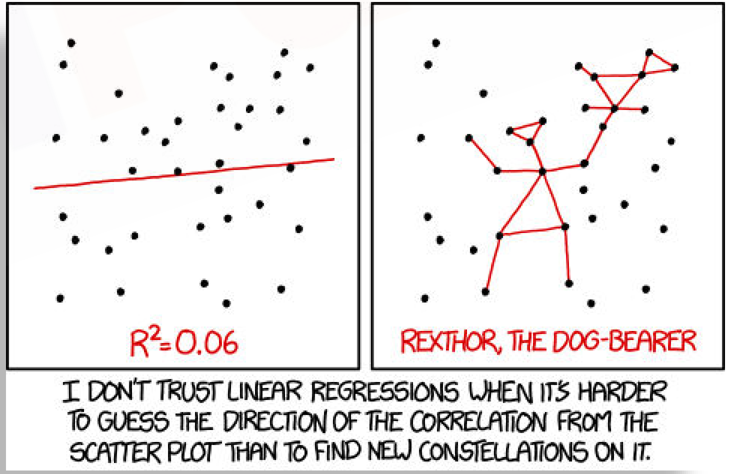
Linear Regression using Python

What is Regression?

- A technique of finding the relationship between two or more variables
- Change in dependent variable is associated with a change in one or more independent variables.



What is Regression?

Regression is a technique that displays the relationship between variable "y" based on the values of variable "x".

For example,.

As the temperature drops people put on more jackets to keep warm

Regression Use Case

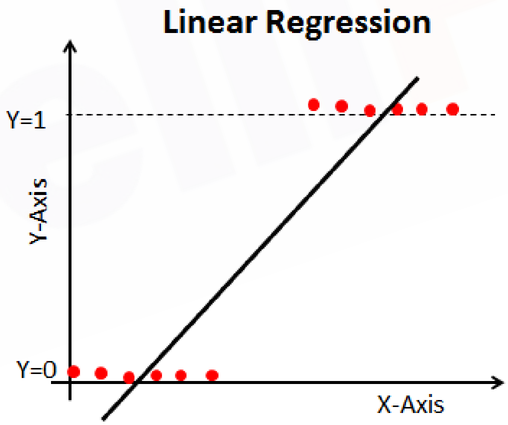
- Temperature vs. Number of cones sold at ice cream store
- Inches of rain vs. new cars sold
- Daily Snowfall vs. number of skier visits

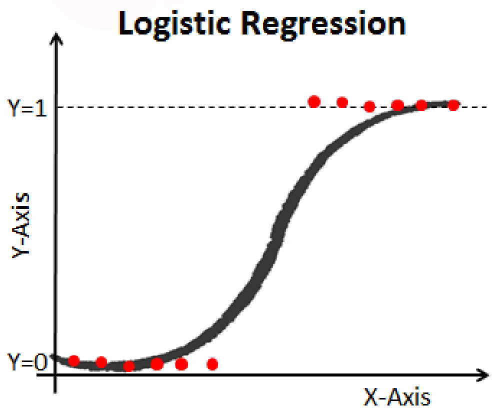
If you think there is a relationship between two things regression would help confirm it!

Types of Regression

There are various types of Regression, but we will focus on:

- Linear Regression
- Logistic Regression



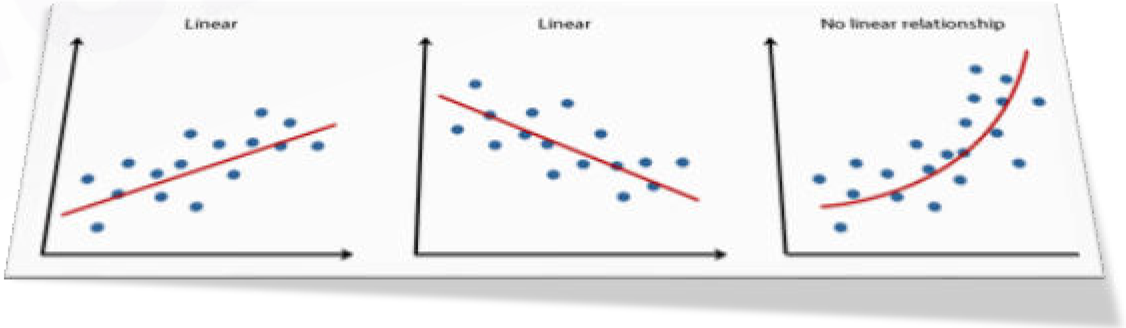


Types of Regression

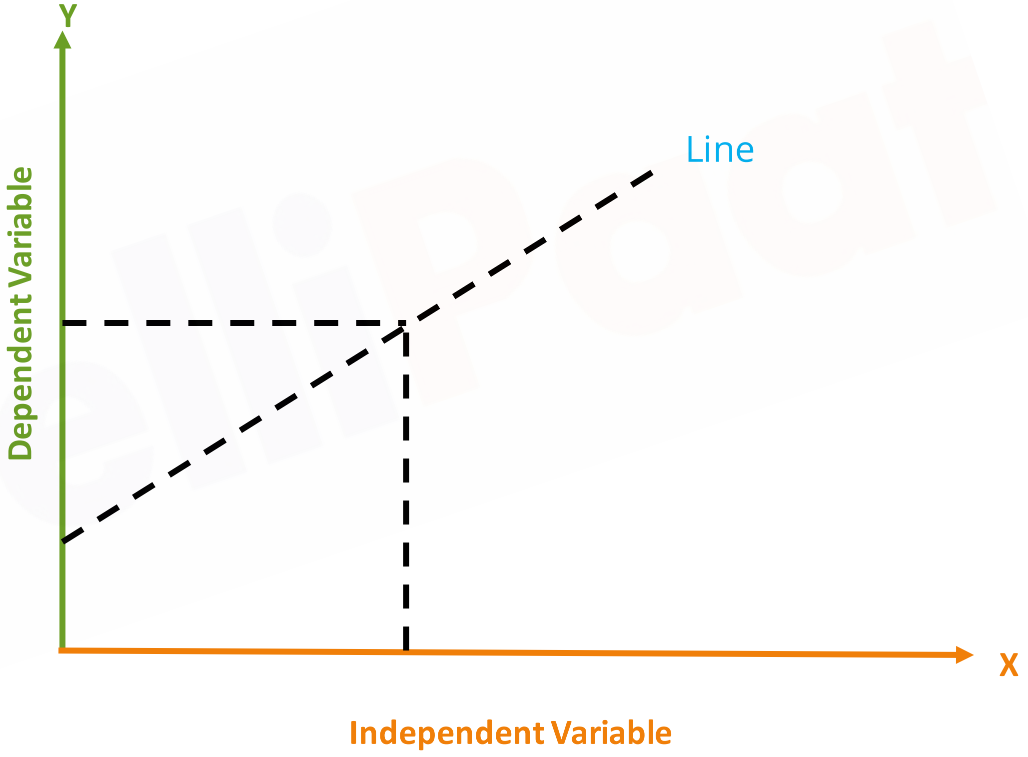
LINEAR REGRESSION	LOGISTIC REGRESSION
Continuous Variables	Categorical Variables
Solves Regression Issue	Solves Classification Issue
Straight Line	S-Curve

What is Linear Regression?

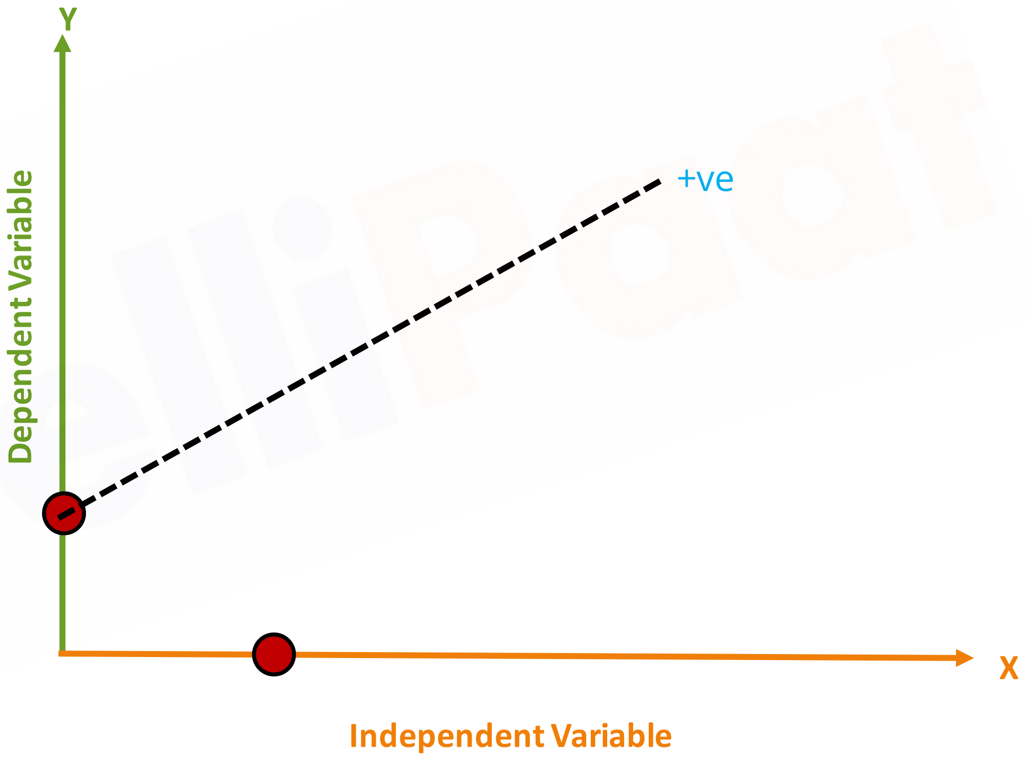
- Simple linear regression is useful for finding relationship between two continuous variables
- One is predictor or independent variable and other is response or dependent variable



