[Math-4] L-1 (Unit-3) By- Monika Mittal (MM) 1

Topic:- Moments

Statistical Techniques-I

1, 2] PDE
3, 4, 5] S.T

Average Deviation:- It is mean of the absolute value of deviations of a given set of numbers from their arithmetic mean.

Mean
Mode
Median

$x: x_1 \quad x_2 \quad x_3 \quad \dots \quad x_n$

$f: f_1 \quad f_2 \quad f_3 \quad \dots \quad f_n$

$\bar{x}$

$\Sigma f_i = N$

$$\text{Mean deviation} = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\Sigma f_i} = \frac{\Sigma f_i |x_i - \bar{x}|}{N}$$

Standard Deviation:- S.D. is defined as square root of mean of the square of the deviation from their arithmetic mean.

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$S.D. = \sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}}$$

$$Var = (S.D.)^2 = \sigma^2$$

Moments:→ Moments of distribution are arithmetic mean of various powers of deviations from some given number.

Moments

Moment about Mean (μ_r) [Central Moment]

Moment about any Arbitrary No (A) (μ_r') (Raw Moment)

Moment about Origin (ν_r)

For freq distribution

$$x : \rightarrow x_1 \quad x_2 \quad x_3 \dots x_n$$
$$f : \rightarrow f_1 \quad f_2 \quad f_3 \dots f_n$$

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

$r = 0, 1, 2, 3, 4 \dots$

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

$$\boxed{\mu_0 = 1}$$

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

$$\boxed{\mu_1 = 0} \qquad \boxed{\bar{x} = \frac{\sum f_i x_i}{\sum f_i}}$$

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

$$\boxed{\mu_2 = Var}$$

$$\mu_3 = \frac{\sum f_i (x_i - \bar{x})^3}{\sum f_i}$$

$$\mu_4 = \frac{\sum f_i (x_i - \bar{x})^4}{\sum f_i}$$

$\vdots \quad \dots$ So on

For freq distribution

$$\mu_r' = \frac{\sum_1^n f_i (x_i - A)^r}{\sum_1 f_i}$$

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

$r = 0, 1, 2, 3, 4 \dots$

$$\mu_0' = \frac{\sum f_i (x_i - A)^0}{N}$$
$$= \frac{\sum f_i}{\sum f_i} = 1$$

$$\boxed{\mu_0' = 1}$$

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

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

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

$$\mu_3' = \frac{\sum f_i (x_i - A)^3}{\sum f_i}$$

$$\mu_4' = \frac{\sum f_i (x_i - A)^4}{\sum f_i}$$

For freq distribution

$$\nu_r = \frac{\sum f_i (x_i - 0)^r}{\sum f_i}$$
$$= \frac{\sum_1^n f_i x_i^r}{\sum_1^n f_i}$$

$r = 0, 1, 2, 3, 4 \dots$

$$\nu_0 = \frac{\sum f_i x_i^0}{\sum f_i}$$

$$\boxed{\nu_0 = 1}$$

$$\nu_1 = \frac{\sum f_i x_i^1}{\sum f_i} = \bar{x}$$

$$\nu_2 = \frac{\sum f_i x_i^2}{\sum f_i}$$

$$\nu_3 = \frac{\sum f_i x_i^3}{\sum f_i}$$

$$\nu_4 = \frac{\sum f_i x_i^4}{\sum f_i}$$

Prob-1 Calculate $\mu_1, \mu_2, \mu_3, \mu_4$ (Moment about mean) for the series 4, 7, 10, 13, 16, 19, 22

Sol for Moments

$$\mu_1 = 0$$

$$\mu_2 = \frac{\sum_1^n (x_i-\bar{x})^2}{n}$$

$$\mu_3 = \frac{\sum_1^n (x_i-\bar{x})^3}{n}$$

$$\mu_4 = \frac{\sum_1^n (x_i-\bar{x})^4}{n}$$

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

$$\bar{x} = \frac{\sum x_i}{n}$$

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

$\sum f_i \to n$

S.No.	x	$x-\bar{x}$ $x-13$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
1	4	−9	81	−729	6561
2	7	−6	36	−216	1296
3	10	−3	9	−27	81
4	13	0	0	0	0
5	16	3	9	27	81
6	19	6	36	216	1296
7	22	9	81	729	6561
	$\sum x_i = 91$	0	$\sum(x-\bar{x})^2 = 252$	$\sum(x-\bar{x})^3 = 0$	$\sum(x-\bar{x})^4 = 15876$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{91}{7} = 13 \qquad \bar{x} = 13$$

$n = 7$

$$\mu_1 = \frac{\sum (x-\bar{x})^1}{n} = 0$$

$$\mu_2 = \frac{\sum (x-\bar{x})^2}{n} = \frac{252}{7} = 36$$

$$\mu_3 = \frac{\sum (x-\bar{x})^3}{n} = 0$$

$$\mu_4 = \frac{\sum (x-\bar{x})^4}{n} = \frac{15876}{7} = 2268$$

$$\boxed{\begin{aligned} \mu_1 &= 0 \\ \mu_2 &= 36 \\ \mu_3 &= 0 \\ \mu_4 &= 2268 \end{aligned}}$$

Prob 2 Calculate the first four moments about mean → central moment

Class-interval	0-10	10-20	20-30	30-40	40-50
f	10	20	40	20	10

$\boxed{\text{M.T.U-2014}}$

$\mu_1, \mu_2, \mu_3, \mu_4 = ?$

$$\mu_r = \frac{\sum f_i (x_i - \bar{x})^r}{\sum f_i} \qquad r = 1, 2, 3, 4$$

$\bar{x} = ?$

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

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

class int	f	x (Mid pt)	$f_i x_i$	$x-\bar{x}$ $x-25$	$f(x-25)$	$f(x-25)^2$	$f(x-25)^3$	$f(x-25)^4$
0-10	10	5	50	-20	-200	4000	-80000	1600000
10-20	20	15	300	-10	-200	2000	-20000	200000
20-30	40	25	1000	0	0	0	0	0
30-40	20	35	700	10	200	2000	20000	200000
40-50	10	45	450	20	200	4000	80000	1600000
	$\sum f_i = 100$		$\sum f_i x_i = 2500$		0	12000	0	3600000

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2500}{100} = 25 \qquad \boxed{\bar{x} = 25}$$

$$\mu_1 = \frac{\sum f_i (x_i - \bar{x})}{\sum f_i} = 0, \quad \mu_2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i} = \frac{120\cancel{00}}{1\cancel{00}} = 120$$

$$\mu_3 = \frac{\sum f_i (x_i - \bar{x})^3}{\sum f_i} = 0, \quad \mu_4 = \frac{\sum f_i (x_i - \bar{x})^4}{\sum f_i} = \frac{36000\cancel{00}}{1\cancel{00}} = 36,000$$

Practice Questions

(1) Find first four central moments

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

✓

Ans:- $\mu_1 = 0$ $\mu_3 = -300$

$\mu_2 = 125$ $\mu_4 = 37625$

(2) Calculate first four moment about $\underline{x=15}$ (Raw Moment)

x	10	11	12	13	14	15	16	17	18	19	20	21
f	9	36	75	105	116	107	88	66	45	30	18	5

✓

[Hint $\mu_r' = \dfrac{\sum f_i (x_i - 15)^r}{\sum f_i}$] $\mu_1' = -.207$ $\mu_3' = .958$

$\mu_2' = 5.54$ $\mu_4' = 75.58$

Topic

Relation between μ_r and μ_r'

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

μ_r, μ_r', v_r

$r = 2, 3, 4 \dots$ $\quad$ $\mu_0' = 1$

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

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

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

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

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

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

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

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

Relation between v_r and μ_r

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

$r = 1, 2, 3, 4$ $\quad$ $\boxed{\mu_1 = 0}$ $\boxed{\mu_0 = 1}$

$$v_1 = \mu_1 + \mu_0 \bar{x} = \bar{x}$$

$$v_2 = \mu_2 + {}^2C_1 \mu_1 \bar{x} + {}^2C_2 \mu_0 \bar{x}^2 = \mu_2 + \bar{x}^2 \Rightarrow \boxed{v_2 = \mu_2 + \bar{x}^2}$$

$$v_3 = \mu_3 + {}^3C_1 \mu_2 \bar{x} + {}^3C_2 \mu_1 \bar{x}^2 + {}^3C_3 \mu_0 \bar{x}^3 \Rightarrow \mu_3 + 3\mu_2\bar{x} + \bar{x}^3$$

