Statistical Techniques - I

Measure of Central Techniques

Mean	Median	Mode

Arithmetic Mean If $x_1, x_2 \dots x_n$ are n nos then

$$AM = \bar{x} = \frac{x_1 + x_2 + \dots x_n}{n}$$

If x_1 occurs f_1 times, x_2 occurs f_2 times and so on

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_i} = \sum_{i=1}^{n} \frac{x_i f_i}{f_i}$$

Q: Find the mean of the following

Numbers	frequency	$x_i f_i$
8	5	40
10	8	80
15	8	120
20	4	80
	25	320

$$\bar{x} = \frac{320}{25} = 12.8$$

Shortcut Method Let a be assumed mean

$$\frac{\Sigma f_i d_i}{\Sigma f_i} = \frac{\Sigma f_i (x_i - a)}{\Sigma f_i} = \frac{\Sigma f_i x_i}{\Sigma f_i} - a \frac{\Sigma f_i}{\Sigma f_i}$$

$$= AM - a$$

$$AM = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

Q) Find mean :

Class	frequency	x_i	$d_i = x_i - a$	$f_i d_i$
0-10	7	5	-20	-140
10-20	8	15	-10	-80
20-30	20	25	0	0
30-40	10	35	10	100
40-50	5	45	20	100

$a = 25$

$$AM = a + \frac{\Sigma f_i d_i}{\Sigma f_i} = 25 + \frac{(-20)}{50} = \frac{125 - 2}{5} = \frac{123}{5}$$

$$= 24.6$$

Step-Deviation Method:

$$u = \frac{x_i - a}{i} \qquad AM = a + \frac{\Sigma f u}{\Sigma f} \times i$$

where i = width of the class interval

Q:

class	freq	x_i	$d_i = x_i - a$	$u = \frac{x_i - a}{10}$	$f_i u_i$
0-10	7	5	-20	-2	-14
10-20	8	15	-10	-1	-8
20-30	20	25	0	0	0
30-40	10	35	10	1	10
40-50	5	45	20	2	10
					-2

$$AM = a + \frac{\Sigma f u}{\Sigma f} \times i$$

$$= 25 + \frac{(-2)}{50} \times 10 = 24.6$$

Q:

Marks	Stud	x	No. of stud (f_i)	x_i	$d_i = x_i - a$	$u = \frac{x_i - a}{10}$	$f_i u_i$
below 10	5	0-10	~~5~~ 5	5	-50	-5	-25
20	9	10-20	~~15~~ 4	15	-40	-4	-16
30	17	20-30	~~25~~ 8	25	-30	-3	-24
40	29	30-40	~~35~~ 12	35	-20	-2	-24
50	45	40-50	~~45~~ 16	45	-10	-1	-16
60	60	50-60	~~55~~ 15	55	0	0	0
70	70	60-70	~~65~~ 10	65	10	1	10
80	78	70-80	~~75~~ 8	75	20	2	16
90	83	80-90	~~85~~ 5	85	30	3	15
100	85	90-100	~~95~~ 2	95	40	4	8
			85				-86

$$AM = a + \frac{\Sigma f u}{\Sigma f i} \times i = 55 + \frac{(-56)}{85} \times 10 = 48.42$$

Q: The mean of 200 items was 50 later on it was discoverd that 2 items were mis read as 92 and 8 instead of 192 and 88 find the correct mean ?

$AM = 50 = \bar{x}$ $\quad n = 200$

$$\bar{x} = \frac{\Sigma x_i}{n} \Rightarrow 50 = \frac{\Sigma x_i}{200}$$

$$\Sigma x_i = 10000$$

$$\Sigma x_i = 10000 - (92+8) + (192+88) = 10180$$

$$\text{Now}, \bar{x}' = \frac{10180}{200} = \cancel{50.9} \ 50.9$$

Q: Find the mean

class interval	f	x_i	$d_i = x_i - a$	$u_i = \frac{d_i}{10}$	$f_i u_i$
15 - 25	6	20	-30	-3	-18
25 - 35	11	30	-20	-2	-22
35 - 45	7	40	-10	-1	-7
45 - ~~50~~ 55	4	50	0	0	0
55 - 65	4	60	10	1	4
65 - 75	2	70	20	2	4
75 - 85	1	80	30	3	3
	35				-36

$$AM = a + \frac{\Sigma f u}{\Sigma f} \times i = 50 + \frac{(-36)}{35} \times 10$$

$$= 39.72$$

Median :

It is defined as the measure of the central item when they are arranged in ascending or descending order of the magnitude.

» when the number of n is odd $\frac{n+1}{2}$ in item gives the median.

» When the total number of the frequency is even say n then there are two middle item and so the $\frac{1}{2}n^{th}$ and $\frac{n}{2}+1$ th item is then the arithemetic mean of both value gives the median

For Grouped Data

$$\boxed{\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i}$$

N - stands for total frequency

C = Cumulative frequency just preceding the median class

i = width of class

f - frequency of median class

Q: Find the median of following

	CI	f	CF
	65 - 85	4	4
	85 - 105	5	9
	105 - 125	13	22
l -	125 - 145	20 f	42
	145 - 165	14	56
	165 - 185	8	64
	185 - 205	4	68

$$\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i$$

$$= 125 + \frac{\frac{68}{2} - 22}{20} \times 22$$

$$= 137$$

Q: Find the median

No. of days	x	No. of student (f) cf	f
less than 5	0-5	29	29
10	5-10	224	195
15	10-15	465	241
20	15-20	582	117
25	20-25	634	52
30	25-30	644	10
35	30-35	650	6
40	35-40	653	3
45	40-45	655	2

c

$$\text{Median} = l + \frac{\frac{N}{2} - c}{f} \times i = 10 + \frac{\frac{655}{2} - 224}{241} \times i$$

Mode

Mode is defined to be the size of variable which occurs most frequently

Eg: Find the mode

0, 1, 6, 7, 2, 3, 7, 6, 6, 2, 6, 0, 5, 6, 0 Ans = 6 mode

For grouped data

$$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \times i$$

l = lower limit of class

f = freq of modal class

f_{-1} = frequency before the modal class

f_1 = frequency after the modal class

i = width of the class

Q: Find the mode:

Age	frequency	
0-6	6	
6-12	11	
12-18	25	→ f_{-1}
18-24	35	f
24-30	18	f_1
30-36	12	
36-42	6	

$$\text{Mode} = 18 + \frac{35-25}{2\times35-25-18} \times 6$$

$$= 18 + \frac{10}{70-25-18} \times 6$$

$$= 18 + \frac{60}{27} = 20.22$$

Q: Obtain the median

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

x	f	cf
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	~~8~~15	105
8	9	114
9	6	120
	120	

$$\frac{120+1}{2} = \frac{121}{2} = 60.5$$

mode = 20

Geometric Mean

If $x_1, x_2, \ldots x_n$ be n value the GM $(x_1, x_2 \cdots x_n)^{1/n}$

Harmonic Mean

If $x_1, x_2 \ldots x_n$ are $\frac{1}{H} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n}\right)$

Q: Calculate the GM and HM of the following

4 8 16

$$GM = (4 \times 8 \times 16)^{1/3} = 8$$

$$HM = \frac{1}{3}\left(\frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right)$$

$$GM = (\ldots x_n)^{\frac{1}{n}}$$

$$HM = \frac{1}{n}\left(\ldots + \frac{1}{n}\right)$$

30/1/23

Q: The mean annual salary of all employee of a company was ₤ 5000. The mean annual salary paid to mail and female employee were ₤ 5200 and ₤ 4200 rupees respectively. Determine the % of male and female employee by the company.

$$p_1 + p_2 = 100$$

$$\bar{x} = \frac{p_1 \bar{x}_1 + p_2 + \bar{x}_2}{p_1 + p_2}$$

$$5000 = \frac{p_1 \times 5200 + p_2 \times 4200}{p_1 + p_2}$$

$$200 p_1 - 800 p_2 = 0$$

$$200 p_1 - 800(100 - p_1) = 0$$

$p_1 = 80$

$p_2 = 20$

» Partition Values

Those are the values of the variate which divide the total frequency into a number of equal parts.

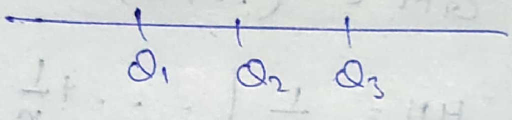
Median be that value of the variate which divide the total frequency into two equal parts.

(i) **Quartiles :** Those values of the variant which divide the total frequency into 4 equal parts.

$Q_1 = l + \frac{i}{f}\left(\frac{N}{4} - c\right) \rightarrow$ Lower Quartiles

$Q_2 = l + \frac{i}{f}\left(\frac{2N}{4} - c\right) \rightarrow$ Median

$Q_3 = l + \frac{i}{f}\left(\frac{3N}{4} - c\right) \rightarrow$ Upper Quartiles



Deciles : Those value of the variate which divide the total frequency into ten equal parts.

$D_1, D_2, \ldots$ denote respectively first, ~~secon~~ second, ... deciles

$D_1 = l + \frac{i}{f}\left(\frac{N}{10} - c\right) \rightarrow$ lower quartiles

$D_2 = l + \frac{i}{f}\left(\frac{2N}{10} - c\right)$ | $D_5 = l + \frac{i}{f}\left(\frac{5N}{10} - c\right)$

Median

$D_7 = l + \frac{i}{f}\left(\frac{7N}{10} - c\right)$

Percentile

Percentile are those values of the variate which divide the total frequency into 100 equal parts.

$P_1, P_2 \ldots$ denote respectively the first, second .... percentile.

The $(P_{50}) \rightarrow 50$th is the Median

Q: Find the median lower, upper Quantiles of the following

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

$$Q_1 = l + \frac{i}{f}\left(\frac{N}{4} - c\right)$$

$$= 60 + \frac{10}{71}\left(\frac{249}{4} - 127\right) = 50.88$$

$$Q_2 = 60 + \frac{10}{71}\left(\frac{249}{2} - 127\right) = 59.64789$$

$$Q_3 = 60 + \frac{10}{71}\left(\frac{3 \times 249}{4} - 127\right) = 68.4155$$

Q: Obtain the median

x	f	CF
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

$$M = \frac{N+1}{2} = \frac{120+1}{2} = \frac{121}{2}$$

$$= 60.5$$

Q: Calculate the arithmetic mean

C I	frequency
0 - 1	8
1 - 3	8
3 - 5	10
5 - 10	12
10 - 15	18
15 - 25	11
25 - 28	10
28 - 30	9
30 - 45	8
45 - 60	6

Q: Calculate the AM

Height	No. of student
219	2
216	4
213	6
210	10
207	11
204	7
201	5
198	4
195	1

31/01/23

Q: Calculate the mean and median of the..

class interval	frequency	CF	x_i	$x_i f_i$
6.5 - 2.5	5	5	7	35
7.5 - 8.5	12	17	8	96
8.5 - 9.5	25	42 c	9	225
9.5 - 10.5	48 f	90	10	480
10.5 - 11.5	32	122	11	352
11.5 - 12.5	6	128	12	72
12.5 - 13.5	1	129	13	13
				1273

$$\text{Median} = l + \frac{\frac{N}{2} - c}{f} \times i = 9.5 + \frac{\frac{129}{2} - 42}{48} = 9.9688$$

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1273}{129} = 9.868$$

Moments

Moments are statistical tool used in statistical investigation. The moment of a distribution are the arithmatic mean of the various power of deviation item from some given no.

