \forall\ i = 1, \ldots n$ then all the points lie on the curve]

The least value of E gives the best fitting curve to the data.
This Method is called method of least squares or principle of least square.

(1) fitting of a Straight line :—

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be n sets of observat'n of related data and

$$y = a + bx \quad \text{—(1)}$$

be the straight line to be fitted.
Now residual / Errors

$$E_i = y_i - Y_i$$

Date: / /
Page:

$\Rightarrow E_i = y_i - a - bx_i$ $\quad \because y_i = a + bx_i$

Let $U = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$ ——— (2)

By Method of least squares, U should be minimum.

$\therefore \frac{\partial U}{\partial a} = 0$ and $\frac{\partial U}{\partial b} = 0$

$\therefore \frac{\partial U}{\partial a} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-1) = 0$

$\Rightarrow \sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} 1 - b\sum_{i=1}^{n} x_i = 0$

$\Rightarrow \Sigma y - an - b\Sigma x = 0$

$\Rightarrow \boxed{\Sigma y = an + b\Sigma x}$ ——— (3)

and $\frac{\partial U}{\partial b} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0$

$\Rightarrow \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0$

$\Rightarrow \boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$ ——— (4)

$\therefore$ U is minimum when

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

These eqns are called Normal equation for a & b.

on solving (3), (4) we get a & b

put these values of a & b in eqn (2)

$y = a + bx$.

we get best fit straight line to the given data.

Date:- / /
Page:-

Ex:- 1 By method of least squares, find the straight line that best fits the following data:

x :	1	2	3	4	5
y :	14	27	40	55	68

Sol:-

Let the straight line of best fit be

$$y = a + bx \quad -(1)$$

Normal eq's are

$$\Sigma y = na + b\Sigma x \quad -(2)$$

and

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad -(3)$$

Here $n = 5$.

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma xy = 748$	$\Sigma x^2 = 55$

Substituting in (2) & (3),

$$204 = 5a + 15b$$
$$748 = 15a + 55b$$

On solving,

$$15a + 45b = 612$$
$$15a + 55b = 748$$
$$-10b = -136$$

$$b = 13.6, \quad a = \frac{204 - 15 \times 13.6}{5}$$

$$= 0$$

So $a = 0$, $b = 13.6$

Hence Required Straight line is

$$y = 13.6x$$

Ex:-2. fit a straight line to the following data by least square Method.

x :	0	1	2	3	4
y :	1	1.8	3.3	4.5	6.3

Sol:-

Let the straight line to be fitted in the given data

$$y = a + bx \quad \text{— (1)}$$

Normal eq^{n}s are

$$\Sigma y = an + b\Sigma x \quad \text{— (2)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{— (3)}$$

Here $n = 5$

x	y	xy	x^2
0	1	0	0
1	1.8	1.8	1
2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16
$\Sigma x = 10$	$\Sigma y = 16.9$	$\Sigma xy = 47.1$	$\Sigma x^2 = 30$

from (1) & (2)

$$16.9 = 5a + 10b$$

$$47.1 = 10a + 30b$$

on solving

Date:- / /
Page:-

$10a + 20b = 33.8$
$10a + 30b = 47.1$

$-10b = -13.3$

$\boxed{b = 1.33}$

$\boxed{a = 0.72}$

Hence Required Straight line is

$\boxed{y = 0.72 + 1.33x}$

(II) **fitting a Second degree parabola :-**

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be the given data and

$$y = a + bx + cx^2 \quad \text{— (1)}$$

be the Second degree parabola to be fitted.

Now Errors,

$$E_i = y_i - Y_i$$

$[\because Y_i = a + bx_i + cx_i^2]$

$$E_i = y_i - a - bx_i - cx_i^2 \quad \text{— (2)}$$

Let $$U = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i)^2 \quad \text{— (3)}$$

By the method of least square, U should be minimum.

For min. value of U

$$\frac{\partial U}{\partial a} = 0, \quad \frac{\partial U}{\partial b} = 0, \quad \frac{\partial U}{\partial c} = 0$$

Date:- / /
Page:-

$$\frac{\partial v}{\partial a} = 2\sum_{i=1}^{n}(y_i - a - bx_i - cx_i^2)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} 1 - b\sum_{i=1}^{n} x_i - c\sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \boxed{\Sigma y = an + b\Sigma x + c\Sigma x^2} \quad -(a)$$

$$\frac{\partial v}{\partial b} = 2\sum_{i=1}^{n}(y_i - a - bx_i - cx_i^2)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 - c\sum_{i=1}^{n} x_i^3 = 0$$

$$\Rightarrow \boxed{\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3} \quad -(b)$$

$$\frac{\partial v}{\partial c} = 2\sum_{i=1}^{n}(y_i - a - bx_i - cx_i^2)(-x_i^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i^2 y - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i^3 - c\sum_{i=1}^{n} x_i^4 = 0$$

$$\Rightarrow \boxed{\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4} \quad -(c)$$

eq^{n}s (a), (b) and (c) are called Normal eq^{n}s. solving these for a, b, c put these values in eq^n $y = a + bx + cx^2$ we get best fitted parabola.

Date:- / /
Page:-

Ex:-1. Fit a second degree parabola in the following data:

x:	0	1	2	3	4
y:	1	4	10	17	30

Sol^{n}:-

Let $y = a + bx + cx^2$ ——— (1) be second degree parabola to be fitted in the given data.

Normal eq^{n}s are

$$\left.\begin{aligned} \Sigma y &= an + b\Sigma x + c\Sigma x^2 \\ \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \text{—— (2)}$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	4	1	1	1	4	4
2	10	4	8	16	20	40
3	17	9	27	81	51	153
4	30	16	64	256	120	480
$\Sigma x = 10$	$\Sigma y = 62$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 195$	$\Sigma x^2y = 677$

using all these value in (2)

$$62 = 5a + 10b + 30c \quad \text{——— (4)}$$
$$195 = 10a + 30b + 100c \quad \text{——— (5)}$$
$$677 = 30a + 100b + 354c \quad \text{——— (6)}$$

on solving (4), (5) & (6), we have

$a = 1.2$, $b = 1.1$, $c = 1.5$

Putting all these values in eq^{n} (1)

$$y = 1.2 + 1.1x + 1.5x^2$$

which is the required eq^{n} of parabola to be fitted in the given data.

Date:- / /
Page:-

(II) fitting of the curve :→ $\boxed{y = ax + bx^2}$

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be the given data and $y = ax + bx^2$ — (1) be the curve fitted to the data.

