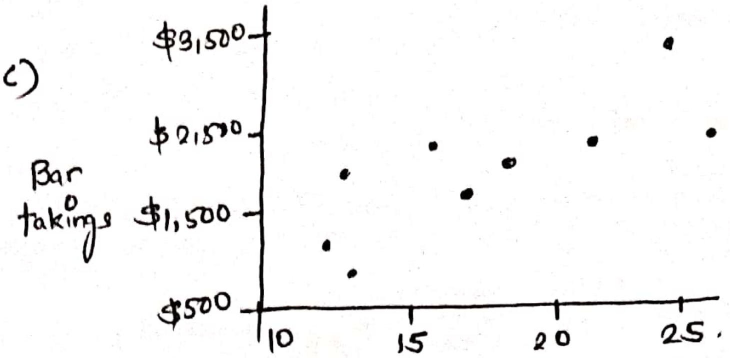
Sum of Square (SS)

Day	Takings	Temp(°C)
3 Jun	$3,213	23
10 Jun	$2;089	21
17 Jun	$2,253	25
24 Jun	$1,801	18
1 Jun	$801	13
8 Jun	$1.934	16
15 Jun	$1,720	13
22 Jun	$1.514	17
29 Jun	$1,017	12

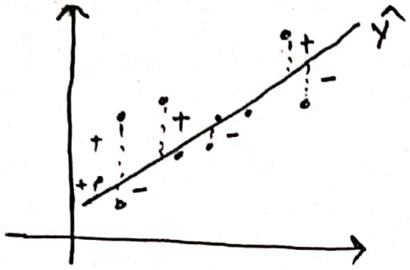


Q) Bar taking given the temperature for the particular day is the dataset.

→ $\hat{y}$ is the predicted value for given value of x.

So, $\hat{y}$ will have a equation, $\hat{y} = -353.11 + 123.54x$

↑ constant term/ y intercept ↑ Gradient/ Coefficient of x

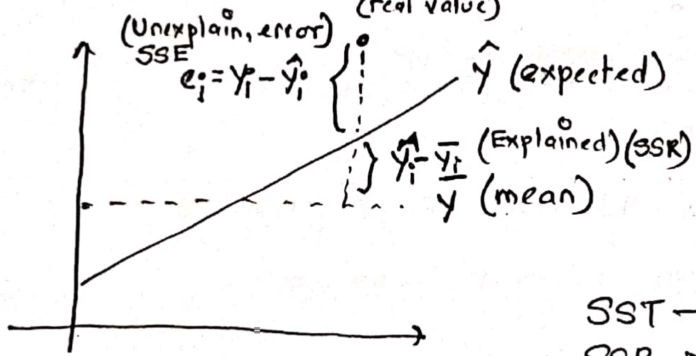


There are 2 kind of error terms - i) +ve error ii) -ve error

But if we sum both +ve and -ve, we will get 0.

So we square, to ignore negative values.

There $\hat{y}$ is the line which decreases the sum of square error (SSE).



Explained component, $\hat{y}_i - \bar{y}_i$

Unexplained component, $y_i - \hat{y}_i$

Total variation = Explained variation + unexplained variation

$$SST = SSR + SSE$$

SST → Sum of square total

SSR → sum of square residual $= \hat{y}_i - \bar{y}$

SSE → Sum of square error $= y_i - \hat{y}_i$

Total deviation = SSR + SSE.

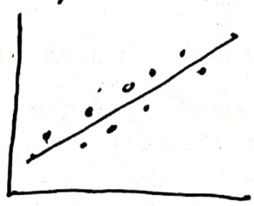
$$= \hat{y}_i - \bar{y} + y_i - \hat{y}_i$$

$$SST = \Sigma(y_i - \bar{y})^2$$

$$R^2 = SSR/SST$$

R^2 is proportion of total variation which is explained.

High SSE
Low R^2
(Scattered point)
Huge error term



Low SSE
High R^2
(Unscattered point)
Low error term

- Regression line is a line which best fit to the observations.