Page No. ______
Date

MODULE-3

Statistical Techniques - I

(I) Measure of Central Tendency :-

(i) Mean :-

for grouped data

$$(\text{Mean})\ \bar{x} = \frac{\Sigma f_i x_i}{N} \quad ; \quad N = \Sigma f_i$$

for ungrouped data

$$\bar{x} = \frac{\Sigma x_i}{n} \quad ; \quad n = \text{no of frequency}$$

(ii) Median :-

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

for ungrouped frequency :

(a) n is odd, the middle value gives the median $= \left(\frac{n+1}{2}\right)^{th}$ value.

(b) n is even, there are two middle values $= \left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$

then the arthematic mean of two middle values gives the median.

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

Page No. ____ Date

for grouped frequency:

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

$l \rightarrow$ lower limit of median class, where median class is the class interval corresponding to cumulative frequency (CF) just $\geq \frac{N}{2}$.

$h \rightarrow$ width of the median class

$f \rightarrow$ frequency of median class

$N \rightarrow \Sigma f$

$C \rightarrow$ cumulative frequency (c.f) of the preceding class from the median class.

⟨iii⟩ **Mode** :- It is the value which occurs most frequently in a set of ~~obs~~ observation. Point of maximum frequency.

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

modal class :- which have highest frequency.

$l \rightarrow$ lower limit of modal-class

$f_m \rightarrow$ frequency of the modal-class

$f_1 \rightarrow$ frequency of pre-modal class.

$f_2 \rightarrow$ frequency of post modal class

$h \rightarrow$ width of modal class

Page No. ____ Date

⟨iv⟩ For a symmetrical distribution mean, mode, median coincide.

⟨v⟩ Empirical mode formula

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

where the method of grouping fails then the mode is calculated by Empirical formula.

Q. Calculate Mean, median, mode from the following:

CI	6.5-7.5	7.5-8.5	8.5-9.5	9.5-10.5	10.5-11.5	11.5-12.5	12.5-13.5
Frequency	5	12	25	48	32	6	1

$= \frac{\text{upper} + \text{lower}}{2}$

	Class Interval (C.I)	mid value of C.I (x_i)	Frequency (f_i)	($f_i x_i$)	cumulative frequency (C.f)
	6.5-7.5	7	5	35	5
	7.5-8.5	8	12	96	17 - (5+12)
	8.5-9.5	9	25 f_1	225	$\boxed{42}$ = C
median class	$\boxed{9.5-10.5}$	10	f = $\boxed{48}$ f_m	480	90
	10.5-11.5	11	32 f_2	352	122
	11.5-12.5	12	6	72	128
	12.5-13.5	13	1	13	129
			$\Sigma f_i = 129$	$\Sigma f_i x_i = 1273$	

always equal.

Page No. ______ Date

Mean :- $\bar{x} = \frac{\Sigma f_i x_i}{N = \Sigma f_i} = \frac{1273}{129} = 9.86$

Median :- Median $= \frac{N}{2} = \frac{129}{2} = 64.5$

so median class is just $\geq \left(\frac{N}{2}\right)$

{ 64.5 is just close to 90 (c.f) so median class is which have closest value of c.f to the median }

median class = 9.5 − 10.5 (lower limit, upper limit)

$l = 9.5$

$h =$ (midvalue of median class) $\Rightarrow 10.5 - 9.5$

$= 1$

$f = 48$

$C = 42$ (c.f of pre-median class)

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

$$= 9.5 + \frac{1}{48}\left(64.5 - 42\right)$$

$\boxed{\text{Median} = 9.96}$

Page No. ______ Date

Mode :- modal class → C.I of highest frequency $= 9.5 - 10.5$

$l = 9.5$

$h = (10.5 - 9.5) = 1$

$f_m = 48$

$f_1 = 25$

$f_2 = 32$

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

$$= 9.5 + \left[\frac{48 - 25}{2 \times 48 - 25 - 32}\right] \times 1$$

$\boxed{\text{Mode} = 10.08}$

Q.

CI	1−5	6−10	11−15	16−20	21−25
Frequency	7	10	16	32	24

$\frac{\text{upper} - \text{lower}}{2} = \frac{6 - 5}{2} = 0.5$

−0.5 to lower limit ; +0.5 to upper limit

C.I	modified C.I	x_i ($\frac{0.5+5.5}{2}$)	f_i	$x_i f_i$	C.F
1−5	0.5−5.5	3	7	21	7
6−10	5.5−10.5	8	10	80	17
11−15	10.5−15.5	13	16 f_1	208	$\boxed{33}$ C
16−20	$\boxed{15.5-20.5}$	18	$\boxed{32}$ f_m	576	65
21−25	20.5−25.5	28	24 f_2	672	89
			$\Sigma f_i = 89$	$\Sigma f_i x_i = 1557$	

i) $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1557}{89} = 17.4$

ii) median $= \frac{N}{2} = \frac{89}{2} = 44.5$

median class $= 15.5 - 20.5$

$l = 15.5$

$h = 20.5 - 15.5 = 5$

$f = 32$

$c = 33$

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

$$= 15.5 + \frac{5}{32}(44.5 - 33)$$

$$= 15.5 + 0.15\,(11.5)$$

$$\boxed{\text{Median} = 17.225}$$

Mode $\Rightarrow$ modal class $= 15.5 - 20.5$

$l = 15.5$, $h = 5$

$f_m = 32$, $f_1 = 16$, $f_2 = 24$

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

$$= 15.5 + \left(\frac{32 - 16}{2 \times 32 - 16 - 24}\right) \times 5$$

$$= 15.5 + \left(\frac{16}{24}\right) \times 5$$

$$= 15.5 + 0.66 \times 5$$

$$= 15.5 + 3.33$$

Mode $= 18.83$

(II) Measure of Dispersion

Moment $\rightarrow$ deviation of item from some given number. (In here deviation in mean value)

'r' $\rightarrow$ no of moments

i) Moment about mean (μ_r) / Central Moment

If x_1, x_2, x_3 --- are the value of a variable 'x' with corresponding frequencies f_1, f_2, f_3 ---
Then r^{th} moment about mean is

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

where
$r = 0, 1, 2, 3$ - -
$N = \Sigma f_i$
$\bar{x}$ = mean

(a) If $r = 0$ then $\mu_0 = \frac{\Sigma f_i}{\Sigma f_i} = 1$

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

(b) If $r = 1$, then $\mu_1 = \frac{\Sigma f_i x_i}{\Sigma f_i} - \frac{\bar{x}\Sigma f_i}{\Sigma f_i}$

$\boxed{\mu_1 = 0}$ always

It is also called the variance = square of standard deviation

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

Similarly we can calculate μ_3, μ_4 - - - .

