①

PICT, PUNE

Assignment No 03

Statistics

Roll. No 21118

Dop: 18/05/2021

Dos: 25/05/2021

Q1 The first four moments of a distribution about the assumed mean ~~are~~ 30 are 0.3, 6, 30 & 400. Calculate the first four moments about the mean, standard deviation, mean & coefficient of skewness & kurtosis.

→

Given: $A = 30$

$\mu_1' = 0.3, \ \mu_2' = 6, \ \mu_3' = 30, \ \mu_4' = 400$

Solution:

$\mu_0 = 1, \ \mu_1 = 0$ .... standard result.

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

$= 6 - (0.3)^2$

$= 5.91$ —— ①

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

$= 30 - 3 \times 6 \times 0.3 + 2(0.3)^3$

$= 24.546$ —— ②

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

$= 567.2157$ —— ③

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

$= 30 + 0.3$

$\bar{x} = 30.03$ = Mean.

Standard deviation $= \sqrt{\mu_2}$

$= \sqrt{5.91}$ ∴ from ①

$= 2.431$

Coefficient of skewness $(\beta_1) = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(24.546)^2}{(5.91)^3}$

$= 17.249$

PICT, PUNE

Coefficient of kurtosis $(\beta_2) = \frac{\mu_4}{\mu_2^2}$

$= \frac{567.21}{(5.91)^2}$ ... Substuting values from eq^n ① & ③

$= 16.24$

Ans: $\boxed{\begin{array}{l} \mu_0 = 0,\ \mu_1 = 1,\ \mu_2 = 5.91,\ \mu_3 = 24.546,\ \mu_4 = 567.21 \\ \beta_1 = 17.249,\ \beta_2 = 16.24 \\ \text{standard deviation} = 2.431 \end{array}}$

Q2 Find coefficient of correlation between X & Y, given that

$n = 25,\quad \Sigma x = 75,\quad \Sigma y = 100,\quad \Sigma x^2 = 250,\quad \Sigma y^2 = 500,\quad \Sigma xy = 325$

→

Coefficient of co-relation between x & y can be given as,

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

Substituting values, we get

$$= \frac{25.325 - 75.100}{\sqrt{24.(250)\ 24.(500)}}$$

$$= \frac{25.325 - 7500}{24.100.\sqrt{10}}$$

$= 0.073$

$\boxed{\therefore \text{Coefficient of corelation between } x \,\&\, y \text{ is } 0.073}$

(2)

Roll. No
21118

PICT, PUNE

Q3 The regression equations are: $8x-10y+66=0$ & $40x-18y=214$
the value of variance of x is 9.
find a) The mean values of x & y
b) The correlation between x & y
c) The standard deviation of y.

→

Given $8x-10y+66=0 \Rightarrow y = \frac{4}{5}x + \frac{33}{5}$ — ①
$40x-18y=214 \Rightarrow y = \frac{20}{9}x - \frac{107}{9}$ — ②

Assume that $\bar{x}$ & $\bar{y}$ are the mean of x & y

$\therefore \bar{y} = \frac{4}{5}\bar{x} + \frac{33}{5}$

$\bar{y} = \frac{20}{9}\bar{x} - \frac{107}{9}$

on solving above equations,
we get
$\boxed{\bar{x} = 13 \text{ \& } \bar{y} = 17}$ Ans a)

Taking combination of line y on x with other line x on y,

$y = \frac{4}{5}x + \frac{33}{5}$
$x = \frac{9}{20} + \frac{107}{20}$ — ③

similarly,
$x = \frac{5}{4}x - \frac{33}{4}$
$y = \frac{20}{9}x - \frac{107}{9}$ → ④

From eqn ③
$b_{xy} = \frac{9}{20}$, & $b_{yx} = \frac{4}{5}$

$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{\frac{9}{20} \cdot \frac{4}{5}} = \pm\frac{6}{10}$

$\therefore r = \pm 0.6$

~~Answer~~, the correlation between x & y should be $+0.6$.

Ans b)

standard deviation of $y = \dfrac{r\,\sigma_x}{b_{xy}}$

$= \dfrac{0.6 \times 3}{9/20}$ ... substituting values from last eqn also $\sigma_x^2 = 9$ $\therefore \sigma_x = 3$

$= 4$

Ans c)

Q4 Fit a straight line to the following data:

x	71	68	73	69	67	65	66	67
y	69	72	70	70	68	67	68	64

→

Assume line as $y = a + bx$ .. slope-intercept-form

Normal equations:

$\sum y = na + b\sum x$ — ①

$\sum xy = a\sum x + b\sum x^2$ — ②

here $n = 8$

X	Y	XY	X^2
71	69	4899	5041
68	72	4896	4624
73	70	5110	5329
69	70	4830	4761
67	68	4556	4489
65	67	4355	4225
66	68	4488	4356
67	64	4288	4489
$\sum x = 546$	$\sum y = 548$	$\sum XY = 37422$	$\sum X^2 = 37314$