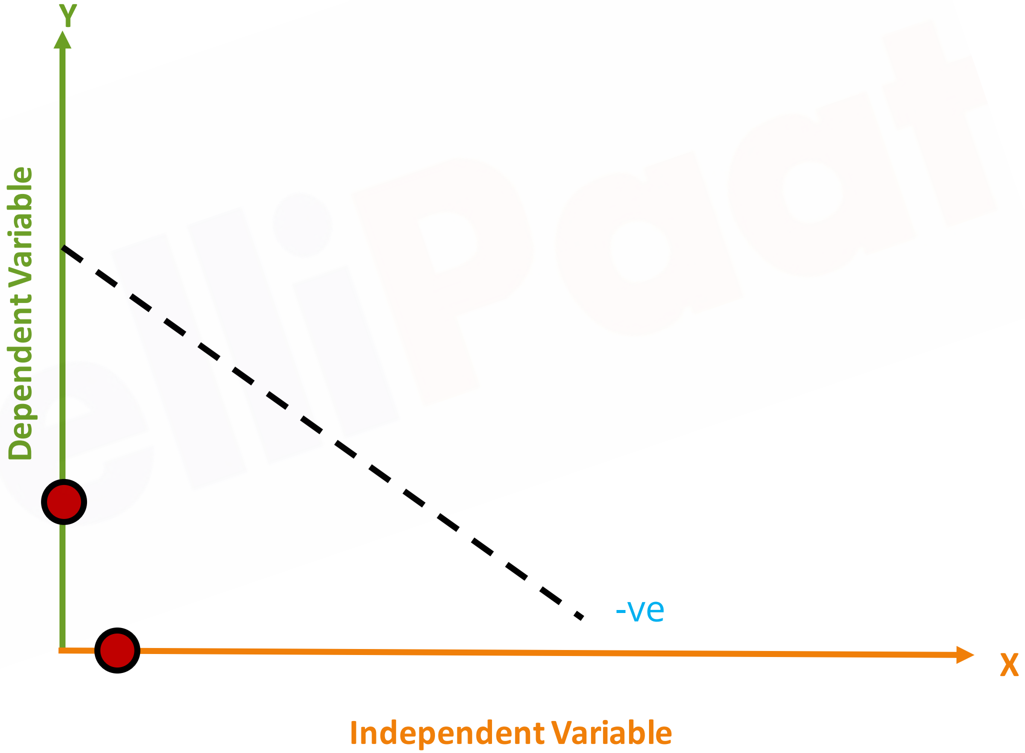
Understanding Linear Regression



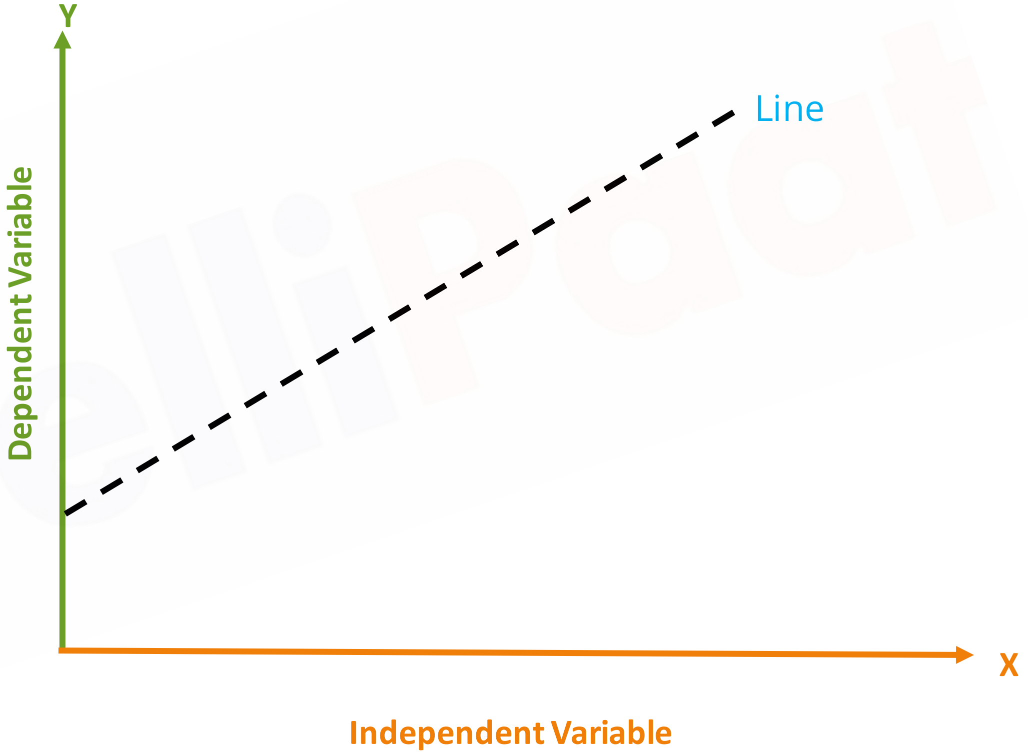
Understanding Linear Regression



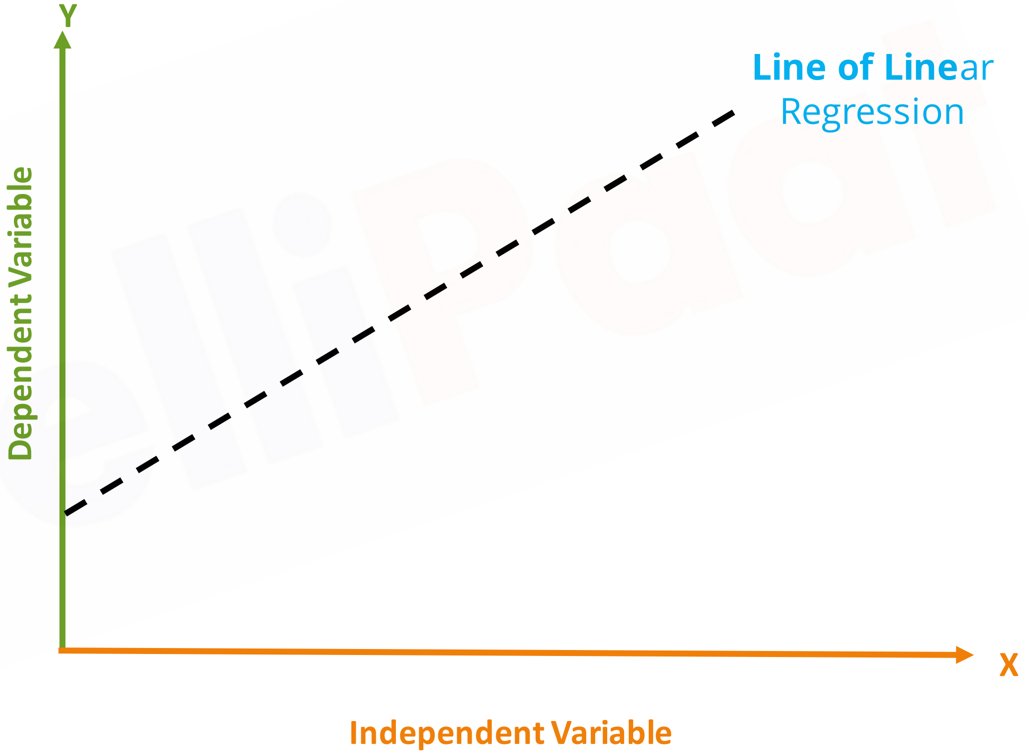
Understanding Linear Regression



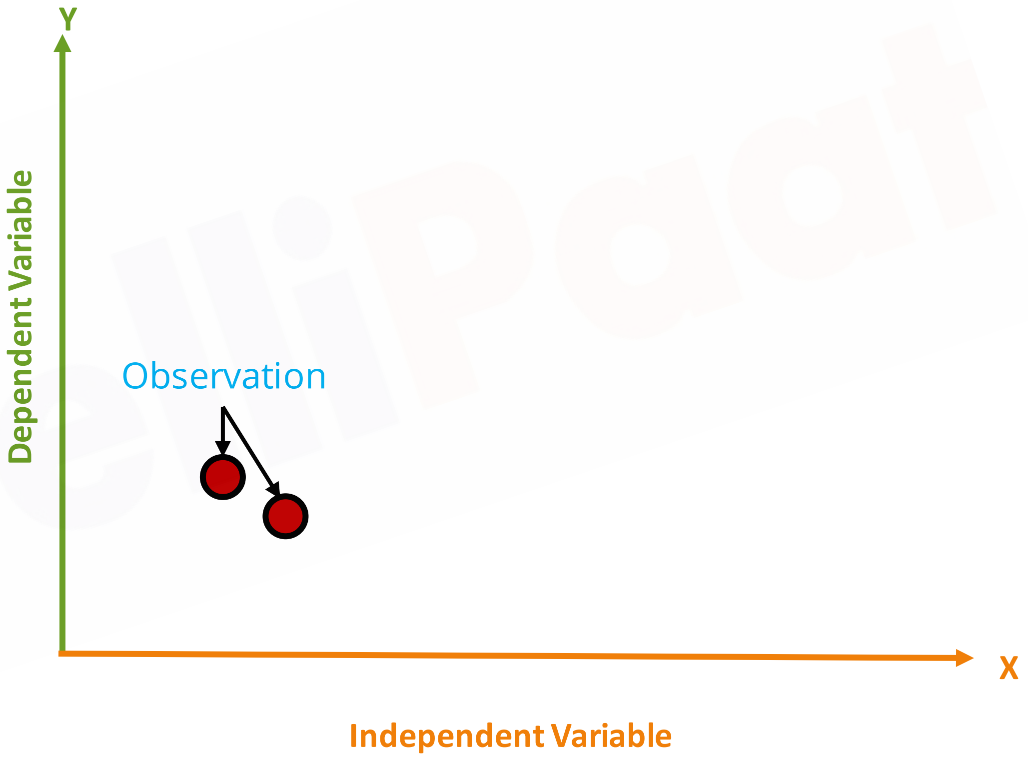
Understanding Linear Regression



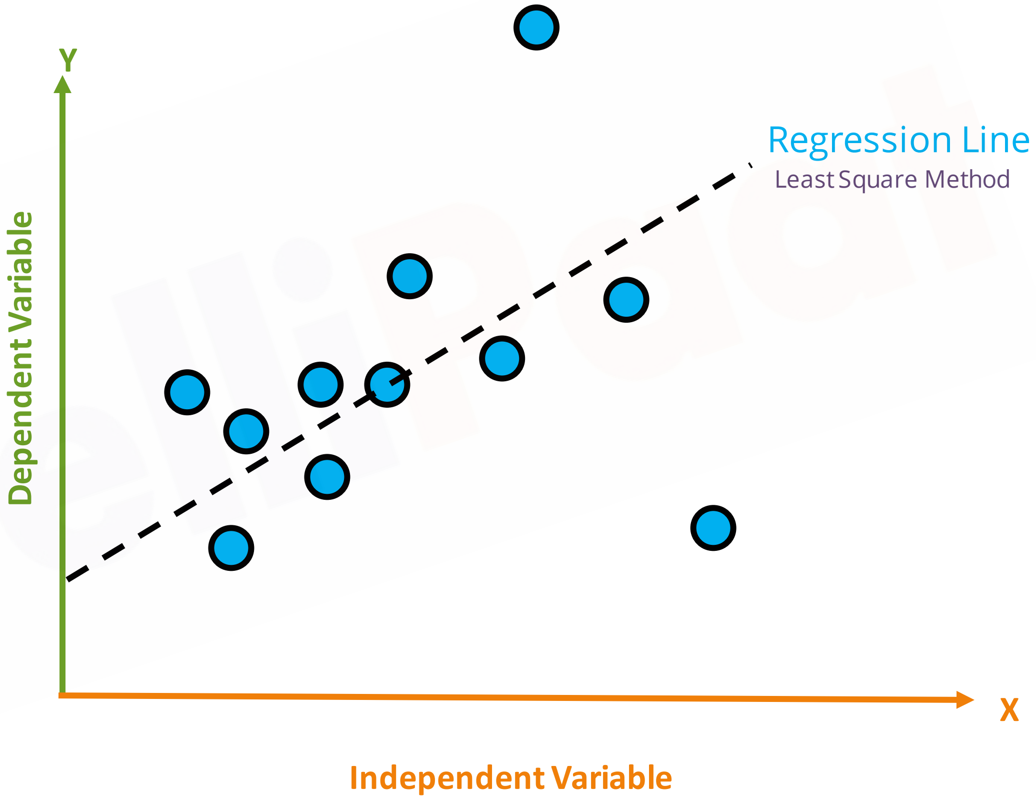
Understanding Linear Regression



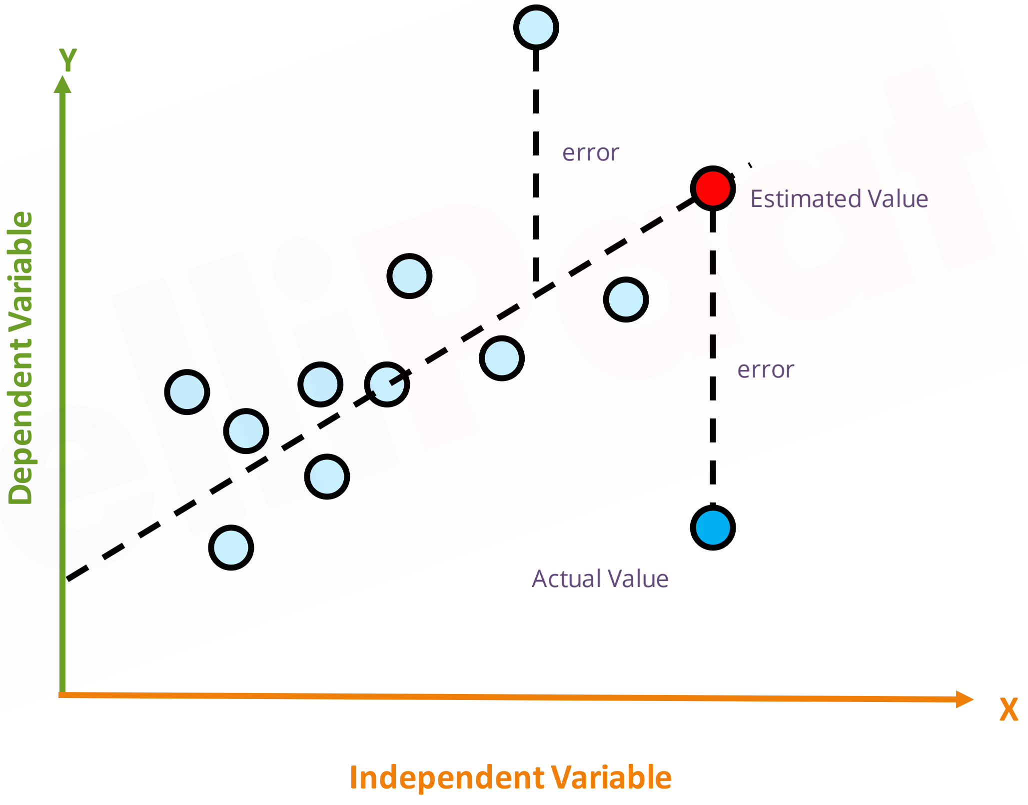
Understanding Linear Regression



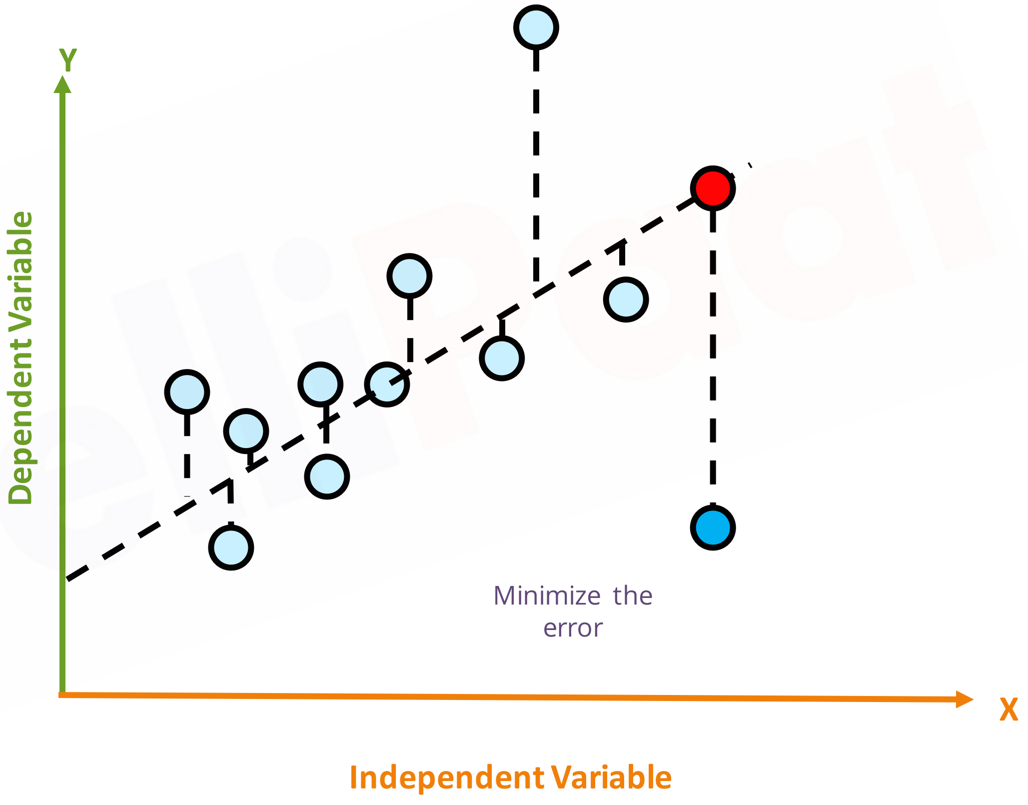
Understanding Linear Regression



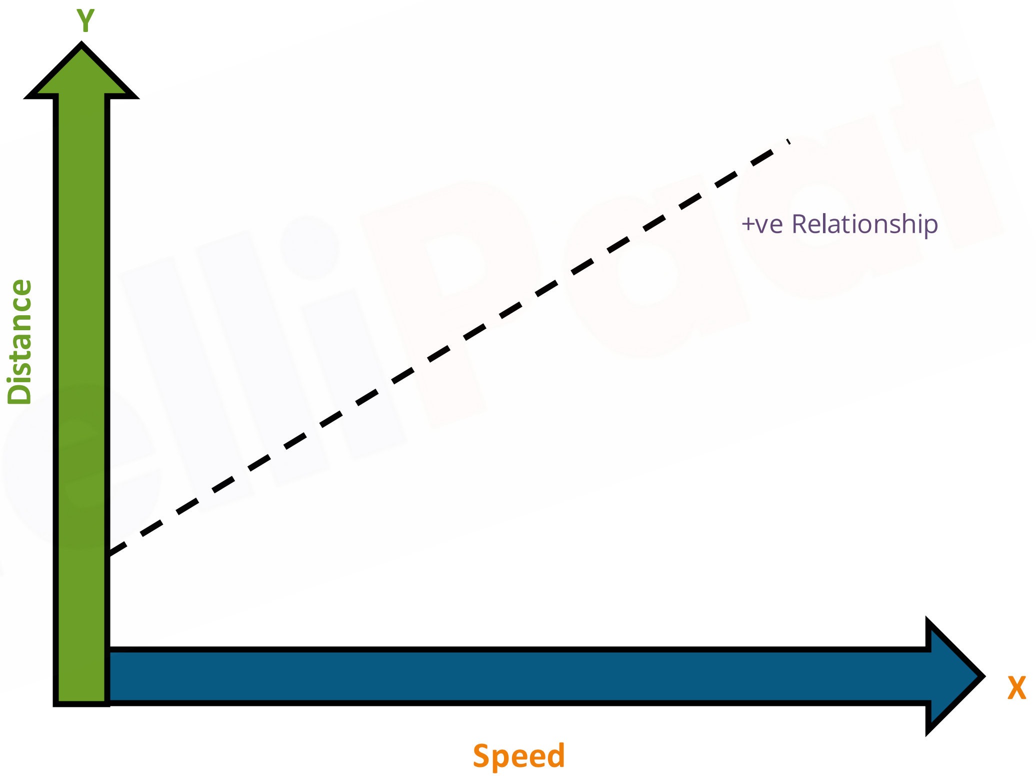
Understanding Linear Regression



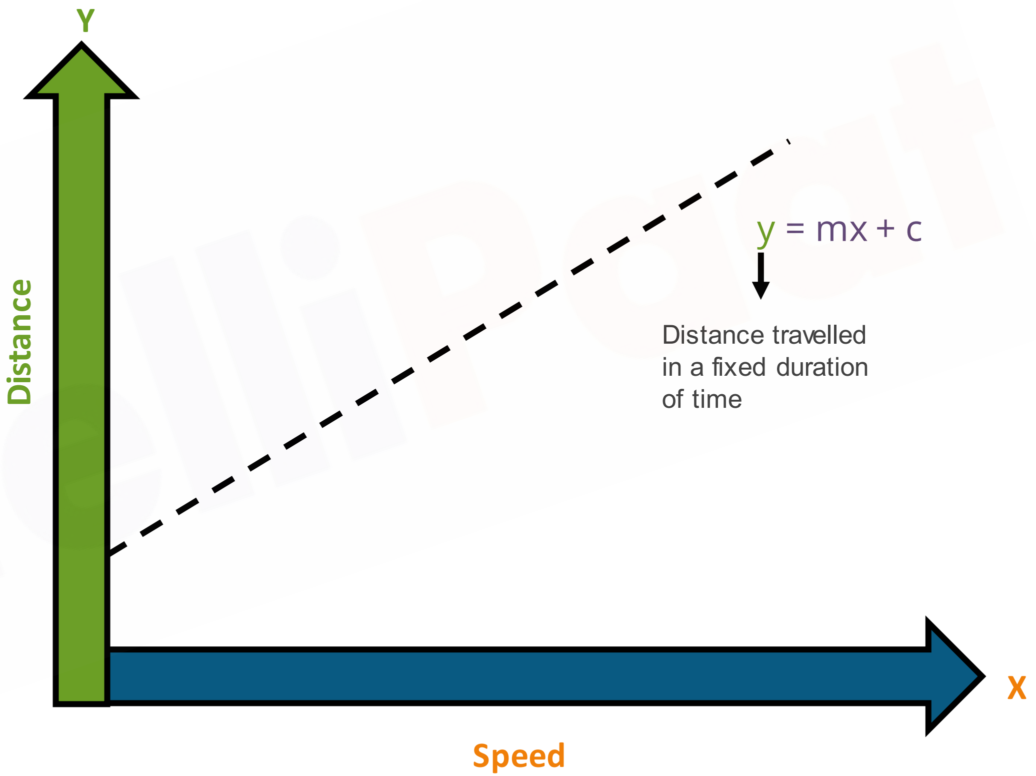
Understanding Linear Regression



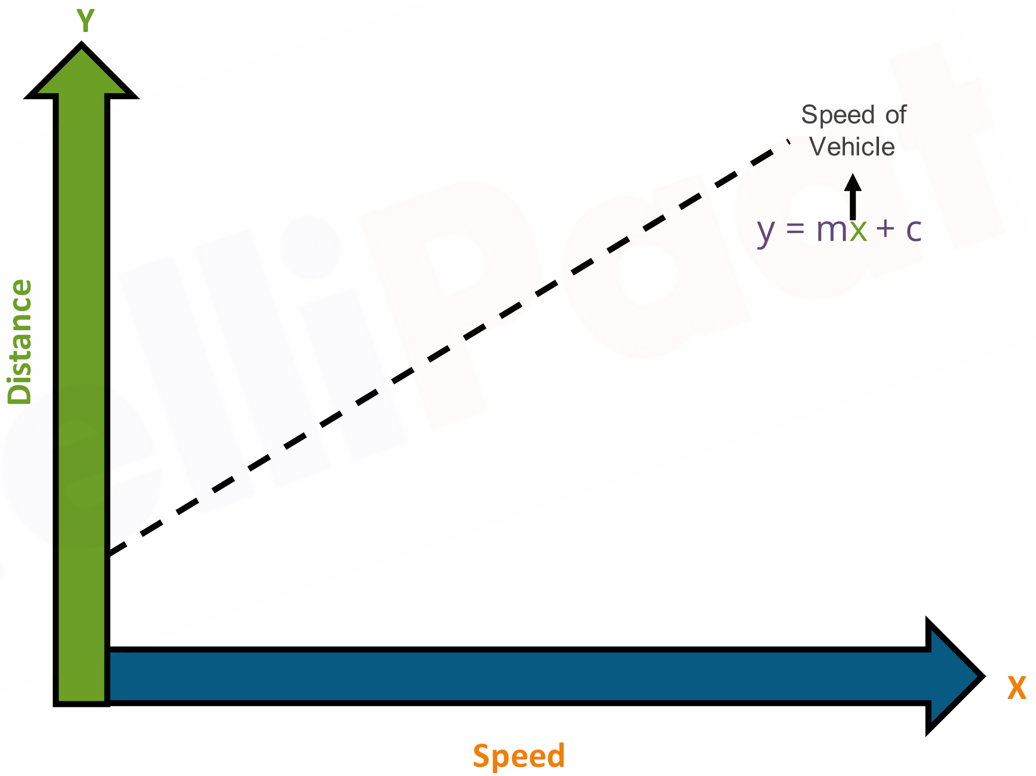
Understanding Linear Regression



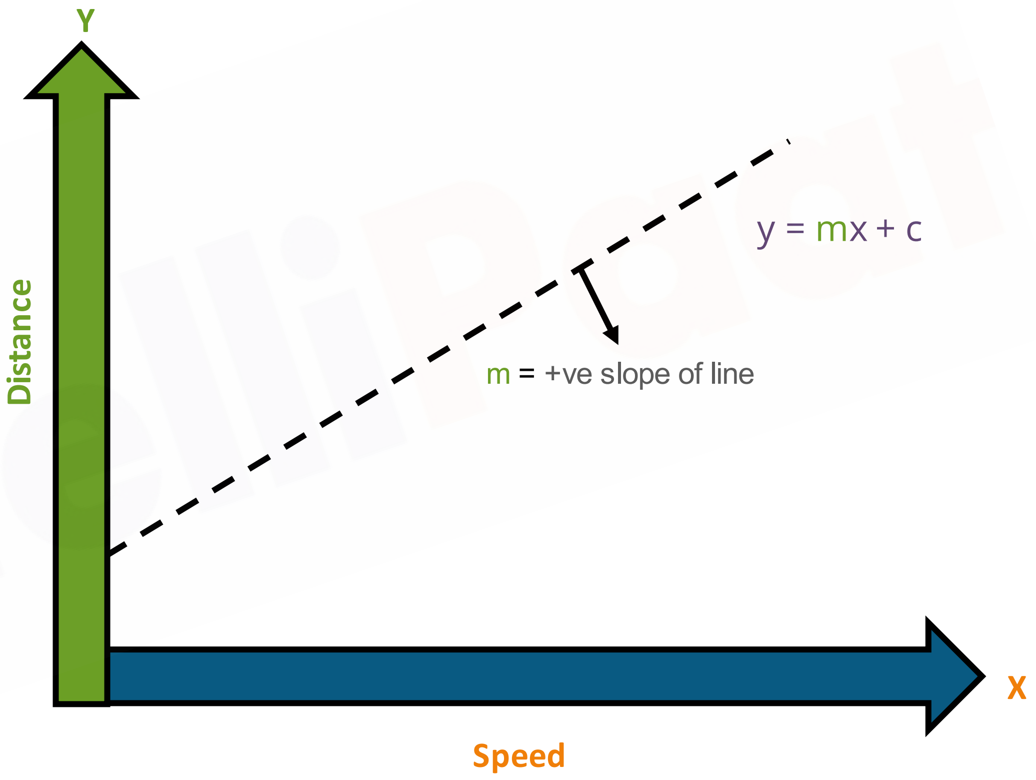
Understanding Linear Regression



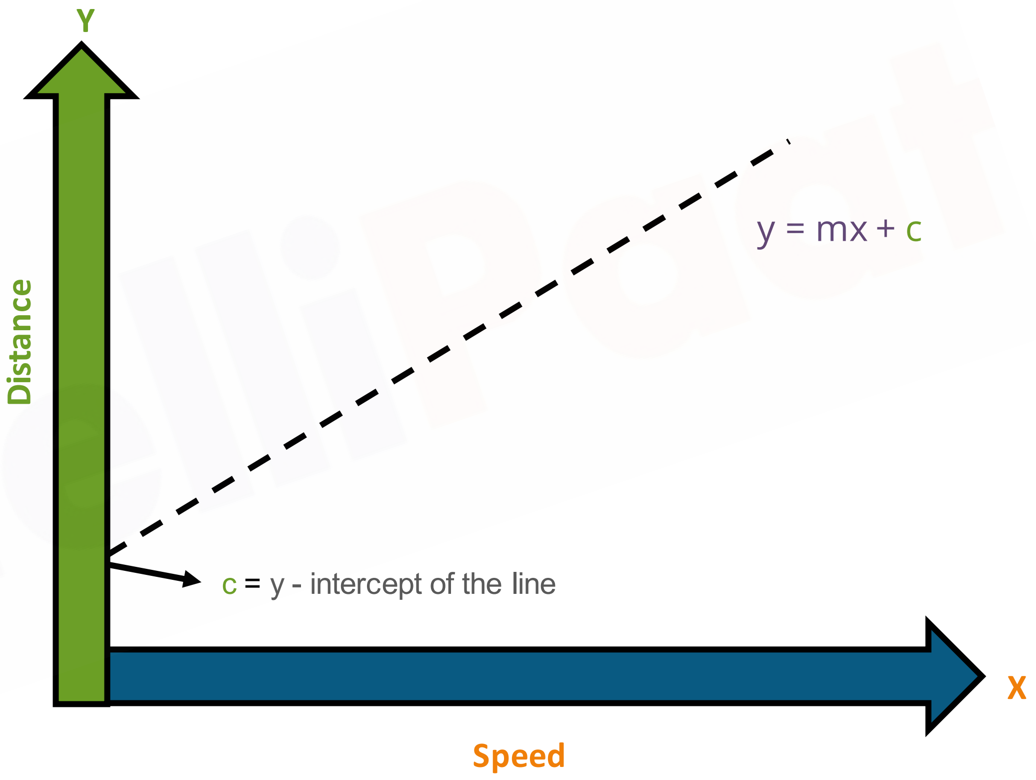
Understanding Linear Regression



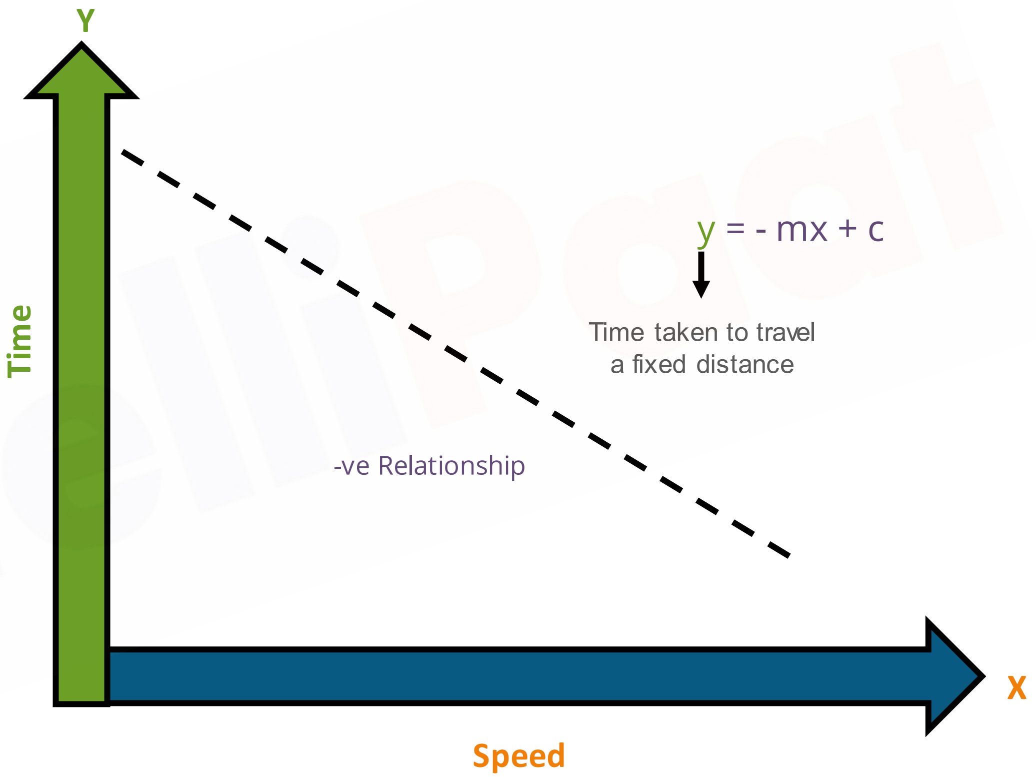
Understanding Linear Regression



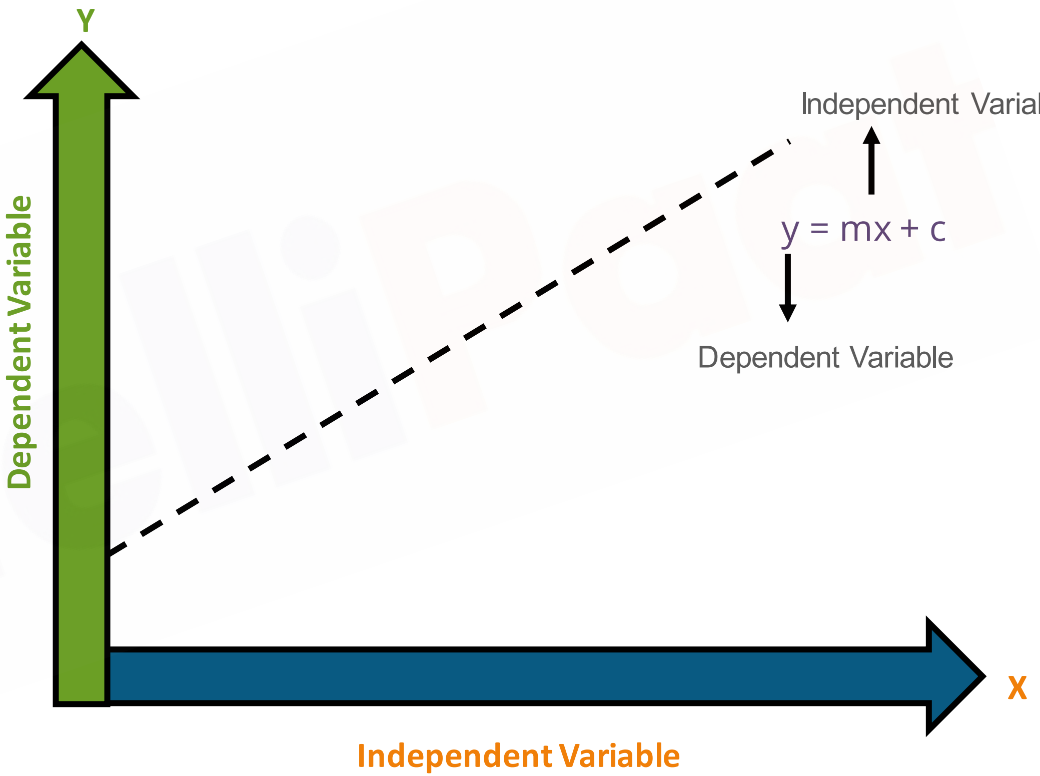
Understanding Linear Regression



Understanding Linear Regression

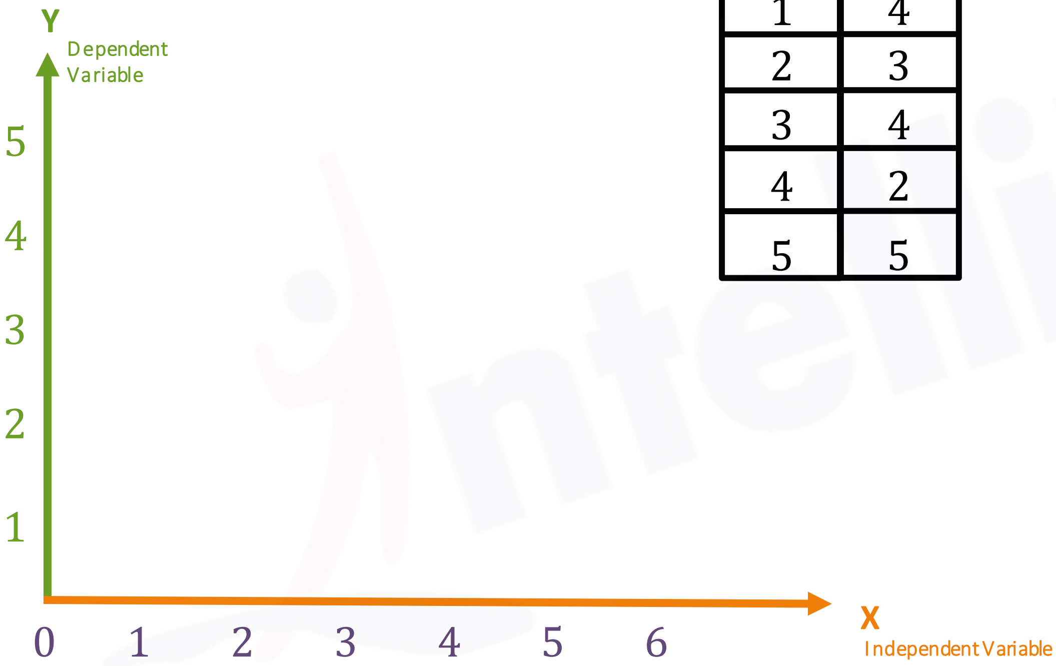


Understanding Linear Regression



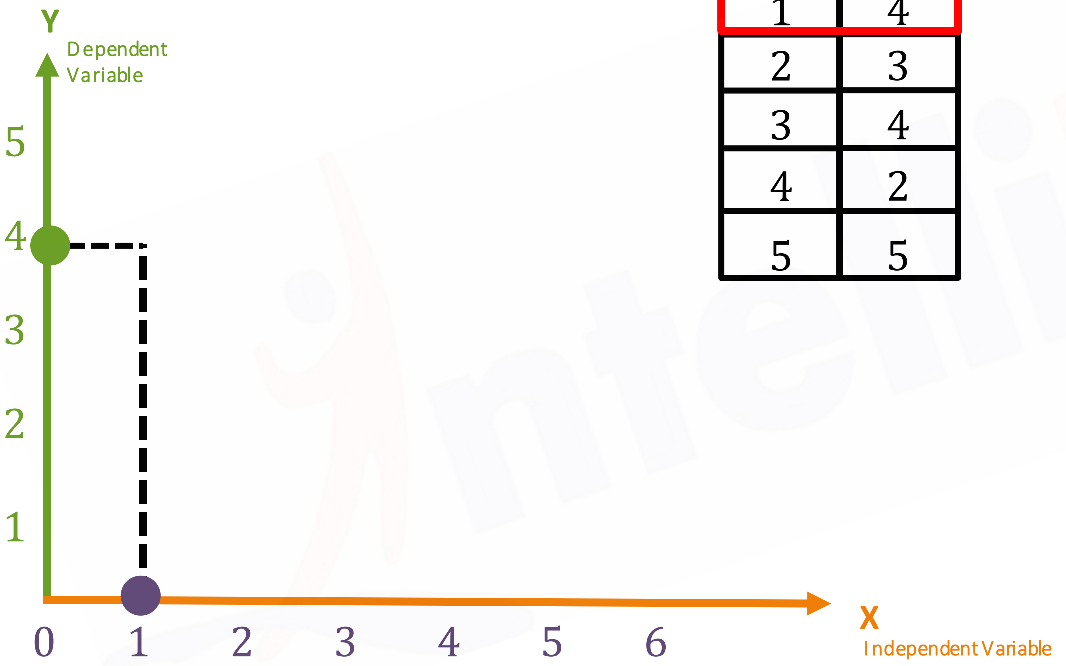
