describing 1.1.6 Benef... Data ...ta

Linear Least Squares:

Definition:

> The least-squares method can be defined as a statistical method that is used to find the equation of the line of best fit related to the given data.

> The line obtained from such a method is called a regression line.

Formula & Steps:

1. Denote the independent variable values as x_i and the dependent ones as y_i.

2. Calculate the average values of x_i and y_i as x and y.

3. Presume the eqn of the line of best fit as $y = mx + c$.

m → slope.

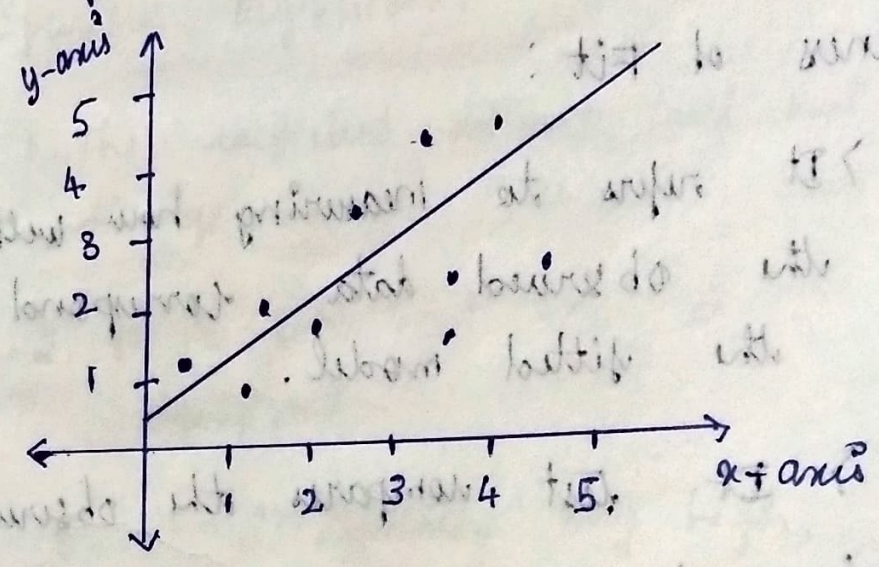
c → Intercept

4. Slope (m) = $\frac{\sum (x - x_i)(y - y_i)}{\sum (x - x_i)^2}$

5. $c = y - mx$

$$y = mx + c$$

Graph:



x – Independent Variables

y - Dependent Variables

Limitations :

> The least square method assumes that the data is evenly distributed and doesn't contain any outliers for deriving a line of best fit.

> But, this method doesn't provide accurate results for unevenly distributed data or for data containing outliers.

Goodness of Fit :

> It refers to measuring how well do the observed data correspond to the fitted model.

> It test compares the observed values to the expected values.

> Goodness-of-fit tests are frequently applied in business decision making.

Testing a Linear Model:

1. co-efficient of determination (R-Square).

2. Hypothesis test for the regression coefficient b_1.

3. Analysis of variance for overall Model

4. Residual Analysis to validate the regression model assumptions.

5. outliers analysis.

Spurious Regression:

1. The coefficient estimate will not converge toward zero.

2. The t value most often is significant.

3. R^2 is typically very high.

4. spurious regression is linked to serially correlated errors.

Linear Least square Method:

a)

x_i	8	3	2	10	11	3	6	5	6	8
y_i	4	12	1	12	9	4	9	6	1	14

Soln:

$$y = ax + b \quad \text{— ①}$$

$$\Sigma y = \Sigma ax + \Sigma b \qquad \boxed{\Sigma_1 = n}$$

$$\Sigma y = a\Sigma x + b\Sigma_1$$

$$\Sigma y = a\Sigma x + nb \quad \text{— ②}$$

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

$$\Sigma xy = a\Sigma x^2 + b\Sigma x \quad \text{— ③}$$

Σx	Σy	Σx^2	Σxy	Σy^2
8	4	64	32	16
3	12	9	36	144
2	1	4	2	1
10	12	100	120	144
11	9	121	99	81
3	4	9	12	16
6	9	36	54	81
5	6	25	30	36
6	1	36	6	1
8	4	64	112	196
62	72	468	503	813
10	10			

$\bar{x} = 6.2$ $\bar{y} = 7.2$

$\Sigma y = a\Sigma x + nb$. — ②

$72 = a\,62 + 10b$ — ④

$503 = 468a + 62b$ — ⑤

$a = 0.67$

$b = 3.04$

$$72 = 62a + 10b \times 62$$

$$503 = 468a + 62b \times 10$$

$$4464 = 3844a + 620b$$

$$5030 = 4680a + 620b$$

$$-566 = -836a$$

$$\frac{283}{418}$$

$$a = \frac{566}{836}$$

$$\boxed{a = 0.67}$$

$$72 = \overset{0.67X}{62a} + 10\overset{b}{(\quad)}$$

$$72 = \underset{41.54}{(\quad)} + 10(b)$$

$$72 - 41.54 = 10b$$

$$30.46 = 10b$$

$$30.46/10 = b \qquad \boxed{b = 3.046}$$

$$y = 0.67(x) + 3.04.$$

Regression:

When a variable is dependent on an independent variable.

Regression of x on y:

→ x is dependent

→ y " independent

Eqn: $X = a + bY$ $\quad \Sigma 1 = n$

Formula:

$$\Sigma x = Na + b\Sigma y \quad \text{—— ①}$$

$$\Sigma xy = a\Sigma y + b\Sigma y^2 \quad \text{—— ②}$$

Regression of y on x:

→ y is dependent

→ x is independent.

Eqn = $y = a + bx$

Formula:

$$\Sigma y = Na + b\Sigma x \longrightarrow (1)$$

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

Q) Find the two lines of regression from the following data:

x	25	22	28	26	35	20	22	40	20	18
y	18	15	20	17	22	14	16	21	15	14

$\Sigma x = 256.$

$\Sigma y = 172.$

$\Sigma x^2 = 7,002$

$\Sigma y^2 = 3,036$

$\Sigma xy = 4,576.$

x on y,

$$\Sigma x = Na + b\Sigma y \quad \text{—— (1)}$$

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

substituting values,

(1) $\Rightarrow$ $256 = 10a + 172b$ —— (3)

(2) $\Rightarrow$ $4576 = 172a + 30366b$ —— (4)

Now, (3) $\times 172$ & $4 \times 10 \Rightarrow$.

$$44032 = 1720a + 29584b. \quad (-)$$

$$45760 = 1720a + 30360b \quad (-)$$

$$-1728 = -776b.$$

$$b = \frac{1728}{776}$$

$$\boxed{b = 2.23}$$

sub b in ③

$$256 = 10a + (172)(2.23)$$

$$10a = -127.56$$

$$a = -12.756.$$

$$\therefore x = 2.23y - 12.756.$$

Regression equation y on x,

$$\Sigma y = Na + b\Sigma x \longrightarrow ⑤$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \longrightarrow ⑥$$

⑤ $\Rightarrow$ $172 = 10a + 256b \longrightarrow ⑦$

$$4576 = 256a + 7002b \longrightarrow ⑧$$

7 x 256 and 8 x 10 $\Rightarrow$

$$44032 = 2560a + 65536b$$

$$45760 = 2560a + 70020b$$

$$-1728 = -4484b$$

$$b = \frac{-1728}{-4484}$$

$$\boxed{b = 0.385}$$

sub b in (7),

$$172 = 10a + (256)(0.385)$$

$$172 = 10a + 98.56$$

$$10a = 73.44$$

$$\boxed{a = 7.344}$$

$$\therefore Y = 7.344 + 0.385X$$