Scanned with CamScanner

Page No. ____ Date ____

ii) **Moment about any point (μ'_r)**

The moment about any given point 'A' is

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

(a) If $r=0$, then $\boxed{\mu'_0 = 1}$

(b) If $r=1$, then $\mu'_1 = (\bar{x} - A)$
$= (\text{mean} - A)$

(c) If $r=2$, then $\mu'_2 = \frac{\sum f_i (x_i - A)^2}{N}$

similarly, we can calculate μ_3, μ_4 ---.

iii) **Moment about Origin (ν_r)**

The r^{th} moment about orgin is

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

(a) If $r=0$, then $\boxed{\nu_0 = 1}$

(b) If $r=1$, then $\nu_1 = \frac{\sum f_i x_i}{\sum f_i} = \bar{x}(\text{mean})$

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

similarly we can calculate ν_2, ν_3 ---.

Page No. ____ Date ____

5 marks question

Relation between μ_r and μ'_r :-

$\mu_1 = 0$

$\mu_2 = \mu'_2 - (\mu'_1)^2$

$\mu_3 = \mu'_3 - 3\mu'_2 \mu'_1 + 2(\mu'_1)^3$

$\mu_4 = \mu'_4 - 4\mu'_3 \mu'_1 + 6\mu_2 (\mu'_1)^2 - 3(\mu'_1)^4$

Relation between μ_r and ν_r :-

$\nu_1 = \bar{x}$ (mean)

★ $\nu_2 = \mu_2 + (\bar{x})^2$ → variance

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

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

Karl Pearson's Coffecient :

β- cofficient :

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

γ- cofficient :

$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$; $\gamma_2 = \beta_2 - 3$

for mean :- $\mu'_1 = \bar{x} - A$

$\Rightarrow \bar{x} = \mu'_1 + A = \nu_1$

Numerical on moments.

5 marks 2014

$\boxed{\text{Type - I}}$

Q. In a distribution, the first four moment about a point $x=4$ are -1.5, 17, -30 and 308. Find the moments about mean and about origin. Also calculate β_1 and β_2.

given $A=4$

$\mu_1' = -1.5$; $\mu_2' = 17$, $\mu_3' = -30$; $\mu_4' = 308$

→ Moment about mean (central moment)

$\boxed{\mu_1 = 0}$

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

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = -30 - 3(17)(-1.5) + 2(-30)^3 = \boxed{39.75}$

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

$\boxed{\mu_4 = 342.31}$

→ Moment about origin $\{\mu_1' = x - A\}$

$\nu_1 = \bar{x}\,(\text{mean}) = \mu_1' + A = -1.5 + 4 = 2.5$

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

$\nu_3 = \mu_3 + 3\mu_2(\bar{x}) + (\bar{x})^3 = 166$

$\nu_4 = \mu_4 + 4\mu_3(\bar{x}) + 6\mu_2(\bar{x})^2 + (\bar{x})^4 = 1332$

calculation of β_1 and β_2 :-

$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.4923$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{342.31}{(14.75)^2} = 1.5733$

10 marks

$\boxed{\text{Type - II}}$

Q. calculate the moment about mean of the given data

C.I	0-10	10-20	20-30	30-40	40-50
frequency (f_i)	5	10	40	20	25

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

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

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

C.I	x_i (mid)	f_i	$x_i f_i$	$(x_i-\bar{x})f_i$	$(x_i-\bar{x})^2 f_i$	$(x_i-\bar{x})^3 f_i$	$(x_i-\bar{x})^4 f_i$
0-10	5	5	25	-25	3125	-78125	1953125
10-20	15	10	150	-15	2250	-33750	506250
20-30	25	40	1000	-5	1000	-5000	25000
30-40	35	20	700	5	500	2500	12500
40-50	45	25	1125	15	5625	84375	1265625
		N=100	$\sum x_i f_i$ =3000		12500	-30000	3762500

$$\bar{x} = \frac{\sum f_i x_i}{N} = \frac{3000}{100} = 30$$

$$\mu_1 = 0$$

$$\mu_2 = \frac{\sum f_i (x_i - \bar{x})^2}{N} = \frac{12500}{100} = 125$$

$$\mu_3 = \frac{\sum f_i (x_i - \bar{x})^3}{N} = \frac{-30000}{100} = -300$$

$$\mu_4 = \frac{\sum f_i (x_i - \bar{x})^4}{N} = \frac{3762500}{100} = 37625$$

(iv) <u>Skewness and Kurtosis</u>

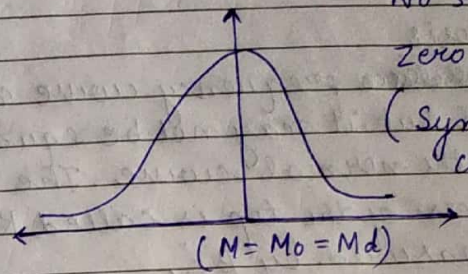
<u>Skewness</u> → Skewness means lack of symmetry or lopsidedness in a frequency distribution

Skewness indicates whether the curve is turned more to one side than to other or whether the curve has a longer tail on one side. Skewness can be positive as well as negative.

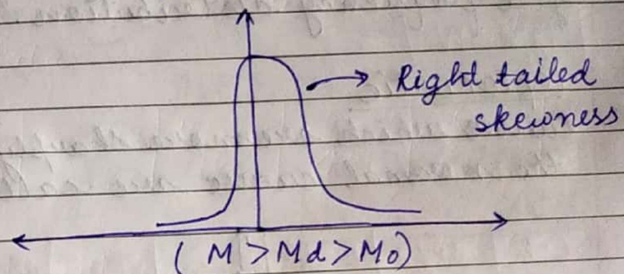
<u>Symmetrical Distribution</u> →

The frequencies are symmetrically distributed about the mean.
Mean, mode and median coincide and median lies half-way between the two quartiles.