Understanding Linear Regression

x	y
1	4
2	3
3	4
4	2
5	5



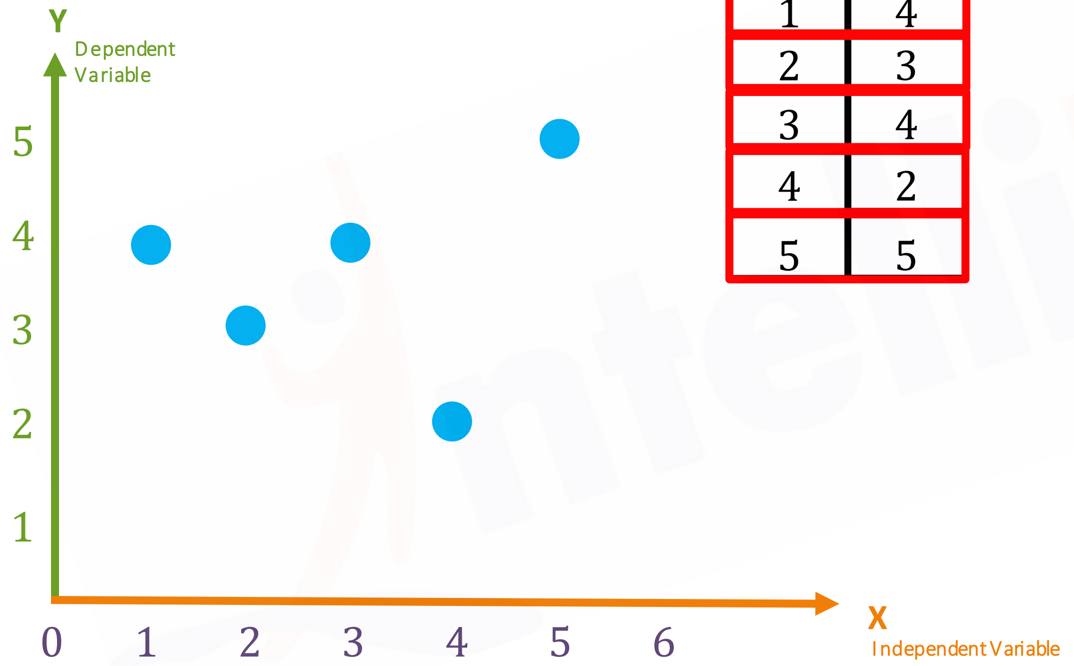
Understanding Linear Regression

x	y
1	4
2	3
3	4
4	2
5	5

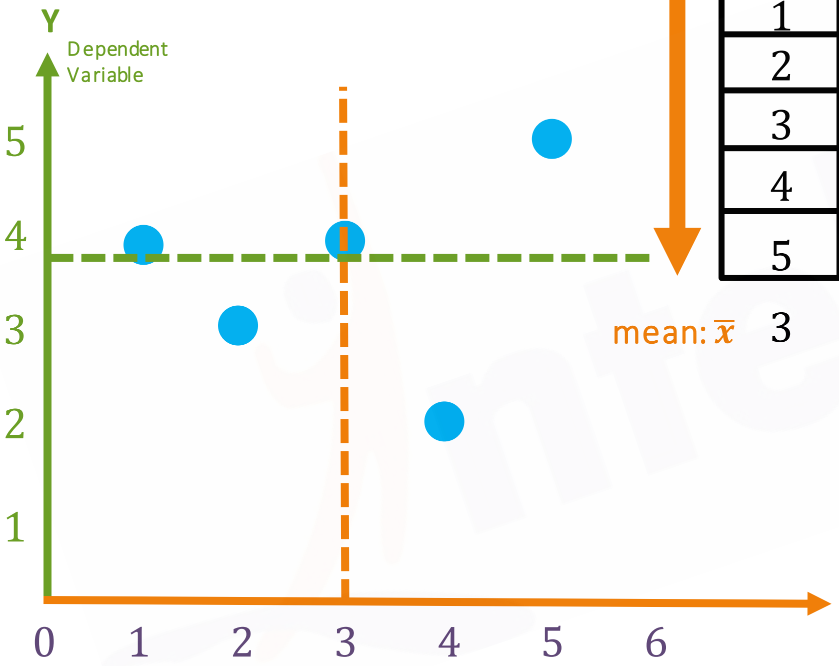


Understanding Linear Regression

x	y
1	4
2	3
3	4
4	2
5	5



Understanding Linear Regression



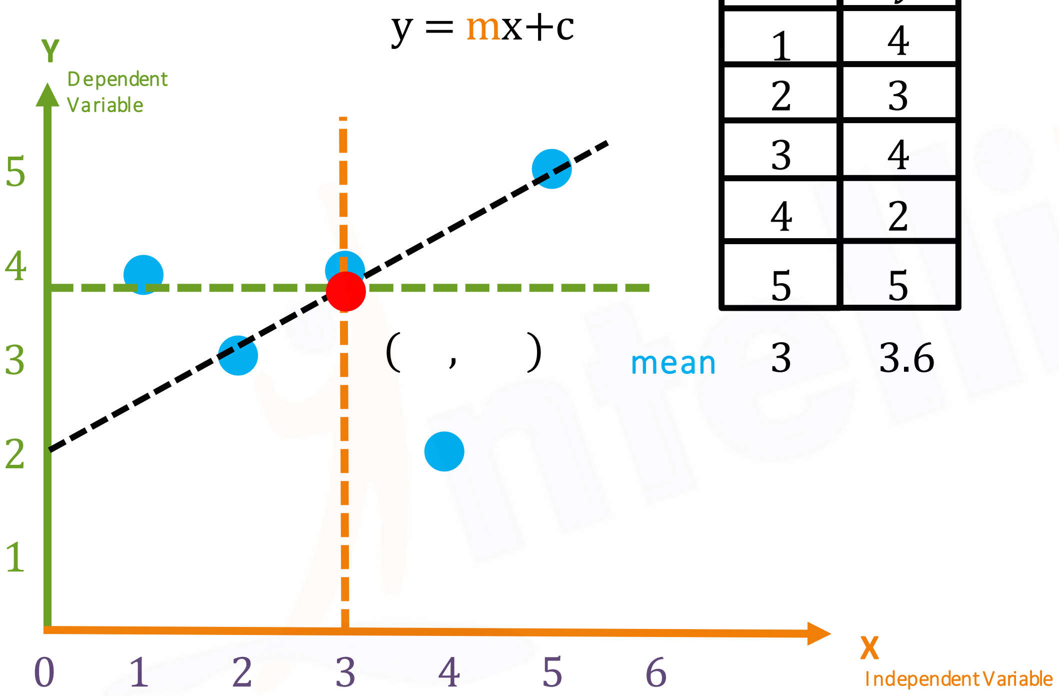
X
Independent Variable

x	y
1	4
2	3
3	4
4	2
5	5

mean: $\bar{x}$ 3

3.6 18/5 mean : $\bar{y}$

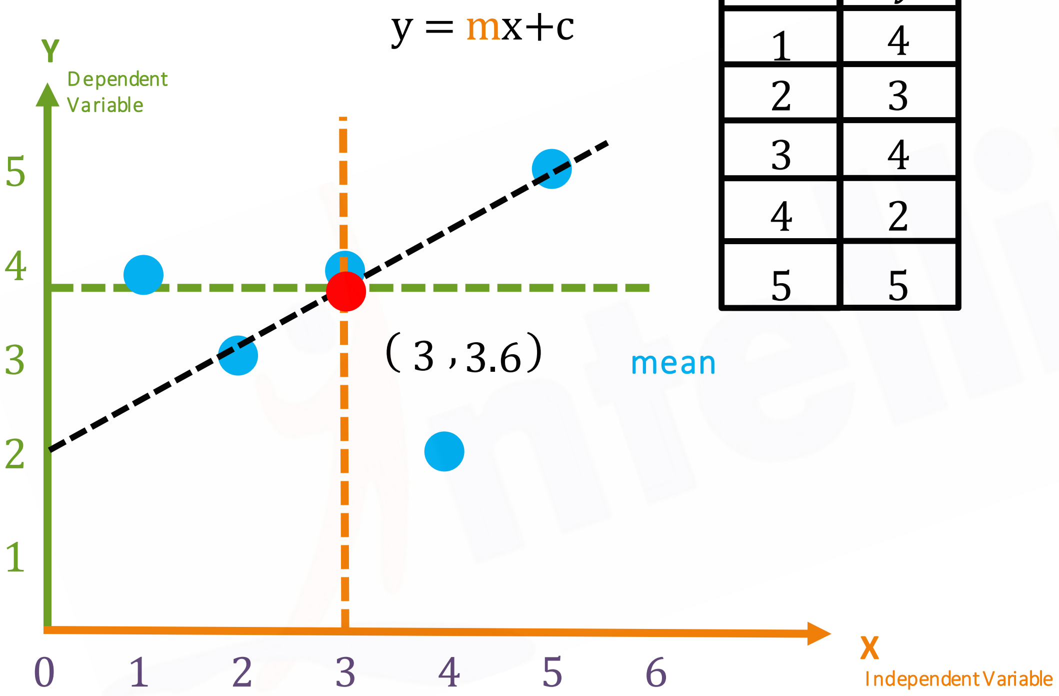
Understanding Linear Regression



x	y
1	4
2	3
3	4
4	2
5	5

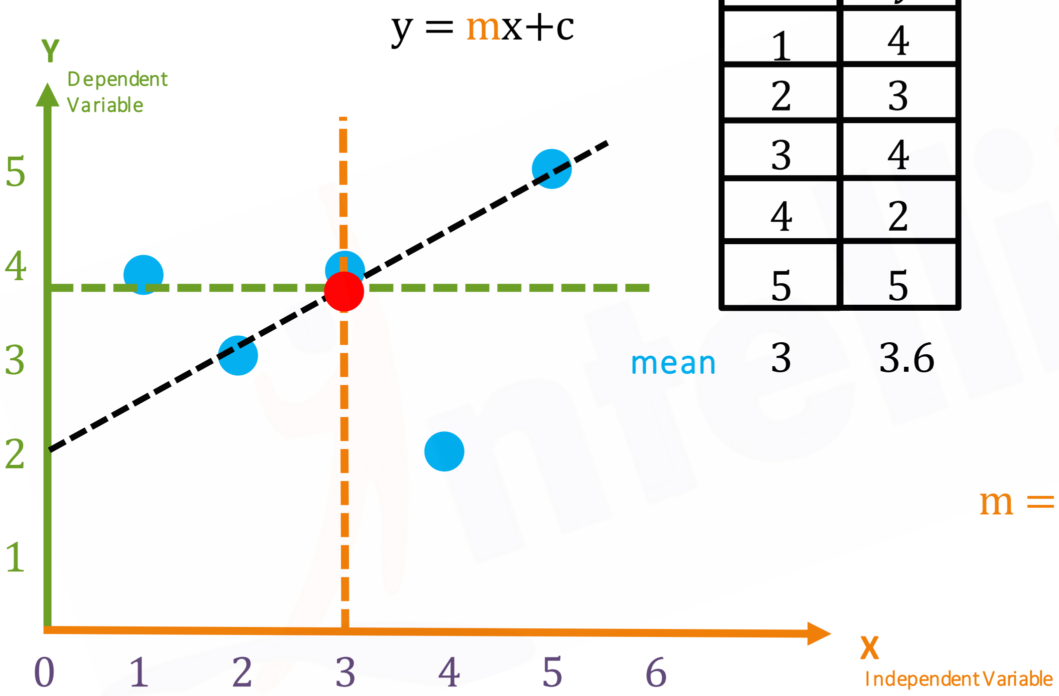
mean 3 3.6

Understanding Linear Regression



x	y
1	4
2	3
3	4
4	2
5	5

Understanding Linear Regression

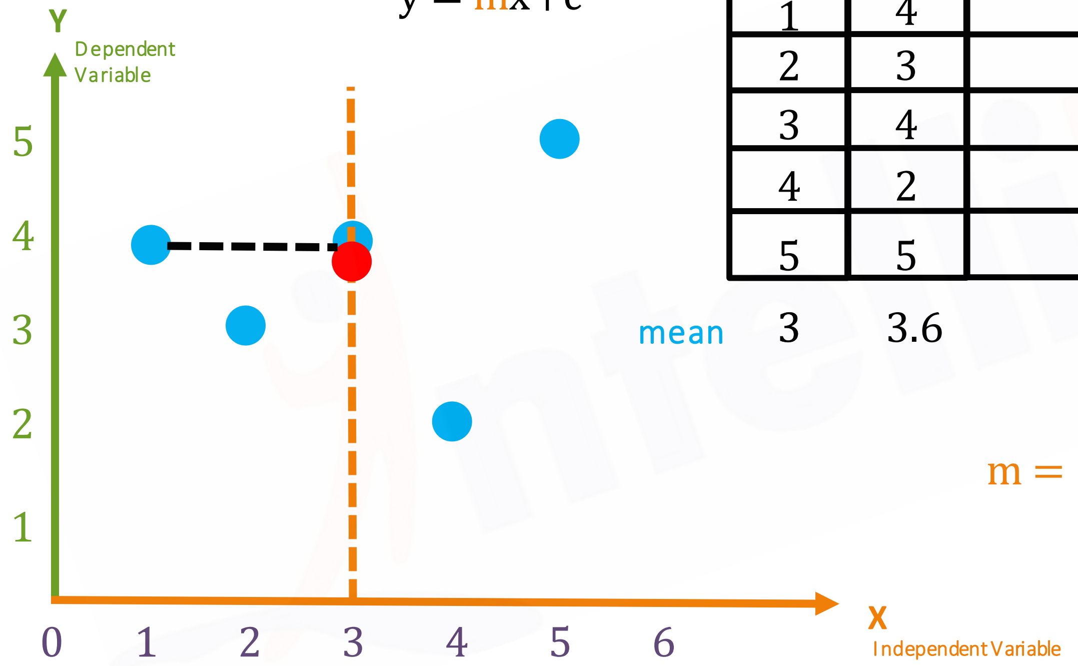


x	y
1	4
2	3
3	4
4	2
5	5
mean 3	3.6

$$m = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma \ (x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

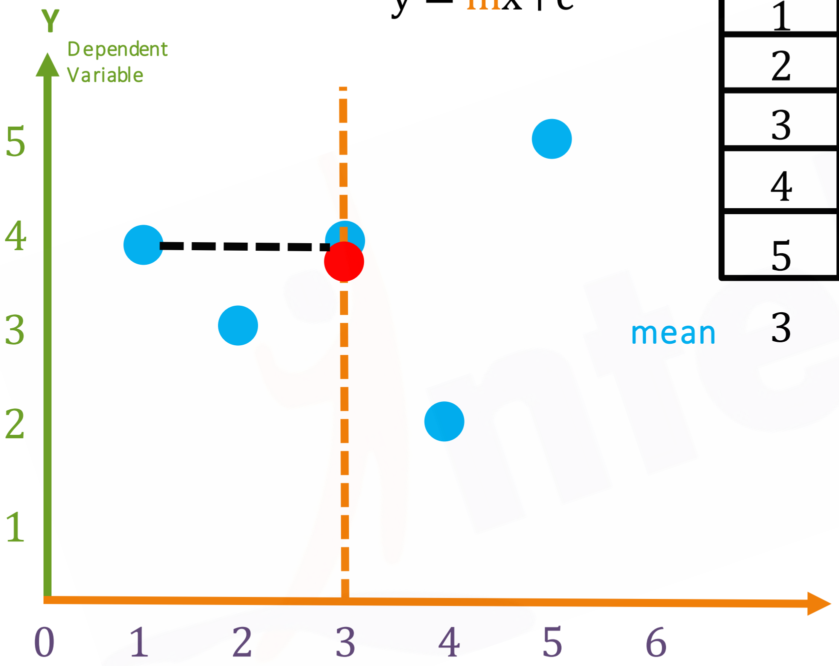


x	y	$x-\bar{x}$
1	4	
2	3	
3	4	
4	2	
5	5	
mean 3	3.6	

$$m = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

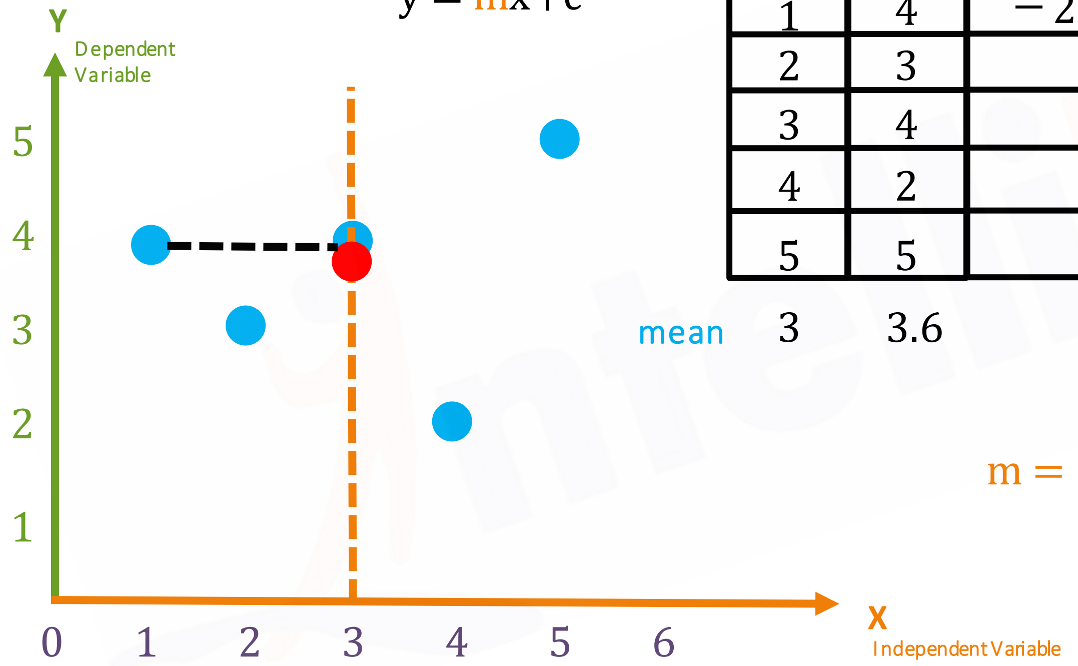


	x	y	$x - \bar{x}$
	1	4	$1 - 3$
	2	3	
	3	4	
	4	2	
	5	5	
mean	3	3.6	