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

$$\boxed{v_4} = \mu_4 + {}^4C_1 \mu_3 \bar{x} + {}^4C_2 \mu_2 \bar{x}^2 + {}^4C_3 \mu_1 \bar{x}^3 + {}^4C_4 \mu_0 \bar{x}^4 \Rightarrow \boxed{\mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4}$$

Relation between v_r and μ_r'

$$v_r = \mu_r' + {}^rC_1 \mu_{r-1}' A + {}^rC_2 \mu_{r-2}' A^2 + \dots + A^r$$

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

$r = 1, 2, 3, 4$

$$v_1 = \mu_1' + {}^1C_1 \mu_0' A = \bar{x} - A + A = \bar{x}$$

$$v_2 = \mu_2' + {}^2C_1 \mu_1' A + {}^2C_2 \mu_0' A^2 = \mu_2' + 2\mu_1' A + A^2$$

Karl Pearson Coefficient

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

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

β-coeff

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

$$\gamma_2 = \beta_2 - 3$$

γ-coeff.

Prob First four moments about the value '4' of the variable are -1.5, 17, -30, 108. Find moment about mean, about origin, β_1, β_2. Also find moments about the point 2

$\mu_1' = -1.5$
$\mu_2' = 17$
$\mu_3' = -30$
$\mu_4' = 108$
Given

$\mu_r = ?$
$\nu_r = ?$
$\beta_1 = ?$
$\beta_2 = ?$

$\mu_r' = ?$ (About pt 2)

Relation b/w μ_r, μ_r'

$$\mu_r = \mu_r' - {}^rc_1\,\mu_{r-1}'\,\mu_1' + {}^rc_2\,\mu_{r-2}'\,\mu_1'^2 + \cdots$$

$\mu_1 = 0$

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

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

$$\boxed{\mu_2 = 17 - (-1.5)^2 = 14.75}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = (-30) - 3(17)(-1.5) + 2(-1.5)^3$$

$$= \underline{\underline{39.75}}$$

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

$$= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 = \underline{\underline{142.312}}$$

Scanned with CamScanner

L-2.3

$A = 4$

$\mu_1' = -1.5$, $\mu_2' = 17$, $\mu_3' = -30$, $\mu_4' = 108$ Given

$\mu_1 = 0$

$\mu_2 = 14.75$

$\mu_3 = 39.75$

$\mu_4 = 142.312$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = .4923$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.312}{(14.75)^2} = .6541$$

Moments about origin (ν_r)

$\boxed{\nu_r, \mu_r}$

$\nu_1 = \bar{x}$

$\boxed{\nu_1 = 2.5}$

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

$-1.5 = \bar{x} - 4$

$\bar{x} = 4 - 1.5 = 2.5$

$\nu_r = \mu_r + {}^rC_1 \mu_{r-1}\bar{x} + {}^rC_2 \mu_{r-2}\bar{x}^2 + \ldots$

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

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

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

$$= 39.75 + 3(14.75)(2.5) + (2.5)^3$$

$$= \boxed{166 = \nu_3}$$

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

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

$$= (142.312) + 4(39.75)(2.5) + 6(14.75)(2.5)^2 + (2.5)^4$$

$$\boxed{\nu_4 = 1132}$$

$\boxed{\nu_1 = 2.5}$ ✓

$\boxed{\nu_2 = 21}$ ✓

$\boxed{\nu_3 = 166}$ ✓

$\boxed{\nu_4 = 1132}$ ✓

Moments about '2' μ_r'

$\mu_1' = \bar{x} - A = 2.5 - 2 = .5$

$\boxed{\mu_1' = .5}$

μ_r, μ_r'

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

$$\mu_2' = \mu_2 + \mu_1'^2 = 14.75 + (.5)^2 = 15$$

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

$$\mu_3' = \mu_3 + 3\mu_2'\mu_1' - 2\mu_1'^3 = 39.75 + 3(15)(.5) - 2(.5)^3$$

$$\boxed{\mu_3' = 62}$$

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

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

$$= 142.312 + 4(62)(.5) - 6(15)(.5)^2 + 3(.5)^4$$

$$= \underline{244}$$

Prob - For a distribution, the mean is 10, Variance is 16, γ_1 is 1, β_2 is 4 find first four moments about origin.

Given

$\bar{x} = 10$

$\sigma^2 = 16$

$\gamma_1 = 1$

$\beta_2 = 4$

$\Rightarrow \underbrace{\nu_1, \nu_2, \nu_3, \nu_4}$ = ??

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

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

$\gamma_1 = \sqrt{\beta_1}$

$\gamma_2 = \beta_2 - 3$

$\boxed{\bar{x} = 10 = \nu_1}$

$\sigma^2 = 16 = \mu_2 = \frac{\Sigma f_i(x_i-\bar{x})^2}{\Sigma f_i} = (S.D)^2$

$\boxed{\mu_2 = 16}$ ✓

$\boxed{\mu_3 = 64}$ ✓

$\boxed{\mu_4 = 1024}$ ✓

$\gamma_1 = \sqrt{\beta_1} \Rightarrow 1 = \beta_1 = \frac{\mu_3^2}{\mu_2^3}$

$\boxed{\mu_3^2 = \mu_2^3} \Rightarrow \mu_3^2 = (16)^3 = (4^2)^3 = (4^3)^2 = (64)^2$

$\boxed{\mu_3 = 64}$

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

$\mu_4 = (16)^2 \times 4 = 1024$

$\boxed{\text{Relation b/w } \nu_r, \mu_r}$

$\mu_1 = 0$

$\mu_2 = 16$

$\mu_3 = 64$

$\mu_4 = 1024$

$\nu_1 = \bar{x} = 10$ ✓

$\nu_2 = \mu_2 + 2\mu_1\bar{x} + 1.\mu_0\bar{x}^2$

$= \mu_2 + \bar{x}^2 = 16 + (10)^2 = 100+16$

$= \boxed{116 = \nu_2}$

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

$= \mu_3 + 3\mu_2\bar{x} + \bar{x}^3 = 64 + 3(16)(10) + (10)^3$

$= 64 + 480 + 1000 = \boxed{1544}$

$\boxed{\nu_3 = 1544}$

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

$= \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4 = 1024 + 4(64)\times 10 + 6(16)(10)^2 + (10)^4$

$= 22184$ ✓

L-2-5

Practice

(1) First four moments about '0' are $-.20$, 1.76, -2.36, 10.88.
find moment about mean. (2009)

Ans $\mu_1 = 0$, $\mu_2 = 1.72$, $\mu_3 = -1.32$, $\mu_4 = 9.409$

$\nu_r \rightarrow \mu_r$

Lecture 3 (Unit 3)