Moments

- Moments about mean (μ_r)
- Moments about an arbitary number (a) (μ_r')
- Moment about an origin (γ_r)

$$\mu_r = \frac{\Sigma f_i (x_i - \bar{x})^r}{\Sigma f_i} \qquad \mu_r' = \frac{\Sigma f_i (x_i - a)^r}{\Sigma f_i} \qquad \gamma_r = \frac{\Sigma f_i x_i^r}{\Sigma f_i}$$

Moment about Mean (μ_r)

≫ $\mu_0 = \frac{\sum f_i (x_i - \bar{x})^0}{\sum f_i} = \frac{\sum f_i}{\sum f_i} = 1$ $\boxed{\mu_0 = 1}$

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

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

≫ $\mu_2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} =$ Variance $= (S.D)^2$

≫ $\mu_3 = \frac{\sum f_i (x_i + \bar{x})^3}{\sum f_i}$ $\mu_r = \frac{\sum f_i (x_i + \bar{x})^r}{\sum f_i}$

Q: Calculate $\mu_1, \mu_2, \mu_3, \mu_4$.

Marks	frequency	x_i	$f_i x_i$	$\bar{x} = 34$	$f_i(x_i - \bar{x})$	$f_i(x_i - \bar{x})^2$
5 - 15	10	10	100	-24	-240	5760
15 - 25	20	20	400	-14	-280	3920
25 - 35	25	30	750	-4	-100	400
35 - 45	20	40	800	6	120	720
45 - 55	15	50	750	16	240	3840
55 - 65	10	60	600	26	260	6760
	100		3400		0	21400

$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{3400}{100} = 34$

$\mu_1 = \frac{\sum f_i (x_i - \bar{x})}{\sum f_i} = \frac{0}{100}$

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

$= \frac{21400}{100} = 214$

NOTE: If $x_1\ x_2 \dots x_n$ are n variable then

$\mu_r = \frac{\sum (x_i - \bar{x})^r}{n}$, n = total number of the variable

Scanned with OKEN Scanner

Q: Find the first four moments.

x	3	6	8	10	18

x	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
3	-6	36	-216	1296
6	-3	9	-27	81
8	-1	1	-1	1
10	1	1	1	1
18	9	81	729	6561
	0	128	486	7940

$\bar{x}$ = AM

$= \frac{\text{Sum of } x_i}{n}$

$= \frac{3+6+8+10+18}{5}$

$= 9$

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

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

$\mu_3 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{486}{5} = 97.2$

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

Moment about mean → Central moment

Moment about arbitary no. → Raw moment

1/2/23

Q: Calculate the first 4 momt about mean

$f_i(x-\bar{x})^3$	No. of flowers x_i	No. of plants f_i	f_ix_i	$x-\bar{x}$	$f_i(x-\bar{x})$	$f_i(x-\bar{x})^2$
-2197	3	1	3	-13	-13	169
-2000	6	2	12	-10	-20	400
-192	12	3	36	-4	-12	144
0	16	4	648	0	0	0
2645	25	5	125	9	45	405
-744		15	240		0	822

$\bar{x} = \frac{\Sigma f_ix_i}{\Sigma f_i} = \frac{240}{15} = 16$

$\mu_1 = 0$

$\mu_2 = \frac{\Sigma f_i(x-\bar{x})^2}{\Sigma f_i} = \frac{822}{15} = 54.8$

$\mu_3 = \frac{\Sigma f_i(x-\bar{x})^2}{\Sigma f_i}$

$= \frac{-744}{15}$

$= -49.6$

Moment about an arbitary number (Raw moments)

$$\mu_r' = \frac{\sum f_i (x_i - a)^r}{\sum f_i} \qquad \mu_0' = \frac{\sum f_i (x_i - a)^0}{\sum f_i} = \frac{\sum f_i}{\sum f_i} = 1$$

$$\mu_1' = \frac{\sum f_i (x_i - a)^1}{\sum f_i} = \frac{\sum f_i x_i}{\sum f_i} = a \frac{\sum f_i}{\sum f_i} = a = \bar{x} - a$$

$$\mu_2' = \frac{\sum f_i (x_i - a)^2}{\sum f_i}$$

Moment about an origin (γ_r)

$$\gamma_r = \frac{\sum f_i x_i^r}{\sum f_i}$$

$$\gamma_0 = \frac{\sum f_i x_i^0}{\sum f_i} = \frac{\sum f_i}{\sum f_i} = 1$$

$$\gamma_1 = \frac{(\sum f_i x_i)^1}{\sum f_i} = \bar{x}$$

$$\gamma_2 = \frac{(\sum f_i x_i)^2}{\sum f_i} =$$

Q) ~~Stat~~ Calculate the moment about $x = 4$ for the following

x_i	f_i	$x_i - a$ ($a = 4$)	$f_i(x_i - a)$	$f_i(x_i - a)^2$	$f_i(x_i - a)^3$
0	1	-4	-4	16	-64
1	9	-3	-27	81	-243
2	26	-2	-52	104	-208
3	59	-1	-59	59	-59
4	72	0	0	0	0
5	52	1	52	52	52
6	29	2	58	116	232
7	7	3	21	67	189
8	1	4	4	16	64
	256		-7	511	-37

$\mu_1 = 0.027$

$\mu_2 = 1.996$

$\mu_3 = -0.1445$

Q: Cal the first 3 moment about mean

CI	f	x_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
0-10	10	5	50	-20	400	4000
10-20	20	15	300	-10	100	2000
20-30	40	25	1000	0	0	0
30-40	20	35	700	10	100	2000
40-50	10	45	450	20	400	4000
	100		2500		1000	12000

Relation b/w μ_r and μ_r'

