ALL BRANCHES

ENGINEERING MATHEMATICS

Lecture-11

Probability

By- Chetan Sir

Topics to be Covered

- FUNDAMENTAL COUNTING
- ADDITION THEOREM
- CONDITIONAL PROBABILITY
- TOTAL PROBABILITY THEOREM
- BAYE'S THEOREM
- STATISTICS – I (PROBABILITY DISTRIBUTIONS)
- STATISTICS – II (CORRELATION AND REGRESSION)

Fitting of parabola:-

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

$$E_i = y_i - (a + bx_i + cx_i^2)$$

$$u = \sum E_i^2 = \sum (y_i - a - bx_i - cx_i^2)^2$$

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

Normal equations of parabola

$$\boxed{\begin{aligned} \sum y &= an + b\sum x + c\sum x^2 \\ \sum xy &= a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y &= a\sum x^2 + b\sum x^3 + c\sum x^4 \end{aligned}}$$

Solve for a, b, c.

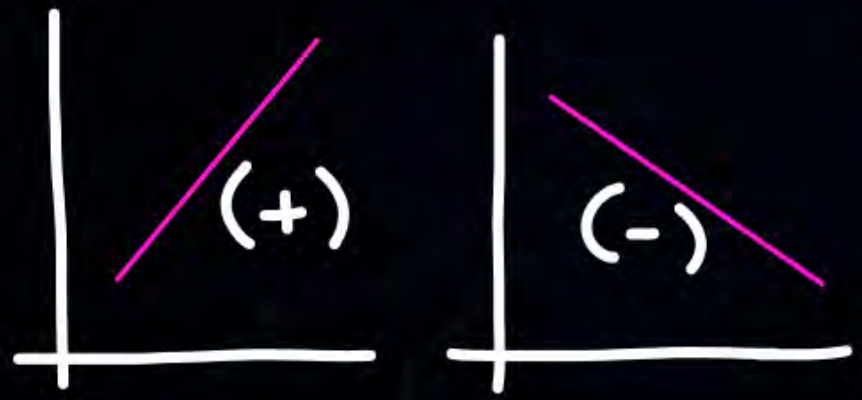
CORRELATION :-

It is statiscal mtd. that measures quantitative/qualitative relationship between different variables.

Types of Correlation

Type I

- Positive correlation
- Negative correlation



Type II

- Simple correlation
(2 variables)
- Multiple correlation
(> 2 variables)
- Partial correlation

Multi. → simultaneous change in variables.

Partial → Variables are changed independently.

Type III

- Linear correlation
- Non-linear correlation

Degree of Correlation:-

$r \rightarrow 0$ to 1 $\quad r \rightarrow -1$ to 0

	Measure of Correlation	Positive	Negative
1.	PERFECT	1	−1
2.	HIGH	0.75 → 1	−0.75 → −1
3.	MODERATE	0.25 → 0.75	−0.75 → −0.25
4.	LOW	0 → 0.25	−0.25 → 0
5.	ZERO CORRELATION	0	0

−1 < Coefficient of correlation < 1

$-1 < r < 1$

METHODS OF ESTIMATING CORRELATION :-

1) Scattered Diagram / Dot Diagram :-

By plotting on graph paper, general trend of dots are observed.

2) Karl Pearson Coefficient of Correlation :-

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

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

$$\left\{ \begin{matrix} u = x - a \\ v = y - b \end{matrix} \right\}$$

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

Ex:-

	Husband age(x)	Wife (Y)age	$u = X-28$	$v = y-23$	u^2	v^2	uv
1.	23	18	-5	-5	25	25	25
2.	27	22	-1	-1	1	1	1
3.	28	23	0	0	0	0	0
4.	29	24	1	1	1	1	1
5.	30	25	2	2	4	4	4
			$\Sigma u = -3$	$\Sigma v = -3$	$\Sigma u^2 = 31$	$\Sigma v^2 = 31$	$\Sigma uv = 31$

$y = x - 5$

$$r = \frac{n\Sigma uv - \Sigma u \Sigma v}{\sqrt{n\Sigma u^2 - (\Sigma u)^2}\sqrt{n\Sigma v^2 - (\Sigma v)^2}} = \frac{5 \times 31 - (-3)(-3)}{\sqrt{5 \times 31 - (-3)^2}\sqrt{5 \times 31 - (-3)^2}} = \frac{146}{146} = 1$$

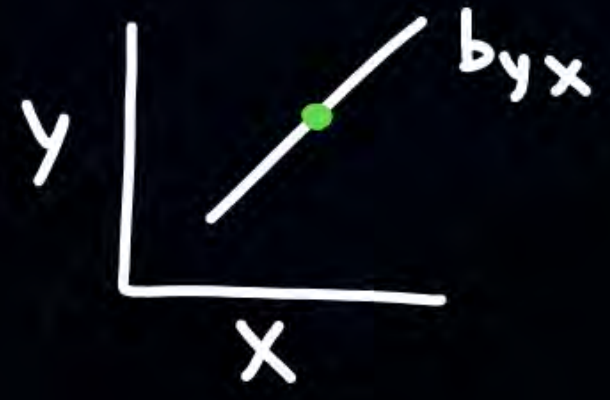
Since $r = 1$; hence perfect positive correlation.