Statistical Techniques-I

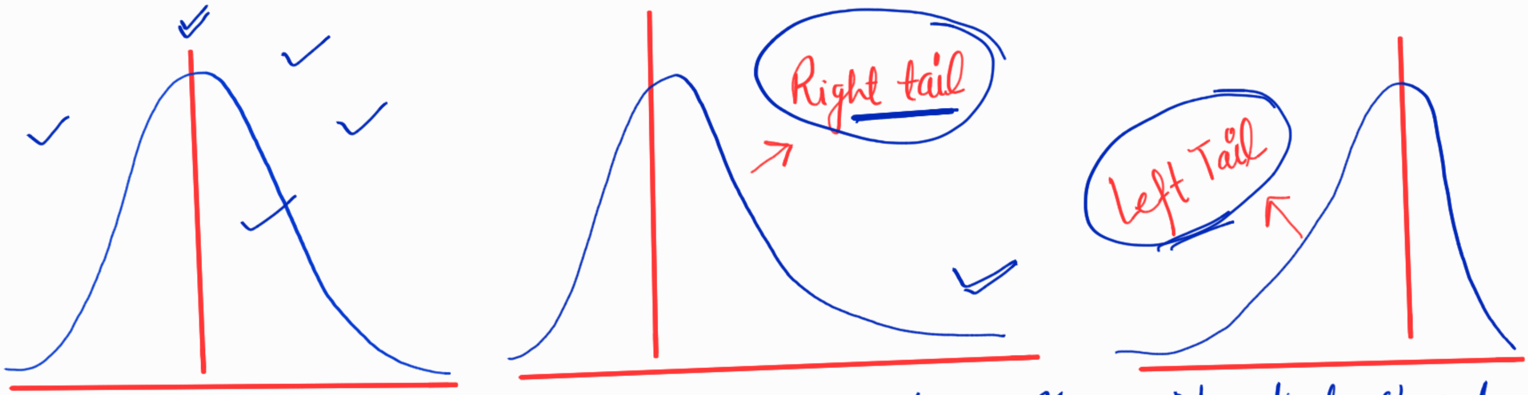
By:- Monika Mittal (MM)

Topic: Skewness and Kurtosis

Moments

Skewness :→ Skewness means lack of symmetry in freq. distribution.

[2 Marks]



M, M_o, M_d — Mean = Mode = Med. — Symmetrical Distribution

Positively Skewed Distribution — A.M. > M_o

Negatively skewed Distribution — A.M < M_o

Moment Coeff. of Skewness $= \dfrac{\mu_3}{\sqrt{\mu_2^3}} = \gamma_1 = \pm\sqrt{\beta_1}$

$$Sk_M = \gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \pm\sqrt{\beta_1}$$

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

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

Note : $\mu_3 \to -ve$ $\quad \gamma_1 \to -ve$

$\mu_3 \to +ve$ $\quad \gamma_1 \to +ve$

This method gives magnitude as well as direction of skewness.

For Symmetrical Distribution, $Sk_M = 0$

Ex 1 :– The first three central moments of distribution are 0, 15, -31, find moment coeff. of skewness

$\mu_1 = 0$

$Sk_M = \gamma_1 = \pm\sqrt{\beta_1}$

$\mu_2 = 15$
$\mu_3 = -31$ Given

$$= \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-31}{\sqrt{(15)^3}}$$

$$= -.53 \text{ (Negatively skewed)}$$

Ex. 2 → The first four moment about the value '5' are 2, 20, 40 and 50, Calculate Moment Coeff of Skewness :

$$Sk_M = \gamma_1 = \pm\sqrt{\beta_1} = \frac{\mu_3'}{\sqrt{\mu_2'^3}}$$

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

Relation μ_r, μ_r'

$\boxed{\mu_1 = 0}$

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

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

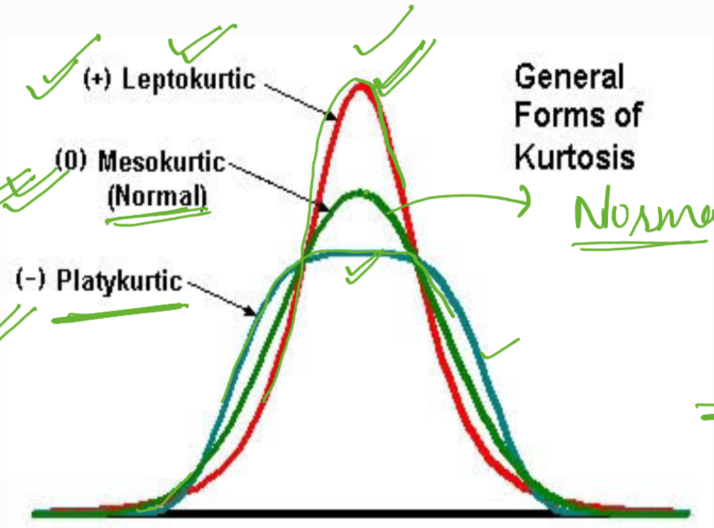
$$= 40 - 3(20)(2) + 2(2)^3 = -64$$

$$Sk_M = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-64}{\sqrt{(16)^3}} = -1$$

Negatively Skewed

Kurtosis : It is degree of peakedness of a distribution, usually taken in relative to Normal distribution.

Kurtosis is a measure of whether the data are heavytailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case.



Normal Distribution

$\left.\begin{matrix} x \\ f \end{matrix}\right\} \rightarrow$ Karl Pearson

Measure of kurtosis

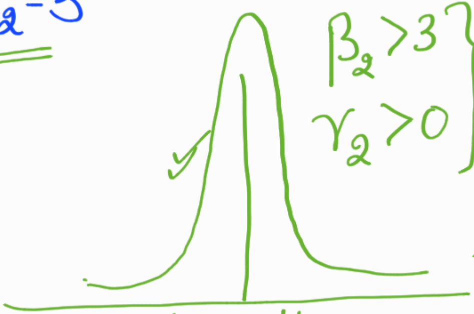
$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

Kurtosis of Distribution is also defined as

$$\gamma_2 = \beta_2 - 3$$

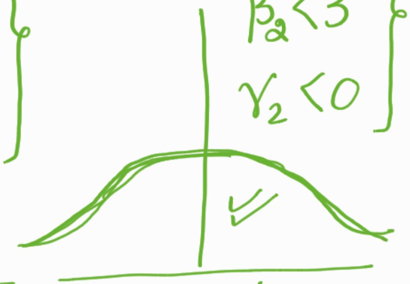
$\left.\begin{matrix} \beta_2 = 3 \\ \gamma_2 = 0 \end{matrix}\right\}$

Mesokurtic



$\left.\begin{matrix} \beta_2 > 3 \\ \gamma_2 > 0 \end{matrix}\right\}$

Leptokurtic



$\left.\begin{matrix} \beta_2 < 3 \\ \gamma_2 < 0 \end{matrix}\right\}$

Platykurtic

Ex.1 → The first four moments of a distribution about '4' are 1, 4, 10, 45, Obtain the various characteristics of distribution on the basis of given info. Comment upon the Nature of distribution [AKTU 2016, 2018]

$A = 4$

$\left.\begin{array}{l} \mu_1' = 1 \\ \mu_2' = 4 \\ \mu_3' = 10 \\ \mu_4' = 45 \end{array}\right\}$ Given

Skewness ✓ Kurtosis ✓

μ_r

μ_r

Relation b/w μ_r and μ_r'

$\boxed{\mu_1 = 0}$

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

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

$= 10 - 3(4)(1) + 2(1)^3 = 0$

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

$= 45 - 4(10)(1) + 6(4)(1)^2 - 3(1)^4$

$= 26$

$\mu_1 = 0$

$\mu_2 = 3$

$\mu_3 = 0$

$\mu_4 = 26$

Skewness $\rightarrow Sk_M = \gamma_1 = \pm\sqrt{\beta_1}$

$= \dfrac{\mu_3}{\sqrt{\mu_2^3}} = 0$

$\hookrightarrow$ Symmetrical

Kurtosis $\quad \beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{26}{9} = 2.89 < 3$

$\gamma_2 = \beta_2 - 3$

$= 2.89 - 3 < 0$

$\left.\begin{array}{l} \beta_2 < 3, \\ \gamma_2 < 0 \end{array}\right\}$ Platykurtic

Lecture 4(Unit 3)

Statistical Techniques-I

By:- Monika Mittal (MM)

Topic: some Important Problems Based upon Moments, Skewness and Kurtosis

Ex 1 The following table represents the height of a batch of 100 students. Find Central Moments, Calculate kurtosis. → β_2

μ_2

Height (cm)	59	61	63	65	67	69	71	73	75	→ x
No of stds	0	2	6	20	40	20	8	2	2	→ f

[AKTU 2018]

About Arbitrary Pt (A) (Raw Moments)

Relations μ_r, μ_r'