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

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

$$= \frac{\sum f_i [(x_i - A) - (\bar{x} - A)]^r}{}$$

$$= \frac{\sum f_i [(x_i - A) - \mu_1']^r}{\sum f_i}$$

$$(a-b)^r = {}^rC_0\, a^r - {}^rC_1\, a^{r-1} b + {}^rC_2\, a^{r-2} b^2 \ldots + (-1)^r b^r$$

$$= \frac{\sum f_i \left({}^rC_0 (x_i - A)^r - {}^rC_1 (x_i - A)^{r-1} \mu_1' + {}^rC_2 (x_i - A)^{r-2} \cdot \mu_1'^2 + \ldots (-1)^r \mu_1'^2 \right)}{\sum f_i}$$

$$= \frac{\sum f_i (x_i - A)^r}{\sum f_i} - \frac{{}^rC_1 \sum f_i (x_i - A)^{r-1} \times \mu_1'}{\sum f_i}$$

$$+ {}^rC_2 \sum f_i (x_i - A)^{r-2} \mu_1'^2$$

$$+ \ldots$$

$\mu_r = \mu'_r - {}^rc_1 \mu'_{r-1}\mu'_1 + {}^rc_2 \mu'_{r-2} \mu_1'^2 - {}^rc_3 \mu'_{r-3}\mu_1'^3 \ldots$

for $r=1$

$\mu_1 = \mu_1' - {}^1c_1 \mu_0' \mu_1' = \mu_1' - \mu_1' = 0$ $\quad \mu_1 = 0$

for $r=2$

$\mu_2 = \mu_2' - {}^2c_1 \mu_1' \mu_1' + {}^2c_2 \mu_0' \mu_1'^2$

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

for $r=3$

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

for $r=4$

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

Relation b/w v_r and μ_r :

$v_1 = \bar{x}$

$v_2 = \mu_2 + \bar{x}^2$

$v_3 = \mu_3 + 3\mu_2\bar{x} + \bar{x}^3$

$v_4 = \mu_4 + 4\mu_3 x + 6\mu_2\bar{x}^2 + \bar{x}^4$

Karl Pearson's β and γ coefficients :

$\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$ $\quad \gamma_1 = +\sqrt{\beta_1}$

$\gamma_2 = \beta_2 - 3$

$\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

Q: The first the moment of a distribution about the value 2 of varible 1, 16, and -40. Show that mean is 3 variance is 15 and $\mu_3 = -86$

$\mu_1' = 1$, $\mu_2' = 16$ $\mu_3' = 40$ $A = 2$

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

$1 = \bar{x} - 2$ $\boxed{\bar{x} = 3}$

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

$16 - 1$

$\mu_2 = 15$

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

$= -40 - 3 \times 16 \times 1 + 1 = -86$

Q: The fits four moment of a distribution about $x = 1$ and 1, 2.5 and 16. Cal fit from moment about mean and about origin.

$u_1' = 1$ $u_2' = 2.5$ $u_3' = 5.5$ $u_4' = 16$

$u_1' = \bar{x} - A$

$1 = \bar{x} - 2$

$\bar{x} = 3$

$u_2 = u_2' - u_1'^2$

$= 2.5 - (1)^2$

$= 1.5$

$u_3 = u_3' - 3u_2' u_1' + 2u_1^3$

$= 0$

$\mu_4 = \mu_4' - 4\mu_3' u_1' + 6\mu_2' u_1^2 - 3u_1^4$

$= 16 - 4 \times 5.5 \times 1 + 6 \times 2.5 \times 1 - 3$

$= 13 - 22 + 15$

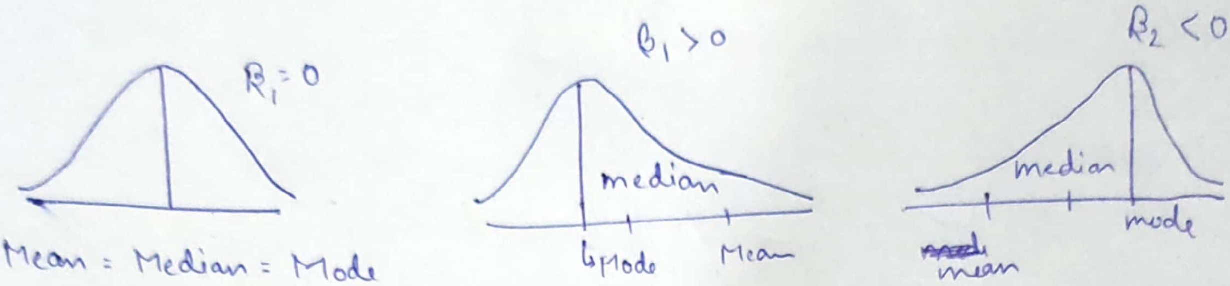
$= 6$

Scanned with OKEN Scanner

6/2/23

Skewness :- Lack of Symmetry

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



Mean = Median = Mode
Symmetrical distribution

+ve Skewed distribution

−ve Skewed distribution

Karl Pearson's Coeff of skewness $= \dfrac{A.M - Mode}{S.D}$

$\because \{Mode = 3\,Median - 2\,A.M\}$

$Skp = A.M - 3\,Median + 2.AM$

$Skp = \dfrac{3(A.M - Median)}{S.D}$

Q: Karl Pearson coeff of skewness of a distribution is 0.32 S.D is 6.5 and mean is 29.6 mode = ?

$Skp = \dfrac{AM - Mode}{SD}$

$\rightarrow 0.32 = \dfrac{29.6 - Mode}{6.5}$

$Mode = -0.32 \times 6.5 + 29.6$

$= 27.52$

Q: The 1st four moment of a distribution about the value 5 of the variable 2, 20, 40, 50. Cal. the moment coeff of Skewness.

$$\boxed{\text{Moment coefficient of Skewness} = \frac{\mu_3}{\sqrt{\mu_2^3}}}$$

Ans = -1

$\pm\sqrt{\beta_1}$

$\mu_1' = 2 \quad \mu_2' = 20 \quad \mu_3' = 40 \quad \mu_4' = 50$

$\mu_1' = \bar{x} - A = \bar{x} - 5$

$2 = \bar{x} - 5 \quad \boxed{\bar{x} = 7}$

$\mu_2 = \mu_2' - \mu_1'^2 = 20 - 4 = 16$

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

$= 40 - 3 \times 20 \times 2 + 2 \times 8$

$= 40 - 120 + 16$

$= -64$

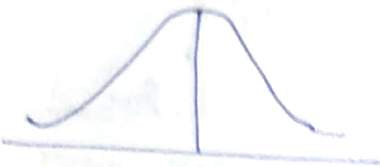
$$\text{Moment Coff} = \frac{64}{\sqrt{16^3}} = \frac{64}{16 \times 4} = -1$$