$$m = \frac{\Sigma\,(x - \bar{x})(y - \bar{y})}{\Sigma\,(x - \bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

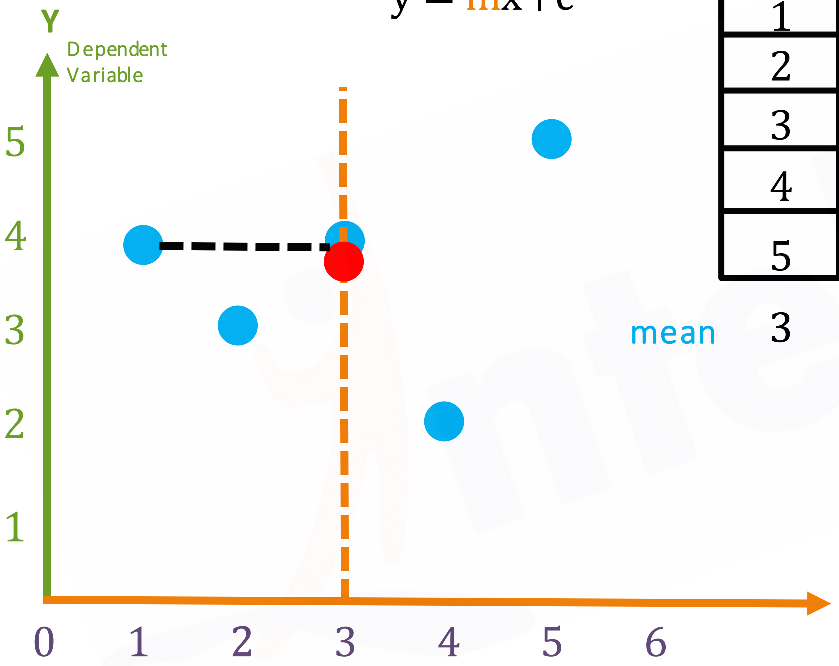


x	y	$x - \bar{x}$
1	4	-2
2	3	
3	4	
4	2	
5	5	
mean 3	3.6	

$$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Understanding Linear Regression

$$y = mx + c$$

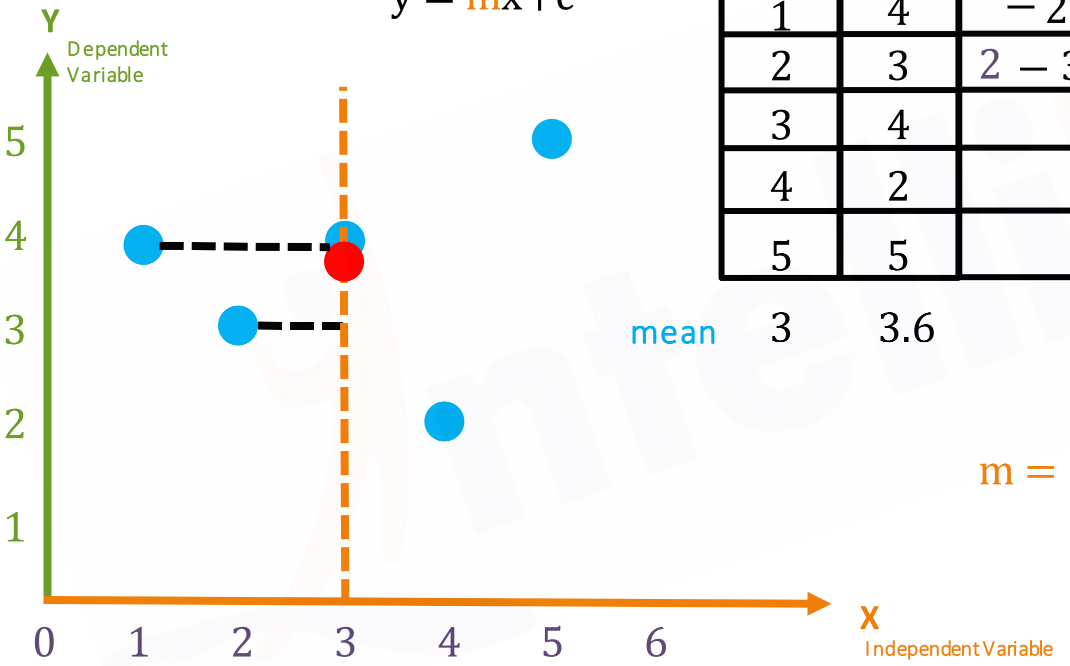


x	y	$x - \bar{x}$
1	4	-2
2	3	
3	4	
4	2	
5	5	
mean 3	3.6	

$$m = \frac{\Sigma\,(x - \bar{x})(y - \bar{y})}{\Sigma\,(x - \bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

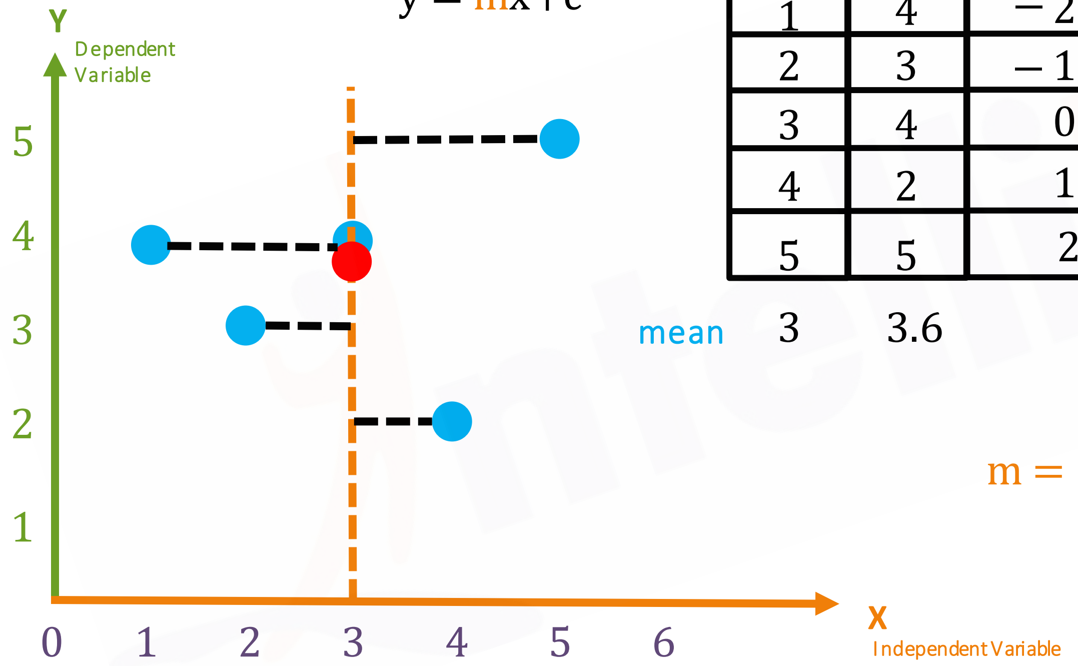


x	y	$x - \bar{x}$
1	4	-2
2	3	$2 - 3$
3	4	
4	2	
5	5	
mean 3	3.6	

$$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Understanding Linear Regression

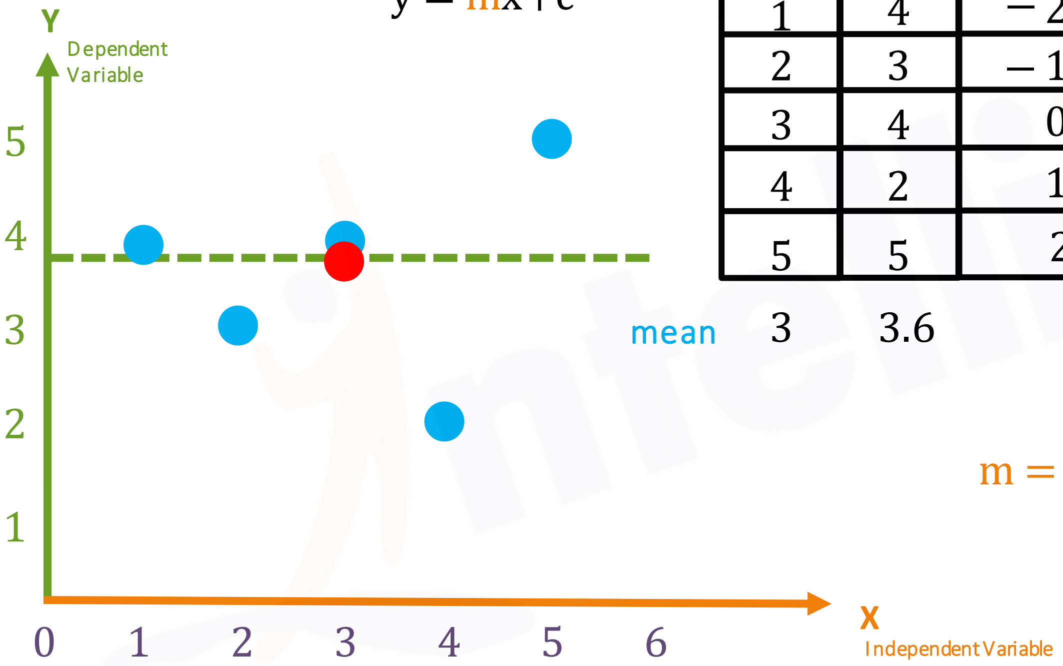
$y = mx + c$



x	y	$x - \bar{x}$
1	4	-2
2	3	-1
3	4	0
4	2	1
5	5	2
mean 3	3.6	

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2}$$

Understanding Linear Regression

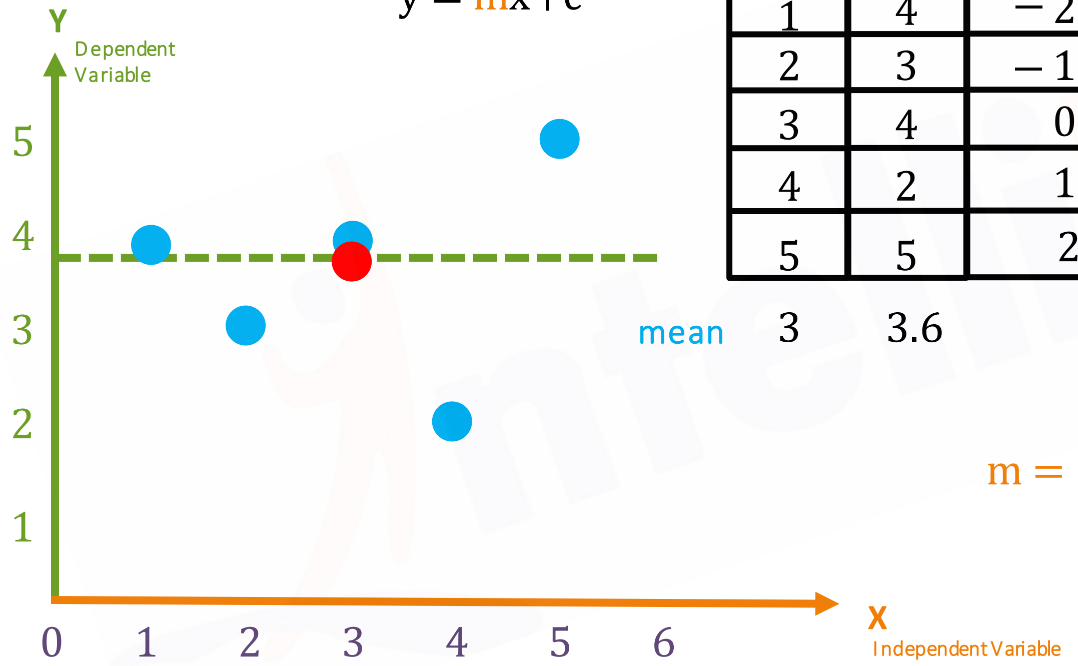


x	y	$x-\bar{x}$	$y-\bar{y}$
1	4	-2	
2	3	-1	
3	4	0	
4	2	1	
5	5	2	
mean 3	3.6		

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

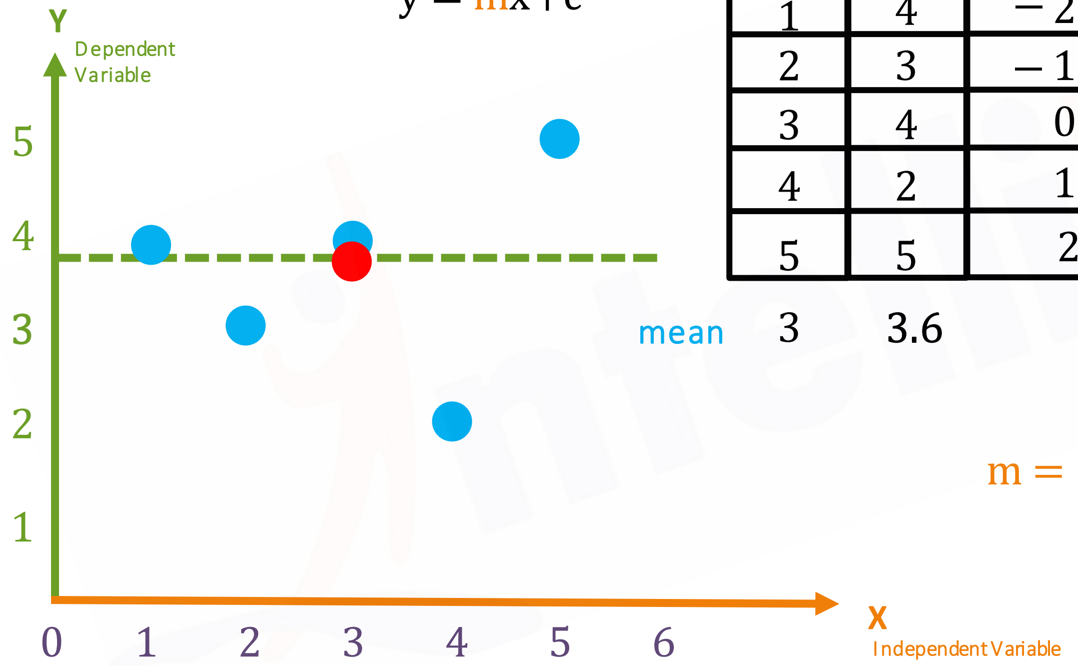


x	y	$x-\bar{x}$	$y-\bar{y}$
1	4	-2	$4-3.6$
2	3	-1	
3	4	0	
4	2	1	
5	5	2	
mean 3	3.6		