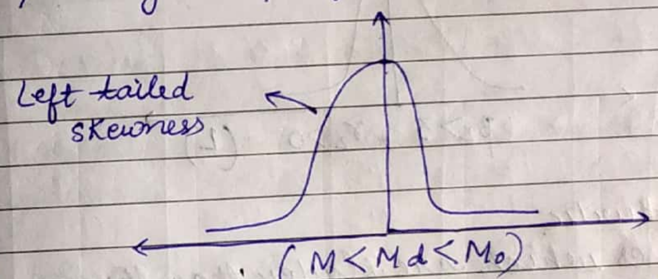
<i> No skewness OR Zero-skewness (Symmetrical curve)



<ii> Positive / Right hand Skewness



<iii> Negative / Left hand Skewness



<u>Moment cofficient of skewness</u> :- It is denoted by 'γ_1'.

coeff of skewness $(\gamma_1) = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}}$

$$\boxed{\text{coff. of skewness } (\gamma_1) = \frac{\mu_3}{\sqrt{\mu_2^3}}}$$

Page No. ____ Date ____

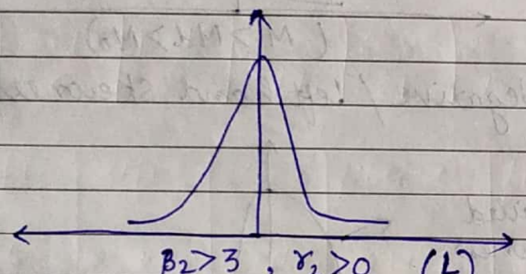
Kurtosis

A ~~[illegible]~~ frequency curve may be symmet-rical but it may not be equally flat topped with the normal curve. The relative flatness of the top is called Kurtosis and is measured by 'β_2'.

Kurtosis refers to the bulginess of the curve of a frequency distribution.

Leptokurtic.

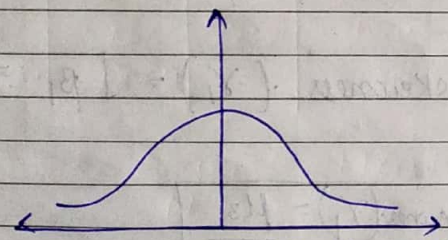
Curve which are more sharply peaked than the normal curve are called leptokurtic.



$\beta_2 > 3$, $\gamma_2 > 0$ (L)

Platykurtic

Curves which are flatter than the normal curves are called platykurtic curve.

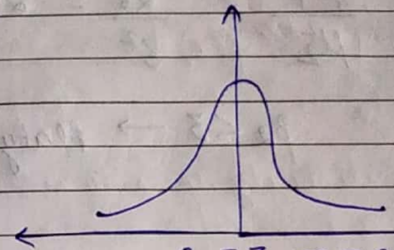


$\beta_2 < 3$, $\gamma_2 < 0$ (P)

Page No. ____ Date ____

Mesokurtic

Curves which are neither flat nor sharply peaked are called normal curves or mesokurtic.



$\beta_2 = 3$, $\gamma_2 = 0$ (M)

Coffecient of Kurtosis $\boxed{(\beta_2) = \frac{\mu_4}{\mu_2^2}}$

$\boxed{\text{Type - III}}$

5 mark Q 2016, 18 The first four moments about $x = 4$ are 1, 4, 10 and 45. find the coff of skewness and Kurtosis also comment upon the nature of curve given $A = 4$, $\mu_1' = 1$, $\mu_2' = 4$, $\mu_3' = 10$, $\mu_4' = 45$

$\boxed{\mu_1 = 0}$

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

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 10 - 3 \times 4 \times 1 + 2 \times 1^3$
$= \boxed{0}$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3\mu_1'^4$
$= 45 - 4 \times 10 \times 1 + 6 \times 4 \times 1 - 3 \times 1$
$= \boxed{26}$

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = 0$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{26}{3^2} = 2.88$$

$\beta_2 < 3 \rightarrow$ Platykurtic curve

$\boxed{\text{Type-IV}}$

10 mark
Q 2013 Calculate first four moments and coeff. of Skewness & Kurtosis

C.I	5-10	10-15	15-20	20-25	25-30	30-35	35-40
fi	6	8	17	21	15	11	2

C.I	xi	fi	xifi	$x_i-\bar{x}$	$f_i(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^3$	$f_i(x_i-\bar{x})^4$
5-10	7.5	6	45	23	3174	73002	1679046
10-15	12.5	8	100	-9.5	-722	-6859	65160.5
15-20	17.5	17	297.5	-4.5	-344.25	-1549.12	6971.04
20-25	22.5	21	472.5	0.5	5.25	2.625	1.3125
25-30	27.5	15	412.5	5.5	453.75	2495.625	13725.93
30-35	32.5	11	357.5	10.5	1212.75	12733.875	133705.63
35-40	37.5	2	75	15.5	480.5	7447.75	115440.12
		$\Sigma f_i = 80$	1760		4260	87273.75	2014050.53

$$\bar{x} = \frac{\Sigma x_i f_i}{N} = \frac{1760}{80} = 22$$

$\mu_1 = 0$

$$\mu_2 = \frac{f_i(x_i-\bar{x})^2}{N} = \frac{4260}{80} = 53.25$$

$$\mu_3 = \frac{f_i(x_i-\bar{x})^2}{N} = \frac{87273.75}{80} \quad 1090.92$$

$$\mu_4 = \frac{2014050.53}{80} = 25175.63$$

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{1090.92}{\sqrt{(53.25)^3}} = \frac{1090.92}{388.57}$$

$\boxed{\gamma_1 = 2.80}$ positive skewness.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{25175.63}{(53.25)^2} = 8.8$$

$\beta_2 > 3 \rightarrow$ Leptokurtic curve

(III) Moment Generating Function (m·g·f)

→ **for continous random variable.**

If x is a continous random variable then m·g·f of 'x' is

$$M_x(t) = \int_a^b e^{tx} f(x)\,dx \quad \text{—— (1)}$$

for $a \le x \le b$

$f(x) \rightarrow$ probability density function (p·d·f)

$t \rightarrow$ is a parameter

→ **for discrete random variable**

If x is a discrete random variable then m·g·f of 'x' is

$$M_x(t) = \sum_{x=a}^{b} e^{tx} P(x) \quad \text{—— (2)}$$

for $a \le x \le b$

$P(x) \rightarrow$ probability mass function (p·m·f)

$t \rightarrow$ is a parameter

Remarks

(1) Mean $(\nu_1) = \left[\frac{d}{dt} M_x(t)\right]_{t=0}$

(2) Variance $(\mu_2) = \nu_2 - (\text{mean})^2$

where $\nu_2 = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0}$

(3) $M_{x+y}(t) = M_x(t) \cdot M_y(t)$

m·g·f of sum of two independent variable is equal to the product of their respective m·g·f.

* 2011, 12, 13

Q obtain the m·g·f of $f(x) = \begin{cases} x, & \text{for } 0<x<1 \\ 2-x, & 1 \le x < 2 \\ 0, & \text{otherwise} \end{cases}$ also calculate mean and variance.