$\mu_1 = 0$

$$\mu_1' = \frac{\Sigma f_i (x_i - A)}{\Sigma f_i}$$

$$\mu_2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} \qquad \mu_2' = \frac{\Sigma f_i (x_i - A)^2}{\Sigma f_i}$$

$$\mu_3 = \frac{\Sigma f_i (x_i - \bar{x})^3}{\Sigma f_i} \qquad \mu_3' = \frac{\Sigma f_i (x_i - A)^3}{\Sigma f_i}$$

$$\mu_4 = \frac{\Sigma f_i (x_i - \bar{x})^4}{\Sigma f_i} \qquad \mu_4' = \frac{\Sigma f_i (x_i - A)^4}{\Sigma f_i}$$

x	f	$x-A$ $x-67$	$u = \frac{x-67}{2}$	fu	fu^2	fu^3	fu^4
59	0	−8	−4	0	0	0	0
61	2	−6	−3	−6	18	−54	162
63	6	−4	−2	−12	24	−48	96
65	20	−2	−1	−20	20	−20	20
67	40	0	0	0	0	0	0
69	20	2	1	20	20	20	20
71	8	4	2	16	32	64	128
73	2	6	3	6	18	54	162
75	2	8	4	8	32	128	512
	100			12	164	144	1100

$x - 67$

$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = ?$ $\boxed{A = 67}$ $u = \frac{x-67}{2}$

$$\mu_1' = \frac{\sum f_i (x_i - 67)}{\sum f_i} \Rightarrow \mu_1' = \frac{\sum f_i u}{\sum f_i} \times 2$$

$$\mu_2' = \frac{\sum f_i (x_i - 67)^2}{\sum f_i} \Rightarrow \mu_2' = \frac{\sum f_i u^2}{\sum f_i} \times 2^2$$

$$\mu_3' = \frac{\sum f_i (x_i - 67)^3}{\sum f_i} \Rightarrow \mu_3' = \frac{\sum f_i u^3}{\sum f_i} \times 2^3$$

$$\mu_4' = \frac{\sum f_i u^4}{\sum f_i} \times 2^4$$

$$\mu_1' = \frac{12}{100} \times 2 = \boxed{\cdot 24 = \mu_1'}$$

$$\mu_2' = \frac{\sum f_i u^2}{\sum f_i} \times 4 = \frac{164}{100} \times 4 = \boxed{6.56 = \mu_2'}$$

$$\mu_3' = \frac{\sum f_i u^3}{\sum f_i} \times 8 = \frac{144}{100} \times 8 = 11.52$$

$$\mu_4' = \frac{\sum f_i u^4}{\sum f_i} \times 16 = \frac{1100}{100} \times 16 = 176$$

$$\boxed{\begin{array}{ll} \mu_1' = \cdot 24 & \\ \mu_2' = 6.56 & \mu_4' = 176 \\ \mu_3' = 11.52 & \end{array}} \Rightarrow \mu_r = ?$$

Relations μ_r, μ_r'

$\mu_1 = 0$

$$\mu_2 = \mu_2' - \mu_1'^2 = 6.56 - (\cdot 24)^2 = 6.5024$$

$\boxed{\mu_2 = 6.5024}$

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

$$= 11.52 - 3(6.56)(\cdot 24) + 2(\cdot 24)^3$$

$$= 6.8244 \Rightarrow \boxed{\mu_3 = 6.8244}$$

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

$$= 176 - 4(11.52)(.24) + 6(6.56)(.24)^2 - 3(.24)^4$$

$$\boxed{\mu_4 = 167.1979}$$

Kurtosis $\quad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{167.1979}{(6.5024)^2}$

$$= 3.9544 > 3$$

$$\beta_2 > 3$$

$$\gamma_2 = \beta_2 - 3 = 3.9544 - 3 > 0$$

$$\left.\begin{matrix}\beta_2 > 3 \\ \gamma_2 > 0\end{matrix}\right\} \text{leptokurtic}$$

Practice Question

Q → Find Skewness & kurtosis

Height	144.5 – 149.5	149.5 – 154.5	154.5 – 159.5	159.5 – 164.5	164.5 – 169.5	169.5 – 174.5	174.5 – 179.5
freq	2	4	13	31	32	15	3

2011

Ans ⇒ $\gamma_1 = -.3834$, $\beta_2 = 3.25$

Target 90+
AKTU
All 5 Units
M S Tomer
One Shot Videos
Just click here
Please open pdf in Adobe acrobat
PDF

Unit-3 (Maths-4) L-5

Topic: **Moment Generating Function** (MGF)

Definition: Let x be Random Var.

$$M_x(t) = E\{e^{tx}\} = \begin{cases} \sum_x e^{tx} p(x) & x \to \text{Discrete} \\ \int_x e^{tx} f(x)\,dx & x \to \text{continuous.} \end{cases}$$

$E(x) \to$ first Moment about origin $= \nu_1$

$E(x^2) \to$ Second " $= \nu_2$

$E(x^3) \to$ Third " $= \nu_3$

$\mu_1 \to$

$\mu_1' \to$

$$E(x) = \nu_1 = \frac{d}{dt} M_x(t) \Big]_{t=0}$$

$$E(x^2) = \nu_2 = \frac{d^2}{dt^2} M_x(t) \Big]_{t=0}$$

$$\nu_r = \frac{d^r}{dt^r} M_x(t) \Big]_{t=0}$$

Ex. Find the moment Generating Function of distribution

$$f(x) = \frac{1}{c} e^{-x/c} \quad ; \quad 0 < x \leq \infty \quad , \quad c > 0.$$

Hence find mean and standard deviation. [2014]

$$M_x(t) = \int_x e^{tx} f(x)\,dt$$

$$= \int_0^{\infty} e^{tx} \frac{1}{c} e^{-x/c} dx$$

$$= \frac{1}{c}\int_0^{\infty} e^{-\left(\frac{1}{c}-t\right)x} dx = \frac{1}{c}\left[\frac{e^{-\left(\frac{1}{c}-t\right)x}}{-\left(\frac{1}{c}-t\right)}\right]_0^{\infty}$$

$$= \frac{1}{c}\left[\frac{1}{-\left(\frac{1}{c}-t\right)}\right]\left[\cancel{e^{-\infty}} \to 0 - e^{-0}\right]$$

$$= \frac{1}{c\left(\frac{1}{c}-t\right)} = \frac{1}{\cancel{c}\left(\frac{1-ct}{\cancel{c}}\right)}$$

$$\boxed{M_x(t) = (1-ct)^{-1} =}$$

$$M_x(t) = 1 + ct + c^2t^2 + c^3t^3 + \cdots$$

$$\boxed{\begin{array}{l}(1-x)^{-1} = 1+x+x^2 \cdots \\ (1+x)^{-1} = 1-x+x^2 \cdots\end{array}}$$

$$\nu_1 = \frac{d}{dt} M_x(t)\Big]_{t=0} = \Big[0 + c + c^2 2t + c^3 3t^2 \cdots\Big]_{t=0}$$

$$= c \qquad \boxed{\nu_1 = c} = \text{Mean} \quad \boxed{\bar{x} = c}$$

$$\nu_2 = \frac{d^2}{dt^2} M_x(t) = \Big[0 + c^2 2 + c^3 6t \cdots\Big]_{t=0} \boxed{= 2c^2}$$

$\mu_2 = \text{Var}$

$S.D = \sqrt{\text{Var}}$

$$\boxed{\nu_2 = 2c^2} \quad \text{Relation} \quad \nu_2 = \mu_2 + \bar{x}^2$$

$$\mu_2 = \nu_2 - \bar{x}^2 = 2c^2 - c^2 = \boxed{c^2 = \text{Var}}$$

$S.D = \sqrt{c^2}$

$\boxed{S.D = C}$

Ex. If P(X=x)= $\frac{1}{2^x}$, $x = 1, 2, 3, 4 \ldots$ Find the moment

Generating function of x. [2014]

$$\boxed{P(x) = \frac{1}{2^x}}$$

$$M_x(t) = \sum_x e^{tx} P(x) = \sum_{x=1}^{\infty} e^{tx} \frac{1}{2^x}$$

$$= \frac{e^t}{2} + \frac{e^{2t}}{2^2} + \frac{e^{3t}}{2^3} + \cdots \infty$$

$$\boxed{\begin{array}{l}G.P. = a + ar + ar^2 \cdots \infty \\ S = \frac{a}{1-r}\end{array}}$$