$$m = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$



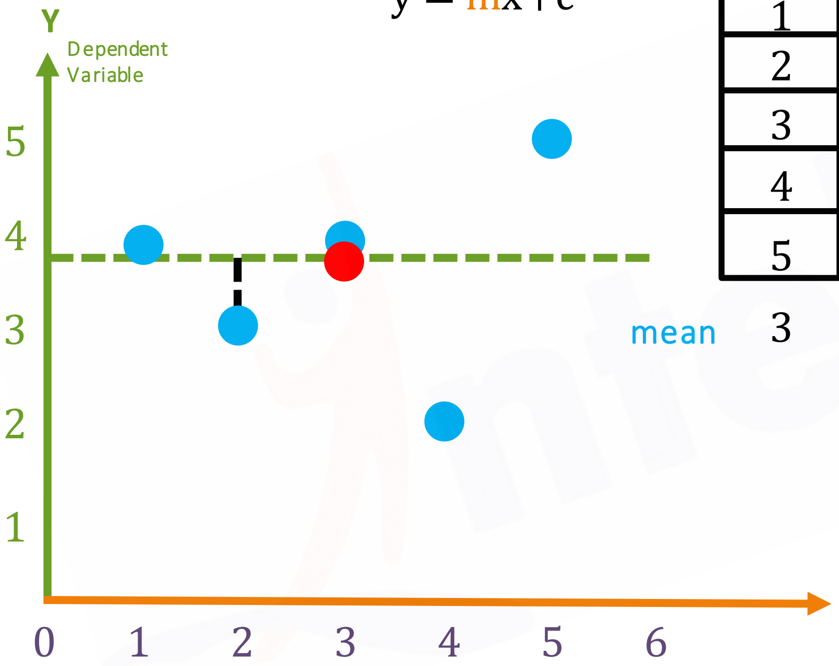
x	y	$x-\bar{x}$	$y-\bar{y}$
1	4	-2	0.4
2	3	-1	
3	4	0	
4	2	1	
5	5	2	

mean 3 3.6

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2}$$

Understanding Linear Regression

y = mx+c

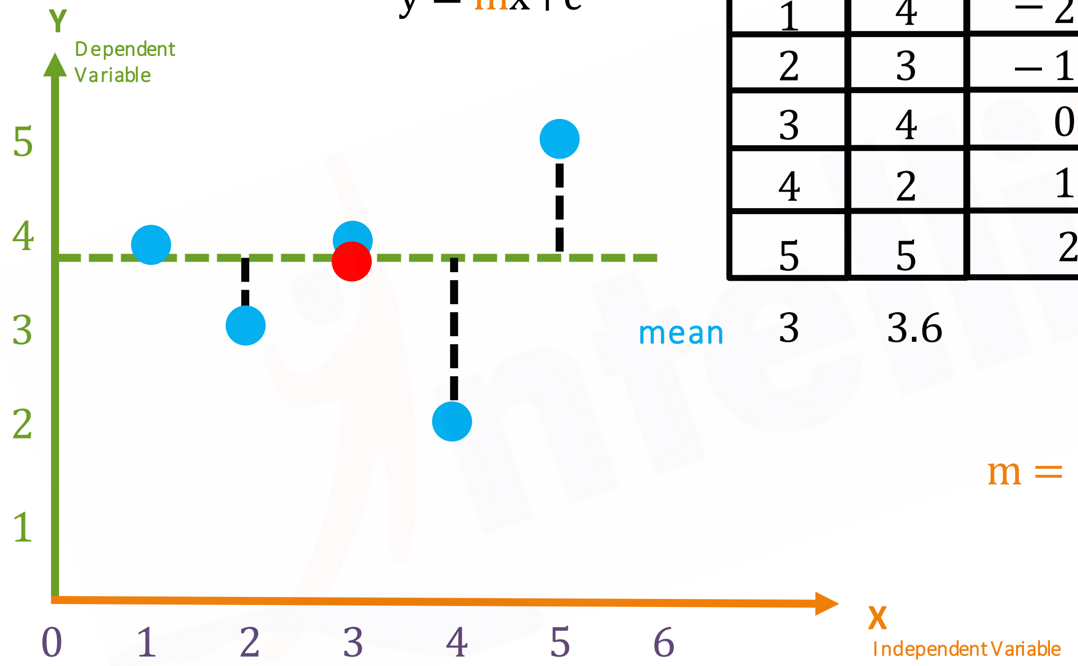


x	y	$x-\bar{x}$	$y-\bar{y}$
1	4	-2	0.4
2	3	-1	3 −3.6
3	4	0	
4	2	1	
5	5	2	
mean 3	3.6		

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\ (x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

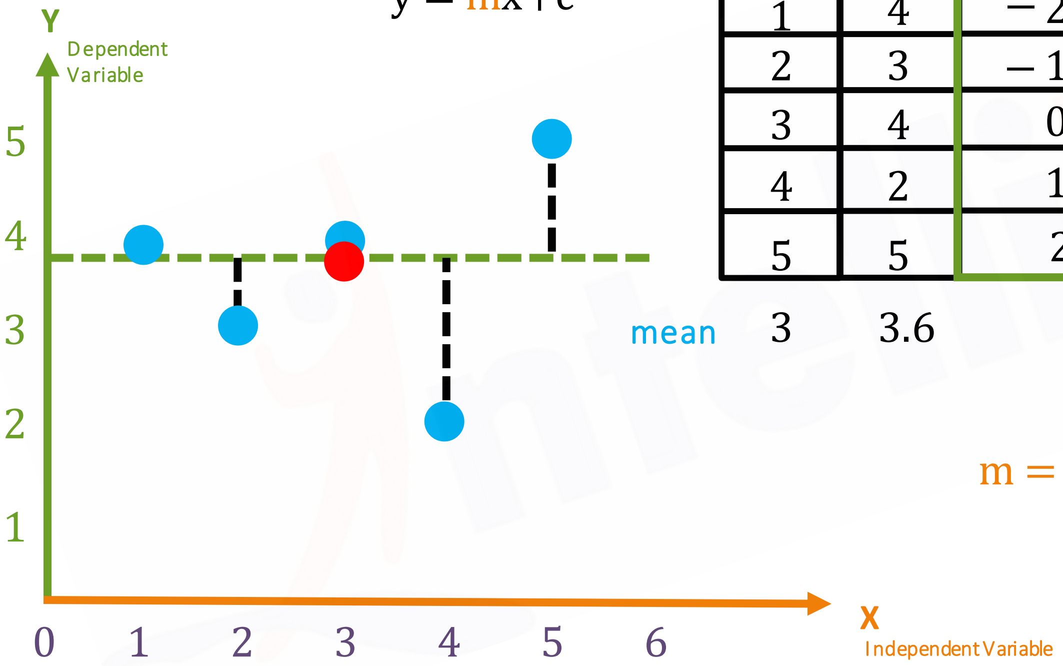


x	y	$x - \bar{x}$	$y - \bar{y}$
1	4	-2	0.4
2	3	-1	-0.6
3	4	0	0.4
4	2	1	-1.6
5	5	2	1.4
mean 3	3.6		

$$m = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

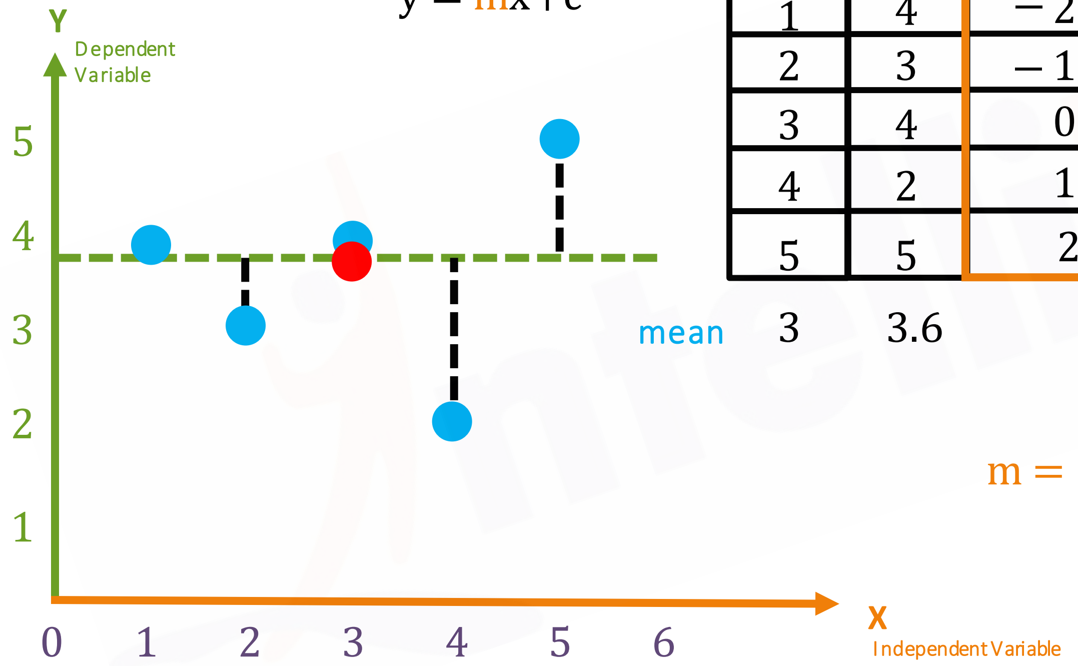


x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$
1	4	−2	0.4	4
2	3	−1	−0.6	1
3	4	0	0.4	0
4	2	1	−1.6	1
5	5	2	1.4	4
mean 3	3.6			

$$m = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

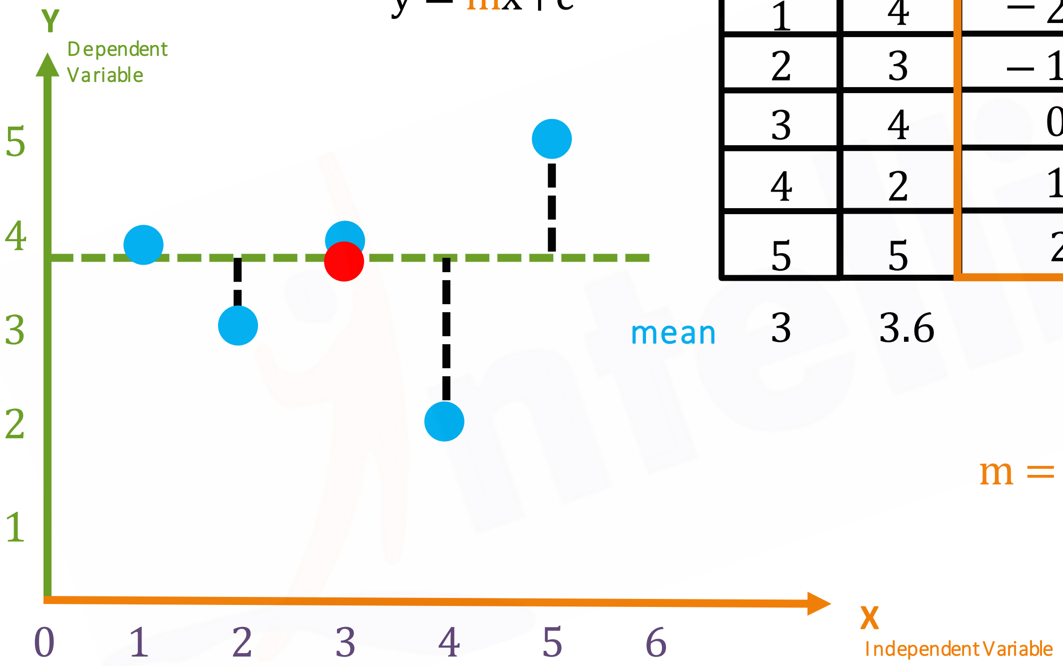


x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	4	-2	0.4	4	-2 x 0.4
2	3	-1	-0.6	1	
3	4	0	0.4	0	
4	2	1	-1.6	1	
5	5	2	1.4	4	
mean: 3	3.6				

$$m = \frac{\Sigma\,(x - \bar{x})(y - \bar{y})}{\Sigma\,(x - \bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