~~soln~~

We know that the m·g·f for continous random variable 'x' is

$$M_x(t) = \int_a^b e^{tx} f(x)\,dx$$

$$M_x(t) = \int_0^1 \underset{v}{e^{tx}} \cdot \underset{u}{x}\,dx + \int_1^2 \underset{v}{e^{tx}} \underset{u}{(2-x)}\,dx + \int_2^\infty e^{tx} \cdot 0\,dx$$

$\Rightarrow \int uv\,dx = uv_1 - u'v_2 + u''v_3 - u'''v_4$

$$= \left[x\frac{e^{tx}}{t} - 1\cdot\frac{e^{tx}}{t^2} + 0\right]_0^1 + \left[(2-x)\frac{e^{tx}}{t} - (-1)\frac{e^{tx}}{t^2} + 0\right]_1^2$$

$$= \left[\left(\frac{e^t}{t} - \frac{e^t}{t^2}\right) - \left(0 - \frac{1}{t^2}\right)\right] + \left[\left(\frac{e^{2t}}{t^2}\right) - \left(\frac{e^t}{t} + \frac{e^t}{t^2}\right)\right]$$

$$= \cancel{\frac{e^t}{t}} - \cancel{\frac{e^t}{t}} + \frac{1}{t^2} + \frac{e^{2t}}{t^2} + \frac{e^t}{t^2} - \frac{e^t}{t^2} - \frac{e^t}{t^2}$$

$$= \frac{1}{t^2}\left[e^{2t} + 1 - 2e^t\right]$$

Page No.
Date

$$= \frac{1}{t^2}\left[(e^t)^2 + 1 - 2e^t\right] = \frac{1}{t^2}\left[e^t - 1\right]^2 \quad (a-b)^2$$

$$= \left[\frac{e^t - 1}{t}\right]^2 \quad \left[\because e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} - - -\right]$$

$$= \left[\frac{\left(\cancel{1} + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + ---\right) - \cancel{1}}{t}\right]^2$$

$$M_x(t) = \left[\frac{\frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + ---}{t}\right]^2$$

m.g.f $\boxed{M_x(t) = \left[\frac{1}{1!} + \frac{t}{2!} + \frac{t^2}{3!} + ---\right]^2}$

Now, mean $(\nu_1) = \left[\frac{d}{dt} M_x(t)\right]_{t=0}$

$$\nu_1 = \left\{2\left[\frac{1}{1!} + \frac{t}{2!} + \frac{t^2}{3!} + --\right] \cdot \left[\frac{1}{2!} + \frac{2t}{3!} + --\right]\right\}_{t=0}$$

$$\nu_1 = 2 \cdot \left(\frac{1}{1!}\right)\left(\frac{1}{2!}\right)$$

$\boxed{\nu_1 = 1}$ mean

Page No.
Date

$$\nu_2 = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0}$$

$$\nu_2 = \left[2 \times \left(\frac{1}{2!} + \frac{2t}{3!} + \frac{3t^2}{4!}\right)\left(\frac{1}{2!} + \frac{2t}{3!} + ---\right)\right.$$
$$\left. + 2\left(\frac{1}{1!} + \frac{t}{2!} + \frac{t^2}{3!}\right)\left(\frac{2}{3!} + \frac{6t}{4!} + --\right)\right]_{t=0}$$

$$\nu_2 = \left[\cancel{2} \times \frac{1}{\cancel{2}} \times \frac{1}{2} + 2 \times 1 \times \frac{\cancel{2}}{1 \times \cancel{2} \times 3}\right]$$

$$\nu_2 = \left[\frac{1}{2} + \frac{2}{3}\right]$$

$$\nu_2 = \frac{7}{6}$$

$$2\left[\frac{1}{1!\,2!} + \frac{2t^2}{2!\,3!} + \frac{t^2\,3t^2}{3!\,4!}\right] - -$$

$\frac{1}{2}$

$$1\ \cancel{2} \times \left[\frac{1}{\cancel{2}} + \left[\frac{4t}{2!\,3!} + \frac{3 \times 4 \times t^3}{3! \times 4!}\right]\right] t = 0$$

$$\nu_2 = \frac{1}{2}\left[+\frac{4t}{3}^{\,0} + \frac{3 \times 4 \times 0}{}\right]$$

$$\mu_2 = \nu_2 - m^2$$

$1 - 1$

$$1\frac{\cancel{2}}{\cancel{2}} + \frac{4t^2}{2!\,3!} + ---$$

$0 + 8t$

2013, 11

Q. find the m.g.f of $P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$. Also find the mean & variance.

m.g.f of discrete random value

$$M_x(t) = \sum_{x=a}^{b} e^{tx} P(x)$$

$$M_x(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \quad ; 0 \le x < \infty$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{e^{tx} \lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} \quad \left[\because e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} \cdots\right]$$

$$= e^{-\lambda} \left[1 + \frac{\lambda e^t}{1!} + \frac{(\lambda e^t)^2}{2!} + \frac{(\lambda e^t)^3}{3!} + \cdots\right]$$

$$= e^{-\lambda} \left[e^{(\lambda e^t)}\right]$$

$$= e^{-\lambda} \cdot e^{\lambda e^t} \quad \left[\because e^x \cdot e^y = e^{x+y}\right]$$

$$= e^{(\lambda e^t - \lambda)}$$

$$\Rightarrow \boxed{e^{\lambda(e^t - 1)} = M_x(t)} \text{ m.g.f} \quad \text{---} (1)$$

PBA

$$\text{mean } (\nu_1) = \left[\frac{d}{dt} M_x(t)\right]_{t=0}$$

$$\nu_1 = \left[e^{\lambda(e^t - 1)} \times \lambda e^t\right]_{t=0}$$

$$\boxed{\nu_1 = \lambda}$$

$$\text{variance } (\nu_2) = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0}$$

$$\nu_2 = \left[\lambda^2 e^{2t} \left(e^{\lambda(e^t - 1)}\right) + \lambda e^{\lambda(e^t - 1)} \times e^t\right]_{t=0}$$

$$\boxed{\nu_2 = \lambda^2 + \lambda}$$

$$\text{Now, variance } (\mu_2) = \nu_2 - (\text{mean})^2$$

$$\mu_2 = \lambda^2 + \lambda - (\lambda)^2$$

$$\boxed{\mu_2 = \lambda}$$

V. Impt

(IV) Method of Least Square [In Curve Fitting]

(a) Fit a Straight line

Let $y = a + bx$ —(1)

be the straight line to be fitted.

or $E = (y - a - bx)$

or $E_i^2 = \sum (y_i - a - bx_i)^2$ —(2)

or [let $E_i^2 = U$]

$U = \sum (y_i - a - bx_i)^2$ —(3)

Now differ (3) w.r.t parameter 'a' & 'b' partially and equate them equal to zero.