Measure of Kurtosis

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

$\beta_2 = 3$

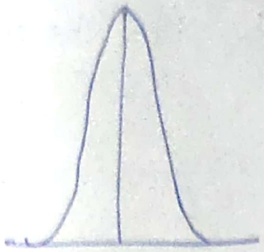
$\gamma_2 = \beta_2 - 3 = 0$



$\beta_2 = 3$

Mesokurtic

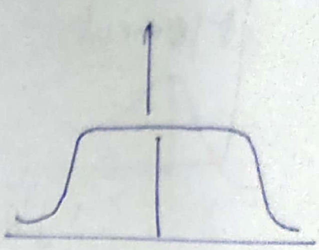
$\gamma_2 = 0 \ldots$



Leptokurtic

$\beta_2 > 3$

$\gamma_2 > 0$



Platykurtic

$\beta_2 < 3$

$\gamma_2 < 0$

Q: The first 4 moment of distribution about $x = 4$ are 1, 4, 10, 45. Obtain the various characteristi of ~~[illegible]~~ distribution on the basis of given info Comnt upon the nature of distribution.

$\mu_1 = 0 \quad \mu_2 = 3 \quad \mu_3 = 0 \quad \mu_4 = 26$

$\beta_2 = 2.89 < 3$ Platykurtic

Q: The following table represent the height of a batch of 100 student. Calculate kurtosis.

$h = 2$

Height	No. of stu (f)	$x-67$	$u = \frac{x-67}{h}$	fu	fu^2	fu^3	fu^4
59	0	-8	-4	0	0		
61	2	-6	-3	-6	18		
63	6	-4	-2	-12	24		
65	20	-2	-1	-20	20		
67	40	0	0	0	0		
69	20	2	1	20	20		
71	8	4	2	16	32		
73	2	6	3	6	18		
75	2	8	4	8	32		
				12			

$$\mu_1' = \frac{\sum f_i (\bar{x} - A)}{\sum f_i} \times h^i \qquad \frac{\sum f_i u^2}{\sum f_i} \times h^i$$

$u_1' = 0.24$ $\quad u_1 = 0$

$u_2' = 6.56$ $\quad u_2 = 6.5024$

$u_3' = 11.52$ $\quad u_3 =$

$u_4' = 176$ $\quad u_4 = 167.192$

$$u_2' = \frac{\sum f_i (x_i - A)^2}{\sum f_i} \times h^2$$

$$u_1' = \frac{12 \times 2}{100} = 0.24$$

$$u_2' = \frac{164}{100} \times 4 = 6.56$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3.4954 > 3$$

Leptokurtic

7/2/22

Q: Calculate the moment of coeff. of skewness

Class	f	x_i	A = 20 $x_i - A$	h = 5 $u = \frac{x_i - A}{h}$	fu	fu^2	fu^3
2.5 - 7.5	8	5	-15	-3	-24	72	-216
7.5 - 12.5	15	10	-10	-2	-30	60	-120
12.5 - 17.5	20	15	-5	-1	-20	20	-20
17.5 - 22.5	32	20	0	0	0	0	0
22.5 - 27.5	23	25	5	1	23	23	23
27.5 - 32.5	17	30	10	2	34	68	136
32.5 - 37.5	5	35	15	3	15	45	135
	120				-2	200	-62

$$\mu_r' = \frac{\Sigma f_i u_i^r}{\Sigma f_i} \times h^r$$

$$\mu_1' = \frac{-2}{120} \times 5 = -\frac{1}{12} = -0.083$$

$$\mu_2' = 60$$

$$\mu_3' = -64.583$$

$$\mu_1 = 0$$

$$\mu_1 = 0$$

$$\mu_2 = 59.993$$

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

$$= -64.583 + 3 \times 60 \times \frac{1}{12} + 2 \times \left(\frac{-1}{12}\right)^3$$

$$= -64.583 + 15 - 0.001143574$$

$$= 49.58$$

Moment coeff of skewness $= \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-49.58}{\sqrt{59.993^3}}$

Curve fitting :

Curve fitting by method of least square :

Curve fitting of straight line :

$xy = ax + bx^2 \longleftarrow y = a + bx \rightarrow$ (i)

$\sum_{i=1}^{n} y = na + b\sum x$ — (ii)

$\sum xy = a\sum x + b\sum x^2 \rightarrow$ (iii)

Find the value a & b from eq^n (ii) & (iii) and put value a & b in eq (i)

this is the best fit of the curve.

~~$y = a + bx + cx^2$~~

Curve fitting of second degree Parabola :

$y = a + bx + cx^2 \longrightarrow$ (i)

$\sum y = na + b\sum x + c\sum x^2 \rightarrow$ (ii)

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \rightarrow$ (iii)

$\sum x^2y = a\sum x^2 + b\sum x^3 + c\sum x^4 \rightarrow$ (iv)

find the value a b & c from eq (ii), (iii), (iv) and put value in eq (i) this is the best fit for the curve.

Q: By the method of least square find the straight line that best fits of the curve.

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

$y = a + bx$

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

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

$204 = 5a + 15b \rightarrow$ (ii)

$748 = 15a + 55b \rightarrow$ (iii)

(iii) − 3(ii)

$$748 = 15a + 55b$$
$$612 = 15a + 45b$$
$$-$$
$$136 = 10b$$
$$b = 13.6$$

The best fit of the curve is $y = 13.6x$

Q: By the method of least sq find the straight line that best fits of the curve:

x	y	x^2	xy
0	1	0	0
1	1.8	1	1.8
2	3.3	4	8
3	4.5	9	27
4	6.3	16	64
10	16.9	30	47.1

$$\Sigma y = na + b\Sigma x$$
$$16.9 = 5a + 10b \rightarrow (i)$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$
$$47.1 = 10a + 30b \rightarrow (ii)$$

$$47.1 = 10a + 30b$$
$$33.8 = 10a + 20b$$
$$(-)$$
$$13.3 = 10b$$
$$b = 1.33$$

$$a = \frac{47.1 - 30 \times 1.33}{10}$$
$$= 0.72$$

$$y = 0.72 + 1.33x$$

Q: By ....