REGRESSION ANALYSIS :- (Regression curve → Regression eqn.)

Regression lines :-

Y on X (Independent)

- $\boxed{(y-\bar{y}) = b_{yx}(x-\bar{x})}$
- $b_{yx} = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$
- $\tan\theta_1 = m_1 = b_{yx}$

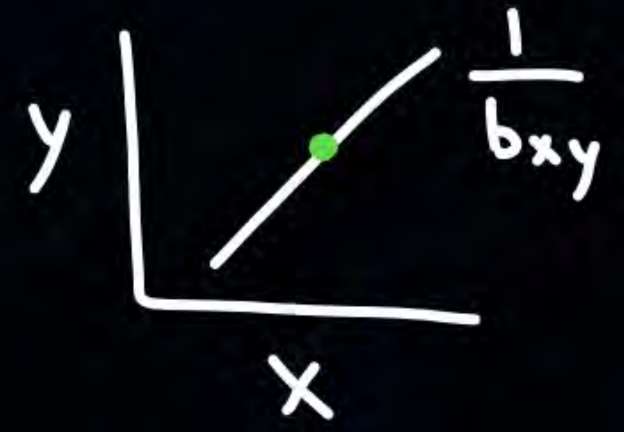


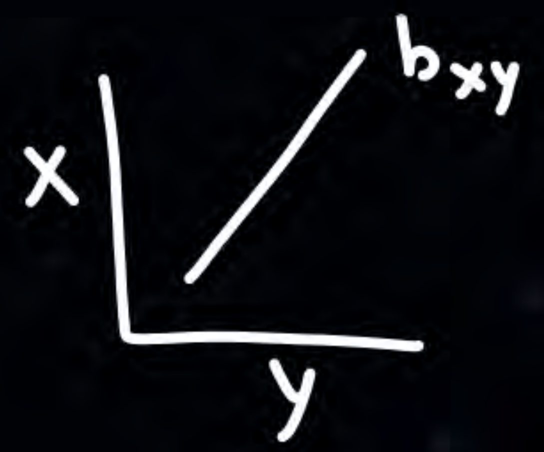
X on Y (Independent)

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

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

$\tan\theta_2 = m_2 = \dfrac{1}{b_{xy}}$





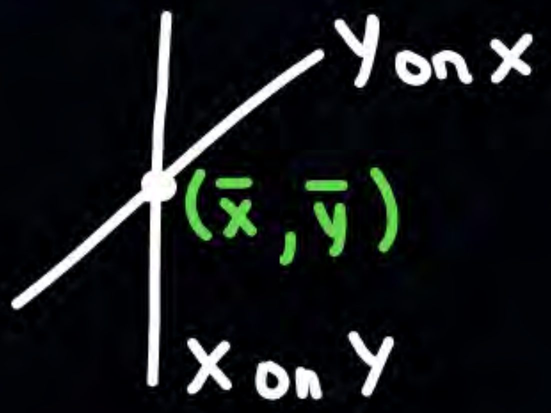
Theorem/ Important points:-

1) $\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}}$ $\boxed{b_{xy} = r\frac{\sigma_x}{\sigma_y}}$

2) In an individual series if we take deviation from actual mean, then the sum of deviations is always zero i.e.

$$\Sigma(x_i - \bar{x}) = 0 \qquad \Sigma(y_i - \bar{y}) = 0$$

3) Both regression lines X on Y and Y on X, intersect at $(\bar{x}, \bar{y})$



Regression coefficients → Y on X (b_{yx})
→ X on Y (b_{xy})

Properties of Regression Coefficients:-

i) G.M. of b_{yx} & b_{xy} $= \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\left(r \cdot \frac{\sigma_y}{\sigma_x}\right)\left(r \cdot \frac{\sigma_x}{\sigma_y}\right)}$