$r = e^t/2$

$$M_x(t) = \frac{\frac{e^t}{2}}{1 - e^t/2} = \frac{e^t/\cancel{2}}{\frac{2-e^t}{\cancel{2}}}$$

$$\boxed{M_x(t) = \frac{e^t}{\cdots}}$$

$2-e^t$

$v_1 = \frac{d}{dt}\left(M_x(t)\right)\Big]_{t=0}$

Prob. Find MGF of random variable x having probability distribution function

$$f(x) = \begin{cases} x & 0 < x < 1 \\ 2-x & 1 \leq x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

$M_x(t)$

Find v_1, v_2, μ_2 (Variance)

$v_1 \rightarrow d/dt$

$v_2 \rightarrow \frac{d^2}{dt^2}$

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

Unit-3 (Maths-4) L-6

By :- Monika Mittal (MM)

Topic: Curve Fitting by Method of Least Square (Part-1)

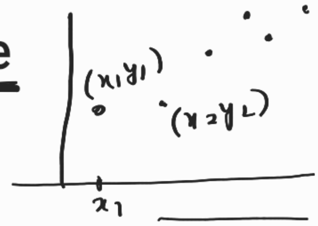
* Fitting of Straight Line

10M

some observations of x and y are given.

Let $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots\ldots (x_n, y_n) \rightarrow n$

By scatter diagram/ dot diagram we can approximate Relationships between x and y.



By Curve Fitting , we can find a mathematical relationship between x and y .

It is useful in study of correlation and regression.

Fitting of Straight Line

$$y = a + bx$$

Let $(x_1, y_1)\ (x_2, y_2) \; ---- \; (x_n, y_n) \rightarrow n$ observations

$$\left.\begin{aligned} y_1' &= a + bx_1 \\ y_2' &= a + bx_2 \\ y_3' &= a + bx_3 \\ &\vdots \end{aligned}\right\}$$ expected Values of y corresponding to $x_1, x_2 \ldots\ldots x_n$

Residual = Error = Observed − expected Value

$$E_1 = y_1 - y_1'$$

$$E_i = y_i - y_i'$$

$$U = \Sigma (y_i - y_i')^2 = \Sigma (y_i - a - bx_i)^2$$

$U \rightarrow$ Minimize , $a, b \rightarrow$

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

$$\frac{\partial U}{\partial a} = 2\Sigma (y_i - a - bx_i)(-1) = 0 \Rightarrow \Sigma y_i - \Sigma a - b\Sigma x_i = 0$$

$$\Sigma y = na + b\Sigma x \quad \text{—— (1)}$$

(Normal Eq^n)

$$\frac{\partial U}{\partial b} = 2\Sigma (y_i - a - bx_i)(-x_i) = 0 \Rightarrow \Sigma (y_i x_i - a x_i - b x_i^2) = 0$$

$$\frac{\partial}{\partial b} \Rightarrow \sum xy = a\sum x + b\sum x^2 \quad \text{—(2)}$$

Normal eq^n

$$\boxed{y = a + bx}$$

Normal Eq^n

$$\left\{ \begin{aligned} \sum y &= na + b\sum x \\ \sum xy &= a\sum x + b\sum x^2 \end{aligned} \right\}$$

$n \rightarrow$ No of observations

Ex. Fit straight line

X:	1	2	3	4	6	8
Y:	2.4	3.1	3.5	4.2	5.0	6.0

$y = a + bx$

Normal Eq^n $\sum y = na + b\sum x$ —(1)

$\sum xy = a\sum x + b\sum x^2$ —(2)

solve $a, b = ?$

$\sum xy = a\sum x + b\sum x^2$ $\boxed{n=6}$

X	y	xy	x^2
1	2.4	2.4	1
2	3.1	6.2	4
3	3.5	10.5	9
4	4.2	16.8	16
6	5.0	30	36
8	6.0	48	64
$\sum X = 24$	$\sum y = 24.2$	$\sum xy = 113.9$	$\sum X^2 = 130$

(1) $\Rightarrow$

$24.2 = 6a + b\,24$

(2) $\Rightarrow$ $113.9 = a\,24 + b\,130$

ON solving –

$$\begin{cases} a = 2.0215 \\ b = .5029 \end{cases}$$

$y = a + bx$

$$\boxed{y = 2.0215 + .5029x}$$

Straight line

$\rightarrow$ Parabola

$\rightarrow$ exponential

By: Monika Mittal (MM)

10 M

Topic: Curve Fitting by Method of Least Square (Part-2)

Fitting of Second degree Parabola ($y = a + bx + cx^2$)

Exponential Curve

Second degree Parabola ($Y = a + bx + cx^2$)

let $(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)$

$y_1' = a + bx_1 + cx_1^2$

$y_2' = a + bx_2 + cx_2^2$

$y_3' = a + bx_3 + cx_3^2$

$\vdots$

$y_1', y_2', y_3' \dots y_n'$ → expected Value of y corresponding, $x_1, x_2 \dots x_n$

$\boxed{y_1 - y_1'}$ = Residual

$U = \Sigma (y_i - y_i')^2$

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

$U = \Sigma (y_i - a - bx_i - cx_i^2)^2$

$\frac{\partial U}{\partial a} = 2\Sigma (y_i - a - bx_i - cx_i^2)(-1) = 0 \Rightarrow \Sigma y_i - \Sigma a - b\Sigma x_i - c\Sigma x_i^2 = 0$

$\frac{\partial U}{\partial b} = 2\Sigma (y_i - a - bx_i - cx_i^2)(-x_i) = 0$

$\frac{\partial U}{\partial c} = 0 \Rightarrow 2\Sigma (y_i - a - bx_i - cx_i^2)(-x_i^2) = 0$

$\Sigma y_i = \Sigma a + b\Sigma x_i + c\Sigma x_i^2 \Rightarrow \Sigma y = na + b\Sigma x + c\Sigma x^2$ — (1)

$\Sigma y_i x_i - \Sigma a x_i - b\Sigma x_i^2 - c\Sigma x_i^3 = 0$

$\Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$ — (2)

$\Sigma y_i x_i^2 - a\Sigma x_i^2 - b\Sigma x_i^3 - c\Sigma x_i^4 = 0$

$\Sigma yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$ — (3)

Ex. Using method of Least Square, Fit second degree parabola $y = a + bx + cx^2$ for following data

X:	2	3	4	5	6
Y:	14	17	20	24	29

$\boxed{n = 5}$

$\rightarrow y = 1.a + bx + cx^2$

Normal Eq^n

$$\begin{cases} ① \Rightarrow \sum y = na + b\sum x + c\sum x^2 \\ ② \Rightarrow \sum xy = a\sum x + b\sum x^2 + c\sum x^3 \\ ③ \Rightarrow \sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \end{cases}$$

Table:-

x	y	x^2	xy	x^3	x^2y	x^4
2	14	4	28	8	56	16
3	17	9	51	27	153	81
4	20	16	80	64	320	256
5	24	25	120	125	600	625
6	29	36	174	216	1044	1296
$\sum x = 20$	$\sum y = 104$	$\sum x^2 = 90$	$\sum xy = 453$	$\sum x^3 = 440$	$\sum x^2y = 2173$	$\sum x^4 = 2274$

$$\sum y = na + b\sum x + c\sum x^2 \Rightarrow 104 = 5a + b20 + c90 \quad ①$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \Rightarrow 453 = a20 + b90 + c440 \quad ②$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \Rightarrow 2173 = a(90) + b(440) + c(2274) \quad ③$$

