4/10/21

Regression.

$y = a + bx$. — (1) $\qquad a + a + a \cdots a_n$.

normal eqns

taking Σ,

$$\Sigma y = \Sigma a + \Sigma bx \quad \text{— I}$$

$$\Sigma y = na + b\Sigma x. \quad \text{— (I)}$$

× by Σx

$$\Sigma xy = \Sigma ax + \Sigma bx^2 =$$

$$= a\Sigma x + b\Sigma x^2 \quad \text{— (II)}$$

for

$$y = a + bx.$$

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

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

for

$$x = a + by \quad \text{— II}$$

$$\Sigma y = \Sigma a + \Sigma by$$

$$\Sigma y = an + b\Sigma y \quad \text{— (3)}$$

$x \Sigma y.$ $\Sigma xy = a\Sigma y + b\Sigma y^2$ — 4

- Construct the simple linear regression equation of Y on X if

$$n = 7, \sum_{i=1}^{n} x_i = 113, \sum_{i=1}^{n} x_i^2 = 1983,$$

$$\sum_{i=1}^{n} y_i = 182 \text{ and } \sum_{i=1}^{n} x_i y_i = 3186.$$

'eqn. y on x.

$y = a + bx$ — I

$\Sigma y = na + b\Sigma x$ — 1

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

Sub in, eqn 1 & 2

$182 = 7a + b(113)$

$7a + 113b = 182$ — 3

$3186 = a(113) + b(1983)$

$113a + 1983b = 3186$ — 4

solving 3 & 4

$a = 0.7985$

$b = 1.56115$

- Number of man-hours and the corresponding productivity (in units) are furnished below. Fit a simple linear regression
- equation $\hat{Y} = a + bx$ applying the method of least squares.

Man-hours	3.6	4.8	7.2	6.9	10.7	6.1	7.9	9.5	5.4
Productivity (in units)	9.3	10.2	11.5	12	18.6	13.2	10.8	22.7	12.7

man hours (x)	prod. (y)	x^2	xy
3.6	9.3		
4.8	10.7		
7.2			

$n = ?$ $\boxed{n = 9}$

Σx, Σy, Σx^2, Σxy

regr eq?

$n = 9$, $\sum x = 62.1$, $\sum y = 121$

$\sum xy = 897.13$ $\sum x^2 = 468.97$

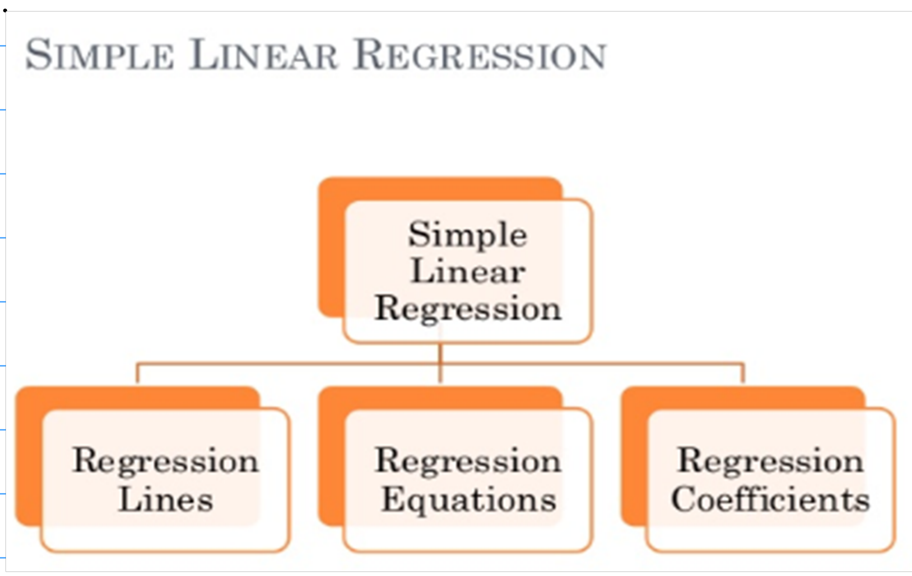
solving.

$a = 2.837$

$b = 1.537$

$y = 2.837 + 1.537x$.

