Regression

Method of estimating the value of one variable when that of the other is known & they are correlated.

b_{xy} → x is dependent on independent y.

b_{yx} → y is dependent on independent x.

- Regression
  - Type 1
    - Simple
    - Multiple
  - Type 2
    - Linear
    - Non-linear

→ Method of studying correlation
1. Scatter diagram
2. Least squares

→ Line of regression eq^n

For y on x

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

This can be written as

$$\boxed{y = b_{yx}x + c}$$

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

For x on y

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

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

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

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

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

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

$$\boxed{b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}}$$

$$\boxed{b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}}$$

[or]

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

$$\sigma_x = \sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$\sigma_y = \sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

$$\boxed{b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}$$

$$\boxed{b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

• using change of origin & change of scale

$$b_{yx} = \frac{\sum dxdy - \frac{\sum dx \sum dy}{n}}{\sum d^2x - \frac{(\sum dx)^2}{n}}$$

$$b_{xy} = \frac{\sum dxdy - \frac{\sum dx \sum dy}{n}}{\sum d^2y - \frac{(\sum dy)^2}{n}}$$

→ Properties of ~~correlation~~ regression co-efficient

1. co-efficient of correlation is the geometric mean of the coefficients of regression. i.e. $\boxed{r = \sqrt{b_{yx} b_{xy}}}$
2. If one of the reg coeff is greater than 1, other must be less than 1.

[or]

$b_{yx} b_{xy} \le 1$

3. The A.M. of reg co-eff is greater than or equal to co-eff of corr.

$\boxed{\text{i.e. } \frac{1}{2}(b_{xy} + b_{yx}) \ge r}$

4. Regression co-eff are independent of change of origin but not of scale.

$\boxed{b_{dxdy} = \frac{k}{h} b_{xy}}$

$\boxed{b_{dydx} = \frac{h}{k} b_{yx}}$

$dx = x-a$, $dy = y-b$ (for reg alone)

5. Both reg coeff's will have the same sign.
6. co-eff of corr has same sign as reg coeff's.

→ Properties of lines of regression

1. The two lines of reg x,y & y,x always intersect at their means.
2. r, b_{yx}, b_{xy} all have same signs.
3. If r=0, reg co-eff are 0.
4. Reg. lines become identical if $r = \pm 1$, if r=0 the lines are ⊥ to each other.

Q] Given :- $x + 6y = 6$

$3x + 2y = 10$

~~sample means x & y~~

$x + 6y = 6$

$6y = 6 - x$

$y = 1 - \frac{1}{6}x$

Compare with $y = a + bx$

$b_{yx} = -\frac{1}{6}$

$3x + 2y = 10$

$3x = 10 - 2y$

$x = \frac{10}{3} - \frac{2}{3}y$

compare with $x = a + by$

$b_{xy} = -\frac{2}{3}$

$b_{xy} b_{yx} \le 1$

$-\frac{2}{3} \cdot -\frac{1}{6} = \frac{1}{9} \le 1$

$x = 6 - 6y$

compare with $x = a + by$

$b_{xy} = 6$

$2y = 10 - 3x$

$y = 5 - 3/2\,x$

$b = -3/2$

$b_{xy} * b_{yx} \leq 1$

$-6 * -3/2 = 9 \nleq 1$

$\therefore$ we take first assumption

ii) coeff of corr b/w x & y

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

$r = \sqrt{-1/6 * -2/3}$

$r = \sqrt{-1/9}$

$\boxed{r = -\frac{1}{3}}$ [$\because$ coeff of corr has same sign as reg coeff]

iii) Solve y given $x = 12$

since y is dependent on x

we solve the eq^n having b_{yx}

$y = 1 - \frac{1}{6} x$

$y = 1 - \frac{1}{6} 12$

$\boxed{y = -1}$

i) Sample means x & y

since lines of reg intersect at ~~thei~~ their means

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

$3\bar{x} + 2\bar{y} = 10 \times 3$

$-9\bar{x} = -30$

$x = \boxed{3}$

$\boxed{\bar{y} = \frac{1}{2}}$

Q] Find b_{yx} & b_{xy} & corr coeff

x	4	2	3	4	2
y	2	3	2	4	4

A:

x	y	x^2	y^2	xy
4	2	16	4	8
2	3	4	9	6
3	2	9	4	6
4	4	16	16	16
2	4	4	16	8
Σ 15	15	49	49	44

$$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$= \frac{44 - \frac{15 \cdot 15}{5}}{49 - \frac{(15)^2}{5}}$$