$a = ?$ $\quad b = ?$ $\quad c = ?$

$a = 11$, $\quad b = .8428$ $\quad c = .3571$

$$\boxed{y = 11 + .8428x + .3571x^2}$$

Ex. Fit exponential Curve for following data. $y = ae^{bx}$

X:	1	5	7	9	12
Y:	10	15	12	15	21

Normal $Eq^n \Rightarrow \quad Y = ae^{bx}$

Taking log both side

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

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

$$Y = A + Bx$$

$y = a + bx$

Let

$\boxed{Y = \log_{10} y}$

$\boxed{A = \log_{10} a}$

$\boxed{B = b \log_{10} e}$

Normal

$$\sum Y = nA + B\sum x \quad \text{—(1)}$$

$$\sum xY = A\sum x + B\sum x^2 \quad \text{—(2)}$$

x	y	$Y = \log_{10} y$	xY	x^2
1	10	1	1	1
5	15	1.1761	5.8805	25
7	12	1.0792	7.5544	49

$\boxed{n = 5}$

9	15	1·1761	10·5849	81
12	21	1·3222	15·8664	144
$\Sigma x = 34$		$\Sigma Y = 5·7536$	$\Sigma xY = 40·8862$	$\Sigma x^2 = 300$

$$\Sigma Y = nA + B\Sigma x \Rightarrow \begin{cases} 5·7536 = 5A + B\,34 & \text{--- (1)} \\ 40·8862 = A\,34 + B\,300 & \text{--- (2)} \end{cases}$$
$$\Sigma xY = A\Sigma x + B\Sigma x^2 \Rightarrow$$

$A = ·9765$

$\log_{10} a = ·9765$

$a = \text{antilog}(·9765)$

$= 9·4733$

$B = ·02561$

$b \log_{10} e = ·02561$

$b = \dfrac{·02561}{\log_{10} e}$

$b = ·05896$

$y = ae^{bx}$

$$\boxed{y = 9·4733\, e^{·05896x}}$$

Ex. The pressure of 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 data to the equation $PV^{\gamma} = C$ [2019]

$PV^{\gamma} = C \Rightarrow P = \dfrac{C}{V^{\gamma}} = CV^{-\gamma}$

Taking log both side, $\underbrace{\log_{10}P}_{Y} = \underbrace{\log_{10}C}_{A} - \underbrace{r}_{B}\underbrace{\log_{10}V}_{X}$

$Y = A + BX$

$\boxed{Y = \log_{10}P}$

$A = \log_{10}C$

$B = -r$

$\boxed{X = \log_{10}V}$

$$\begin{cases} \Sigma Y = nA + B\Sigma X & \text{— (1)} \\ \Sigma XY = A\Sigma X + B\Sigma X^2 & \text{— (2)} \end{cases}$$

Table:

V	P	$X = \log_{10}V$	$Y = \log_{10}P$	XY	X^2
50	64.7	1.69897	1.81090	3.07666	2.88650
60	51.3	1.77815	1.71012	3.04085	3.16182
70	40.5	1.84510	1.60746	2.96592	3.40439
90	25.9	1.95424	1.41330	2.76193	3.81905
100	78.0	2.00000	1.89209	3.78418	4
		$\Sigma X = 9.27646$	$\Sigma Y = 8.43387$	$\Sigma XY = 15.62954$	$\Sigma X^2 = 17.27176$

Put the values in Normal Eq^n's $\boxed{n = 5}$

$\Sigma Y = nA + B\Sigma X \Rightarrow 8.43387 = 5A + B\,9.27646$

$\Sigma XY = A\Sigma X + B\Sigma X^2 \Rightarrow 15.62954 = A\,9.27646 + B\,17.27176$

$A = ?$ $B = ?$

After solving

$A = 2.22476$

$B = -0.28997$

$Y = \log_{10}P$

$\boxed{A = \log_{10}C}$

$\boxed{B = -r}$

$\checkmark\ \gamma = -B$

$\Rightarrow B =$

$\boxed{X = \log_{10} V}$

$A = \log_{10} C = 2{\cdot}22476$

$C = \text{antilog}\,(2{\cdot}22476) = 167{\cdot}78765$

$\gamma = -B = -(-0{\cdot}28997)$

$\gamma = {\cdot}28997$

$PV^{\gamma} = C \Rightarrow \boxed{PV^{\cdot 28997} = 167{\cdot}78765}$

EX. Normal eqn $y = \dfrac{C_0}{x} + C_1\sqrt{x} = ?$

EX write Normal eqn

$y = ab^{x}$

10 M $\leftarrow$ straight line

$y = a + bx + cx^2$

exponential curve

By- Monika Mittal (MM)

Unit-3(Lec-8) Math-4

Topic: Correlation and Correlation Coefficient (r_{xy})

Correlation: correlation is a statistical tool by which we can describe affect on one variable on the change of other variable.

* if two variables move in same direction then correlation is said to be positive correlation.

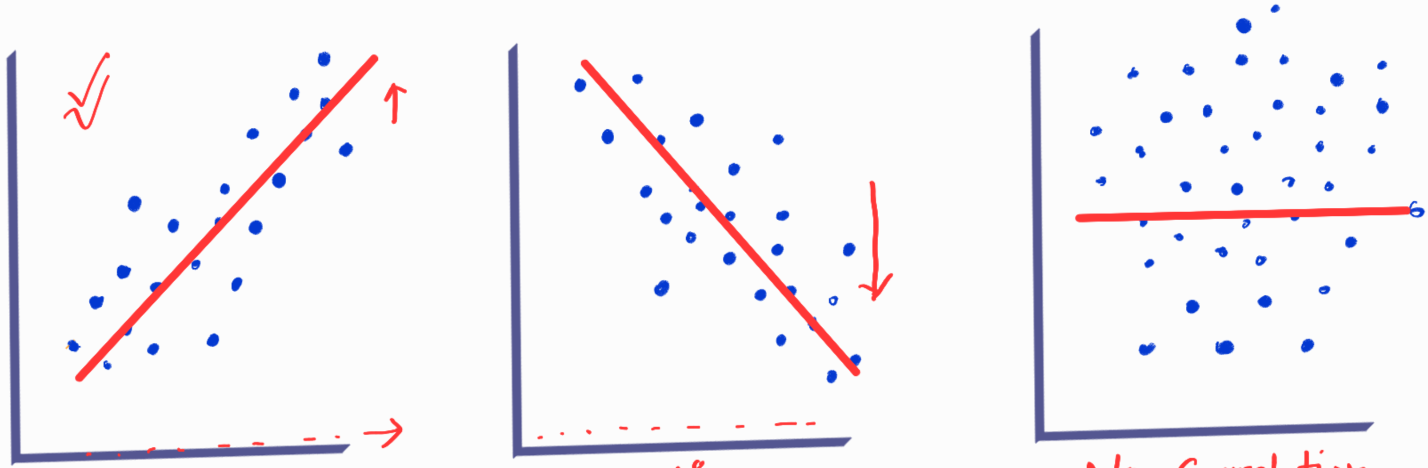
x, y

$x\uparrow \rightarrow y\uparrow \rightarrow$ +ve correlation | $x\downarrow \rightarrow y\downarrow$ ↳ +ve correlation

$x\uparrow \rightarrow y\downarrow \rightarrow$

* if two variables move in opposite direction then correlation is said to be negative correlation.