Residual at $x = x_i$

$E_i = y_i - Y_i$ $\quad [\because Y_i = ax_i + bx_i^2]$

$\Rightarrow E_i = y_i - ax_i - bx_i^2$

Let $U = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (y_i - ax_i - bx_i^2)^2$ — (2)

Now by Method of least Square, U should be minimum.

for Min. value U,

$\frac{\partial U}{\partial a} = 0, \quad \frac{\partial U}{\partial b} = 0$

$\frac{\partial U}{\partial a} = 0 \Rightarrow 2\sum_{i=1}^{n} (y_i - ax_i - bx_i^2)(-x_i) = 0$

$\Rightarrow \sum_{i=1}^{n} (x_i y_i - ax_i^2 - bx_i^3) = 0$

$\Rightarrow \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3$

$\Rightarrow \boxed{\Sigma xy = a\Sigma x^2 + b\Sigma x^3}$ — (3)

$\frac{\partial U}{\partial b} = 0 \Rightarrow 2\sum_{i=1}^{n} (y_i - ax_i - bx_i^3)(-x_i^2) = 0$

$\Rightarrow \sum_{i=1}^{n} (x_i^2 y_i - ax_i^3 - bx_i^4) = 0$

$\Rightarrow \boxed{\Sigma x^2 y = a\Sigma x^3 + b\Sigma x^4}$ — (4)

eqn (3) and (4) are Normal equations.

Date:- / /
Page:-

Ex:- By Method of least square, find the curve $y = ax + bx^2$ that best fit the following data: (2014)

x:	1	2	3	4	5
y:	1.8	5.1	8.9	14.1	19.8

Sol^n:-

Let $y = ax + bx^2$ be the curve that best fits to the given data

Here Normal equations are

$$\left.\begin{aligned} \Sigma xy &= a\Sigma x^2 + b\Sigma x^3 \\ \Sigma x^2 y &= a\Sigma x^3 + b\Sigma x^4 \end{aligned}\right\} \quad -(2)$$

Here $n = 5$.

x	y	x^2	x^3	x^4	xy	x^2y
1	1.8	1	1	1	1.8	1.8
2	5.1	4	8	16	10.2	20.4
3	8.9	9	27	81	26.7	80.1
4	14.1	16	64	256	56.4	225.6
5	19.8	25	125	625	99	495
		$\Sigma x^2 = 55$	$\Sigma x^3 = 225$	$\Sigma x^4 = 979$	$\Sigma xy = 194.1$	$\Sigma x^2y = 822.9$

Substituting these values in equation (2), we get

$$194.1 = 55a + 225b$$
$$822.9 = 225a + 979b$$

on solving, we get

$$a = \frac{83.85}{55} = 1.52$$

$$b = \frac{317.4}{664} = 0.49$$

Hence Required eq^n of curve

$$y = 1.52x + 0.49x^2$$

Date:- / /
Page:-

(IV) Fitting of an exponential curve $y = ae^{bx}$

Given $y = ae^{bx}$ — (1)

Taking logarithm on both sides of eq^n (1)

$$\log_{10} y = \log_{10}(a \cdot e^{bx}) = \log_{10} a + \log_{10} e^{bx}$$

$$\log_{10} y = \log_{10} a + bx \log_{10} e$$

$$\Rightarrow \log_{10} y = \log_{10} a + (b \cdot \log_{10} e) \cdot x$$

$$\boxed{Y = A + BX}$$

where $A = \log_{10} a$, $B = b \cdot \log_{10} e$, $X = x$

$Y = \log_{10} y$

Normal equations are -

$$\Sigma Y = A \cdot n + B \Sigma X$$

and $$\Sigma XY = A \Sigma X + B \Sigma X^2$$

Solving on these, we get A and B

Then $\boxed{a = \text{Antilog } A}$, $\boxed{b = \frac{B}{\log_{10} e}}$

Ex:- 1. Find the curve of best fit of the type $y = ae^{bx}$ to the following data by the method of least Squares:

x:	1	5	7	9	12
y:	10	15	12	15	21

Sol^n:- The curve to be fitted is $y = ae^{bx}$ — (1)

or $Y = A + BX$ — (2)

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = b \log_{10} e$, $X = x$

Date:- / /
Page:-

∴ The normal equations are

$$\Sigma Y = A \cdot n + B \Sigma X \quad \text{—(3)}$$

$$\Sigma XY = A \Sigma X + B \Sigma X^2 \quad \text{—(4)}$$

Here $n = 5$

$X = x$	y	$Y = \log_{10} y$	X^2	YX
1	10	1	1	1
5	15	1.1761	25	5.8805
7	12	1.0792	49	7.5544
9	15	1.1761	81	10.5849
12	21	1.3222	144	15.8664
$\Sigma X = 34$		$\Sigma Y = 5.7536$	$\Sigma X^2 = 300$	$\Sigma XY = 40.8862$

Substituting these values in eqn (3) & (4), we get

$$5.7536 = 5A + 34B$$
$$40.8862 = 34A + 300B$$

on solving, we get

$$A = 0.9766 \,, \quad B = 0.02561$$

∴

$$a = \text{Antilog}_{10} A \,, \quad b = \frac{B}{\log_{10} e} = \frac{0.02561}{\log_{10} e}$$

$$\Rightarrow a = 9.4754 \,, \quad b = 0.059.$$

Hence The Required curve is $\boxed{y = 9.4754 e^{0.059x}}$

(V.) Fitting of the curve $y = ax^b$.

given $y = ax^b$

Taking logarithm on both sides, we get

$$\log_{10} y = \log_{10} a + b \log_{10} x$$

or $\boxed{Y = A + BX}$

where $Y = \log_{10} y$, $A = \log_{10} a$, $B = b$, $X = \log_{10} x$

Date:- / /
Page:-

Normal eqⁿs are

$$\Sigma y = A \cdot n + B \Sigma x$$
$$\Sigma xy = A \Sigma x + B \Sigma x^2$$

on solving we get A and B.

$a = \text{Antilog } A, \quad b = B$

(VI) Fitting of the curve $y = ab^x$

given curve is $y = ab^x$ —①

taking logarithm on both sides

$$\log y = \log_{10}(ab^x)$$

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

or $\boxed{Y = A + BX}$ —②

where

$Y = \log_{10} y, \quad A = \log_{10} a, \quad B = \log_{10} b, \quad X = x$

Normal eqⁿ are

$$\Sigma Y = A \cdot n + B \Sigma X$$
$$\Sigma XY = A \Sigma X + B \Sigma X^2$$

where n is the no. of pairs of values of x and y.

