Simple Linear Regression (Least Square Error) (Analytical Approach)

Approach - 1

Algorithm

step 1 :- Understand the data by visualization or finding correlation.

step2 :- Derive the linear model by finding its coefficient θ_1 - slope and θ_2 - y intercept. Equation of line (model) will be :

$$y = \theta_1 x + \theta_2$$

To generate the model follow steps given below.

(a.) find the centroid (The average model) where slope is 0.

Centroid = $(\bar{x}, \bar{y})$

$\bar{x}$ = mean of independent variable

$\bar{y}$ = mean of dependent variable

(b.) find the slope θ_1 by following statistical equation from the data.

$$\theta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where, (x_i, y_i), i^{th} values in the data set.

(c.) Calculate the value of θ_2 from the equation by putting θ_1 and $(\bar{x}, \bar{y})$

$$\bar{y} = \theta_1 \bar{x} + \theta_2 \Rightarrow \theta_2 = \bar{y} - \theta_1 \bar{x}$$

It is assumed that the model will pass from the centroid $(\bar{x}, \bar{Y})$.
Final eqn will be (Final model) :-

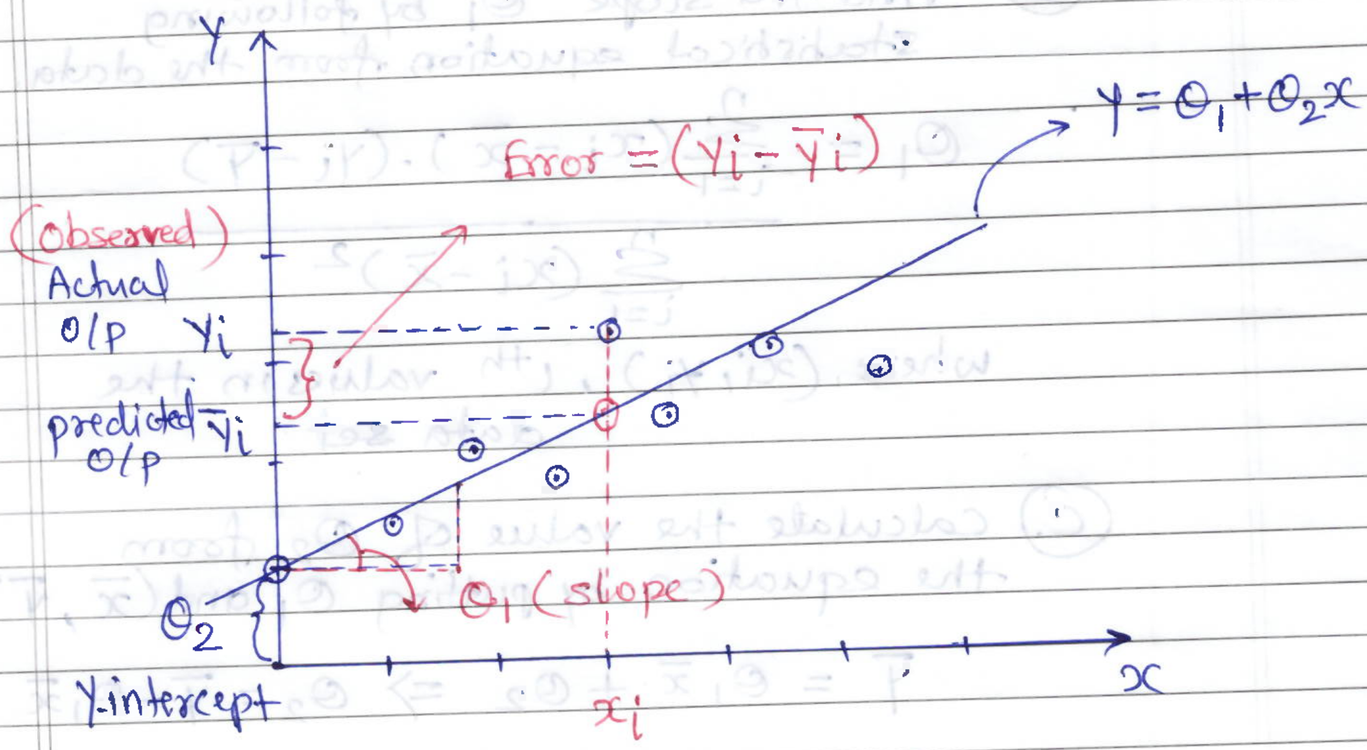
$$\boxed{Y = \theta_1 x + \theta_2}$$

<u>Step 3</u> :- Calculate the error between actual data points and predicted values (lease square method) from the devised model.