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

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

$\Rightarrow \sum y_i = na + b\sum x_i$ {$n \to$ total no of data}

$\Rightarrow \boxed{\sum y = na + b\sum x}$ —(4)

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

$\Rightarrow \sum (x_i y_i - x_i a - bx_i^2) = 0$

$\Rightarrow \sum x_i y_i - a\sum x_i - b\sum x_i^2 = 0$

$\Rightarrow \sum x_i y_i = a\sum x_i + b\sum x_i^2$

$\Rightarrow \boxed{\sum xy = a\sum x + b\sum x^2}$ —(5)

equation (4) and (5) are called Normal Equation.

Q. Fit a straight line for the following data:–

x	2	3	4	5	6	7
y	3.0	5.0	5.5	6.0	8.0	9.5

We know that Normal Equations are:–

$\sum y = na + b\sum x$ —(1)

$\sum xy = a\sum x + b\sum x^2$ —(2)

{where $n = 6$}

x	y	xy	x^2
2	3	6	4
3	5	15	9
4	5.5	22	16
5	6	30	25
6	8	48	36
7	9.5	66.5	49
$\sum x = 27$	$\sum y = 37$	$\sum xy = 187.5$	$\sum x^2 = 139$

from normal equation

$37 = 6a + 27b$

$187.5 = 27a + 139b$

$a = 0.767$

$b = 1.2$

Equation of straight line

$y = a + bx$

Ans. $\boxed{y = 0.767 + (1.2)x}$

(b) Fit a Curve

Q. for the given data fit a curve $y = a + bx + cx^2$ (second degree Parabola)

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

given equation $y = a + bx + cx^2$ — ①

a, b and c are parameters

$E = y - a - bx - cx^2$ or $E_i^2 = \Sigma(y_i - a - bx_i - cx_i^2)^2$

or

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

differenciating eq ② w.r.t a, b & c partially and equating equals to zero

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

$\Rightarrow \Sigma y_i - na - b\Sigma x_i - c\Sigma x_i^2 = 0$

$\Sigma y_i = na + b\Sigma x_i + c\Sigma x_i^2$

$\boxed{\Sigma y = na + b\Sigma x + c\Sigma x^2}$ — ③

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

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

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

$\boxed{\Sigma xy = xa + b\Sigma x^2 + c\Sigma x^3}$ — ④

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

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

$\Rightarrow \Sigma x_i^2 y_i = a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4$

$\Rightarrow \boxed{\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4}$ — ⑤

equation ③, ④ and ⑤ are the required Normal equation.

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	99.0	495.0
15	49.7	55	225	979	194.1	822.9

From Normal Equation $[n = 5]$

$5a + 15b + 55c = 49.7$

$15a + 55b + 225c = 194.1$

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

$a = -0.46$ $\quad c = 0.442$

$b = 1.842$

curve $\Rightarrow y = a + bx + cx^2$

Ans $\boxed{y = -0.46 + 1.842x + 0.442x^2}$

2019, 2010

Q fitting the curve $y = ax + \frac{b}{x}$ by least sq method

x	1	2	3	4	5	6	7	8
y	5.4	6.2	8.2	10.3	12.6	14.8	17.2	19.5

equation : $y = ax + \frac{b}{x}$ —— (1)

$$E = y - ax - \frac{b}{x}$$

$$E_i^2 = \left(y_i - ax_i - \frac{b}{x_i}\right)^2$$

$$U = \Sigma\left(y_i - ax_i - \frac{b}{x_i}\right)^2 \quad \text{—— (2)}$$

differenciating eq (2) partially w.r.t a & b and equating equal to zero

$$\frac{\partial U}{\partial a} = 2\Sigma\left(y_i - ax_i - \frac{b}{x_i}\right)(-x_i) = 0$$

$$\Rightarrow \Sigma y_i x_i - a\Sigma x_i^2 - bn = 0$$

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

now $\frac{\partial u}{\partial b} = 2\Sigma\left(y_i - ax_i - \frac{b}{x_i}\right)\left(-\frac{1}{x_i}\right) = 0$

x_i^{-1}

$-\frac{1}{x_i^2}$

$$\frac{\Sigma y_i}{x_i} - \Sigma a - \frac{\Sigma b}{x_i^2} = 0$$

$\cancel{\frac{1}{x_i}\left(\Sigma \frac{y_i}{x_i} - an - b\Sigma\frac{1}{x_i^2}\right) = 0}$

$$\frac{\Sigma y_i}{x_i} = an + b\Sigma\frac{1}{x_i^2}$$

$$\boxed{\Sigma\left(\frac{y}{x}\right) = an + b\Sigma\left(\frac{1}{x^2}\right)} \quad \text{—— (4)}$$

x	y	xy	$\frac{y}{x}$	x^2	$\frac{1}{x^2}$
1	5.4	5.4	5.4	1	1
2	6.2	12.4	3.1	4	0.25
3	8.2	24.6	2.7	9	0.11
4	10.3	41.2	2.5	16	0.062
5	12.6	63	2.5	25	0.04
6	14.8	88.8	0.9	36	0.027
7	17.2	120.4	2.4	49	0.020
8	19.5	156	2.4	64	0.015
36	94.2	511.8	21.9	204	1.52

Normal eqn will be

$$204a + 8b = 511.8$$

$$8a + 1.52b = 21.9$$

$$a = -9.98, \quad b = 66.97.$$

Ans. $\boxed{y = (-9.98)x + \frac{66.97}{x}}$

Scanned with CamScanner

I

Impt.

Q. $PV^{\gamma} = (\text{constant})\ C$, find the curve using least sq method :-

$V(cm^3)$	50	60	70	90	100
$P(Kg/cm^2)$	64.7	51.3	40.5	25.9	78

$PV^{\gamma} = C$ is the given equation.

Let $P = CV^{-\gamma}$

$\log P = \log C + (-\gamma)\log V$ —— (1)

comparing with

$Y = A + BX$

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

and so the normal equations will be

$\Sigma Y = nA + B\Sigma X$ —— (2)

$\Sigma XY = A\Sigma X + B\Sigma X^2$ —— (3)

V	P	$Y = \log P$	$X = \log V$	XY	X^2
50	64.7	1.81090	1.69897	3.07666	2.88650
60	51.3	1.71012	1.77815	3.04085	3.16182
70	40.5	1.60746	1.84510	2.96592	3.40439
90	25.9	1.41330	1.95424	2.76193	3.81905
100	78	1.89209	2	3.78418	4
		= 8.43387	= 9.27646	= 15.6295	= 17.2717

Normal eqns :-

$5A + 9.27646B = 8.43387$ —— (4)

$9.27646A + 17.2717B = 15.6295$ —— (5)