x	y	x^2	xy	x^3	x^4
0	-4	0	0	0	0
1	-1	1	-1	1	1
2	4	4	8	8	16
3	11	9	33	27	81
4	20	16	80	64	256
30	30	30	120	~~389~~ 100	354

x^2y

$$\Sigma y = na + b\Sigma x$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$
$$30 = 5a + 10b \rightarrow (i)$$
$$120 = 10a + 30b \rightarrow (ii)$$
$$60 = 10b$$
$$b = 6 \qquad a = -6$$

$$\Sigma y = na + b\,\Sigma x + c\,\Sigma x^2$$

$$80 = 5a + 10b + 30c \rightarrow (1)$$

$$\Sigma xy = a\,\Sigma x + b\,\Sigma x^2 + c\,\Sigma x^3$$

$$120 = 10a + 30b + 100c \rightarrow (11)$$

$$\Sigma x^2 y = a\,\Sigma x^2 + b\,\Sigma x^3 + c\,\Sigma x^4$$

$$146 = 30a + 100b + 354c \rightarrow (111)$$

$a = -4 \qquad b = 2 \qquad c = 1$

8/2/23

Q: Given the following experimental values – fit by the method of least square a parabola of the type

x : 0 1 2 3

y : 2 4 10 5

$$y = a + bx^2$$

Let $x^2 = t$

$y = a + bt$

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

$\Sigma ty = a\Sigma t + b\Sigma t^2$

x	y	t = x^2	t^2 = x^4	ty = x^2y
0	2	0	0	0
1	4	1	1	4
2	10	4	16	40
3	5	9	81	45
	21	14	98	89

$\Sigma y = na + b\Sigma t \rightarrow 21 = 4a + 14b$

$\Sigma ty = a\Sigma t + b\Sigma t^2 \rightarrow 89 = 14a + 98b$

$$① \rightarrow \quad 147 = 28a + 98b$$
$$- \quad 89 = 14a + 98b$$
$$58 = 14a$$

$a = \cancel{8}\ \frac{29}{7} = 4.1$

$b = 0.31$

$y = 4.1 + 0.31x^2$

Fitting of an exponential curve :

$$\boxed{y = ae^{bx}}$$

taking log both sides

$$\log_{10} y = \log (ae^{bx})$$

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

$$\underbrace{\log_{10} y}_{Y} = \underbrace{\log a}_{A} + \underbrace{bx \log_{10} e}_{BX}$$

where, $Y = \log_{10} y$

$A = \log a$

$B = b \log_{10} e$

$X = x$

$$Y = A + BX \longrightarrow (1)$$

$$\Sigma Y = nA + B\Sigma X \longrightarrow (11)$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2 \longrightarrow (111)$$

Solution

> Find the value A & B from (11) & (111) and

$a =$ Antilog A and $b = \dfrac{B}{\log_{10} e}$

Q: Find the curves of best fit of the type $y = ae^{bx}$

x	y	$Y = \log_{10} y$	~~$X = x$~~	xY	x^2
1	10	1		1	1
5	15	1.1766	5	5.883	25
7	12	1.071		7.497	49
9	15	1.1766		10.589	81
12	21	1.322		15.8664	144
		5.7		40.8862	300

$\Sigma X = 34$

$\Sigma Y = 5.7$

$5.7 = 5A + 34B \rightarrow (1)$

$40.886 = 34A + 300B \rightarrow (11)$

$A = 0.9766$

$B = 0.02561$

$a = 10^A = 9.475$

$b = 0.05896$

So the best fit of the curve

$y = 9.745\, e^{0.5896 e}$

Scanned with OKEN Scanner

Q: Obtain a relation of the form $y = ab^x$

x	y	$Y = \log_{10} y$	xY	x^2
2	8.3	0.9190	1.838	4
3	15.4	1.187	3.561	9
4	33.1	1.519	6.076	16
5	65.2	1.814	9.07	25
6	127.4	2.105	12.63	36
20	249.4	7.604	33.175	90

$$\Sigma Y = nA + B\Sigma X$$

$$7.604 \ \cancel{249.4} = 5A + 20B \longrightarrow (1)$$

$$33.175 = 20A + 90B$$

$$\cancel{964.425 = 10B} \qquad 10B = 2.759$$

$$\cancel{B = 96.44} \qquad B = 0.2759$$

$$5A = 7.604 - 20 \times 0.3[illegible]$$

$$= 7.604 - [illegible]6$$

$$= \frac{1.6}{5} = 0.3$$

9/2/23

Q: The pressure of gas corresponding to various volume V is measured

V	P	$X = \log V$	$Y = \log P$	XY	X^2
50	64.7	1.6989	1.8109	3.0133	2.8865
60	51.3	1.7781	1.7101	3.0408	3.1618
70	40.5	1.8451	1.6074	2.9659	3.4044
90	25.9	1.9542	1.4133	2.7619	3.8190
100	78	2	1.8920	3.7841	4

Fit of the data to the eqn $PV^{r} = C$

$PV^{r} = C$

$P = CV^{-r}$

$$\underbrace{\log P}_{Y} = \underbrace{\log C}_{A} - \underbrace{r \log V}_{+\ B X}$$

where, $Y = \log P$

$A = \log C$

$B = -r$

$X = \log V$

$\sum Y = nA + B \sum X$

$8.4338 = 5A + 9.2764 B$

$15.6295 = 9.2764A + 17.2717 B$

$\sum Y = 8.4338$

$\sum X = 9.2764$

$\sum X^2 = 17.2717$

$\sum XY = 15.6295$

$A = 2.2247$ $\quad B = -0.2899$

$r = 0.2899$

$C = 10^{2.2247} = \cancel{167.7876}$ 167.7876

Required Curve $= PV^{0.2899} = 167.7876$

Method of least Square :