$$\text{Error} = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

where, $y_i \rightarrow$ Actual value (observed value)

$\bar{y}_i \rightarrow i^{th}$ predicted value (using devised ~~value~~ model)



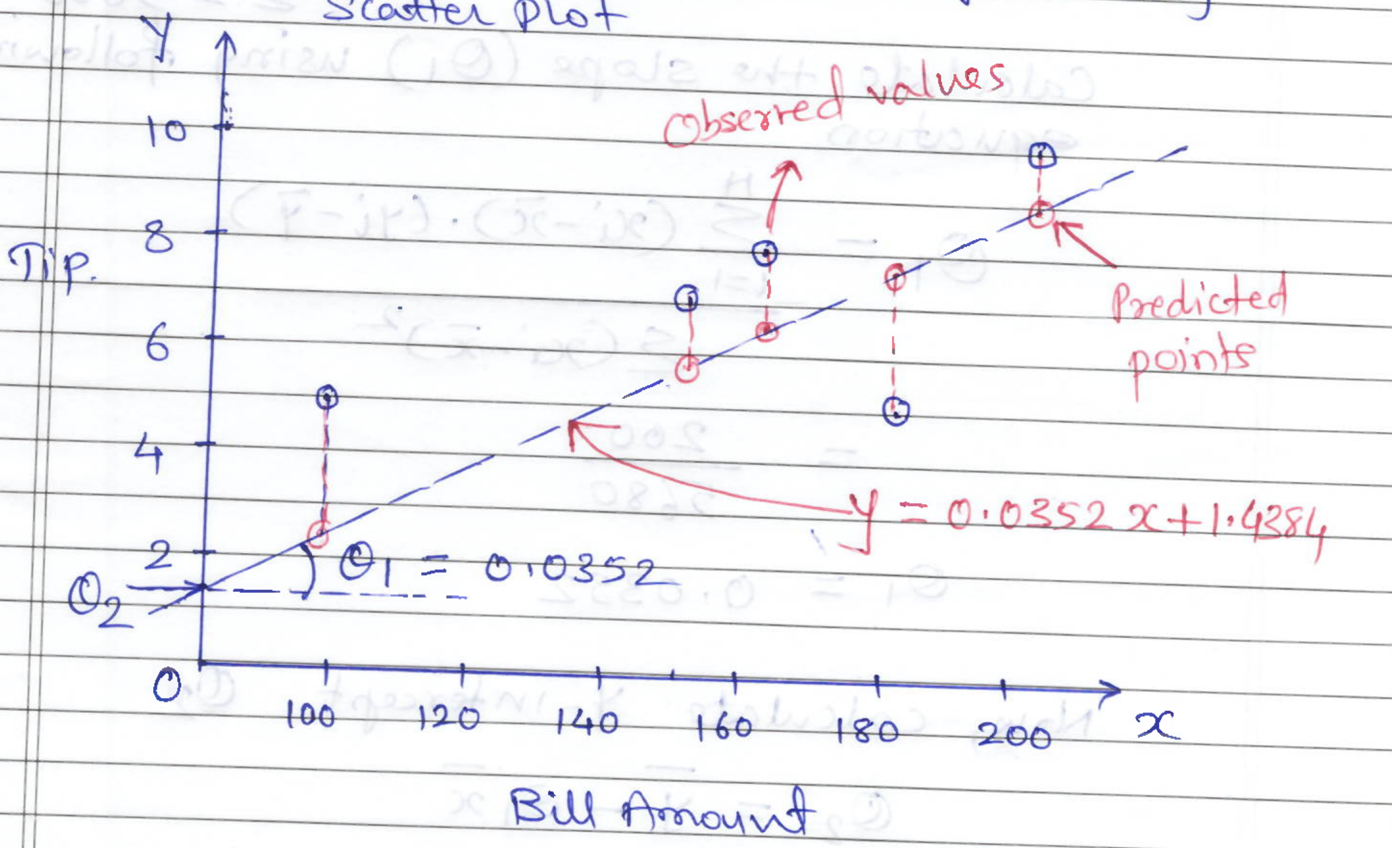
Problem-1 – Consider following data of a restaurant for a tip received by a waiter based on bill amount in Rs.