After solving Normal eqⁿs we get A and B and Hence

$$a = \text{Anti}\log_{10} A, \quad b = \text{Anti}\log_{10} B$$

Ex:- Obtain a relation of the form $y = ab^x$ for the following data by the Method of least squares

x	2	3	4	5	6
y	8.3	15.4	33.1	65.2	127.4

Date:- / /
Page:-

Sol^n The curve to be fitted is $y = ab^x$ —①

taking log on both sides

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

or $Y = A + BX$ —②

where

$Y = \log_{10} y$, $A = \log_{10} a$, $B = \log_{10} b$, $X = x$

Normal equations are -

$$\Sigma Y = A \cdot n + B \Sigma X$$

$$\Sigma XY = A \Sigma X + B \Sigma X^2 \quad , n = 5$$

$X = x$	y	$Y = \log_{10} y$	X^2	XY
2	8.3	0.9191	4	1.8382
3	15.4	1.1872	9	3.5616
4	33.1	1.5198	16	6.0792
5	65.2	1.8142	25	9.0710
6	127.4	2.1052	36	12.6312
$\Sigma X = 20$		$\Sigma Y = 7.5455$	$\Sigma X^2 = 90$	$\Sigma XY = 33.1812$

Substituting these values in Normal equations

$$7.5455 = 5A + 20B$$

$$33.1812 = 20A + 90B$$

on solving we get,

$$A = 0.31 \quad , B = 0.3$$

$\therefore$

$$a = \text{antilog } A = \text{antilog } 0.31 = 2.04$$

$$b = \text{antilog } B = \text{antilog } 0.3 = 1.995$$

Hence the required curve is

$$\boxed{y = 2.04 (1.995)^x}$$

Date:- / /
Page:-

(VII) **Fitting of the curve $PV^r = C$:-**

given $PV^r = C \Rightarrow V^r = \frac{C}{P} \Rightarrow P = CV^{-r}$

Taking logarithm on both sides

$$\log P = \log C - r \log V$$

or $Y = A + BX$

where $Y = \log P$, $A = \log C$, $B = -r$, $X = \log V$

Normal eqns are

$$\Sigma Y = A \cdot n + B \Sigma X$$
$$\Sigma XY = A \Sigma X + B \Sigma X^2$$

Imp

Ex:- The pressure of the gas corresponding to various volumes V is measured given by the following data

V (cm^3):	50	60	70	90	100
P ($kg\,(cm^{-2})$):	64.7	51.3	40.5	25.9	78

fit the data to the eqn $PV^r = C$ (2019)

Sol^n:-

Given curve $PV^r = C$ — (1)

$\Rightarrow P = CV^{-r}$

Taking logarithm on both sides, we get

$$\log P = \log C - r \log V$$

or $Y = A + BX$ — (2)

where

$Y = \log P$, $A = \log C$, $B = -r$, $X = \log V$

Normal eqns are

$$\Sigma Y = A \cdot n + B \Sigma X$$
$$\Sigma XY = A \Sigma X + B \Sigma X^2$$

Here $n = 5$.

Date:- / /
Page:-

V	P	$X = \log V$	$Y = \log P$	X^2	XY
50	64.7	1.69897	1.81090	2.88650	3.07666
60	51.3	1.77815	1.71012	3.16182	3.04085
70	40.5	1.84510	1.60746	3.40439	2.96592
90	25.9	1.95424	1.41330	3.81905	2.76193
100	78	2	1.89209	4	3.78418
		$\Sigma X = 9.27646$	$\Sigma Y = 8.43387$	~~$\Sigma XY = 15.62954$~~	$\Sigma XY = 15.62954$

$\Sigma X^2 = 17.27176$

Putting these values in Normal equations,

$$8.43387 = 5A + 9.27646B$$
$$15.62954 = 9.27646A + 17.27176B$$

Solving these eqⁿs, we get

$$A = 2.22476, \quad B = -0.28997$$

$$\therefore \quad \gamma = -B = 0.28997$$

$$C = \text{antilog } A = \text{antilog } 2.22476$$
$$= 167.78765$$

Hence, the required eqⁿ of curve is

$$\boxed{PV^{0.28997} = 167.78765}$$

(VIII) **Fitting of the curve** $y = \frac{c_0}{x} + c_1\sqrt{x}$

Given curve to be fitted $y = \frac{c_0}{x} + c_1\sqrt{x}$ ——— (1)

Error at point $x = x_i$ is

$$E_i = y_i - Y_i = y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i}$$

By Method of least squares, the values of c_0, c_1 are such that

$\because Y_i = \frac{c_0}{x} + c_1\sqrt{x}$

$$U = \sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} \left(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i}\right)^2$$

is Minimum

Date: / /
Page:

So Normal equations are given by

$$\frac{\partial U}{\partial c_0} = 0 \quad \text{and} \quad \frac{\partial U}{\partial c_1} = 0$$

Now $\frac{\partial U}{\partial c_0} = 0$

$$\Rightarrow 2\sum_{i=1}^{n}\left(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i}\right)\left(\frac{-1}{x_i}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{y_i}{x_i} - \frac{c_0}{x_i^2} - \frac{c_1}{\sqrt{x_i}}\right) = 0$$

$$\Rightarrow \boxed{\sum \frac{y}{x} = c_0 \sum \frac{1}{x^2} + c_1 \sum \frac{1}{\sqrt{x}}}$$

and $\frac{\partial U}{\partial c_1} = 0$

$$\Rightarrow 2\sum_{i=1}^{n}\left(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i}\right)\left(-\sqrt{x_i}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(y_i\sqrt{x_i} - \frac{c_0}{\sqrt{x_i}} - c_1 x_i\right) = 0$$

$$\Rightarrow \boxed{\sum y\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x}$$

So Normal eqns are

$$\sum \frac{y}{x} = c_0 \sum \frac{1}{x^2} + c_1 \sum \frac{1}{\sqrt{x}}$$

and $$\sum y\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x$$

Date:- / /
Page:-

Ex: Use Method of least squares to fit the curve $y = \frac{c_0}{x} + c_1\sqrt{x}$ to the following table of values

x :	0.1	0.2	0.4	0.5	1	2
y :	21	11	7	6	5	6

Soln:-

The given curve to be fitted is $y = \frac{c_0}{x} + c_1\sqrt{x}$ —— (1)

The Normal equations are

$$\sum \frac{y}{x} = c_0 \sum \frac{1}{x^2} + c_1 \sum \frac{1}{\sqrt{x}} \quad \text{—— (2)}$$

and $$\sum y\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x \quad \text{—— (3)}$$

x	y	y/x	$y\sqrt{x}$	$1/\sqrt{x}$	$1/x^2$
0.1	21	210	6.64078	3.16228	100
0.2	11	55	4.91935	2.23607	25
0.4	7	17.5	4.42719	1.58114	6.25
0.5	6	12	4.24264	1.41421	4
1	5	5	5	1	1
2	6	3	8.48528	0.70711	0.25
$\sum x = 4.2$		$\sum \frac{y}{x} = 302.5$	$\sum y\sqrt{x} = 33.71524$	$\sum \frac{1}{\sqrt{x}} = 10.10081$	$\sum \frac{1}{x^2} = 136.5$

from eqn (2) and (3), we have

$$302.5 = 136.5\,c_0 + 10.10081\,c_1$$
$$33.71524 = 10.10081\,c_0 + 4.2\,c_1$$

Solving these, we get

$$c_0 = 1.97327, \quad c_1 = 3.28182$$

Hence the required eqn of curve is

$$\boxed{y = \frac{1.97327}{x} + 3.28182\,x}$$

Date:- / /
Page:-

(IX) fitting of the curve $xy = b + ax = ax + b$

$$xy = b + ax$$
$$y = \frac{b}{x} + a$$
$$y = bX + a \quad , \quad X = \frac{1}{x} \; , \; Y = y$$

Normal eq^{n}s are

$$\Sigma y = a.n + b\Sigma X$$
$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

(X) Fitting of the curve $y = ax + \frac{b}{x}$.

