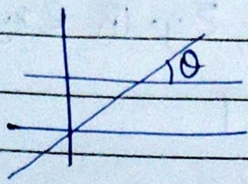
_/_/_

Stats



X on Y

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

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

$b_{xy}=\dfrac{r\sigma_x}{\sigma_y}$ ← regression coeff

Y on X

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

$b_{yx}=r\dfrac{\sigma_y}{\sigma_x}$ ← regression coeff

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

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

$\Rightarrow$ if $b_{xy}\geq 0$, $b_{yx}\geq 0$
$r\geq 0$

$\Rightarrow$ if $b_{xy}\leq 0$, $b_{yx}\leq 0$
$r<0$

Simple Regression : Univariate
Multiple " : Multivariate
Linear " : two variable st line relationship

eg.

$\boxed{x}$	1	2	3	4	5	6	7	8	9
$\boxed{y}$	9	8	10	12	11	13	14	16	15

Ans **X on Y**

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

$r=\dfrac{cov(x,y)}{\sigma_x\sigma_y}=\dfrac{\frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{\Sigma(x-\bar{x})}{n}}\sqrt{\frac{(y-\bar{y})}{n}}}$

$\bar{x}=\dfrac{\Sigma x}{n}=\dfrac{45}{9}=5$

$b_{xy}=0.95\dfrac{\sqrt{\frac{60}{9}}}{\sqrt{\frac{60}{9}}}=0.95=0.95$

$\bar{y}=\dfrac{108}{9}=12$

$(x-5)=0.95(y-12)$

$x=0.95y-6.4$

~~do Y on X similarly~~

Eg $n=18$ $\Sigma x^2=60$ $\Sigma y^2=96$, $\Sigma x=12$, $\Sigma y=18$ $\Sigma xy=48$

Ans $\bar{x}=\frac{\Sigma x}{n}=\frac{12}{18}=\frac{2}{3}$ $\bar{y}=\frac{18}{18}=1$

$r=\frac{cov(x,y)}{\sigma_x\sigma_y}=\frac{E(xy)-E(x)E(y)}{\sqrt{E(x^2)-E(x)^2}\sqrt{E(y^2)-E(y)^2}}$

$=0.597$

$\sigma_x=\sqrt{\frac{60}{18}-\left(\frac{12}{18}\right)^2}=1.69$ $\sigma_y=2.98$

$b_{xy}=0.57\left(\frac{1.69}{2.98}\right)$

$(x-2/3)=b_{xy}(y-1)$

Eg In partially destroyed lab,

$Var(x)=9$ $8x-10y+66=0$, $40x-18y=214$

i) Mean$_{x,y}$ ii) σ_y iii) r

Ans x on y and y on x has point of intersection $(\bar{x},\bar{y})$

$8x-10y=-66$
$40x-18y=214$ $(\bar{x}=13, \bar{y}=17)$

$x=\frac{10}{8}y-\frac{66}{8}$ $y=\frac{40}{18}x-\frac{214}{18}$

$b_{xy}=\frac{10}{8}$ $b_{yx}=\frac{40}{18}$

$r=\sqrt{\frac{10}{8}\cdot\frac{40}{18}}>1$ $\boxed{\text{wrong}}$ $\sigma_x^2=9$

So, $b_{xy}=\frac{18}{40}$ & $b_{yx}=\frac{8}{10}$ $\sigma_x=3$

$r=\sqrt{\frac{18}{40}\cdot\frac{8}{10}}=\frac{4.3}{203}=\frac{3}{5}$

$b_{yx}=\frac{4}{5}=\frac{\sigma_y}{3}\times\frac{3}{5}\Rightarrow\sigma_y=4$

Angle of regression

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

$m_1=\frac{1}{r}\frac{\sigma_y}{\sigma_x}$

y on x

$m_2=r\frac{\sigma_y}{\sigma_x}$

$$\tan\theta = \frac{m_1 - m_2}{1 + m_1 m_2} = \frac{\left(\frac{1}{r} - r\right)\frac{\sigma_y}{\sigma_x}}{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2}} = \frac{1 - r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$$

if $r = 0$, $\theta = \tan^{-1}(\infty) = \pi/2$ perpendicular

if $r = 1$, $\theta = \tan^{-1}(0) = 0$ coincide

if $r = -1$, $\theta = \tan^{-1}(0) = \pi$ opposite

eg if $r = 0.5$ and angle is $\tan^{-1} 3/5$

prove $\sigma_x = \frac{1}{2}\sigma_y$

Ans $\frac{3}{5} = \frac{1 - \frac{1}{4}}{1/2} \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$ $\Rightarrow 2\sigma_x^2(\sigma_x - 2\sigma_y) - \sigma_y(\sigma_x - 2\sigma_y) = 0$

$\Rightarrow \sigma_x = \sigma_y \frac{1}{2}$

Curve fitting

↳ n observations

$(x_1, y_1), \ldots, (x_n, y_n)$ $y = a + bx$

$$\sum_{n=1}^{n} y_i = na + b\sum_{n=1}^{n} x_i$$

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

eg

x	y
1	3
2	4
3	5
4	6
5	7
6	8
Σ 21	3

Σxy ↳ 133 Σx^2 ↳ 91

$y = a + bx$

1) $\Sigma y = na + b\Sigma x$

2) $\Sigma xy = a\Sigma x + b\Sigma x^2$

$33 = 6a + 21b$

$133 = 21a + 91b$

$a = 2 \quad b = 1$

$y = 2 + x$

Parabolic fitting

$y = a + bx + cx^2$

(1) $\Sigma y = na + b\Sigma x + c\Sigma x^2$

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

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