Sr.No.	Bill amount (x)	Tip received (y)
1	100	5
2	150	7
3	200	10
4	180	5
5	160	8
6.	120	? (Need to predict)

Devise a linear regression model to predict the tip for bill amount Rs. 120. (Use statistical approach)

Solution :–

Step 1 plot the data in 2-D space using Scatter Plot



Step.2 Find the Linear Regression model by finding Θ_1 & Θ_2 (slope & y intercept)

→ calculate $\bar{x}$, $\bar{y}$, $(x_i-\bar{x})$, $(y_i-\bar{y})$.

(i)	x	y	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$
1	100	5	-58	-2	116
2	150	7	-8	0	0
3	200	10	42	3	126
4	180	5	22	-2	-44
5	160	8	2	1	2
					$\Sigma = 200$

$N=5$ $\quad \overline{\Sigma x_i/N} \quad \overline{\Sigma y_i/N}$

$\bar{x} = \frac{790}{5}$ $\quad \bar{y} = \frac{35}{5}$

$\bar{x} = 158$ $\quad \bar{y} = 7$

i	$(x_i-\bar{x})^2$
1	3364
2	64
3	1764
4	484
5	4
	$\Sigma = 5680$

Calculate the slope (Θ_1) using following equation

$$\Theta_1 = \frac{\sum_{i=1}^{N} (x_i-\bar{x}) \cdot (y_i-\bar{y})}{\Sigma (x_i-\bar{x})^2}$$

$$= \frac{200}{5680}$$

$$\Theta_1 = 0.0352$$

Now, calculate y-intercept Θ_2

$$\Theta_2 = \bar{y} - \Theta_1 \bar{x}$$

$$\Theta_2 = 7 - 0.0352 \cdot (158)$$

$\theta_2 = 1.4384$

Final equation of the line (model) is

~~$\bar{y} = 0.0352\,\theta_1 + 1.$~~

$$\boxed{\bar{y} = 0.0352\,\bar{x} + 1.4384} \longrightarrow A$$

Now predict the value of y for new value of x using equation A.

Test set $\Rightarrow x = 120$.

$y = 0.0352\,(120) + 1.4384$

$\boxed{y = 5.6624} \longrightarrow$ predicted value

Step 3 Calculate the error with respect to all the values (actual) of x for the devised model. (equation A)

i	Actual x	Actual y	Predicted $\bar{y} = 0.0352x + 1.4384$	$(y_i - \bar{y})$
1	100	5	4.95	0.05
2	150	7	6.71	0.29
3	200	10	8.47	1.53
4	180	5	7.77	−2.77
5	160	8	7.07	0.93
				$\Sigma =$

Least Squared Error $= \sum_{i=1}^{N} (y_i - \bar{y})^2$

$= (0.05)^2 + (0.29)^2 + (1.53)^2 + (-2.77)^2 + (0.93)^2$

$= 10.96$

Least Squared Error with the devised model is

$$\therefore \boxed{\text{Error} = 10.96}$$

You can also calculate the RMSE

$$\text{Root mean squared error} = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}}$$

$$\text{RMSE of the model} = \sqrt{\frac{10.96}{5}}$$

$$\boxed{\text{RMSE} = 1.48}$$

~~Approach - 2~~

(Additional Problem Statements)

① Consider the following data to predict the open price of a share based on its previous day closing price using simple linear Regression.

i	closing Price (x)	opening price (Y)
1	8	8.5
2	10	9
3	9.5	7
4	6.5	7
5	7.3	7.8
6	8	8.2

What will be the opening price if previous day closing price is 8.8 ?