Normal eq^{ns} are

$$\Sigma xy = a\Sigma x^2 + b.n$$

and $$\Sigma \frac{y}{x} = a.n + b\Sigma \frac{1}{x^2}$$

Date: / /
Page:

(11) **Correlation :** →

In a Bivariate distribution, if the change in one variable affects a change in other variable, the variables are said to be correlated.

→ if the two variables deviate in the same direction i.e. if the increase (or decrease) in one results in a corresponding increase (decrease) in the other, correlation is said to be direct or positive correlation

e.g. the correlation between income and expenditure is positive.

→ if the two variables deviate in opposite direction i.e if the increase (or decrease) in one results in a corresponding decrease (or increase) in the other, correlation is said to be inverse or negative correlation.

e.g. The correlation between volume and the pressure of a perfect gas or the correlation b/w price and demand is negative.

→ correlation is said to be perfect if the deviation in one variable is followed by a corresponding proportional deviation in the other.

Date:- / /
Page:-

(#) Measure of correlation :→

Karl Pearson's coefficient of correlation (Product Moment correlation coefficient) :→

Consider two variables x and y then we know

$$\text{Var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sigma_x^2$$

σ_x is S.D for x

$$\text{Var}(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sigma_y^2$$

σ_y is S.D for y

& $$\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Karl Pearson's coefficient of correlation b/w two variables x and y, denoted by $r(x,y)$ or r_{xy} is a numerical measure of linear relationship b/w them and is defined as

$$r(x,y) = r_{xy} = \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x)\cdot\text{Var}(y)}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum(y_i - \bar{y})^2}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \cdot \sigma_y}$$

$$\boxed{r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{N\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}}$$

Date:- / /
Page:-

Alternative form of $r(x,y)$:

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

n = No. of pairs of values of x and y.

if $X = x_i - \bar{x}$, $Y = y_i - \bar{y}$ deviation from A.M

$$r_{xy} = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}}$$

Note:- $-1 \le r_{xy} \le 1$

→ if $u = \frac{x-a}{h}$, $v = \frac{y-b}{k}$

Then $r_{xy} = r_{uv}$

$$r(u,v) = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

Ex:-1 from the data given below, find the no. of items n:
$r_{xy} = 0.5$, $\Sigma XY = 120$, $\Sigma X^2 = 90$, $\sigma_y = 8$
where X and Y are deviations from the arithmetic mean. (2018)

Sol:- given $r_{xy} = 0.5$, $\Sigma XY = 120$, $\Sigma X^2 = 90$, $\sigma_y = 8$
we have

$$r_{xy} = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \cdot \Sigma Y^2}}$$ where $X = x_i - \bar{x}$, $Y = y_i - \bar{y}$

Date:- / /
Page:-

Now $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

$8^2 = \frac{1}{n}\Sigma y^2 = \frac{\Sigma y^2}{n}$

$\Rightarrow \Sigma y^2 = 8^2 n = 64n$

So

$$r_{xy} = \frac{120}{\sqrt{90 \times 64n}} = 0.5$$

Squaring both sides

$$\frac{14400}{90 \times 64n} = 0.25$$

$$n = \frac{14400}{90 \times 0.25 \times 64}$$

$$\boxed{n = 10}$$

Ex:-2 Calculate the coefficient of correlation b/w the age of husband & wife from the following data:

Age of husband : 35 34 40 43 56 20 38

Age of wife : 32 30 31 32 53 20 33

Sol:- $\bar{x} = \frac{\Sigma x_i}{n} = \frac{266}{7} = 38$, $\bar{y} = \frac{\Sigma y_i}{n} = \frac{231}{7} = 33$

x_i	y_i	$X = x_i - \bar{x}$	$Y = y_i - \bar{y}$	X^2	Y^2	XY
35	32	-3	-1	9	1	3
34	30	-4	-3	16	9	12
40	31	2	-2	4	4	-4
43	32	5	-1	25	1	-5
56	53	18	20	324	400	360
20	20	-18	-13	324	169	234
38	33	0	0	0	0	0
$\Sigma x_i = 266$	$\Sigma y_i = 231$			$\Sigma X^2 = 702$	$\Sigma Y^2 = 584$	$\Sigma XY = 600$

Date:- / /
Page:-

∴ Karl Pearson's coefficient of correlation

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2}\sqrt{\Sigma(y_i - \bar{y})^2}}$$

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

$$= \frac{600}{\sqrt{702 \times 584}} = \frac{600}{640.287}$$

$$= 0.937$$

Ex:-3. Find the coefficient of correlation b/w the values of x and y:

x: 1 3 5 7 8 10

y: 8 12 15 17 18 20

Sol^n:-

x	y	x^2	y^2	xy
1	8	1	64	8
3	12	9	144	36
5	15	25	225	75
7	17	49	289	119
8	18	64	324	144
10	20	100	400	200
$\Sigma x = 34$	$\Sigma y = 90$	$\Sigma x^2 = 248$	$\Sigma y^2 = 1446$	$\Sigma xy = 582$

Karl Pearson's coefficient of correlation is

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

Date:- / /
Page:-

$$r_{xy} = \frac{6 \times 582 - 34 \times 90}{\sqrt{6 \times 248 - 34^2}\sqrt{6 \times 1446 - 90^2}}$$

$$= 0.9879$$

Ex:- 4. Find the coefficient of correlation for the following table

x:	10	14	18	22	26	30
y:	18	12	24	6	30	36

(2011)

Sol^n:-

Let $u = \frac{x-22}{4}$, $v = \frac{y-24}{6}$

$a = 22$, $h = 4$ $\quad b = 24$, $k = 6$

x	y	$u = \frac{x-22}{4}$	$v = \frac{y-24}{6}$	u^2	v^2	uv
10	18	-3	-1	9	1	3
14	12	-2	-2	4	4	4
18	24	-1	0	1	0	0
22	6	0	-3	0	9	0
26	30	1	1	1	1	1
30	36	2	2	4	4	4
		$\Sigma u = -3$	$\Sigma v = -3$	$\Sigma u^2 = 19$	$\Sigma v^2 = 19$	$\Sigma uv = 12$

Here $n = 6$, $\bar{u} = \frac{\Sigma u}{n}$, $\bar{v} = \frac{\Sigma v}{n}$

So $\bar{u} = \frac{-3}{6} = \frac{-1}{2}$, $\bar{v} = \frac{-3}{6} = \frac{-1}{2}$

$$r_{uv} = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}}$$

$$= \frac{6 \times 12 - (-3)(-3)}{\sqrt{6 \times 19 - (-3)^2}\sqrt{6 \times 19 - (-3)^2}}$$

$$= \frac{63}{\sqrt{105}\sqrt{105}} = 0.6$$

$\boxed{r_{xy} = r_{uv} = 0.6}$

Date:- / /
Page:-

(*) Rank correlation : →

Sometimes we have to deal with problems in which data can not be quantitatively measured but qualitative assessment is possible.