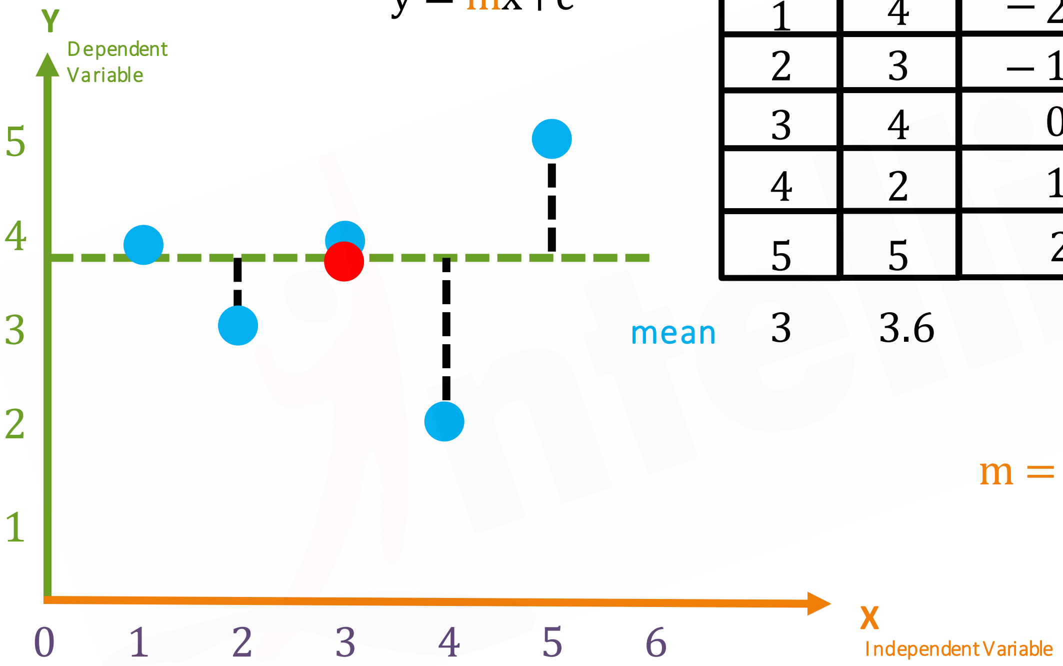


x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6				

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

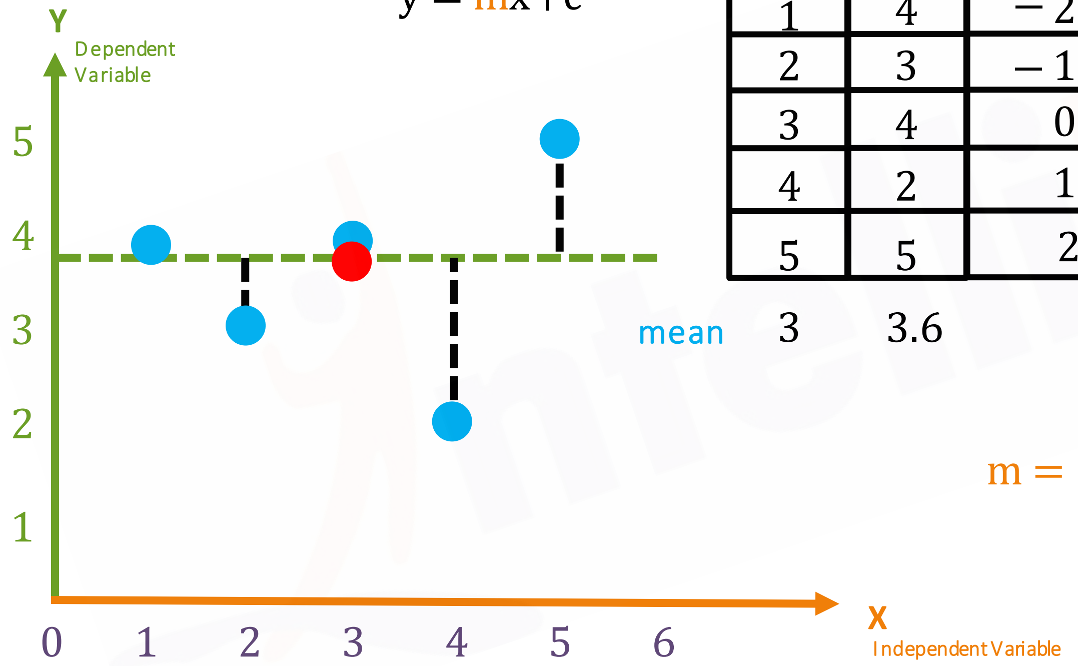


x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2}$$

Understanding Linear Regression

$y = mx + c$

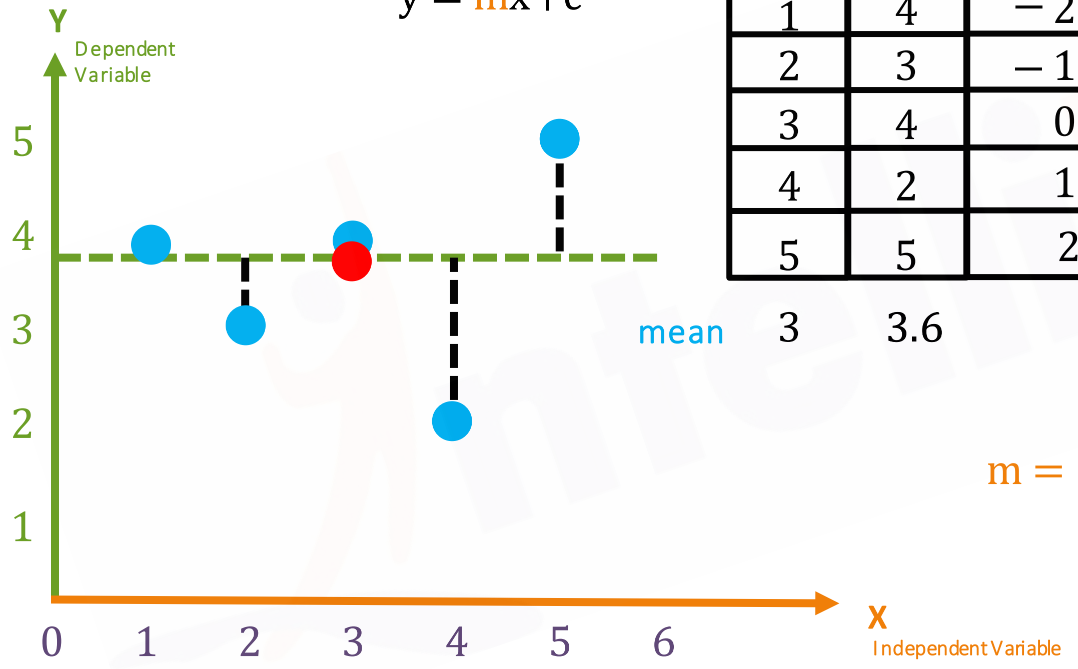


x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = \frac{1}{10}$$

Understanding Linear Regression

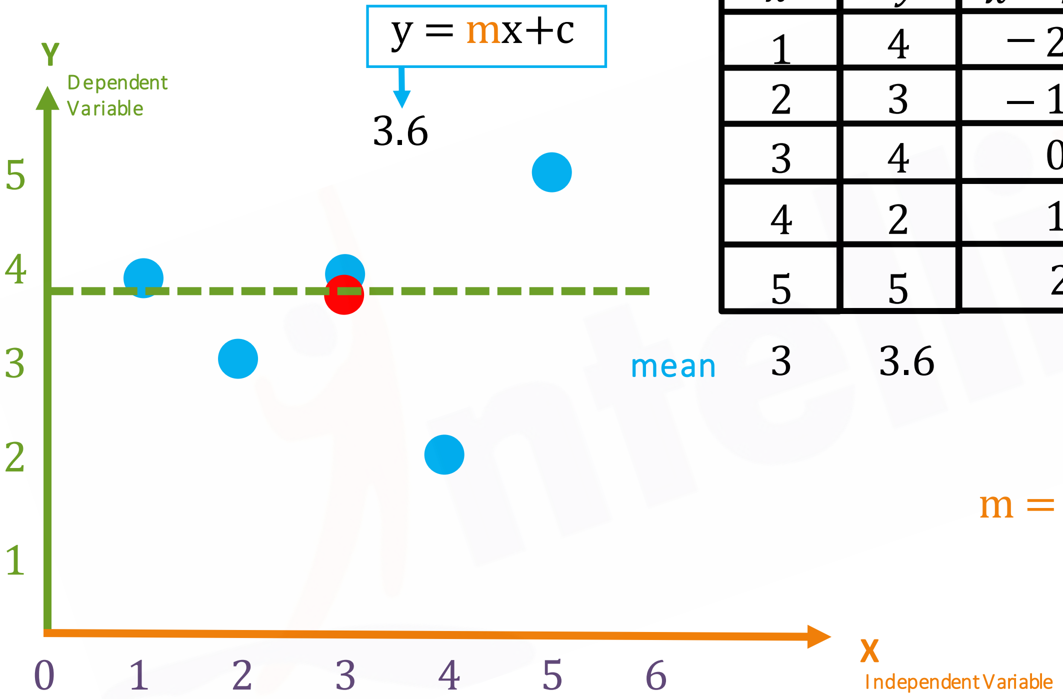
$y = mx + c$



x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2} = \mathbf{0.1}$$

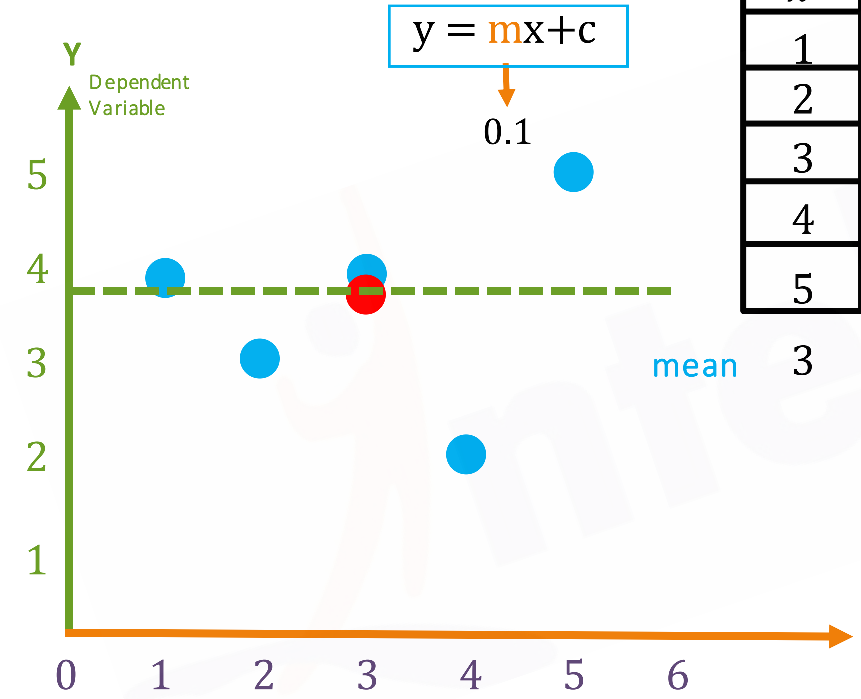
Understanding Linear Regression



x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

Understanding Linear Regression

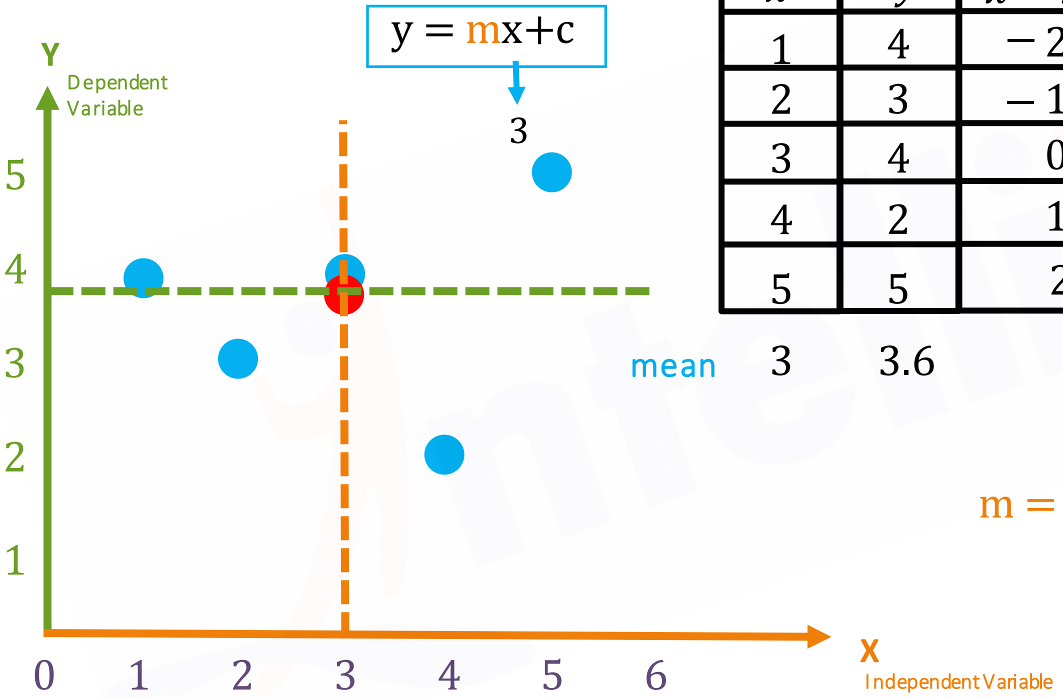


X
Independent Variable

	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
	1	4	-2	0.4	4	-0.8
	2	3	-1	-0.6	1	0.6
	3	4	0	0.4	0	0
	4	2	1	-1.6	1	-1.6
	5	5	2	1.4	4	2.8
mean	3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

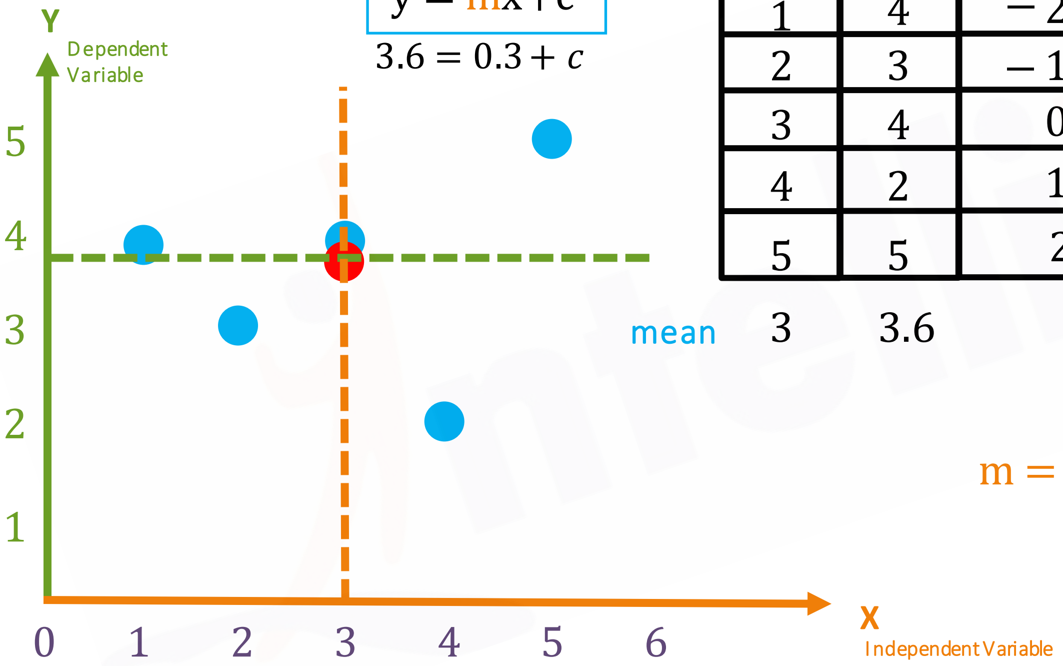
Understanding Linear Regression



x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

Understanding Linear Regression



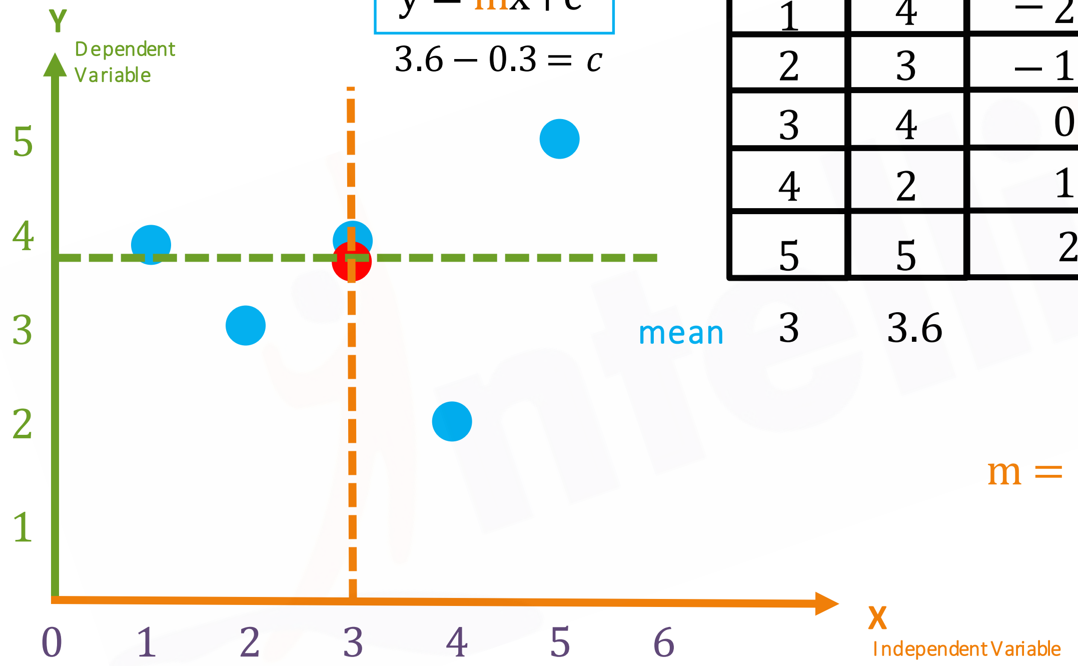
$$y = mx + c$$

$$3.6 = 0.3 + c$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

Understanding Linear Regression



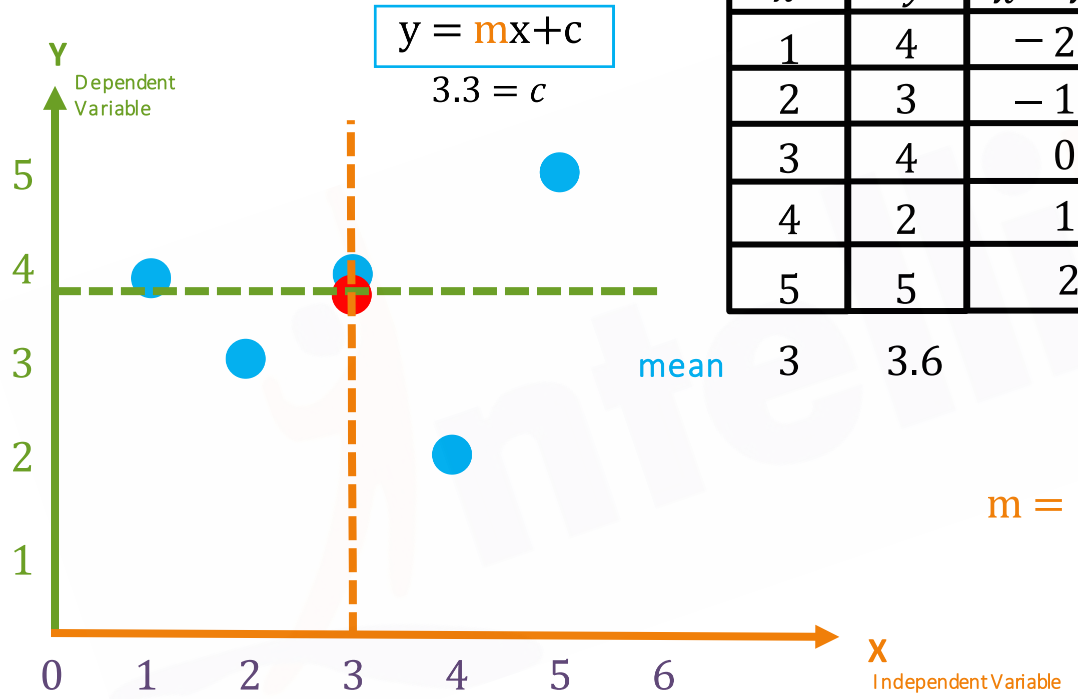
$$\boxed{y = mx + c}$$

$$3.6 - 0.3 = c$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

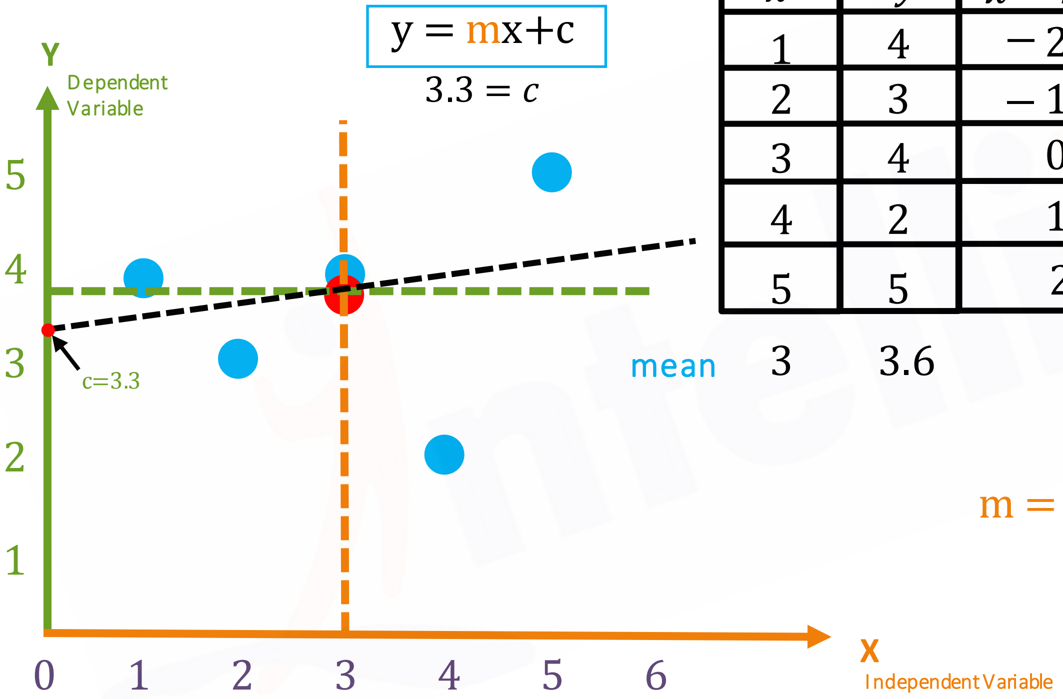
Understanding Linear Regression



x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
1	4	-2	0.4	4	-0.8
2	3	-1	-0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	-1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2} = 0.1$$

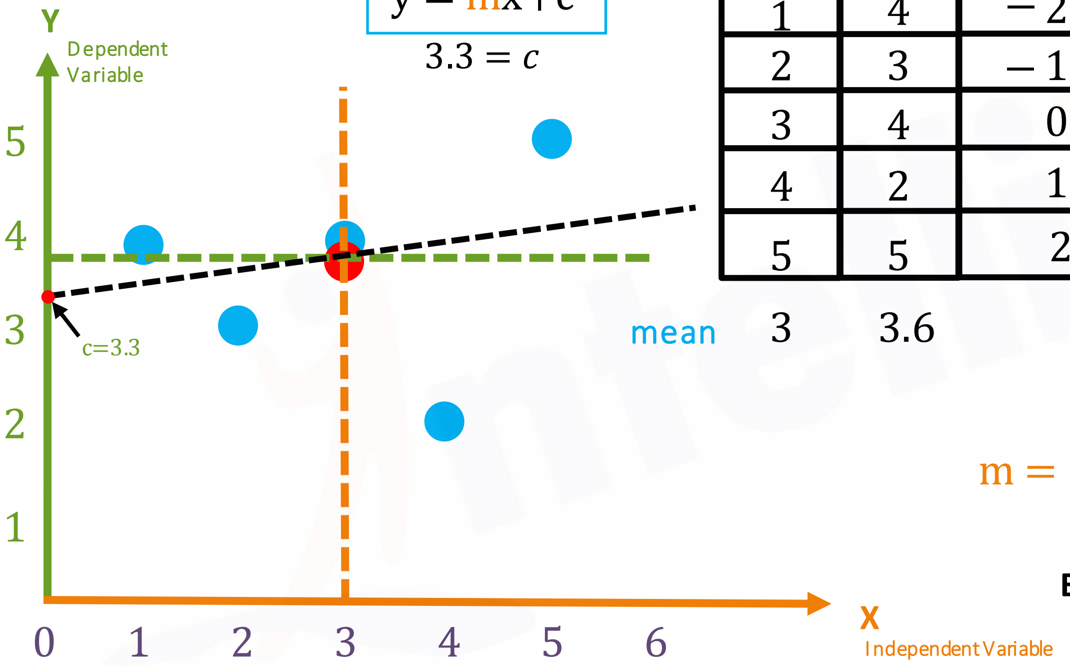
Understanding Linear Regression



	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
	1	4	-2	0.4	4	-0.8
	2	3	-1	-0.6	1	0.6
	3	4	0	0.4	0	0
	4	2	1	-1.6	1	-1.6
	5	5	2	1.4	4	2.8
mean	3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma\,(x-\bar{x})(y-\bar{y})}{\Sigma\,(x-\bar{x})^2} = 0.1$$

