

Regression Line Analysis:

Regression Analysis is the process of constructing a mathematical model or function that can be used to predict or determine one variable by another variable or other variables. The most elementary regression model is called **simple regression** or **bivariate regression** involving two variables in which one variable is predicted by another variable. In simple regression, the variable to be predicted is called the **dependent variable** and is designated as y . The predictor is called the **independent variable**, or explanatory variable, and is designated as x .

In simple regression analysis, only a straight-line relationship between two variables is examined.

Equation of Regression Line:

$$\hat{y} = b_0 + b_1x \text{ (} \hat{\text{cap}} \text{)}$$

(note : Slope intercept equation of line is given by $y = mx + c$)

Where **Slope (b_1)** = $\frac{\sum[(x-\bar{x}) \cdot (y-\bar{y})]}{\sum(x-\bar{x})^2}$

And **Intercept (b_0)** = $\bar{y} - (b_1 \cdot \bar{x})$

Example 1 : Find Least Square Regression Line for the following data:

x	5	12	9	15	7
y	16	6	8	4	7

Solution:

Here $n = 5$

$$\bar{x} = \frac{\sum x}{n} = \frac{5+12+9+15+7}{5} = \frac{48}{5} = 9.6$$

$$\text{And } \bar{y} = \frac{\sum y}{n} = \frac{16+6+8+4+7}{5} = \frac{41}{5} = 8.2$$

x	y	$(x - \bar{x})$	$(x - \bar{x})^2$	$y - \bar{y}$	$(x - \bar{x}) \cdot (y - \bar{y})$
5	16	$5 - 9.6 = -4.6$	21.16	$16 - 8.2 = 7.8$	-35.88
12	6	2.4	5.76	-2.2	-5.28
9	8	-0.6	0.36	-0.2	0.12
15	4	5.4	29.16	-4.2	-22.68
7	7	-2.6	6.76	-1.2	3.12
$\sum x = 48$	$\sum y = 41$		$\sum (x - \bar{x})^2 = 63.200$		$\sum [(x - \bar{x}) \cdot (y - \bar{y})] = -60.60$

Now,

$$\text{Slope } (b_1) = \frac{\sum[(x-\bar{x}) \cdot (y-\bar{y})]}{\sum(x-\bar{x})^2} = \frac{-60.60}{63.200} = -0.959$$

$$\text{And Intercept } (b_0) = \bar{y} - (b_1 \cdot \bar{x}) = 8.200 - ((-0.959) \cdot 9.60) = 8.2 - (-9.206)$$

$$\text{Therefore } b_0 = 17.406$$

Therefore equation of regression line can be given by,

$$\hat{y} = b_0 + b_1x$$

$$\hat{y} = 17.406 - 0.959x$$

Example 2: Find Least Square Regression Line for the following data:

x	140	119	103	91	65	29	24
y	25	29	46	70	88	112	128

Here n = 7

$$\bar{x} = 81.57$$

$$\bar{y} = 71.14$$

x	y	(x - \bar{x})	(x - \bar{x}) ²	y - \bar{y}	(x - \bar{x}) · (y - \bar{y})
140	25	58.43	3414.064	-46.14	-2696.07
119	29	37.43	1401.004	-42.14	-1577.375
103	46	21.43	459.245	-25.14	-538.793
91	70	9.43	88.925	-1.14	-10.76
65	88	-16.57	274.56	16.86	-279.33
29	112	-52.57	2763.60	40.86	-2147.90
24	128	-57.57	3314.30	56.86	-3273.31
$\sum x = 571$	$\sum y = 498$		$\sum (x - \bar{x})^2 = 11715.71$		$\sum [(x - \bar{x}) \cdot (y - \bar{y})] = -10523.571$

• RESIDUAL ANALYSIS:

1) Residual = $y - \hat{y}$

2) Sum of Square Due to Error (SSE) = $\sum (y - \hat{y})^2$

3) Standard Error of Estimation (Se) = $\sqrt{\frac{SSE}{n-2}}$

4) Total Sum of Square (SST) = $\sum (y - \bar{y})^2$

Also, SST = SSE + SSR

5) Sum of Square of Regression (SSR) = $\sum (\hat{y} - \bar{y})^2$

6) Coefficient of Determination (r^2) = $\frac{SSR}{SST}$

Here \bar{y} = Mean of y and \hat{y} = Predicated value of y from regression line

Example: Consider the following data:

x	5	12	9	15	7
y	16	6	8	4	7

Regression Line of this equation is $\hat{y} = 17.406 - 0.959x$. Then find the following parameters

(i) Residual of each observation (ii) Standard Error of Estimation(Se) (iii) Sum of Square due to Error (SSE) (iv) Coefficient of Determination (v) Total Sum of Squares (vi) Sum of Square of Regression

Solution:

x	y	$\hat{y} = 17.406 - 0.959x$	Residual $y - \hat{y}$	SSE $(y - \hat{y})^2$	$y - \bar{y}$	SST $(y - \bar{y})^2$
5	16	$17.406 - ((0.959)(5))$ =12.610	$16 - 12.610$ =3.390	11.492	$16 - 8.2 = 7.8$	60.840
12	6	5.898	0.102	0.010	-2.2	4.840
9	8	8.775	-0.775	0.601	-0.2	0.040
15	4	3.021	0.979	0.958	-4.2	17.640
7	7	10.693	-3.693	13.638	-1.2	1.440
$\sum x$ = 48	$\sum y$ = 41			$\sum (y - \hat{y})^2$ = 26.699		$\sum (y - \bar{y})^2$ = 84.800

x	y	$\hat{y} = 17.406 - 0.959x$	$\hat{y} - \bar{y}$	SSR $(\hat{y} - \bar{y})^2$
5	16	$17.406 - ((0.959)(5))$ =12.610	$12.610 - 8.2 = 4.410$	19.448
12	6	5.898	-2.302	5.299
9	8	8.775	0.575	0.331
15	4	3.021	-5.179	26.822
7	7	10.693	2.493	6.215

$\sum x$ = 48	$\sum y$ = 41			$\sum (\hat{y} - \bar{y})^2$ = 58.116
------------------	------------------	--	--	--

From the table:

$$1) \text{ Standard Error of Estimation (Se)} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{26.699}{5-2}} = \sqrt{\frac{26.699}{3}} = 2.983$$

$$2) SSE = \sum (y - \hat{y})^2 = 26.699$$

$$3) SST = \sum (y - \bar{y})^2 = 84.800$$

$$4) SSR = \sum (\hat{y} - \bar{y})^2 = 58.116$$

$$5) \text{ Coefficient of Determination } (r^2) = \frac{SSR}{SST} = \frac{58.116}{84.800} = 0.685$$

Example 2 : Consider the following data:

x	12	21	28	8	20
y	17	15	22	19	24

Regression line of this data is given by $\hat{y} = 16.516 + 0.162x$. Then determine the following parameters:

(1) $y(18)$ (2) Residual of each observation (3) Standard Error of Estimation(Se) (4) Sum of Square due to Error (SSE) (5) Coefficient of Determination (6) Total Sum of Squares (7) Sum of Square of Regression

Solution :

$$(1) y(18) = 16.516 + 0.162(18) = 16.516 + 2.916 = 19.432$$

Example 3: Consider the following data:

x	61	63	67	69	70	74	76	81	86	91	95	97
y	4.28	4.08	4.42	4.17	4.48	4.30	4.82	4.70	5.11	5.13	5.64	5.56

Regression Line of this data is given by $\hat{y} = 1.57 + 0.0407x$. Evaluate Standard Error of estimation of this data:

Solution:

Here $n = 12$

x	y	$\hat{y} = 1.57 + 0.0407x$	Residual $y - \hat{y}$	SSE $(y - \hat{y})^2$
61	4.28	$1.57 + 0.0407(61) = 4.053$	$4.28 - 4.053 = 0.227$	0.052
63	4.08	4.134	-0.054	0.003
67	4.42	4.297	0.131	0.017
69	4.17	4.378	-0.208	0.043
70	4.48	4.419	0.061	0.004
74	4.30	4.582	-0.282	0.080
76	4.82	4.663	0.157	0.025
81	4.70	4.867	-0.167	0.028
86	5.11	5.070	0.040	0.002
91	5.13	5.274	-0.144	0.021
95	5.64	5.437	0.203	0.041
97	5.56	5.518	0.042	0.002
				$\sum (y - \hat{y})^2 = 0.318$

$$SSE = \sum (y - \hat{y})^2 = 0.318$$

$$\text{Standard Error of Estimation(Se)} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{0.318}{12-2}} = \sqrt{\frac{0.318}{10}} = 0.178$$

Example1 : Find equation of Regression Line and Coefficient of determination of the data:

x	10	14	28	18	22
y	17	19	20	23	27

Example 2 : Find equation of Regression Line and Se of the data:

x	11	15	27	19	20
y	13	21	29	23	25

Example 3: Find Karl Pearson Coefficient and equation of Regression Line of the data:

x	5	8	11	15	18
y	7	12	14	19	21

Example 4: For given data find values of y for given values of x.

x	12	21	28	8	20
y(x)	17	15	22	19	24

Find $y(17)$, $y(25.6)$, $y(2.4)$.

1) Which of the following is not a possible value of Karl – Pearson correlation coefficient?

a) 0.928 b) 0 c) 1 d) 1.0001

2) Which of the following is a possible value of Karl – Pearson correlation coefficient?

a) 0.9278 b) 1.00001 c) -1.0001 d) 1.5

3) Which of the following is correct formula of slope in regression line?

a) $\frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sum(x-\bar{x})^2}$ b) $\frac{\sum[(x+\bar{x})(y+\bar{y})]}{\sum(x+\bar{x})^2}$ c) $\frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sum(x-\bar{x})^3}$ d) $\frac{\sum[(x-\bar{x})-(y-\bar{y})]}{\sum(x-\bar{x})^2}$

4) Which of the following is correct formula of Karl – Pearson coefficient?

a) $\frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sqrt{\sum(x-\bar{x})}}$ b) $\frac{\sum[(x-\bar{x})(y-\bar{y})]}{\sum\sqrt{(x-\bar{x})^2-(y-\bar{y})^2}}$ c) $\frac{\sum[(x+\bar{x})(y+\bar{y})]}{\sum\sqrt{(x-\bar{x})^2\cdot(y-\bar{y})^2}}$ d) None of these

5) What is the value of $y(5)$ in the equation $\hat{y} = 17.203 + 0.123x$

a) 0.123 b) 17.203 c) 17.818 d) 0.615

6) Which of the following is correct formula of S_e ?

a) $\sqrt{\frac{SSR}{n-2}}$ b) $\sqrt{\frac{SST}{n-2}}$ c) $\sqrt{\frac{SSE}{n-2}}$ d) $\sqrt{\frac{SSR}{n-1}}$

7) Which of the following is correct formula of SSR?

a) $\sum(\hat{y} + \bar{y})^2$ b) $\sum(\hat{y} - \bar{y})^2$ c) $\sum(\hat{y} / \bar{y})^2$ d) $\sum(\hat{y} \cdot \bar{y})^2$

8) $SST = \underline{\hspace{2cm}} + SSR$

a) σ_e b) r^2 c) SSE d) None of these

9) Which of the following is correct formula of coefficient of determination?

a) $\frac{S_e}{SSR}$ b) $\frac{SST}{SSR}$ c) $\frac{SSR}{SSE}$ d) $\frac{SSR}{SST}$

10) Which of the following is correct formula of total sum of square?

a) $\sum(y - \bar{y})^2$ b) $\sum(y + \bar{y})^2$ c) $\sum(y / \bar{y})^2$ d) $\sum(y \cdot \bar{y})^2$