$A = 2.22476 \qquad B = -0.29091$

$\gamma = -B = 0.29091$

$\log C = A$

$C = \text{Antilog}(A) = \text{anti}(2.22476)$

$\boxed{C = 167.7876}$

$\Rightarrow PV^{\gamma} = C$

Ans. $\boxed{PV^{0.29091} = 167.7876}$

Q. fit a curve $y = C_0 x + \frac{C_1}{\sqrt{x}}$, for the data

x	0.2	0.3	0.5	1	2
y	16	14	11	6	3

$E \Rightarrow y = C_0 x + \frac{C_1}{\sqrt{x}}$ —— (1)

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

$E_i = y_i - C_0 x_i + \frac{C_1}{\sqrt{x_i}}$

$E_i^2 = \left(y_i - C_0 x_i + \frac{C_1}{\sqrt{x_i}}\right)^2$

$U = \Sigma\left(y_i - C_0 x_i + \frac{C_1}{\sqrt{x_i}}\right)^2$ —— (2)

differenciating eq (2) w.r.t C_0 & C_1 partially

$$\frac{\partial U}{\partial C_0} = 0$$

$$\frac{\partial U}{\partial C_0} = 2\Sigma\left(y_i - C_0x_i + \frac{C_1}{\sqrt{x_i}}\right) \times (-x_i) = 0$$

$$\Rightarrow \Sigma y_i x_i - \Sigma C_0 x_i^2 + \Sigma C_1\sqrt{x_i} = 0$$

$$\Rightarrow \Sigma y_i x_i = C_0\Sigma x_i^2 - C_1\Sigma\sqrt{x_i}$$

$$\boxed{\Sigma xy = C_0\Sigma x^2 - C_1\Sigma\sqrt{x}} \quad (3)$$

$$\frac{\partial U}{\partial C_1} = 0$$

$$\frac{\partial U}{\partial C_1} = 2\Sigma\left(y_i - C_0x_i + \frac{C_1}{\sqrt{x_i}}\right)\left(\frac{1}{\sqrt{x_i}}\right) = 0$$

$$\Rightarrow \Sigma\frac{y_i}{x_i} - \Sigma C_0\sqrt{x_i} + \frac{\Sigma C_1}{x_i} = 0$$

$$\Rightarrow \Sigma\left(\frac{y_i}{x_i}\right) = C_0\Sigma\sqrt{x_i} - C_1\Sigma\frac{1}{x_i}$$

$$\Rightarrow \boxed{\Sigma\left(\frac{y}{x}\right) = C_0\Sigma\sqrt{x} - C_1\Sigma\left(\frac{1}{x}\right)} \quad (4)$$

eq (3) and (4) are normal equations

x	y	xy	y/x	x^2	$\sqrt{x}$	$1/x$
0.2	16	3.2	80	0.04	0.447	5
0.3	14	4.2	46.6	0.09	0.547	3.3
0.5	11	5.5	22	0.25	0.707	2
1	6	6	6	1	1	1
2	3	6	1.5	4	1.414	0.5
4	50	24.9	156.1	5.38	4.1152	11.8

−7.71

−7.48

16.13

−15.84

from 3 and 4 we get normal equation

$$C_0\, 5.38 - C_1\, 4.1152 = 24.9$$
$$C_0\, 4.1152 - C_1\, 11.8 = 156.1$$

$$C_0 = -7.71 \qquad C_1 = 16.13$$

$$y = C_0 x + C_1\left(\frac{1}{\sqrt{x}}\right)$$

Ans. $$\boxed{y = -7.71x + 16.13\left(\frac{1}{\sqrt{x}}\right)}$$

Page No. ____ Date

(V) **Correlation**

In any distribution when change in one variable affect a change in other variable, then variable are said to be correlated. It is measured by correlation coefficient or coefficient of correlation.

for eg:

x	x_1	x_2	---	x_n
y	y_1	y_2	---	y_n

x and y are correlated.

(a) Karl-Pearson's coefficient of correlation
OR
Moment Product correlation-coefficient

for different values of $x = x_1, x_2, x_3 ---$, the value of y are $y_1, y_2, y_3 ---$. Then Coefficient of correlation between x & y are defined as

$$\rho = \rho(x,y) = \gamma(x,y) = \gamma_{xy} \Rightarrow \gamma$$

$$\boxed{\gamma = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\,\sigma_y}}$$

$$\boxed{\text{or } \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2}}}$$

where n = no of data

$$\bar{x} = \text{mean of } x = \frac{\Sigma x}{n} = \frac{\Sigma f_i x_i}{N}$$

Page No. ____ Date

$$\bar{y} = \text{mean of } y = \frac{\Sigma y}{n} = \frac{\Sigma f_i y_i}{N}$$

$$\sigma_x = \text{S.D (standard deviation) of } x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\sigma_y = \text{S.D of } y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

Remark

1. Generally γ lies between -1 to 1
   i.e $-1 \le r \le 1$
2. For $0 < \gamma \le 1$, Positive-correlation
3. For $-1 \le \gamma < 0$, Negative-correlation
4. For $\gamma = 0$, no correlation

Q.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
10	18	-10	-11	121	110	100
14	12	-6	-7	49	42	36
18	24	-2	-3	9	6	4
22	6	2	1	1	2	4
26	30	6	5	25	30	36
30	36	10	9	81	90	100
120	126			286	280	280

$$\bar{x} = \frac{\Sigma x}{n} = \frac{120}{6} = 20$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{126}{6} = 21$$

$$\gamma = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\,\sigma_y}$$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{280}{6}} = \sqrt{46.6} = 6.826$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\frac{286}{6}} = \sqrt{47.6} = 6.904$$

$$\gamma = \frac{280}{6 \times 6.826 \times 6.904}$$

$$\gamma = \frac{280}{282.76} = 0.99$$

Ans $\boxed{\gamma = 0.99}$ (much closer to 1)

i.e strong-positive correlation

(b) Rank Correlation Coefficients

$$\gamma = 1 - \left[\frac{6\,\Sigma D_i^2}{n(n^2-1)}\right]$$

where $D_i = R_1 \sim R_2$
(difference of ranks)

$n \rightarrow$ no of data

Remark

① Since we know that Karl's Pearson coefficient of corr. is :-

$$\gamma = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\,\sigma_y} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2-(\Sigma x)^2}\cdot\sqrt{n\Sigma y^2-(\Sigma y)^2}}$$

also

② $$\gamma = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\,\sigma_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2}}$$

2016

Q. find the rank correlation coeff. of following data.

Marks A (R_1)	Marks B (R_2)	$D_i = R_1 - R_2$	D_i^2
15	50	−35	1225
20	30	−10	100
27	55	−28	784
13	30	−17	289
45	25	20	400
60	10	50	2500
20	30	−10	100
75	70	5	25
			$\Sigma D_i^2 = 5423$