Spearman's Coefficient of Rank correlation

Let a group of n individuals be arranged in order of merit or proficiency in possession of two characteristics A and B. The ranks in two characteristics are, in general, different.
Let x_i, y_i, $i = 1, 2, \ldots, n$ be ranks of n indi. in the group for char A and B respectively
Then

Rank correlation coefficient

$$\boxed{r = 1 - \left[\frac{6\sum D_i^2}{n(n^2-1)}\right]}$$

where n is no. of observation
$D_i = x_i - y_i =$ Difference in Ranks.

Note :- $-1 \le r \le 1$

Ex. 1 Calculate the Rank correlation coefficient.

Marks in chem :	78	36	98	25	75	82	90	62	65	39
Marks in Maths :	84	51	91	60	68	62	86	58	63	47

Sol^n :-

Date:- / /
Page:-

X (chem)	Y (Maths)	R_x	R_y	$D = R_x - R_y$	D^2
78	84	4	3	1	1
36	51	9	9	0	0
98	91	1	1	0	0
25	60	10	7	3	9
75	68	5	4	1	1
82	62	3	6	-3	9
90	86	2	2	0	0
62	58	7	8	-1	1
65	63	6	5	1	1
39	47	8	10	-2	4
					$\Sigma D^2 = 26$

So rank correlation coefficient

$$r = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

$(n = 10)$

$$r = 1 - \frac{6 \times 26}{10(10^2-1)}$$

$$= 1 - \frac{156}{990}$$

$$r = .8424$$

Ex:-2. Ten competitors in a beauty contest were ranked by three judges in the following order

First judge:	1	6	5	10	3	2	4	9	7	8
Second judge:	3	5	8	4	7	10	2	1	6	9
Third judge:	6	4	9	8	1	2	3	10	5	7

Use method of rank correlation to determine which pair of judges has the nearest approach to

Date:- / /
Page:-

common taste in beauty?

Sol:-

Competitors	R_1	R_2	R_3	$D_{12}=R_1-R_2$	$D_{13}=R_1-R_3$	$D_{23}=R_2-R_3$	D_{12}^2	D_{13}^2	D_{23}^2
A	1	3	6	-2	-5	-3	4	25	9
B	6	5	4	1	2	1	1	4	1
C	5	8	9	-3	-4	-1	9	16	1
D	10	4	8	6	2	-4	36	4	16
E	3	7	1	-4	2	6	16	4	36
F	2	10	2	-8	0	8	64	0	64
G	4	2	3	2	1	-1	4	1	1
H	9	1	10	8	-1	-9	64	1	81
I	7	6	5	1	2	1	1	4	1
J	8	9	7	-1	1	2	1	1	4
$n=10$							$\Sigma D_{12}^2=200$	$\Sigma D_{13}^2=60$	$\Sigma D_{23}^2=214$

Rank correlation coefficient b/w first and second judges

$$r_{12} = 1 - \frac{6\Sigma D_{12}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 200}{10(10^2-1)} = 1 - \frac{1200}{990}$$

$$= -0.212$$

Ily. $$r_{13} = 1 - \frac{6\Sigma D_{13}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 60}{10(10^2-1)} = 1 - \frac{360}{990}$$

$$= 0.636$$

and $$r_{23} = 1 - \frac{6\Sigma D_{23}^2}{n(n^2-1)} = 1 - \frac{6 \times 214}{10(10^2-1)}$$

$$= -0.297$$

Date:- / /
Page:-

correlation b/w first and second judges is negative i.e. their opinions regarding beauty test are opposite to each other. Similarly, opinions of second and third judges are opposite to each other, but opinions of first and third judges are of similar type as their correlation is positive. It means their likings and dislikings are very much common.

if the Ranks are Repeated

Then Rank correlation coefficient

$$\gamma = 1 - \frac{6[\Sigma D^2 + E]}{n(n^2-1)}$$

Where

$$E = \frac{1}{12} m_1(m_1^2-1) + \frac{m_2(m_2^2-1)}{12} + \cdots$$

where $m_1, m_2 \cdots$ stands for the number of times different items repeats

Ex:- Obtain the Rank correlation coefficient of the following data:

X:	68	64	75	50	64	80	75	40	55	64
Y:	62	58	68	45	81	60	68	48	50	70

Sol^n:-

Date:- / /
Page:-

X	y	R_x	R_y	$D = R_x - R_y$	D^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\Sigma D^2 = 72$

Here $n = 10$

So

$$r = 1 - \frac{6[\Sigma D^2 + E]}{n(n^2-1)}$$

$$E = \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}$$

$m = 2, 3, 2$

$$= \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}$$

$$= \frac{2\times3 + 3\times8 + 2\times3}{12} = \frac{36}{12}$$

$$= 3$$

$$r = 1 - \frac{6[72+3]}{10(10^2-1)}$$

$$= 1 - \frac{6\times75}{99\times10}$$

$$= \frac{6}{11} = 0.545$$

Date:- / /
Page:-

Regression :→

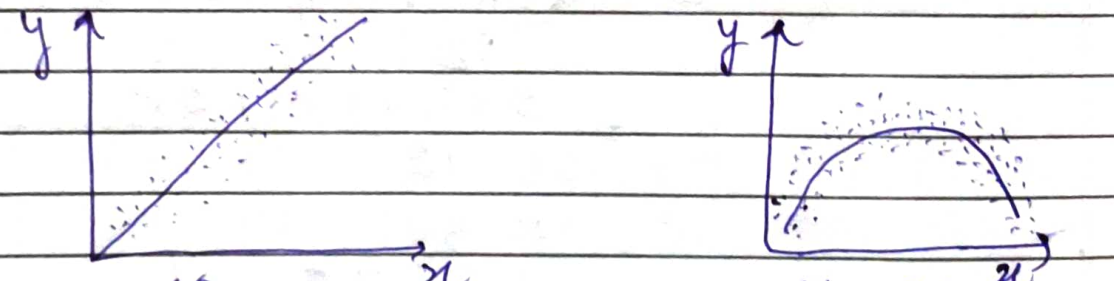
Regression means functional relationship between two or more related variables.