$x\uparrow \rightarrow y\downarrow \rightarrow$ −ve correlation, $x\downarrow \rightarrow y\uparrow \rightarrow$ −ve correlation



Positive Correlation

Negative Correlation

No Correlation

Scatter /dot diagram

Karl Pearson 's Coefficient of Correlation $[r_{xy}$ OR $r(x,y)]$

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

$\bar{x}$ = Mean of x

$\bar{y}$ = Mean of y

$x_i-\bar{x} \Rightarrow$ deviation from Mean

$y_i-\bar{y}$ = Deviation from Mean

OR

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

$S.D = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}}$

$$\sqrt{\frac{\sum(x_i-\bar{x})^2}{n}}\sqrt{\frac{\sum(y_i-\bar{y})^2}{n}} \qquad = \sigma_x$$

$$\boxed{r_{xy} = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x\sigma_y}}$$

OR ✓

$r \to [-1 \quad 1]$

$r=0$

$\underbrace{r=-1 \text{ OR } 1}_{\text{Perfect correlation}}$

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

$r=-1$ Negative

$r=1$ +ve

Ex. Find Coefficient of correlation for following data

X:	1	3	5	7	8	10
Y:	8	12	15	17	18	20

$\boxed{n=6}$

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
$\sum X = 34$	$\sum Y = 90$	$\sum XY = 582$	$\sum X^2 = 1248$	$\sum Y^2 = 1446$

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

$$r_{xy} = \frac{6\times 582-(34)(90)}{\sqrt{6\times 1248-(34)^2}\sqrt{6\times 1446-(90)^2}}$$

$$\boxed{r_{xy} = .9879}$$

+ve

=

Ex. From the following data , find no. of items n $\quad n = ?$

$r_{xy} = .5$, $\Sigma X^2 = 90$, $\Sigma XY = 120$, $\sigma_y = 8$
where X and Y are deviations of x and y from its mean.

[2016]

$r = .5$

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

deviation

$X = x_i - \bar{x}$

$Y = y_i - \bar{y}$

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

$$.5 = \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n} \times n}}$$

$$.5 = \frac{120}{\sqrt{90} \times 8\sqrt{n}} \Rightarrow .5 = \frac{120}{3\sqrt{10} \times 8\sqrt{n}}$$

$$\Rightarrow \frac{5}{\sqrt{10}\sqrt{n}} = .5$$

$$\Rightarrow \frac{25}{10n} = .25 \Rightarrow 25 = 2.5n$$

$$n = \frac{25}{2.5} = 10$$

$$n = 10$$

Ex. A computer while calculating correlation Coefficient between two variables X and Y from 25 pairs of observations obtained the following results;

$n = 25$, $\Sigma X = 125$, $\Sigma X^2 = 650$, $\Sigma Y = 100$, $\Sigma Y^2 = 460$

$\Sigma XY = 508$

It was, later discovered that he had copied down two pairs as

X	Y
6	14
8	6

While the correct values were

X	Y
8	12
6	8

Obtain correct value of correlation Coefficient.

Coorected $\Sigma X = 125 - 6 - 8 + 8 + 6$
$= 125$

Corrected $\Sigma Y = 100 - (14+6) + (12+8)$
$= 100$

Corrected $\Sigma X^2 = 650 - (6^2 + 8^2) + (8^2 + 6^2) = 650$

" $\Sigma Y^2 = 460 - (14^2 + 6^2) + (12^2 + 8^2) = 436$

" $\Sigma XY = 508 - (6 \times 14 + 8 \times 6) + (8 \times 12 + 6 \times 8) =$
520

$$r_{xy} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{n \Sigma X^2 - (\Sigma X)^2} \sqrt{n \Sigma Y^2 - (\Sigma Y)^2}}$$

$$r_{xy} = \frac{(25 \times 520) - (125 \times 100)}{\sqrt{25 \times 650 - (125)^2} \sqrt{25 \times 436 - (100)^2}}$$

$= .6666 \ldots$

≈ 0.67

Unit-3(Lec 9) Maths-4

By-Monika Mittal (MM)

Topic: Rank Correlation

Correlation coeff

Rank correlation:

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

n = No of observation

$D = R_1 - R_2$

.

Ex. Ten students got following marks in Maths and Chemistry. Calculate rank correlation coefficient.

Marks(Chem.) : 78 36 98 25 75 82 90 62 65 39

Marks (Maths) : 84 51 91 60 68 62 86 58 63 47

Sol.

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
					26

$r = 1 - \frac{6\sum D^2}{}$

$n = 10$

$$n(n^2-1)$$

$$= 1 - \frac{6 \times 26}{10(10^2-1)} = 1 - \frac{6 \times 26}{10 \times 99}$$

$$= .8424$$

Ex. 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 in beauty contest ?

Sol

Competitor	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
							200	60	214

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

$$= 1 - \frac{6(200)}{10(99)} =$$

$$= -.212$$

×

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

$$= 1 - \frac{6 \times 60}{10 \times 99}$$

↑↑ $= .636$ +ve

↑ ↓

$$r_{23} = 1 - \frac{6 \Sigma D_{23}^2}{n(n^2-1)} = 1 - \frac{6(214)}{10 \times 99}$$

$= -.297 \uparrow\downarrow \times$

$=$

Unit-3(Lec 10) Maths-4

By- Monika Mittal (MM)

Topic: Rank Correlation (Repeated rank)

$$\rho = 1 - \frac{6\left[\Sigma d^2 + F\right]}{n(n^2-1)}$$

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

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

Ex. Obtain the rank correlation Coefficient for 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

X	R(x)	Y	R(y)	$D=R_x-R_y$	D^2
68 - 4	4	62 —	5	-1	1
64 (5)	6	58 —	7	-1	1
75 2	2.5	68 3	3.5	-1	1
50 - 9	9	45	10	-1	1
64 (6)	6	81 - 1	1	5	25
80 1	1	60 - 6	6	-5	25
75 3	2.5	68 4	3.5	-1	1
40	10	48	9	1	1
55 - 8	8	50 —	8	0	0
64 (7)	6	70 - 2	2	4	16
					72

$\frac{2+3}{2} = \frac{5}{2} = 2.5$

$75 \to 2 \to m_1$
$64 \to 3 \to m_2$
$68 \to 2 \to m_3$

$\frac{5+6+7}{3} = 6$

$\frac{3+4}{2} = 3.5$

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

$$= 1 - 6\left[\Sigma D^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}\right]$$

$n(n^2-1)$

$$= 1 - \frac{6\left[72 + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12} + \frac{2(2^2-1)}{12}\right]}{10(10^2-1)}$$

$$= 1 - \frac{6 \times 75}{990} = .545$$

By - Monika Mittal (MM)

Topic: **Regression Analysis (Part 1)**

Regression means functional relationship between two or more related variables.

If two variables are correlated then there exists a relation between them.
Regression Analysis is processes of estimating relationship between dependent variable and independent variables.

Linear Regression

When points in scatter diagram are concentrated around a straight line then it is Linear Regression otherwise Non - Linear Regression.

Linear

$\bar{x}, \bar{y}$ } mean

There are two line of regression

Y ON X

$\boxed{y = a+bx}$ $\Rightarrow$ $(y-\bar{y}) = b_{yx}(x-\bar{x})$, $\boxed{b_{yx} = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}}$

Eqn

Reg. coeff y on x

OR $\boxed{b_{yx} = r\dfrac{\sigma_y}{\sigma_x}}$

$r \to$ corr. coeff

X ON Y

x=A+BY

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

$\boxed{b_{xy} = r\dfrac{\sigma_x}{\sigma_y}}$ OR $b_{xy} = \dfrac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$

$\boxed{b_{xy} = r\dfrac{\sigma_x}{\sigma_y}}$

Both line of regression passes through $(\bar{x}, \bar{y})$ $\leftarrow$ Imp.

i.e $\left\{ \begin{array}{l} \bar{y} = a + b\bar{x} \\ \bar{x} = A + B\bar{y} \end{array} \right\}$ (y on x) $y = a+bx$; (x on y) $x = A+By$

Regression can be used for Prediction and forecasting.

Ex. Find coefficient of Correlation (r) and obtain the equation of lines of Regression for the following data. $\rightarrow$ y on x, x on y [2017]

x: 6 2 10 4 8

y: 9 11 5 8 7

Estimate the value of y when $x = 5$ after finding regression Equation.

Sol $r = \dfrac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$ (Coeff of corr.)

x	y	xy	x^2	y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\Sigma x = 30$	$\Sigma y = 40$	$\Sigma xy = 214$	$\Sigma x^2 = 220$	$\Sigma y^2 = 340$

$n = 5$

$$r = \frac{n \times \Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{5 \times 214 - 30 \times 40}{\sqrt{5 \times 220 - (30)^2}\sqrt{5 \times 340 - (40)^2}}$$

$$\boxed{r = -.9192}$$

Y ON X $\quad y - \bar{y} = b_{yx}(x - \bar{x})$

b_{yx} → Reg. coeff.