Step 1: Let (x_i, y_i) $i = 1, 2, \ldots n$ be a set of observation

$y = a + bx$ and first estimate error

$E_i = y_i - (a + bx_i)$

Step 2: Introduce a new quantity 'U' such that

$$U = \sum E_i^2 = \sum (y_i - (a + bx_i))^2$$

By the principle of least square U is minimum

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

$$\frac{\partial U}{\partial a} = 2\sum (y_i - (a + bx_i))(-1) = 0$$

$$\sum (y_i - (a + bx_i)) = 0$$

$$\sum y_i = na + b\sum x_i \longrightarrow (ii)$$

$$\frac{\partial U}{\partial b} = 2\sum (y_i - (a + bx_i))(-x_i) = 0$$

$$= \boxed{\sum x_i y_i = a\sum x_i + b\sum x_i^2} \longrightarrow (iii)$$

Eq (2) and (3) are "Normal Eqn" of the curve

$y = a + bx$

Q: $y = ax + bx^2$

$E_i = y_i - (ax + bx^2)$

$U = \sum E_i^2 = \sum (y_i - (ax + bx_i^2))^2$

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

$\frac{\partial U}{\partial a} =$ $\quad y = a + bx + cx^2$

$\sum xy = a\sum x^2 + b\sum x^3$

$\sum x^2 y = a\sum x^3 + b\sum x^4$

Q: By the method of least square ~~fitted~~ fits the curve

$y = ax + bx^2$

x	y	x^2	x^3	x^4	x^2y	xy
1	1.8	1	1	1	1.8	1.8
2	5.1	4	8	16	20.4	10.2
3	8.9	9	27	81	80.1	26.7
4	14.1	16	64	256	225.6	56.4
5	19.8	25	125	625	495	99.
15		55	225	979	822.9	194.1

Q: $194.1 = 55a + 225b$

$822.9 = 225a + 979b$

$a = 1.52 \qquad b = 0.49$ so the best fit

$y = 1.52x + 0.49x^2$

Q: Find the normal eq of curve

$y = \frac{C_0}{x} + C_1\sqrt{x}$

~~?~~ 7/2/23

Q: Use the method of least square to fit the curve $y = \frac{c_0}{x} + c_1\sqrt{x}$ to the following table

x	y	y/x	$1/x^2$	$1/\sqrt{x}$	$y\sqrt{x}$
0.1	21	210	100	3.1622	6.6407
0.2	11	55	25	2.2360	4.9193
0.4	7	17.5	6.25	1.5811	4.4271
0.5	6	12	4	~~0.7071~~ 1.414	4.2426
1	5	5	1	1	5
2	6	3	0.25	0.7071	8.4850

$E_i = y - (a + bx)$

$E_i = y_i - \left(\frac{c_0}{x_i} + c_1\sqrt{x_i}\right)$

$U = \sum E_i^2 = \sum\left(y_i - \left(\frac{c_0}{x_i} + c_1\sqrt{x}\right)\right)^2$

$\frac{\partial U}{\partial c_0} = 2\sum\left(y_i - \left(\frac{c_0}{x} + c_1\sqrt{x}\right)\right)\left(-\frac{1}{x}\right) = 0$

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

$\frac{\partial U}{\partial c_1} = 2\sum\left(y_i - \left(\frac{c_0}{x} + c_1\sqrt{x}\right)\right)(-\sqrt{x}) = 0$

$$\boxed{\sum y_i\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x_i}$$

$$302.5 = 136.5\,C_0 + 10.10081\,C_1$$

$$33.7152 = 10.10081\,C_0 + 4.2\,C_1$$

CORRELATION :

Whenever two variable ~~and~~ say x, y are so related that increase in one is accomplained by an increase or decrease in the other then the variable are said to be ~~correla~~ correlated.

» If an increas in one variable correspond to increase in the other the correlation is said to be Positive Correlation

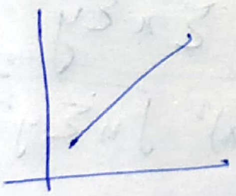
» Same $(\uparrow) \rightarrow (\downarrow)$ Negative Correlation.

» If there is no relationship between 2 relation then no correlation

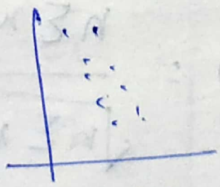
Perfect Correlation :

If 2 variable vary in such a way that their ratio is always constant then the correlation is said to be perfect

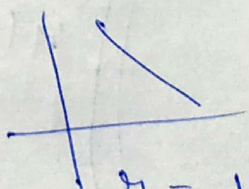
$r < 1$



$r = +1$

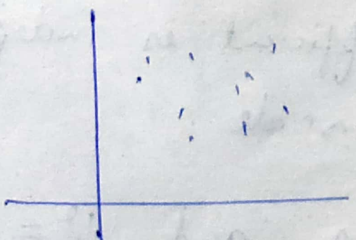


$-1 < r$



$r = -1$

$$\boxed{-1 \leq r \leq 1}$$



$r = 0$

$$\text{Variance} = \frac{\Sigma(x-\bar{x})^2}{n}$$

$$\text{Co variance} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

Correlation (r):

$$r(x,y) = \frac{\text{Covariance}(x,y)}{\sqrt{\text{variance } x}\sqrt{\text{variance } y}}$$

$$= \frac{\frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n}}}$$

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

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

NOTE: Correlation coefficient is independent of change of origin and scale.