The only fundamental difference, if any, between problems of curve fitting and regression is that in regression, any of the variables may be considered as independent or dependent while in curve fitting, one variable can not be dependent.

Regression measures the nature and extent of correlation. Regression is the estimation or prediction of unknown values of one variable from known values of another variable.

Linear Regression :-

when points in Scatter diagram are concentrated around a straight line then it is Linear regression otherwise Non-linear regression.



Linear Regression

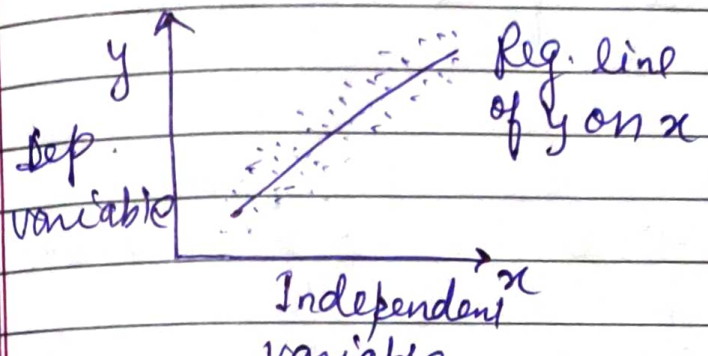
Non linear Regression

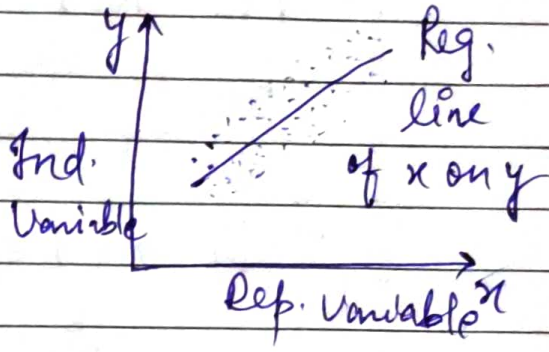
Scatter diagram

Date:- / /
Page:-

Suppose that the Scatter diagram indicates some relationship b/w two variables x & y, the dots of the Scatter diagram will be more or less concentrated around the curve
This curve is called the curve of regression

When curve is a Straight line, it is called a line of regression, & regression is said to be linear.





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

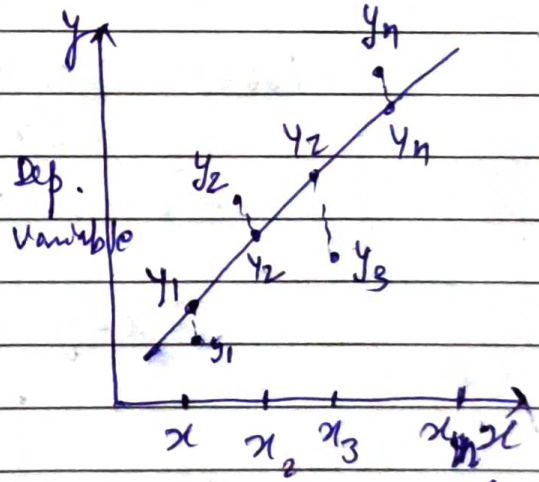
Derivation of line of Regression :—

I. y on x :-

Let $y = a + bx$ be the eqn of regression line of y on x

given value $\rightarrow y_i^o$

$Y_i \rightarrow$ expected value



Error $E_i = y_i^o - Y_i^o$

$E_i^2 = (y_i^o - Y_i^o)^2$

$= (y_i^o - a - b x_i^o)^2$ $\quad \because Y_i^o = a + bx_i^o$

Date:- / /
Page:-

Now consider Total Error

$$U = \sum_{i=1}^{n} E_i^2$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i)^2 \quad \text{---(1)}$$

By Method of least squares, the constants a & b are chosen in such a way that the sum of the square of errors / residuals is minimum.

for Min. value of U

$$\frac{\partial U}{\partial a} = 0$$

$$2\sum_{i=1}^{n} (y_i - a - bx_i)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} 1 - b\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \boxed{\Sigma y = a \cdot n + b\Sigma x} \quad \text{---(2)}$$

$$\frac{\partial U}{\partial b} = 0$$

$$2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0$$

$$\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2$$

$$\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2} \quad \text{---(3)}$$

(2) & (3) are Normal eqns of $y = a + bx$

Now $\Sigma x \Sigma y = na\Sigma x + b(\Sigma x)^2$ from (2)

$n\Sigma xy = na\Sigma x + bn\Sigma x^2$ from (3)

$$\Sigma x \Sigma y - n\Sigma xy = b\left[(\Sigma x)^2 - n\Sigma x^2\right]$$

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

Date:- / /
Page:-

from (2) $\Sigma y = na + b\Sigma x$

$\Rightarrow a = \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$

$\Rightarrow a = \bar{y} - b\bar{x}$

$\Rightarrow \boxed{\bar{y} = a + b\bar{x}}$

$\Rightarrow$ line $y = a + bx$ passes through the point $(\bar{x}, \bar{y})$

Putting $a = \bar{y} - b\bar{x}$ in eqn $y = a + bx$

$y = \bar{y} - b\bar{x} + bx$

$y - \bar{y} = b(x - \bar{x})$

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

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

called Regression coefficient of y on x

In eqn (3) shifting the origin to $(\bar{x}, \bar{y})$, we get

$\Sigma(x - \bar{x})(y - \bar{y}) = a\Sigma(x - \bar{x}) + b\Sigma(x - \bar{x})^2$

$\because \Sigma(x - \bar{x}) = \mu_1 = 0$

$\Rightarrow n\sigma_x\sigma_y = a \cdot 0 + b \cdot n\sigma_x^2$ $\quad \sigma_x^2 = \frac{\Sigma(x - \bar{x})^2}{n}$

$\Rightarrow b = \boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}}$ $\quad r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y}$

called Regression coef.

Date:- / /
Page:-

Hence line of regression of y on x is

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

where

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

or $$b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

II. Line of regression of x on y →

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

where

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

or $$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

Note:-

if $r=0$, the two line of regression become $y=\bar{y}$ and $x=\bar{x}$ which are two straight lines parallel to x and y axes respectively and passing through their means $\bar{y}$ and $\bar{x}$. They are mutually perpendicular.
if $r=\pm 1$, the two line of regression will coincide.