6/10/2021.



Recall:

- Regression Equation of Y on X
- Y=a+bX where Y is the dependent variable to be estimated and X is the independent variable.
- a and b are two unknown constants which determine the position of the line.

$$\sum Y = aN + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2$$

- Regression Equation of X on Y
- The regression equation of X on Y is X=a+bY.

$$\sum X = a'N + b'\sum Y$$

$$\sum XY = a'\sum Y + b'\sum Y^2$$

Regression line y on x.

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

Regression line x on y.

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

Regression coefficients. b_{xy}, b_{yx}.

Regression coefficients measures the average change in the values of one variable for a unit change in the value of another variable. These represent the slope of regression lines

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

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

Correlation coeff:

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

Problems

- The following data gives the experience of machine operators and their performance ratings

Operators	1	2	3	4	5	6	7	8
Experience (*X*)	8	11	7	10	12	5	4	6
Ratings (*Y*)	11	30	25	44	38	25	20	27

- Obtain the regression equations and estimate the ratings corresponding to the experience x=15.

x	y	xy	x^2	y^2
8	11	88		
11	30	330		
7	25			
10	44			
12	38			
5	25			
4	20			
6	27			
$\Sigma x = 63$	$\Sigma y = 220$	$\Sigma xy = 1856$	$\Sigma x^2 = 555$	$\Sigma y^2 = 6780$

reg. eqn : Y on x

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

$$\bar{x} = \frac{\Sigma x}{n} = \frac{63}{8} = 7.875$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{220}{8} = 27.5$$

\#

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

$$= \frac{8 \times 1856 - 63 \times 220}{8 \times 555 - (63)^2}$$

$$= 2.0976$$

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

$$y - 27.5 = 2.0976(x - 7.875)$$

$$y = 2.0976x - (2.0976)(7.875) + 27.5$$

$$y = 2.0976x + 10.98$$

When $x = 15$? $y = ?$

$$y = 2.0976(15) + 10.98$$
$$= 42.44$$

Reg. lines x on y.

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

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

$$= \frac{8 \times 1856 - 63 \times 220}{8 \times 6780 - (220)^2}$$

$$= 0.169$$

reg. eqn. x on y.

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 7.875 = 0.169(y - 27.5)$$
$$x = 0.169y + 3.222$$

2) two regression lines

$$y = -1.5x + 7 \quad \text{——— (1)}$$

$$x = 0.6y + 9 \quad \text{——— (2)}$$

(1) $\rightarrow$ y on x.

(2) $\rightarrow$ x on y.

Is there any mistake in the data.

(1) $\Rightarrow$ reg coeff. $= b_{yx} = -1.5$ ×

(2) $\Rightarrow$ reg coef. $= b_{xy} = 0.6$

both reg coeff are of different sign

So the given eqn. cannot be regression line.

$\Rightarrow$ mistake in data.

3) The reg. coeff. are $b_{yx} = \frac{5}{6}$, $b_{xy} = \frac{9}{20}$.

What is the value of r_{xy}?

corr. coeff $r = \pm\sqrt{b_{xy} \cdot b_{yx}}$

$= \pm\sqrt{\frac{\cancel{9}}{\cancel{6}_2} \times \frac{1\cancel{9}3}{\cancel{24}_4}} = 0.\cancel{547}$

0.612

$= \pm\sqrt{\frac{3}{8}} = 0.612$

$= +\sqrt{3/8}$ why?

Since both signs of b_{xy} & b_{yx} are positive,
the correlation coefficient between x and y is positive

HW

12 data

1

Hours, x	0	1	2	3	3	5	5	5	6	7	7	10
Test score, y	96	85	82	74	95	68	76	84	58	65	75	50
xy	0	85	164	222	285	340	380	420	348	455	525	500
x^2	0	1	4	9	9	25	25	25	36	49	49	100
y^2	9216	7225	6724	5476	9025	4624	5776	7056	3364	4225	5625	2500

12 different students watching TV during weekends and scores of a test conducted on monday.

(a) find the eqn. of regression line.

(b) use the eqn. to find the expected test score for a student who watches 9 hours of TV.

Ans. ?
next class