G.M. of b_{yx} & b_{xy} $= r = \sqrt{b_{yx} \cdot b_{xy}}$

ii) $0 < \overbrace{b_{xy} \cdot b_{yx}}^{r^2} < 1$

$\Rightarrow$ If $b_{yx} > 1$, then $b_{xy} < 1$

$\Rightarrow$ If $b_{xy} > 1$, then $b_{yx} < 1$

$-1 < r < +1$

$0 < r^2 < 1$

iii) A.M. of b_{yx} and b_{xy} > Coefficient of correlation

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

iv) b_{xy}, b_{yx} and r have same sign

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

Angle b/w 2 lines of regression :-

$$\tan\theta = \pm\left(\frac{m_2 - m_1}{1 + m_1 m_2}\right) = \boxed{\frac{\frac{1}{b_{xy}} - b_{yx}}{1 + b_{yx}\cdot\frac{1}{b_{xy}}}} = \frac{\frac{\sigma_y}{r\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + \left(r\frac{\sigma_y}{\sigma_x}\right)\left(\frac{\sigma_y}{r\sigma_x}\right)}$$

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

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

$$\boxed{r = \frac{Cov(x,y)}{\sqrt{Var(x)\,Var(Y)}}}$$

Ex:- i) Find regression line Y on X for the following data:-

ii) Estimate the value of Y when $x = 8$

	X	Y	X^2	XY
1.	1	1	1	1
2.	3	2	9	6
3.	4	4	16	16
4.	6	4	36	24
5.	8	5	64	40
6.	9	7	81	63
7.	11	8	121	88
8.	14	9	196	126
	$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma x^2 = 524$	$\Sigma xy = 364$

Regression line y on x,

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

$\bar{x}, \bar{y}, b_{yx}$

$\rightarrow \bar{x} = \frac{\Sigma x}{n} = \frac{56}{8} = 7$

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

$\rightarrow b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{8 \times 364 - 56 \times 40}{8 \times 524 - 56^2}$

$= \frac{7}{11} = 0.636$

$$y - 5 = \frac{7}{11}(x - 7)$$

$$y = \frac{7}{11}x - \frac{49}{11} + 5$$

$$\boxed{y = \frac{7}{11}x + \frac{6}{11}}$$

ii) When $x = 8$; $y = \frac{7}{11}(8) + \frac{6}{11}$

$$y = 5.635$$

Ex: Two variables x and y have regression lines :-

$$3x+2y-26=0 \longrightarrow Y \text{ on } X \quad | \quad X \text{ on } Y$$
$$6x+y-31=0 \longrightarrow X \text{ on } Y \quad | \quad Y \text{ on } X$$

Calculate :-

i) Mean $\bar{x}$ and $\bar{y}$

ii) Regression coefficients b_{xy} and b_{yx}.

iii) Correlation coefficient

iv) Find the angle b/w two lines.

v) the variance of y, if the variance of x is 25.

$$Var(x) = \sigma_x^2 = 25$$
$$\sigma_x = 5$$

i) Regression line intersect at $(\bar{x}, \bar{y})$

$$\bar{x} = 4 \; ; \; \bar{y} = 7$$

ii) let Y on X be $3x + 2y = 26$ X on Y be $6x + y = 31$

$$y = -\frac{3}{2}x + 13 \qquad x = -\frac{y}{6} + \frac{31}{6}$$

$$\boxed{b_{yx} = -3/2} \qquad \boxed{b_{xy} = -1/6}$$

$(0, 1) \leftarrow r^2 = b_{yx} \cdot b_{xy} = \left(-\frac{3}{2}\right)\left(-\frac{1}{6}\right) = \frac{1}{4} < 1$ (We have assumed correctly)

iii) $r = \pm \frac{1}{2}$; $\boxed{r = -1/2}$ $\because$ b_{yx}, b_{xy} and r have same sign.

iv) $\tan\theta = \pm\left(\dfrac{\frac{1}{b_{xy}} - b_{yx}}{1 + \frac{1}{b_{xy}}\cdot b_{yx}}\right)$, $\dfrac{1-r^2}{r}\left(\dfrac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$

$\tan\theta = \pm\left(\dfrac{\frac{1}{-1/6} - \left(-3/2\right)}{1 + \frac{1}{-1/6}\left(-3/2\right)}\right)$ $= \dfrac{1-\left(-1/2\right)^2}{-1/2}\left[\dfrac{5\times 15}{5^2+15^2}\right]$

v) $Var(x) = 25$

$\sigma_x^2 = 25$

$\{\sigma_x = 5\}$

$\sigma_y = 15$

$Var(Y) = \sigma_y^2 = 15^2$

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

$\boxed{Var(Y) = 225}$

$\left(-\dfrac{3}{2}\right) = \left(-\dfrac{1}{2}\right)\cdot\dfrac{\sigma_y}{5}$

Thank you

GW

Soldiers !