Page No. ____ Date ____

Rank-Correlation cofficient $\quad n = 8$

$$\gamma = 1 - \left[\frac{6 \Sigma Di^2}{n(n^2-1)}\right]$$

$$= 1 - \left[\frac{6 \times 5423}{8 \times (8^2-1)}\right]$$

$$= 1 - \left[\frac{32538}{504}\right] = 1 - 64.559$$

$$\boxed{\gamma = -63.559}$$

Q. given that $n = 25$, $\Sigma x = 125$, $\Sigma x^2 = 650$

$\Sigma y = 100$, $\Sigma y^2 = 460$ and $\Sigma xy = 508$

find the value of corr coffec γ :-

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

$$\gamma = \frac{25 \times 508 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2} \sqrt{25 \times 460 - (100)^2}}$$

$$\gamma = \frac{12700 - 12500}{\sqrt{16250 - 15625} \cdot \sqrt{11500 - 10000}}$$

$$\gamma = \frac{200}{\sqrt{625} \times \sqrt{1500}} = \frac{200}{25 \times 38.72}$$

$$\gamma = 200/968.2 = \boxed{0.206} = \gamma$$

Page No. ____ Date ____

Q.

R_1	R_2	R_3	D_{12} R_1-R_2	D_{23} R_2-R_3	D_{31} R_1-R_3	D_{12}^2	D_{23}^2	D_{31}^2
1	3	6	-2	-3	-5	4	25	9
6	5	4	1	1	2	1	4	1
5	8	9	-3	-1	-4	9	16	1
10	4	8	6	-4	2	36	4	16
3	7	1	-4	6	2	16	4	36
2	10	2	-8	8	0	64	0	64
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	1	81
7	6	5	1	1	2	1	4	1
8	9	7	1	2	1	1	1	4
						200	60	214

$n=10$

Rank correlation coeff. b/w first & second

$$\gamma_{12} = 1 - \left[\frac{6 \Sigma D_{12}^2}{n(n^2-1)}\right] = 1 - \left[\frac{6 \times 200}{10(99)}\right] = -0.212$$

$$\gamma_{23} = 1 - \left[\frac{6 \times 214}{10 \times 99}\right] = -0.297$$

$$\gamma_{13} = 1 - \left[\frac{6 \times 60}{10 \times 99}\right] = 0.636$$

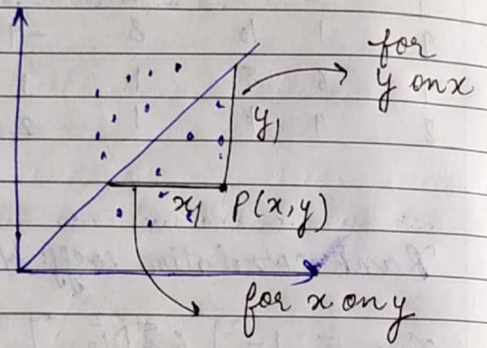
★ Derivation in exam

Page No. Date

Impt.

(VI) Regression Analysis

Line of Regression :– A straight line which is passing through the data at which maximum numbers of data lies around or on it, then straight line is called line of Regression.



Two types of line of Regression

⟨1⟩ y on x :– (line parallel to y axis)

$$(y-\bar{y}) = \cancel{x}\, \frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

OR

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

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

(coff of regression)

Page No. Date

⟨2⟩ x on y :– (line parallel to x-axis)

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

OR

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

where $\bar{x}$ = mean of x , $\bar{y}$ = mean of y

r = coff of correlation

σ_x = S.D of x , σ_y = S.D of y

Remark

1. When distance of point to the line is taken along parallel to y-axis then line of regression is called y on x and vice-versa.

2. $b_{yx} \times b_{xy} = r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y}$

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

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

(GM)

i.e. coeffi. of correlation is Geometric mean between regression coefficient b_{xy} and b_{yx}.

2015, 17, 18, 19

V. Impt

Page No. ____ Date ____

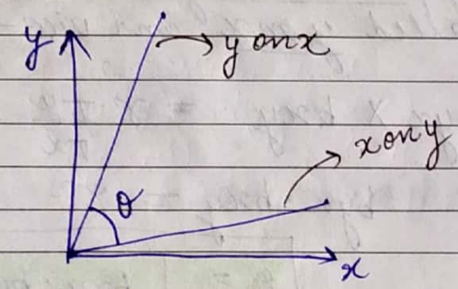
Derivation of Angle between two lines of Regression

OR

Ques:→ If θ is acute angle between lines of regression then show that $\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \cdot \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$

where symbols have their usual meaning.
Also discuss the significance of formula at $r=0$ and $r=\pm 1$.

Let θ be the acute angle between two lines of regression.
Then slopes of lines of regression will be calculated by the equations of line of regressions.



y on x :- $(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$ — ①

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

$\frac{1}{r}\cdot\frac{\sigma_y}{\sigma_x}(x-\bar{x}) = (y-\bar{y})$ — ②

comparing eq ① & ② with $y = mx+c$

Page No. ____ Date ____

hence slope of line ① $m_1 = r\frac{\sigma_y}{\sigma_x}$ — ③

and slope of line ② $m_2 = \frac{1}{r}\frac{\sigma_y}{\sigma_x}$ — ④

and we know that $\tan\theta = \frac{m_2-m_1}{1+m_1m_2}$

$$\tan\theta = \frac{\frac{1}{r}\frac{\sigma_y}{\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1+\frac{1}{r}\frac{\sigma_y}{\sigma_x}\times r\frac{\sigma_y}{\sigma_x}}$$

$$= \frac{\left(\frac{1}{r}-r\right)\frac{\sigma_y}{\sigma_x}}{1+\frac{(\sigma_y)^2}{(\sigma_x)^2}} = \frac{\left(\frac{1-r^2}{r}\right)\frac{\sigma_y}{\sigma_x}}{\frac{(\sigma_x)^2+(\sigma_y)^2}{(\sigma_x)^2}}$$

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

$$\boxed{\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x\cdot\sigma_y}{\sigma_x^2+\sigma_y^2}\right)}$$ Hence Proved.

→ If $r=0$, then $\tan\theta = \infty$
$\theta = \frac{\pi}{2}$

i.e for $r=0$ both lines are $\perp^r$ to each other

→ If $r=\pm 1$, then $\tan\theta = 0$
$\theta = 0$

i.e for $r=\pm 1$ both lines are parallel to each other.

Page No. ____ Date ____

Numericals

2017

Q. For two random variable x & y with same mean the two regression equations are :-
$y = ax+b$ and $x = \alpha y + \beta$, then show that
$\frac{b}{\beta} = \left(\frac{1-a}{1-\alpha}\right)$ also find common mean.