Understanding Linear Regression



$y = mx + c$

$3.3 = c$

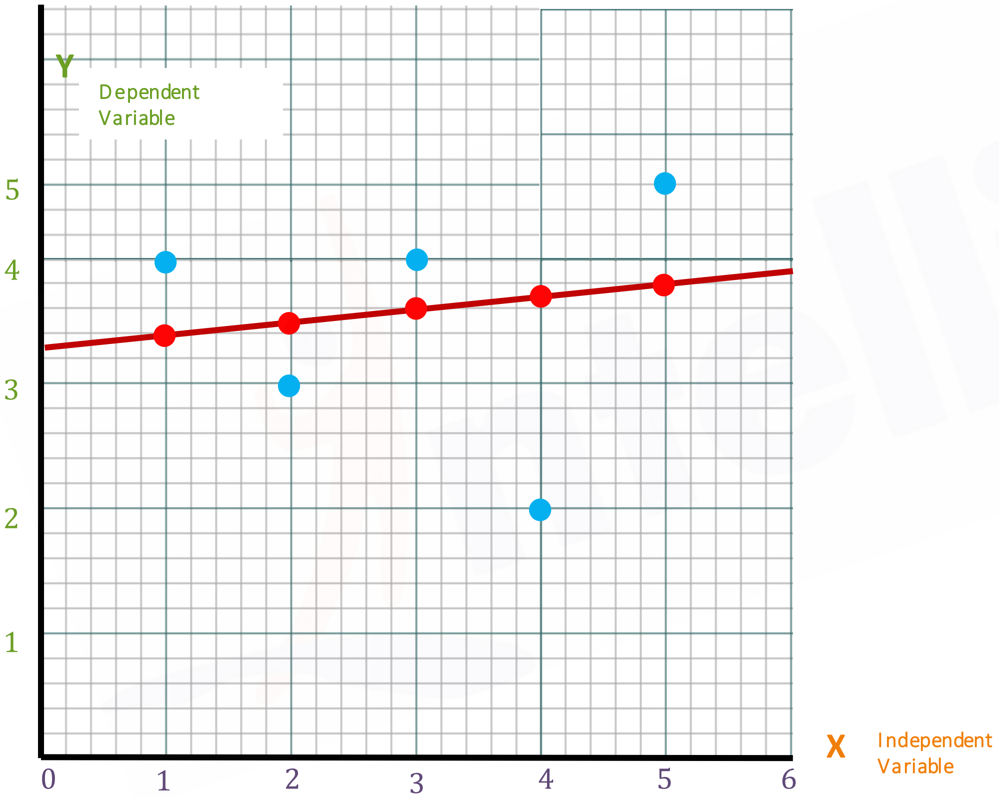
x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$
1	4	−2	0.4	4	-0.8
2	3	−1	−0.6	1	0.6
3	4	0	0.4	0	0
4	2	1	−1.6	1	-1.6
5	5	2	1.4	4	2.8
mean 3	3.6			$\Sigma = 10$	$\Sigma = 1$

$$m = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2} = 0.1$$

Equation of Regression line for m = 0.1, c = 3.3 is:

$$\boxed{y = 0.1x + 3.3}$$

Mean Square Error

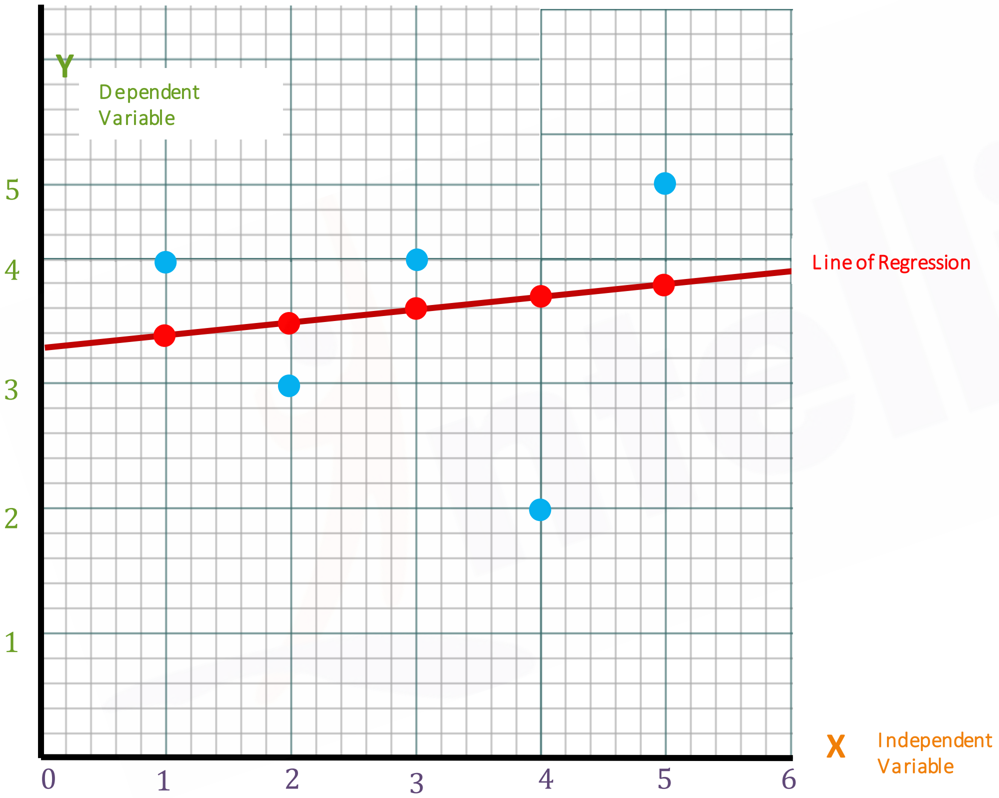


$m = 0.1$
$c = 3.3$
$y = 0.1x + 3.3$

For given $m = 0.1$ & $c = 3.3$,
Lets predict values for y when
$x = \{1,2,3,4,5\}$

$y = 0.1 \text{ x } 1 + 3.3 = 3.2$
$y = 0.1 \text{ x } 2 + 3.3 = 3.1$
$y = 0.1 \text{ x } 3 + 3.3 = 3.0$
$y = 0.1 \text{ x } 4 + 3.3 = 2.9$
$y = 0.1 \text{ x } 5 + 3.3 = 2.8$

Mean Square Error

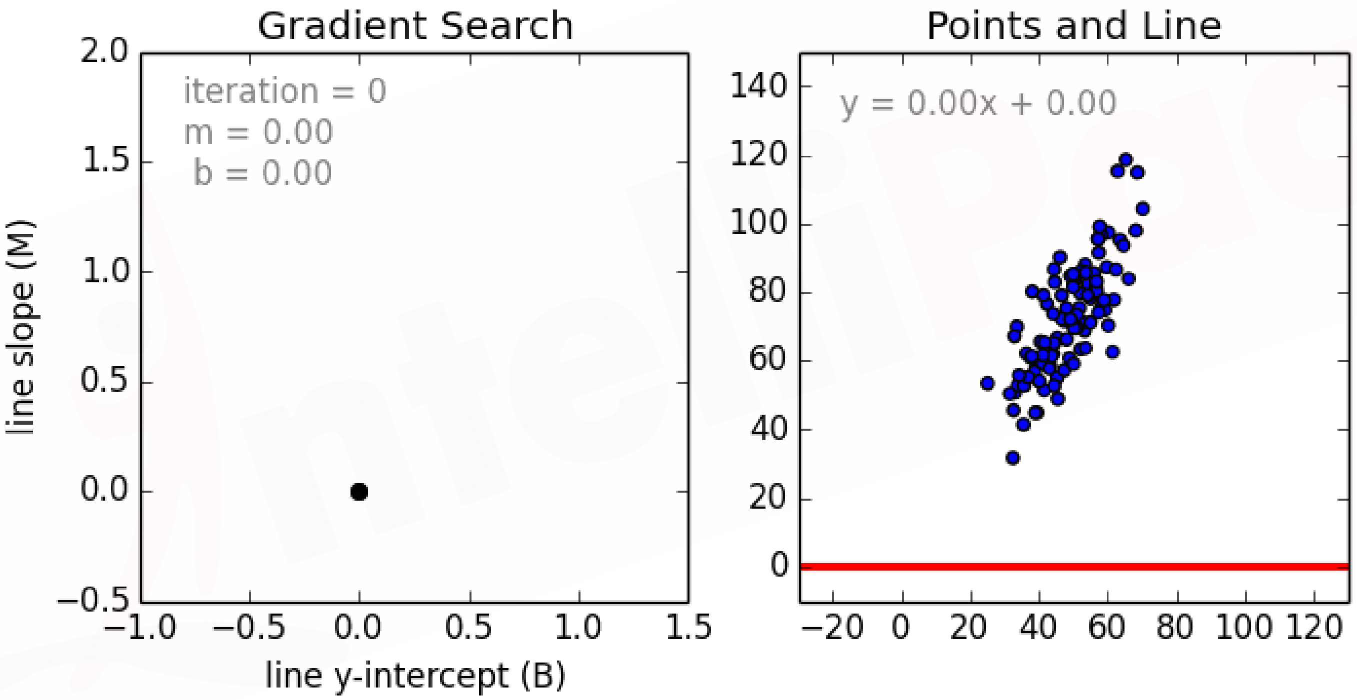


$m = 0.1$
$c = 3.3$
$y = 0.1x + 3.3$

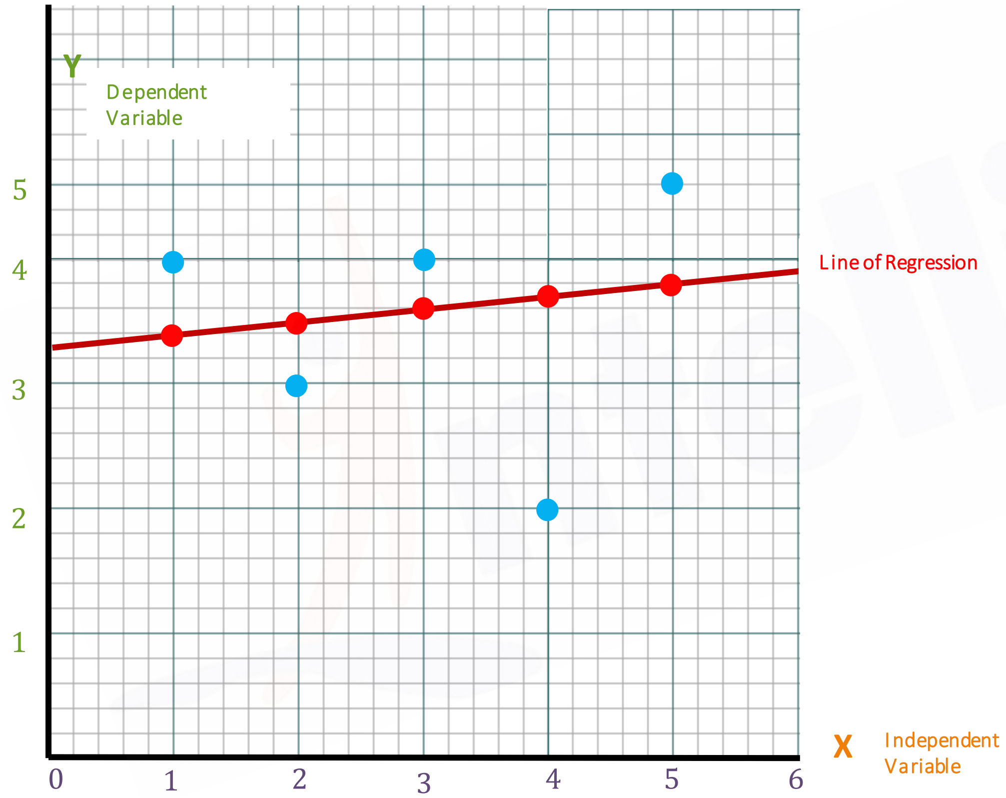
For given $m = 0.1$ & $c = 3.3$,
Lets predict values for y when
$x = \{1,2,3,4,5\}$

$y = 0.1 \text{ x } 1 + 3.3 = 3.2$
$y = 0.1 \text{ x } 2 + 3.3 = 3.1$
$y = 0.1 \text{ x } 3 + 3.3 = 3.0$
$y = 0.1 \text{ x } 4 + 3.3 = 2.9$
$y = 0.1 \text{ x } 5 + 3.3 = 2.8$

Finding the best Fit Line



Goodness of Fit – R^2

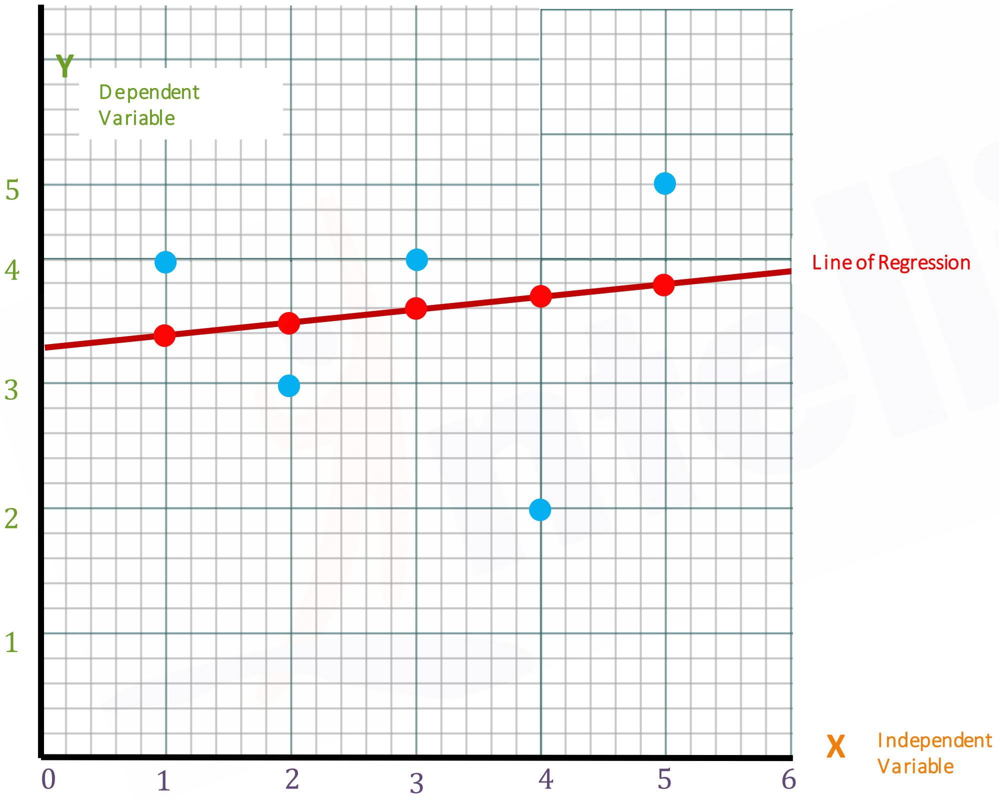


x	y_p
1	3.2
2	3.1
3	3.0
4	2.9
5	2.8

$$R^2 = \frac{\sum (\text{Predicted Distance} - \text{Mean})^2}{\sum (\text{Actual Distance} - \text{Mean})^2}$$

$$R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2}$$

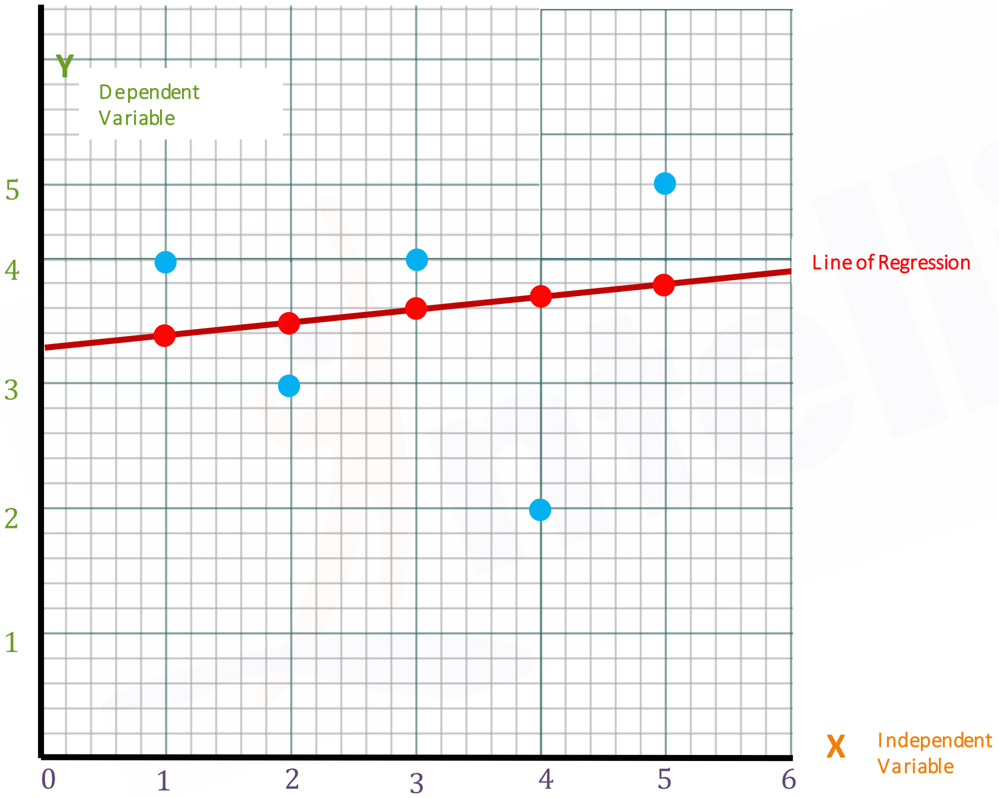
Goodness of Fit – R^2



x	y	y_p	$(y_p - \bar{y})$
1	4	3.2	3.2-3.6
2	3	3.1	
3	4	3.0	
4	2	2.9	
5	5	2.8	

$$R^2 = \frac{\Sigma\ (y_p - \bar{y})^2}{\Sigma\ (y - \bar{y})^2}$$

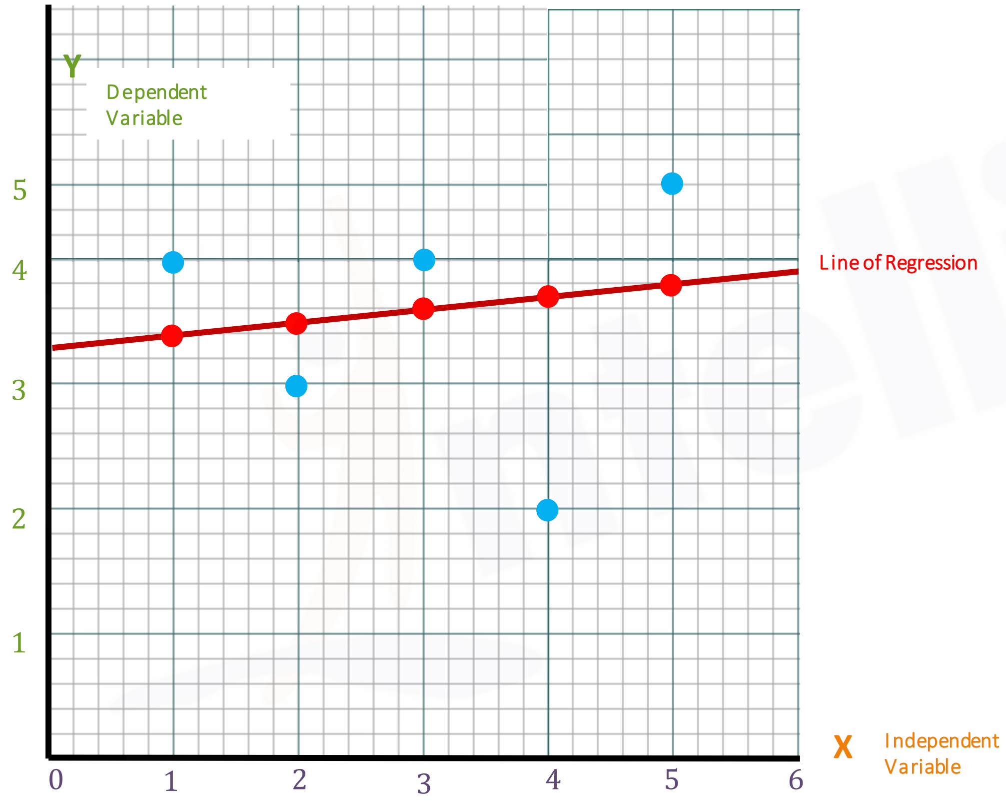
Goodness of Fit – R^2



x	y	y_p	$(y_p - \bar{y})$	$(y - \bar{y})$
1	4	3.2	-0.4	4 – 3.6
2	3	3.1	-0.5	
3	4	3.0	-0.6	
4	2	2.9	-0.7	
5	5	2.8	-0.8	