Date:- / /
Page:-

(#) Properties of Regression :—

(1.) correlation coefficient is the geometric mean btw the regression coefficients

Regression coefficients $b_{yx} = r\frac{\sigma_y}{\sigma_x}$, $b_{xy} = r\frac{\sigma_x}{\sigma_y}$

G.M of b_{yx} and b_{xy}

$$= \sqrt{b_{yx} \times b_{xy}}$$

$$= \sqrt{r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y}}$$

$$= \sqrt{r^2}$$

$$= r$$

(2.) Arithmatic mean of regression coefficient is greater than the correlation coefficient.

we have $\frac{b_{yx} + b_{xy}}{2} > r$

Now $\frac{1}{2}\left(r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y}\right) > r$

$\Rightarrow r\left(\frac{\sigma_y^2 + \sigma_x^2}{2\sigma_x\sigma_y}\right) > r$

$\Rightarrow \frac{\sigma_y^2 + \sigma_x^2}{2\sigma_x\sigma_y} > 1$

$\Rightarrow \sigma_y^2 + \sigma_x^2 - 2\sigma_x\sigma_y > 0$

$\Rightarrow (\sigma_x - \sigma_y)^2 > 0$

which is true

(3) b_{yx}, b_{xy} and r have same sign.

Date:- / /
Page:-

Imp

(#) Angle between two lines of Regression :-

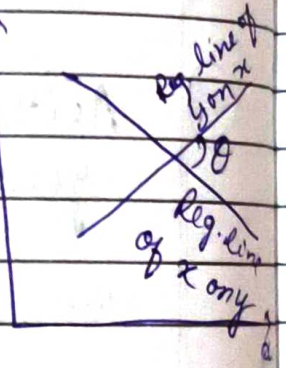
if θ is the acute angle between the two Regression lines in the case of two variables x and y, Show that

$$\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

where r, σ_x, σ_y are correlation coef. and Regression coefficients

explain the significance of the formula $r=0$ and $r = \pm 1$. (2015, 17)

Proof:- Eqn of line of Regression y on x



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

$$\Rightarrow y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

its slope $m_1 = r\frac{\sigma_y}{\sigma_x}$ — (1)

eqn of line of Regression of x on y

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

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

So $$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

or $$y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x})$$

Date:- / /
Page:-

its slope $m_2 = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$ — (2)

for angle $\tan\theta = \pm \frac{m_2 - m_1}{1 + m_1 m_2}$

$$= \pm \frac{\frac{\sigma_y}{r\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{\sigma_y}{r\sigma_x}\cdot\frac{r\sigma_y}{\sigma_x}}$$

$$= \pm \frac{\sigma_y}{\sigma_x}\cdot\frac{1-r^2}{r}\cdot\frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}$$

$$= \pm \frac{1-r^2}{r}\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

Since $-1 \le r \le 1 \Rightarrow r^2 \le 1$
and σ_x, σ_y are positive

∴ +ve sign gives the acute angle b/w the lines

Hence $$\boxed{\tan\theta = \frac{1-r^2}{r}\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}}$$

(1) When $r = 0 \Rightarrow \tan\theta = \infty \Rightarrow \theta = \pi/2$
∴ two lines of regression are perpendicular to each other

(2) When $r = \pm 1 \Rightarrow \tan\theta = 0 \Rightarrow \theta = 0$ or π
Hence the lines of regression coincide & there is perfect correlation b/w x & y.

Date:- / /
Page:-

Ex:-1. if the Regression coefficients are 0.8 and 0.2 what would be the value of coefficient of correlation ?

Sol^n:- we Know that $r = \sqrt{byx \cdot bxy}$

or $r^2 = byx \cdot bxy$

$= 0.8 \times 0.2$

$= 0.16$

Since r has the same sign as both the regression coefficients byx and bxy.

Hence

$r = \sqrt{0.16}$

$= 0.4$

Ex:-2. Calculate linear regression coefficients from the following data :

x:	1	2	3	4	5	6	7	8
y:	8	7	10	12	14	17	20	24

Sol^n :-

Linear regression coefficients are given by

$$byx = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

and $$bxy = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

Here $n = 8$

$$byx = \frac{8 \times 599 - 36 \times 107}{8 \times 204 - (36)^2} = \frac{940}{331} = 2.7796$$

and $$bxy = \frac{8 \times 599 - 36 \times 107}{8 \times 1763 - (107)^2} = \frac{940}{2655} = 0.3540$$

Date:- / /
Page:-

table :-

x	y	x^2	y^2	xy
1	3	1	9	3
2	7	4	49	14
3	10	9	100	30
4	12	16	144	48
5	14	25	196	70
6	17	36	289	102
7	20	49	400	140
8	24	64	576	192
$\Sigma x = 36$	$\Sigma y = 107$	$\Sigma x^2 = 204$	$\Sigma y^2 = 1763$	$\Sigma xy = 599$

2012,19 Ex:-3 The following data regarding the heights (y) and the weights (x) of 100 college students are given
$\Sigma x = 15000$, $\Sigma x^2 = 2272500$, $\Sigma y = 6800$, $\Sigma y^2 = 463025$, $\Sigma xy = 1022250$
find the Regression line of height on weight.

Sol^n:-
given $\Sigma x = 15000$, $\Sigma x^2 = 2272500$
$\Sigma y = 6800$, $\Sigma y^2 = 463025$
$\Sigma xy = 1022250$, $n = 100$
we have to find Regression line y on x which is

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

where

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

$$= \frac{100 \times 1022250 - 15000 \times 6800}{100 \times 2272500 - 15000^2}$$

$$b_{yx} = 0.1$$

Date:- / /
Page:-

also $\bar{x} = \frac{\Sigma x}{n} = \frac{15000}{100} = 150$

and $\bar{y} = \frac{\Sigma y}{n} = \frac{6800}{100} = 68$

∴ Regression line of height (y) on weight (x)

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

$$y - 68 = 0.1(x - 150)$$

$$y = 0.1x - 15 + 68.$$

$$\boxed{y = 0.1x + 53}$$

Ex°4. The following table gives ~~ex~~ age (x) in years of cars and annual maintenance cost (y) in hundred rupees:

X:	1	3	5	7	9
Y:	15	18	21	23	22

Estimate the maintenance cost for a 4-years car after finding the regression eqⁿ.

Sol^n:-

x	y	xy	x^2
1	15	15	1
3	18	54	9
5	21	105	25
7	23	161	49
9	22	198	81
$\Sigma x = 25$	$\Sigma y = 99$	$\Sigma xy = 533$	$\Sigma x^2 = 165$

Here $n = 5$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{25}{5} = 5, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{99}{5} = 19.8$$

Date:- / /
Page:-

Regression coefficient

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

$$= \frac{5 \times 533 - 25 \times 99}{5 \times 165 - (25)^2}$$

$$= 0.95$$

Hence Maintenance cost on age of cars (x) is given by Regression line

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

$$y - 19.8 = 0.95(x - 5)$$

$$y - 19.8 = 0.95x - 5 \times 0.95$$

$$y = 0.95x + 15.05$$

when $x = 4$ years

$$y = 0.95 \times 4 + 15.05$$

$$= 18.85 ₹.$$

$$\boxed{y = 18.85 ₹}$$

2018

Exo 5 In a partially destroyed laboratory record of an analysis of a correlation data the following results are only legible:

Variance of $x = 9$

Regression eqn: $8x - 10y + 66 = 0$

$40x - 18y = 214$

What were (a) the mean values of x and y

(b) the standard deviation of y

(c) the coeff of correlation b/w x & y.