From given equations $y = ax+b$ —(1)
$x = \alpha y + \beta$ —(2)

we get $b_{yx} = a$
$b_{xy} = \alpha$

Let 'm' be the common mean for x & y both
i.e $\bar{x} = \bar{y} = m$

Now Eqns of regression lines
$(y-\bar{y}) = b_{yx}(x-\bar{x})$

y on x :- $\Rightarrow (y-m) = a(x-m) \Rightarrow$ ~~y = ax +~~
$\Rightarrow y = ax + m(1-a)$ —(3)

x on y :- $(x-m) = \alpha(y-m)$
$\Rightarrow x = \alpha y + m(1-\alpha)$ —(4)

By comparing eq (1) (2), (3) and (4)

$b = m(1-a)$ and $\beta = m(1-\alpha)$

$\therefore \frac{b}{\beta} = \frac{m(1-a)}{m(1-\alpha)}$

$\boxed{\frac{b}{\beta} = \left(\frac{1-a}{1-\alpha}\right)}$ Hence proved.

Page No. ____ Date ____

Now for the mean, Since 'm' is common mean for both x & y then it satisfy the given eqn of regression lines.

$b = m(1-a)$
$b = m - am$
$m = b + am$ and $m = \beta + \alpha m$

By equating them
$b + am = \beta + \alpha m$
$am - \alpha m = \beta - b$
$m(a-\alpha) = \beta - b$

Ans. $\boxed{m = \frac{\beta - b}{a-\alpha}}$ is the mean

2016

Q. given eqn of regression lines : $x = -0.4y + 6.4$ and $y = -0.6x + 4.6$. Find (i) Mean of x & y
ii) coeff. of correlation.

$x = -0.4y + 6.4$ —(1)
$y = -0.6x + 4.6$ —(2)

let $(\bar{x}, \bar{y})$ be the mean of (x, y), then it satisfy the eqns of regression.

i.e $\bar{x} + 0.4\bar{y} - 6.4 = 0$
$0.6\bar{x} + \bar{y} - 4.6 = 0$

Ans. $\boxed{\bar{x} = 6 \quad ; \quad \bar{y} = 1}$

Scanned with CamScanner

Variance $\rightarrow (S.D)^2$

Page No. ____ Date ____

(b) now given eq of regression

$x = -0.4y + 6.4 \longrightarrow x$ on y

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

$y = -0.6x + 4.6 \longrightarrow y$ on x

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

$\therefore$ By comparing above equations we get

$b_{xy} = -0.4$ & $b_{yx} = -0.6$

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

$\sqrt{-0.4 \times -0.6} = r$

$\sqrt{-0.024} = r$

$\boxed{-0.489 = r}$

coeff. of corregression

2017,19

Q. In Partially destroyed lab records of an analysis of corr. data, variance of $x = 9$, regression eqns :- $8x - 10y + 66 = 0$

$40x - 18y = 214$

Find. S.D of y and corr. coeff of x & y

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

x on $y \Rightarrow 40x - 18y = 214$

Page No. ____ Date ____

regression eqn of y on x $(y-\bar{y}) = b_{yx}(x-\bar{x})$

$\Rightarrow 10y = 8x + 66$

$y = 0.8x + 6.6$

on comparing

$\therefore b_{yx} = 0.8$

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

$40x = 214 + 18y$

$x = \frac{18}{40}y + \frac{214}{40}$

$x = 0.45y + 5.35$

$\therefore b_{xy} = 0.45$

now $r^2 = b_{xy} \times b_{yx}$

$r^2 = 0.45 \times 0.8$

$r = \sqrt{0.36}$

$\boxed{r = 0.6} \rightarrow$ coeff of corre.

now given $(\sigma x)^2 = 9 \rightarrow \sigma x = 3$

$b_{yx} = 0.8$

$r \frac{\sigma y}{\sigma x} = 0.8$

$(0.6) \frac{\sigma y}{\sigma x} = 0.8$

$\sigma y = \frac{0.8 \times 3}{0.6} = 4$

$\sigma y = 4$ (S.D of $y = 4$)

$\boxed{(\sigma y)^2 = 16} \leftarrow$ variance of y

Page No.
Date

2019.

Q. The equation of two regression lines, obtained in a correlation analysis of 60 observation are :- $5x = 6y + 24$ and $1000y = 768x - 3608$ then what is the correlation cofficient? Also show that ratio of coeff. of variability of x to y is $\left(\frac{5}{24}\right)$ & what is the ratio of variance of x and y.

$$5x = 6y + 24 \qquad (x-\bar{x}) = b_{xy}(y-\bar{y})$$

$$x = 1.2y + 4.8$$

$$\boxed{b_{xy} = 1.2}$$

$$1000y = 768x - 3608$$

$$y = 0.768x - 36.08$$

$$\boxed{b_{yx} = 0.768}$$

$$r = \sqrt{b_{xy} \times b_{yx}} = \sqrt{1.2 \times 0.768} = \sqrt{0.9216}$$

$$\boxed{r = 0.96} \longrightarrow \text{coff of correlation}$$

$$b_{xy} = r\frac{\sigma x}{\sigma y} \quad \text{and} \quad b_{yx} = r\frac{\sigma y}{\sigma x}$$

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

$$\frac{1.2}{0.768} = \left(\frac{\sigma x}{\sigma y}\right)^2$$

$$\sqrt{1.5625} = \frac{\sigma x}{\sigma y}$$

$$\boxed{\frac{\sigma x}{\sigma y} = 1.25} = 5/4$$

Page No.
Date

coeff of variability of $x = \frac{\sigma x}{\bar{x}}$

coeff of variabily of $y = \frac{\sigma y}{\bar{y}}$

let $\bar{x}$ & $\bar{y}$ are means, then from eqn of Regression lines

$$5\bar{x} = 6\bar{y} + 24$$

$$1000\bar{y} = 768\bar{x} - 3608$$

$$\bar{x} = 6 \text{ and } \bar{y} = 1$$

now coff of variabily of $x = \frac{\sigma x}{\bar{x}} = \frac{\sigma x}{6}$

$$\frac{\sigma y}{\bar{y}} = \frac{\sigma y}{1}$$

$$\text{Ratio of variability} = \frac{\left(\frac{\sigma x}{6}\right)}{\left(\frac{\sigma y}{1}\right)}$$

$$= \frac{\sigma x}{\sigma y} \times \frac{1}{6}$$

$$= \frac{5}{4} \times \frac{1}{6}$$

$$\boxed{\text{Ratio of Variability} = \frac{5}{24}}$$

Hence Proved.