Let us $u = \frac{x-a}{h}$ and $v = \frac{y-b}{k}$

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

14/2/23

Q: Find the coefficient correlation b/w the values of x and y

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

$$r(x,y) = \frac{(6\times 582) - 34\times 90}{\sqrt{6\times 248 - (34)^2}\sqrt{6\times 1446 - (90)^2}}$$

$$= 0.9879$$

Q: Find the correlation b/w x and y

x	y	$a = 22$, $x-22$	$u = \frac{x-22}{4}$	$y-24$	$v = \frac{y-24}{6}$	uv	u^2	v^2
10	18	-12	-3	-6	-1	3	9	1
14	12	-8	-2	-12	-2	4	4	4
18	24	-4	-1	0	0	0	1	0
22	6	0	0	-3	-3	0	0	9
26	30	4	1	1	1	1	1	1
30	36	8	2	2	2	4	4	4
			-3		-3	12		

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

$$= \frac{6\times12 - (-3)(-3)}{\sqrt{6\times19-(-3)^2}\sqrt{6\times19-(-3)^2}} = 0.6$$

$$r = 0.789$$

Q:

x	y
78	84
36	51
98	91
25	60
75	68
82	62
90	86
62	58
65	53
39	47

Rank Correlation

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

where D_i = difference b/w the ranks

Q: Calculation rank correlation
The marks secured by the rascals in the selection (R) test (x) and in the proficiency test (y)

$r = 0.789$

SN	x	y	Rx	Ry	$D_i = R_x - R_y$	D_i^2
1	10	30	9	9	0	0
2	15	42	5	3	2	4
3	12	45	8	2	6	36
4	17	46	3	1	2	4
5	13	33	7	8	-1	1
6	16	34	4	7	-3	9
7	24	40	1	4	-3	9
8	14	35	6	6	0	0
9	22	39	2	5	-3	9

$$r = 1 - \frac{6 \times 72}{9(9^2-1)} = 0.4$$

Scanned with OKEN Scanner

Q: Ten competitors in a sexy contest were ranked by the three judges in the following -

	I	II	III
A	1	3	6
B	6	5	4
C	5	8	9
D	10	4	8
E	3	7	1
F	2	10	2
G	4	2	3
H	9	1	10
I	7	6	5
J	8	9	7

$r_{x12} = 0.212$

$r_{x23} = -0.297$

$r_{31} = 0.636$

$D_i = R_1 - R_2$	D_i^2	$D_i' = R_3 - R_1$	$D_i'^2$
-2	4	5	25
1	1	-2	4
-3	9	4	16
6	36	-2	4
-4	16	-2	4
-8	64	0	0
2	4	-1	1
8	64	1	1
1	1	-2	4
-1	1	-1	1
			60

$$r = 1 - \frac{n \sum D_i^2}{n(n^2-1)} = 1 - \frac{1\cancel{0} \times 60}{99\cancel{0}} = \frac{99-60}{99}$$

$$= \frac{39}{99}$$

Tide Rank

If any two or more individuals have same rank or same value rank in the series of marks.

$$r = 1 - \frac{6(\Sigma D_i^2 - f)}{n[(n^2)-1]}$$ where $f = f_{m_1} + f_{m_2} + \ldots$

$$f = \frac{m(m^2-1)}{12}$$

Q:

X	y	Rx	Ry	di	di^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	74	6	2	4	16

$$f = f_{75} + f_{64} + f_{68}$$

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

$$r = 1 - \frac{6(\Sigma di^2 - f)}{n(n^2-1)} = 1 - \frac{6(72-3)}{10(10^2-1)}$$

$$= 0.545$$

Scanned with OKEN Scanner

19|2|25

Q: Ten student gets the following % marks in chemistry and physics calculate marks correlation.

Students	Marks in Chem	Marks in Phy	$D_i = R_1 - R_2$
1	78	84	
2	36	51	
3	98	91	
4	25	60	
5	75	68	
6	82	62	
7	90	86	
8	62	58	
9	65	63	
10	39	47	

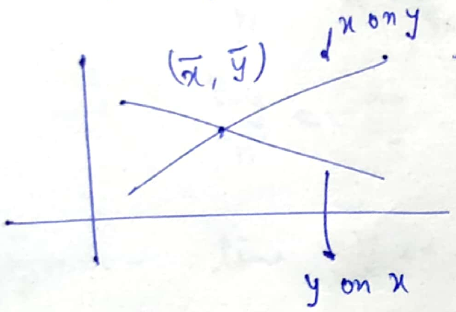
Regression Analysis :

The term 'regression' stands for some sorts of functional relationship b/w two or more related variable. The only fundamental difference b/w if any between problem of curve fiting and regression any of the variable may be considered as independent or dependent while in curve fitting one variable cannot be dependent.

» If the curve of regression is a straight line than it is said to be "Line of Regression".

» If the curve of regression is not straight line then it is said to be "Non-Linear regression".

Line of regression → x on y, y on x



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

$$b_{xy} = r\frac{\sigma x}{\sigma y}$$

y on x

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

$$b_{yx} = r\frac{\sigma y}{\sigma x}$$

» where b_{xy} and b_{yx} are regression coeff and σx and σy are the standard deviation of x and y.

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

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

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

$$\sigma y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

Q. Calculate the linear regression of coeff

x	y	xy	x^2	y^2
1	3	3	1	9
2	7	14	4	49
3	10	30	9	100
4	12	48	16	144
5	14	70	25	196
6	17	102	36	289
7	20	140	49	400
8	24	152	64	571
36	107	599	204	

$(\Sigma x)^2 = 1296$

$(\Sigma y)^2 = 11449$

$$b_{yx} = \frac{n\,\Sigma xy - \Sigma x\,\Sigma y}{n\,\Sigma x^2 - (\Sigma x)^2} = \frac{(8 \times 599) - (36 \times 107)}{(8 \times 208) - 1296}$$

$$= 2.7976$$

$$b_{xy} = \frac{n\,\Sigma xy - \Sigma x\,\Sigma y}{n\,\Sigma y^2 - (\Sigma y)^2} = 0.354$$

Q: The following table give a ge(x) in years cars annual maintainence (y) in hundred rupees.

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
25	99	533	165

Estimate the maintainence cost for a 4 year old car after finding the regression.

y on x

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{99}{5} = 19.8$$

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

$$= 0.95$$

Regression line y on x

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

$$(y-19.8) = 0.95(x-5)$$

$$y = 0.95x + 15.05$$

when $x=4$, $y = 0.95 \times 4 + 15.05$ $= 18.85$ hundred rupees.

$= ₹1885$