$\boxed{b_{yx} ~~= 10.24~~ -0.25}$

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

$$= \frac{44 - \frac{15 \cdot 15}{5}}{49 - \frac{(15)^2}{5}}$$

$\boxed{b_{xy} = -0.25}$

$r = \sqrt{-0.25 \cdot -0.25}$

$\boxed{r = -0.25}$

DOMS Page No. Date / /

Q] Obt 2 reg lines & corr coeff

Sales	100	98	78	85	110	93	80
Purchase	85	90	70	72	95	81	74

A: $a = 93$, $b = 81$

dx	dy	dx^2	dy^2	$dxdy$
7	4	49	16	28
5	9	25	81	45
-15	-11	225	121	165
-8	-9	64	81	72
17	14	289	196	238
0	0	0	0	0
-13	-7	169	49	91
Σ -7	0	821	544	639

$\Sigma x = 644$ $\Sigma y = 567$

$\bar{x} = 92$ $\bar{y} = 81$

$$b_{yx} = \frac{\Sigma dxdy - \frac{\Sigma dx \Sigma dy}{n}}{\Sigma dx^2 - \frac{(\Sigma dx)^2}{n}}$$

$$= \frac{639 - 0}{821 - 49/7}$$

$= 0.785$

$$b_{xy} = \frac{\Sigma dxdy - \frac{\Sigma dx \Sigma dy}{n}}{\Sigma dy^2 - \frac{(\Sigma dy)^2}{n}}$$

$$= \frac{639}{544}$$

$= 1.174$

x on y

$(x - \bar{x}) = 1.1746(y - 81)$

$\boxed{x = 1.174y - 3.142}$

y on x

$(y - \bar{y}) = 0.785(x - 92)$

$\boxed{y = 0.785x + 8.78}$

→ Curve Fitting

It is an exact relationship b/w two variables by algebraic eq^n.

~~There are 4 ways of~~

Fitting a straight line.

$\boxed{\Sigma y = na + b\Sigma x}$

[0,1]

$\boxed{\Sigma xy = a\Sigma x + b\Sigma x^2}$

Q] Find straight line of best fit

x	1	2	3	4	5
y	14	27	40	55	68

A: St line: $y = a + bx$

Normal eq^n

$\Sigma y = an + b\Sigma x$ — ①

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

x	y	x^2	yx
1	14	1	14
2	27	4	44
3	40	9	120
4	55	16	220
5	68	25	125
Σ 15	204	55	748 ~~423~~

$204 = 5a + 15b$ ×3

$748 = 15a + 55b$

$a = 0$ $y = 13.6$

$y = a + bx$

~~13.6~~ = $\boxed{y = 0 + 13.6x}$

Q] Fit a straight line

x	1	2	3	4	6	8
y	2.4	3.1	3.5	4.2	5.0	6.0

A: Given eq^n : $y = a_0 + a_1 x$

$\Sigma y = a_0 n + a_1 \Sigma x$

$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2$

	x	y	x^2	xy
	1	2.4	1	2.4
	2	3.1	4	6.2
	3	3.5	9	10.5
	4	4.2	16	16.8
	6	5.0	36	30.0
	8	6.0	64	48
Σ	24	24.2	130	113.9

$24.2 = 6a_0 + a_1 24$

$113.9 = a_0 24 + a_1 130$

$a_0 = 2.021$

$a_1 = 0.503$

$\boxed{y = 2.021 + 0.503x}$

Fitting of curve $y = ax + bx^2$

$\boxed{\Sigma xy = a\Sigma x^2 + b\Sigma x^3}$

[or]

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

Q] Fit the curve

x	0	1	2	3
y	2	4	10	15

A: For curve $y = a + bx^2$

$\Sigma y = an + b\Sigma x^2$ — (1)

$\Sigma yx^2 = a\Sigma x^2 + b\Sigma x^4$ — (2)

	x	y	x^2	x^4	yx^2
	0	2	0	0	0
	1	4	1	1	4
	2	10	4	16	40
	3	15	9	81	135
Σ	6	31	14	98	179

$\Sigma y = a$

$31 = 4a + b14$

$179 = a14 + b98$

$a = \cancel{1.731}\ 2.714$

$b = \cancel{1.719}\ 1.438$

$y = a + bx^2$

$\boxed{y = 2.714 + 1.438x^2}$

→ Fitting an exponential curve ($y = ae^{bx}$)

$y = ae^{bx}$

taking log on both sides

$\log_{10} y = \log_{10} a + bx\log_{10} e$

$\boxed{Y = A + BX}$

Normal equations are

$\boxed{\Sigma Y = nA + B\Sigma X}$

&

$\boxed{\Sigma XY = A\Sigma X + B\Sigma X^2}$

Solving these we get A & B.

$a = \text{Antilog } A$

$b = \dfrac{B}{\log_{10} e}$

Q]

x	1	5	7	9	12
y	10	15	12	15	21

A: Given eq^n : $y = ae^{bx}$

Taking log on both sides.