$\bar{x} = ?$

$\bar{x} = \frac{\Sigma x}{n} = \frac{30}{5} = 6$ $\bar{y} = ?$

$\bar{y} = \frac{\Sigma y}{n} = \frac{40}{5} = 8$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{5 \times 214 - 30 \times 40}{5 \times 220 - (30)^2}$$

$$b_{yx} = -.65$$

Y ON X

$$y - 8 = -.65(x - 6)$$

$$\boxed{y = -.65x + 11.9}$$ ✓

X ON Y

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

$$b_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2} = \frac{5 \times 214 - 30 \times 40}{5 \times 340 - (40)^2}$$

$$b_{xy} = -1.3$$

$$x - 6 = -1.3(y - 8)$$

$$\boxed{x = -1.3y + 16.4}$$ X ON Y

b_{xy}, b_{yx}, r

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

y on x $\quad y = -.65x + 11.9$

$x = 5 \quad y = ?$

$$y = -.65 \times 5 + 11.9$$

$y =$ [illegible]

$$\boxed{y = 8.65}$$

Topic: Regression Analysis (Part 2)

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

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

Reg coeff

Properties of Regression coefficients

1) imp. Correlation Coeff. is geometric mean of regression coefficients.

Reg. Coeff

$b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$ (y on x)

$b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$ (x on y)

$\sqrt{b_{yx}\,b_{xy}} = \sqrt{r\cancel{\dfrac{\sigma_y}{\sigma_x}}\, r\cancel{\dfrac{\sigma_x}{\sigma_y}}} = r$

$$\boxed{r=\sqrt{b_{yx}\,b_{xy}}}$$

2) If one regression coeff is greater than unity then other regression coeff must be less than unity.

Let $\boxed{b_{yx}>1}$ then $\dfrac{1}{b_{yx}}<1$

$r=\sqrt{b_{xy}\,b_{yx}}$

we know that $b_{yx}\,b_{xy}=r^2$

$r^2\leq 1$

$-1\leq r\leq 1$

$b_{yx}\,b_{xy}\leq 1$

$b_{xy}\leq \dfrac{1}{b_{yx}}<1 \Rightarrow \boxed{b_{xy}<1}$

v.v. imp

Angle between two regression lines

If θ is acute angle b/w two regression lines, then

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

[AKTU 2015, 2017]

r → corr. coeff

Proof :→ y on x $\Rightarrow y-\bar{y}=b_{yx}(x-\bar{x}) \Rightarrow (y-\bar{y}) = r\dfrac{\sigma_y}{\sigma_x}(x-\bar{x})$ — (1)

$\underline{X \text{ ON } Y} \Rightarrow x-\bar{x} = b_{xy}(y-\bar{y}) \Rightarrow (x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$ — (2)

Line (1) $\Rightarrow m_1 = r\frac{\sigma_y}{\sigma_x}$,

Line (2) $\Rightarrow m_2 = \frac{\sigma_y}{r\sigma_x}$,

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

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

$$\tan\theta = \pm \frac{(1-r^2)}{r}\frac{\sigma_y}{\cancel{\sigma_x}}\frac{\sigma_x^{\cancel{2}}}{\sigma_x^2+\sigma_y^2} = \pm \frac{1-r^2}{r}\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

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

$r^2 \leq 1$

$\left.\begin{matrix}\sigma_x \\ \sigma_y\end{matrix}\right\}$ +ve

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

$\theta \to$ acute angle

when $r=0$ $\tan\theta = \infty$

$\theta = \pi/2$

$r = \pm 1$ $\tan\theta = 0 \Rightarrow \theta = 0$

reg. lines coincide $\rightarrow$ perfect correlation

Ex. In a partially destroyed lab record of a analysis of a correlation data, the following results only are legible

✓ Variance of $x = 9$

✓ Regression Eq^n: $8x - 10y + 66 = 0$

$40x - 18y = 214$

[AKTU 2018]

What are (a) Mean Value of x and y ✓$(\bar{x}, \bar{y})$

(a) Mean value of ... y

(b) the Standard deviation of y ✓(σ_y)

(c) Coeff of correlation ✓ (r)

Sol^n → Both Reg lines passes through $(\bar{x}, \bar{y})$

$$\left\{\begin{array}{l} 8x - 10y + 66 = 0 \Rightarrow \\ 40x - 18y = 214 \Rightarrow \end{array}\right\} \quad \left.\begin{array}{l} 8\bar{x} - 10\bar{y} + 66 = 0 \\ 40\bar{x} - 18\bar{y} - 214 = 0 \end{array}\right\} \text{ solve}$$

After solving $\boxed{\bar{x} = 13 \qquad \bar{y} = 17}$

(b) (c) σ_y, r

$\sigma_x^2 = 9$

$\boxed{\sigma_x = 3}$

$r = \sqrt{b_{xy}\, b_{yx}}$

$$\left\{\begin{array}{l} b_{yx} = r\dfrac{\sigma_y}{\sigma_x} \\ b_{xy} = r\dfrac{\sigma_x}{\sigma_y} \end{array}\right\}$$

$8x - 10y + 66 = 0 \Rightarrow$

y ON x

$10y = 8x + 66$

$\boxed{y = .8x + 6.6}$ — (1)

$40x - 18y = 214 \Rightarrow$ X ON Y

$40x = 18y + 214$

$\boxed{x = .45y + 5.35}$

y ON x $\quad b_{yx} = .8$

x ON y $\quad b_{xy} = .45$

$r = \sqrt{.8 \times .45}$

$\boxed{r = 0.6}$

$\boxed{\sigma_y = ?}$

$b_{yx} = \dfrac{r\sigma_y}{\sigma_x}$

$.8 = .6\dfrac{\sigma_y}{3}$

$\sigma_y = \dfrac{.8 \times 3}{.6} = 4$

$\boxed{\sigma_y = 4}$

Non-Linear Regression

Let $y = a + bx + cx^2$ be second degree parabolic curve of regression y on x

By Using Normal Eq^{n}s

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad \text{—(1)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad \text{—(2)}$$
$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad \text{—(3)}$$

3

$a, b, c = ?$

$y \quad x$

Multiple Linear Regression

where dependent Variable is a function of two or more independent Variables.

Consider linear function

$$y = a + bx + cz$$

[this is called Regression Plane]

$$\Sigma y = ma + b\Sigma x + c\Sigma z \quad \text{—(1)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz \quad \text{—(2)}$$
$$\Sigma yz = a\Sigma z + b\Sigma xz + c\Sigma z^2 \quad \text{—(3)}$$

a, b, c

$m \rightarrow$ No of observati

Ex. Obtain Multiple linear Regression y on x & z

x	1	2	3	4
z	0	1	2	3
y	12	18	24	30

$\boxed{m = 4}$

Sol

$$y = a + bx + cz$$

$$\begin{cases} \Sigma y = ma + b\Sigma x + c\Sigma z \\ \Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma zx \\ \Sigma yz = a\Sigma z + b\Sigma xz + c\Sigma z^2 \end{cases}$$

x	z	y	x^2	z^2	xy	xz	yz
1	0	12	1	0	12	0	0
2	1	18	4	1	36	2	18
3	2	24	9	4	72	6	48
4	3	30	16	9	120	12	90
$\Sigma x = 10$	$\Sigma z = 6$	$\Sigma y = 84$	$\Sigma x^2 = 30$	$\Sigma z^2 = 14$	$\Sigma xy = 240$	$\Sigma xz = 20$	$\Sigma yz = 156$

$$\begin{cases} 84 = 4a + b10 + c6 & \text{--- (1)} \\ 240 = a10 + b30 + c20 & \text{--- (2)} \\ 156 = a6 + b20 + c14 & \text{--- (3)} \end{cases}$$

On solving $\Rightarrow$ $a = 10$

$b = 2$

$c = 4$

Thank You