REGRESSION

> Regression analysis is a method to predict the value of a variable based on the value of the another variable.

> prediction of dependent variable with the help of independent variable.

Regression Equation

y on x	x on y
$y = a + bx$	$x = a + by$
$y - \bar{y} = byx(x - \bar{x})$	$x - \bar{x} = bxy(y - \bar{y})$

• Regression line y on x ⟶ $y = a + bx$ ⟶ $b = byx$ (regression coefficient) ⟶ Approx change in y due to one unit change in x

\# −ll− x on y ⟶ $x = a + by$ ⟶ $bxy = b$ ⟶ Approx change in x due to change in one unit of y

\# line are derived using least square method.

formulas :- calculation of regression co-efficient

\# $byx = \frac{cov(x,y)}{\sigma_x^2}$ # $bxy = \frac{cov(x,y)}{\sigma_y^2}$

\# $byx = \frac{\Sigma(x-\bar{x}).(y-\bar{y})}{\Sigma(x-\bar{x})^2}$ # $bxy = \frac{\Sigma(y-\bar{y}).(x-\bar{x})}{\Sigma(y-\bar{y})^2}$

\# $byx = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}$ # $bxy = \frac{\Sigma xy - \frac{\Sigma x \times \Sigma y}{N}}{\Sigma y^2 - \frac{(\Sigma y)^2}{N}}$

\# $byx = r\frac{\sigma y}{\sigma x}$ # $bxy = r\frac{\sigma x}{\sigma y}$

Note:-

* $byx \times bxy = r \times \frac{\sigma y}{\sigma y} \times r\frac{\sigma x}{\sigma y} = r^2$

\# $byx \times bxy = r^2$

\# $r = \pm\sqrt{byx \times bxy}$

\# $r^2 \leq 1$

\# $byx \times bxy \leq 1$

\# r is the geometric mean of two regression co-efficients.

\# two regression lines intersects each other at the point $(\bar{x}, \bar{y})$ that is their Arithmetic means.

> How to indentify which Equation is 'y on x' 'x on y'

y on x	x on y
$A_1x + B_1y + C_1 = 0$	$A_2x + B_2y + C_2 = 0$
convert standard form	−ll−
$(y = a + bx)$	$(x = a + by)$
$by = -C_1 - A_1x$	$A_2x = -C_2 - B_2y$
# $b = byx = \frac{-A_1}{B_1}$	# $b = bxy = \frac{-B_2}{A_2}$

→ Change of origin $(+K, -K)$
Regression co-efficient are independent of ~~ori~~ change in origin.

→ change of scale $(\times K, \div K)$ it affect.
To make them Equal

$buv = \frac{\text{scale of } y}{\text{scale of } x} \times byx$

$bvu = \frac{\text{scale of } x}{\text{scale of } y} \times bxy$

→ Angle between two lines

$byx = \text{slope}(M_1)$ $bxy = \text{slope}(M_2)$

$\text{Angle } \theta = \left|\frac{M_1 + M_2}{1 + M_1M_2}\right|$

\# when two regression lines are perpendicular correlation $(r) = 0$

\# when two lines are co-incident to each other then $(r) = r = \pm 1$

- If downward $r = -1$
- If upward $r = 1$
- not mentioned direction $r = \pm 1$

\# $r < \frac{byx + bxy}{2}$

$r <$ AM of two co-efficient of regre..

\# **Bivariate data**: are the data collected for two variable at the same point of time.

\# for bivariate data the frequency table having $(p+q)$ classification the total number of cells is → $(p \times q)$

#7

(#) for the $p \times q$ bivariate frequency table the maximum number of marginal distribution is $\rightarrow 2$

(#) for $P \times Q$ classification of bivariate data the maximum number of condition distribution is $\rightarrow P+q$

(Refer module for deep understanding ↑)

⟹ concept of probable Error.

$P.E = 0.6745\left(\frac{1-r^2}{\sqrt{N}}\right)$

$S.E = \frac{1-r^2}{\sqrt{n}}$ [correlation of population lies in the internal $r - P.E$, $r + P.E$]

$P.E = 0.6745 \times (S.E)$

$r < 6(P.E)$ no evidence of correlation.

$r > 6(P.E)$ evidence of correlation exists.

∴ SE = standard Error.