Date:- / /
Page:-

Soln:- given, variance of $x = \sigma_x^2 = 9$ — (1)

Regression eqn:

$8x - 10y + 66 = 0$ — (2)

$40x - 18y - 214 = 0$ — (3)

Since regression lines (2) and (3) passes through the central point $(\bar{x}, \bar{y})$ then

$8\bar{x} - 10\bar{y} + 66 = 0$ — (4)

$40\bar{x} - 18\bar{y} - 214 = 0$ — (5)

from (4) & (5),

$40\bar{x} - 50\bar{y} = -330$

$40\bar{x} - 18\bar{y} = 214$

$+ \quad - \quad +$

$32\bar{y} = 544$

$\bar{y} = \frac{544}{32} = 17$

$\boxed{\bar{y} = 17}$

using in eqn (4)

$8\bar{x} - 10 \times 17 = -66$

$8\bar{x} = -66 + 170$

$8\bar{x} = 104$

$\boxed{\bar{x} = 13}$

eqn (2) and (3) can be written as

$y = 0.8x + 6.6$ (R.L of y on x)

$\therefore$ Reg. Coeff. of y on x

$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.8$ — (6)

Date:- / /
Page:-

eqn (3) can be written as

$$40x = 18y + 214$$

Reg. line of x on y

$$x = 0.45y + 5.35$$

∴ Reg. coef. of x on y

$$b_{xy} = \rho \frac{\sigma_x}{\sigma_y} = 0.45 \quad \text{—(7)}$$

Multiplying eqn (6) & (7).

$$r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y} = 0.8 \times 0.45$$

$$r^2 = 0.36$$

Correlation coef. $\boxed{r = 0.6}$

from (6), $r\frac{\sigma_y}{\sigma_x} = 0.8$

$$\sigma_y = \frac{0.8 \times \sigma_x}{r}$$

from (1)

$$= \frac{0.8 \times 3}{0.6}$$

$$= 4$$

Hence S.D of y $\boxed{\sigma_y = 4}$

& variance $\sigma_y^2 = 16$.

Date:- / /
Page:-

2019

Ex:- The equations of two regression lines, obtained in a correlation analysis of 60 observations are

$$5x = 6y + 24$$
$$1000y = 768x - 3608$$

What is the correlation coefficient? Show that the ratio of coefficient of variability of x to that of y is $\frac{5}{24}$. What is the ratio of variances of x and y?

Soln:-

Regression line of x on y

$$5x = 6y + 24$$
$$x = \frac{6}{5}y + \frac{24}{5}$$

∴ Regression coef. of x on y is

$$b_{xy} = \frac{6}{5} = r\frac{\sigma_x}{\sigma_y} \quad \text{— (1)}$$

Regression line of y on x

$$1000y = 768x - 3608$$
$$y = 0.768x - 3.608$$

∴ Regression coef. of y on x is

$$b_{yx} = 0.768 = r\frac{\sigma_y}{\sigma_x} \quad \text{— (2)}$$

Multiplying eqn (1) & (2),

$$b_{xy} \times b_{yx} = \frac{6}{5} \times 0.768$$

$$r\frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x} = 0.9216$$

$$r^2 = 0.9216$$

correlation coef. $\boxed{r = 0.96}$

Date:- / /
Page:-

Now dividing eq^n (1) and (2)

$$\frac{\sigma_x^2}{\sigma_y^2} = \frac{6}{5 \times 0.768} = 1.5625$$

Taking square root on both sides,

$$\frac{\sigma_x}{\sigma_y} = 1.25 = \frac{5}{4}$$

Ration of variances of x and y

$$\frac{Var(x)}{Var(y)} = \frac{\sigma_x^2}{\sigma_y^2} = \frac{5^2}{4^2} = \frac{25}{16}$$

Now Since the Regression line passes through the point $(\bar{x}, \bar{y})$

$$5\bar{x} = 6\bar{y} + 24 \quad \text{— (3)}$$

$$1000\bar{y} = 768\bar{x} - 3608 \quad \text{— (4)}$$

Solving eq^n (3) & (4), we get

$$\bar{x} = 6, \quad \bar{y} = 1$$

Coef. of variability of $x = \frac{\sigma_x}{\bar{x}}$

coef. of variability of $y = \frac{\sigma_y}{\bar{y}}$

Ratio of coef. of variability of x & y

$$= \frac{\sigma_x}{\bar{x}} \times \frac{\bar{y}}{\sigma_y}$$

$$= \frac{\bar{y}}{\bar{x}} \times \left(\frac{\sigma_x}{\sigma_y}\right) = \frac{1}{6} \times \frac{5}{4} = \frac{5}{24}$$

Date:- / /
Page:-

⊛ Polynomial fit: Non-linear Regression :-

Let $y = a + bx + cx^2$ —①
be a second degree parabolic curve of regression of y on x to be fitted for the data (x_i, y_i) $i = 1, 2, \ldots, n$.
Normal eq^{n}s of ① are

$$\Sigma y = a \cdot n + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

we find a, b and c from these eq^{n}s and put in eq^n ①

Ex:- fit a second degree parabolic curve of regression of y on x to the following data;

x:	1	2	3	4
y:	6	11	18	27

Sol^n:-

Let $y = a + bx + cx^2$ be the IInd degree parabolic curve to be fitted in the data its normal eq^{n}s are

$$\Sigma y = a \cdot n + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

Here $n = 4$

Date:- / /
Page:-

x	y	x^2	x^3	x^4	xy	x^2y
1	6	1	1	1	6	6
2	11	4	8	16	22	44
3	18	9	27	81	54	162
4	27	16	64	256	108	432
$\Sigma x=10$	$\Sigma y=62$	$\Sigma x^2=30$	$\Sigma x^3=100$	$\Sigma x^4=354$	$\Sigma xy=195$	$\Sigma x^2y=644$

∴ Normal eqns are

$$62 = 4a + 10b + 30c$$
$$195 = 10a + 30b + 100c$$
$$644 = 30a + 100b + 354c$$

Solving these eqns we get

$$a = 3,\ b = 2,\ c = 1$$

Hence $y = 3 + 2x + x^2$ is required regression parabolic curve of y on x.