3

Roll. No
21118

substitute these values in eqn ① & ②
we get
$$548 = 8a + 546b$$
$$37422 = 546a + 37314b$$
On solving above eqns,
we get
$$a = \frac{435}{11} \quad \& \quad b = \frac{14}{33}$$
∴ equation of straight line is
$$\boxed{\begin{aligned} y &= a + bx \\ y &= \frac{435}{11} + \frac{14}{33}x \end{aligned}}$$
Ans

Q5 find the second degree parabola to the following:

x	10	15	20	25	30	35	40
y	11	13	16	20	27	34	38

→ Assume equation of parabola as $y = a + bx + cx^2$
$n = 7$

x	y	x^2	xy	x^2y	x^3	x^4
10	11	100	110	1100	1000	10000
15	13	225	195	2925	3375	50625
20	16	400	320	6400	8000	160000
25	20	625	500	12500	15625	390625
30	27	900	810	24300	27000	810000
35	34	1225	1190	41650	42875	1500625
40	38	1600	1520	60800	64000	2560000
$\Sigma x = 175$	$\Sigma y = 159$	$\Sigma x^2 = 5075$	$\Sigma xy = 4645$	$\Sigma x^2y = 149675$	$\Sigma x^3 = 161875$	$\Sigma x^4 = 5481875$

PICT, PUNE

Normal equations are:

$\Sigma y = an + b\Sigma x + c\Sigma x^2$

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

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

substitute values from table in the above equations

we get

$159 = a(7) + b(175) + c(5075)$

$4645 = a(175) + b(5075) + c(161875)$.

$149675 = a(5075) + b(161875) + c(5481875)$

on solving above equations

we get,

$a = 7.78,\ b = 0.1,\ c = 0.0171$.

$\boxed{\begin{array}{l}\therefore \text{Equation of parabola will be} \\ y = 7.78 + 0.1x + 0.0171x^2.\end{array}}$

Ans

Q6 The following are the scores of the two batsman A & B in a series of innings:

A	12	115	6	73	7	19	119	36	84	29
B	47	12	16	42	4	51	37	48	13	0

who is the better score getter & who is more consistent.

→ The batsman with less variation will ~~ba~~ be more consistent.

∴ Calculating coefficient of variation:

here $\Sigma A = 500$, $n = 10$

∴ mean $\overline{X}_A = \frac{\Sigma X_A}{n} = \frac{500}{10} = 50$

Roll No
21118

(4)

PICT, PUNE

similarly $\bar{x}_B = \frac{\Sigma x_B}{n} = \frac{270}{10} = 27$

x_A	x_B	$x_A - \bar{x}_A$	$(x_A - \bar{x}_A)^2$	$x_B - \bar{x}_B$	$(x_A - \bar{x}_B)^2$
12	47	-38	1444	20	400
115	12	65	4225	-15	225
6	16	-44	1936	-11	121
73	42	23	529	15	225
7	4	-43	1849	-23	529
19	51	-31	961	24	576
119	37	69	4761	10	100
36	48	-14	196	21	441
84	13	34	1156	-14	196
29	0	-21	441	-27	729
$\Sigma x_A = 500$	$\Sigma x_B = 270$		$\Sigma(x_A - \bar{x}_A)^2 = 17498$		$\Sigma(x_B - \bar{x}_B)^2 = 3542$

for A:

standard deviation $(\sigma_A) = \sqrt{\frac{\Sigma(x_A - \bar{x}_A)^2}{n}}$

$= \sqrt{\frac{17498}{10}}$

$\sigma_A = 41.83$

coefficient of variation $= \frac{\sigma_A}{\bar{x}_A} = \frac{41.83}{50}$

$= 0.8366$

for player B:

standard deviation $(\sigma_B) = \sqrt{\frac{\Sigma(x_B - \bar{x}_B)^2}{n}}$

$= \sqrt{\frac{3542}{10}}$

$\sigma_B = 18.82$

$$\text{coefficient of variation} = \frac{\sigma_B}{\bar{x}_B} = \frac{82.82}{27}$$

$$= 0.697$$

$\therefore$ By comparing coefficient of variation, we have

$$\left(\begin{matrix}\text{coeff. of} \\ \text{variation of A}\end{matrix}\right) > \left(\begin{matrix}\text{coefficient of variation} \\ \text{of player B}\end{matrix}\right)$$

$\therefore$ Player B is more consistent.

Now, $\left(\begin{matrix}\text{the mean of} \\ \text{player A}\end{matrix}\right) > \left(\begin{matrix}\text{Mean of player} \\ \text{B}\end{matrix}\right)$

$\therefore$ Player A is better score getter.