$$R^2 = \frac{\Sigma\ (y_p - \bar{y})^2}{\Sigma\ (y - \bar{y})^2}$$

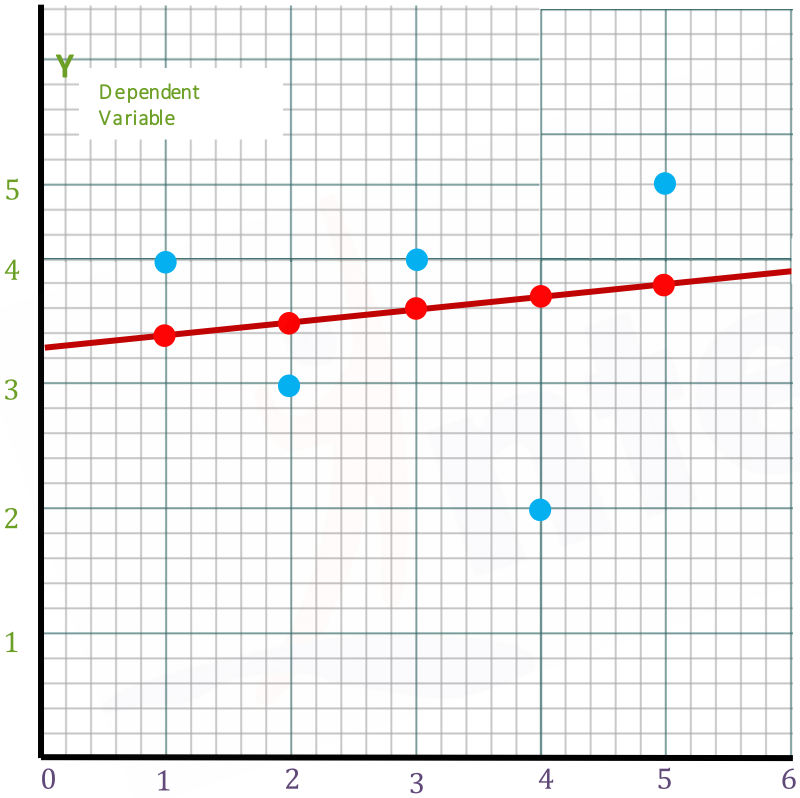
Goodness of Fit – R^2



x	y	y_p	$(y_p - \bar{y})$	$(y - \bar{y})$
1	4	3.2	-0.4	0.4
2	3	3.1	-0.5	-0.6
3	4	3.0	-0.6	0.4
4	2	2.9	-0.7	-1.6
5	5	2.8	-0.8	1.4

$$R^2 = \frac{\Sigma \ (y_p - \bar{y})^2}{\Sigma \ (y - \bar{y})^2}$$

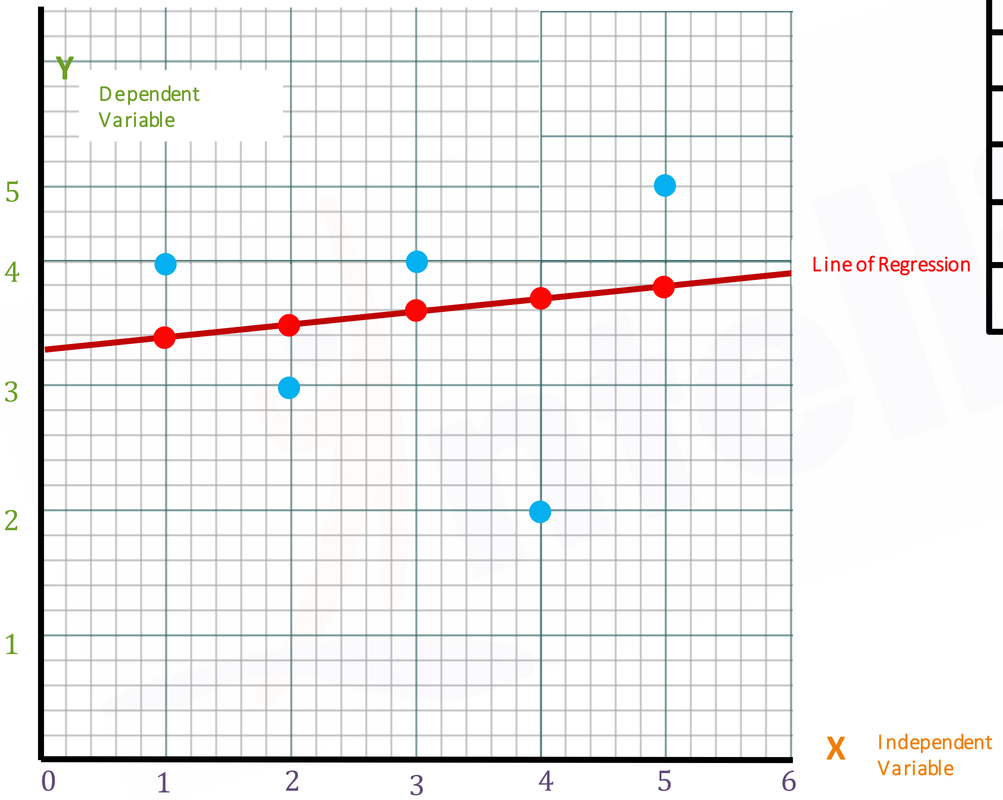
Goodness of Fit – R^2



y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$
3.2	-0.4	0.4	$(-0.4)^2$
3.1	-0.5	-0.6	
3.0	-0.6	0.4	
2.9	-0.7	-1.6	
2.8	-0.8	1.4	

$$R^2 = \frac{\Sigma \ (y_p - \bar{y})^2}{\Sigma \ (y - \bar{y})^2}$$

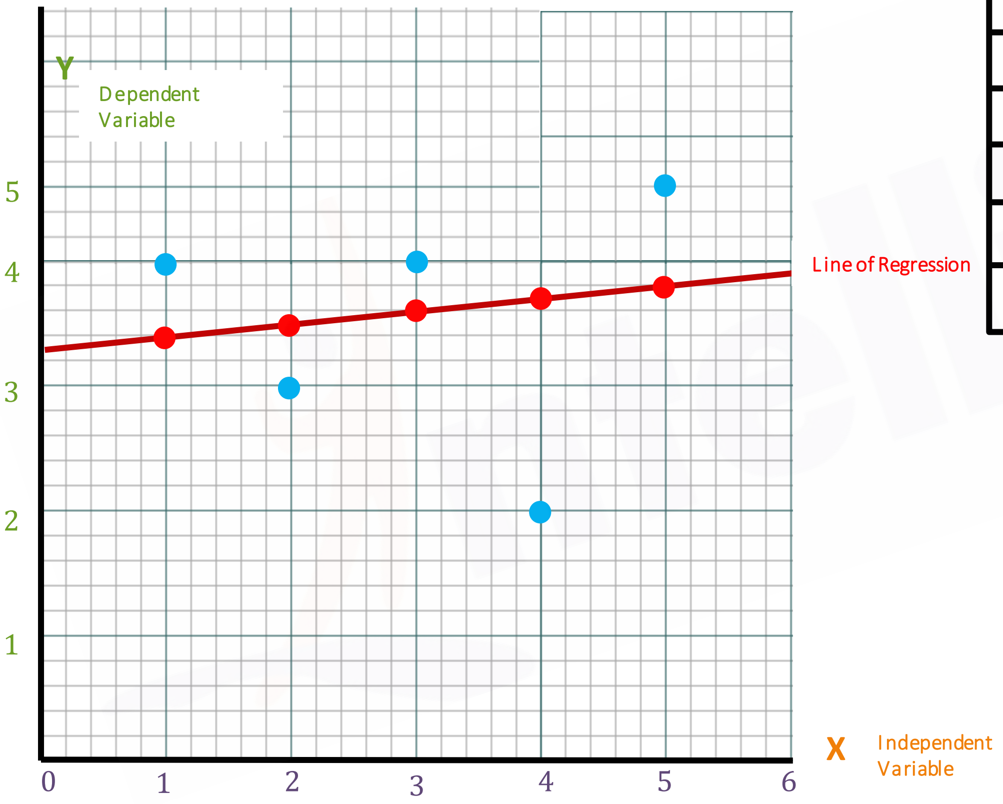
Goodness of Fit – R^2



y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$
3.2	-0.4	0.4	0.16
3.1	-0.5	-0.6	0.25
3.0	-0.6	0.4	0.36
2.9	-0.7	-1.6	0.49
2.8	-0.8	1.4	0.64

$$R^2 = \frac{\Sigma\ (y_p - \bar{y})^2}{\Sigma\ (y - \bar{y})^2}$$

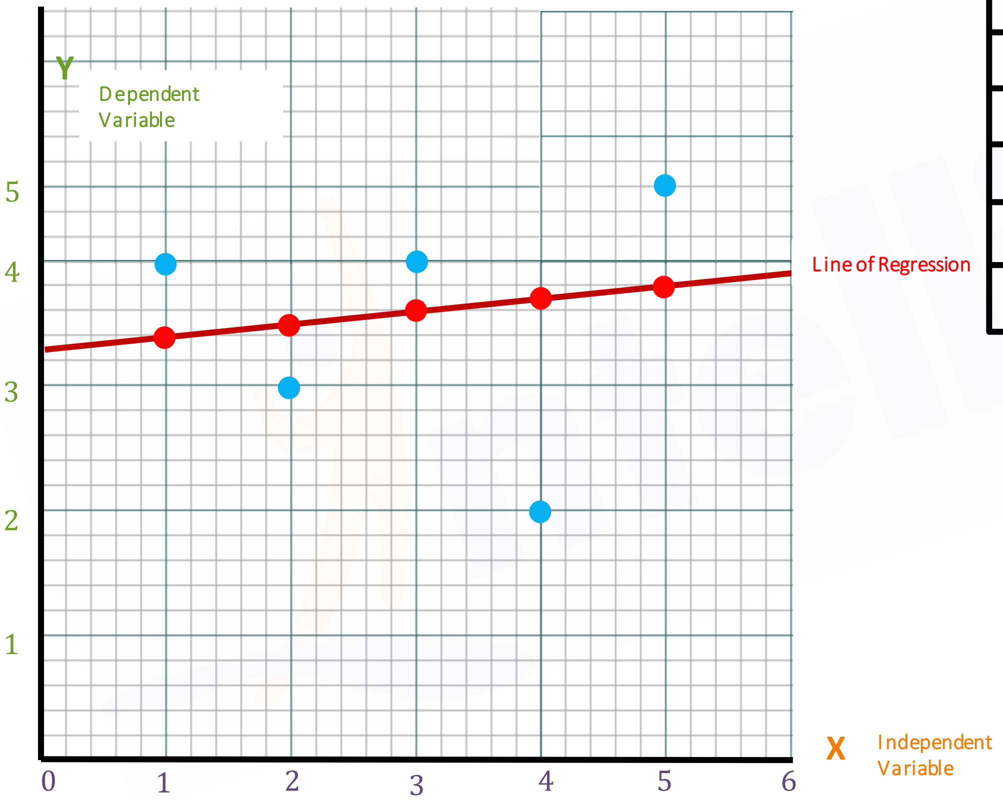
Goodness of Fit – R^2



y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$	$(y - \bar{y})^2$
3.2	-0.4	0.4	0.16	$(0.4)^2$
3.1	-0.5	-0.6	0.25	
3.0	-0.6	0.4	0.36	
2.9	-0.7	-1.6	0.49	
2.8	-0.8	1.4	0.64	

$$R^2 = \frac{\Sigma\ (y_p - \bar{y})^2}{\Sigma\ (y - \bar{y})^2}$$

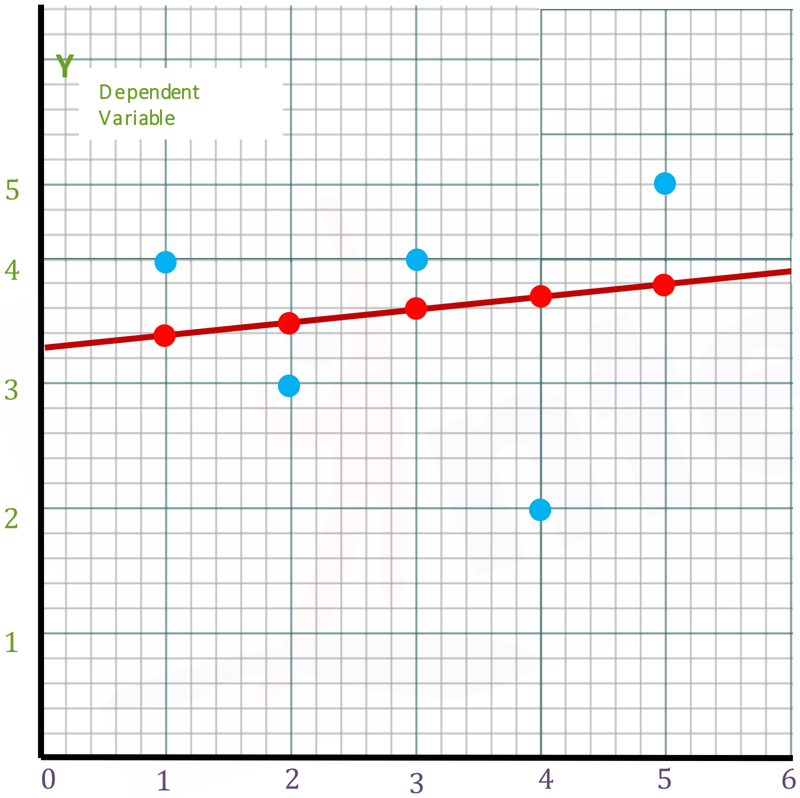
Goodness of Fit – R^2



y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$	$(y - \bar{y})^2$
3.2	-0.4	0.4	0.16	0.16
3.1	-0.5	-0.6	0.25	0.36
3.0	-0.6	0.4	0.36	0.16
2.9	-0.7	-1.6	0.49	2.56
2.8	-0.8	1.4	0.64	1.96

$$R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2}$$

Goodness of Fit – R^2

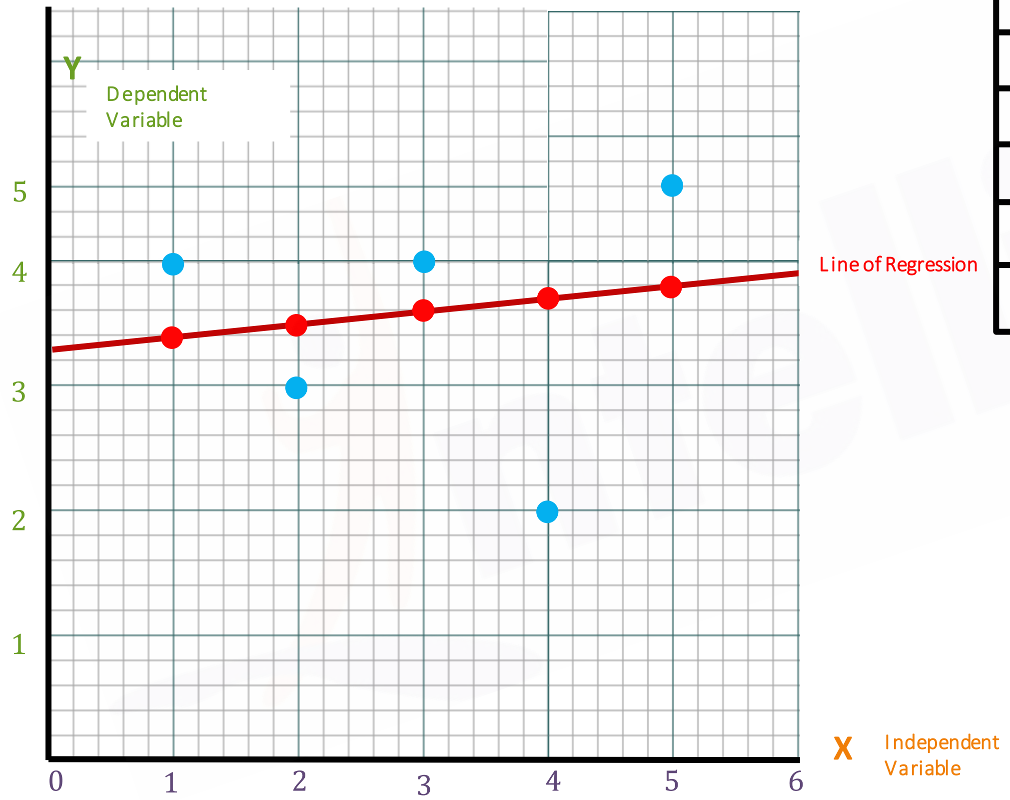


y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$	$(y - \bar{y})^2$
3.2	-0.4	0.4	0.16	0.16
3.1	-0.5	-0.6	0.25	0.36
3.0	-0.6	0.4	0.36	0.16
2.9	-0.7	-1.6	0.49	2.56
2.8	-0.8	1.4	0.64	1.96

$\sum$ 1.9 $\sum$ 5.2

$$R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{1.9}{5.2}$$

Goodness of Fit – R^2

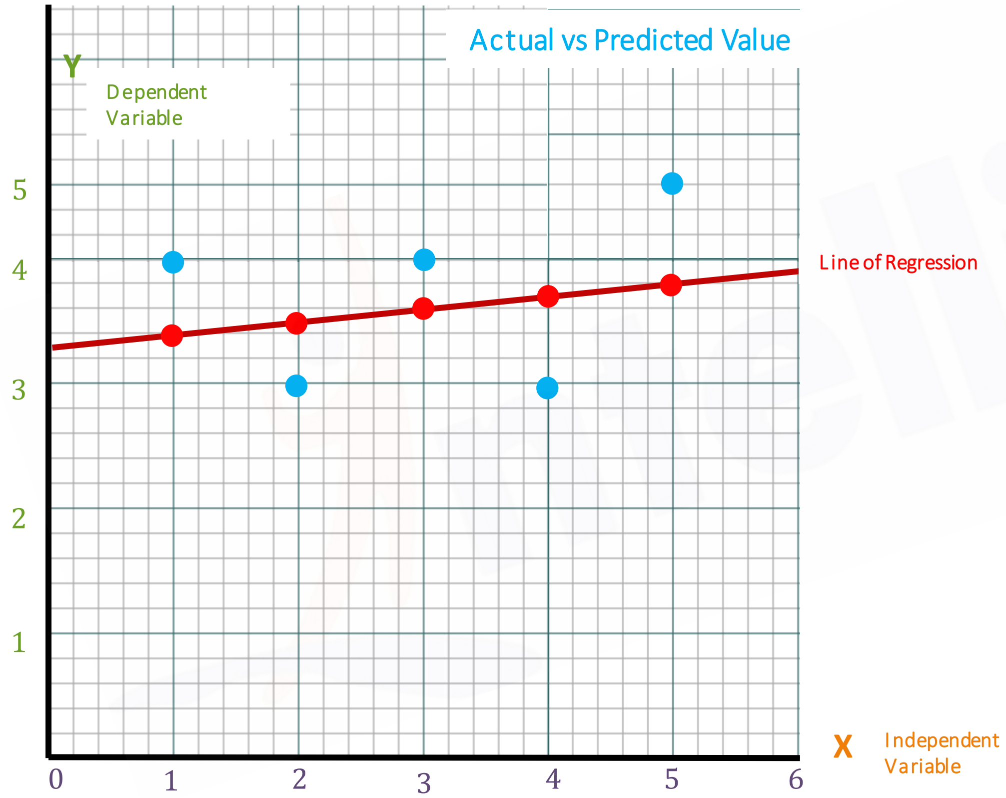


y_p	$(y_p - \bar{y})$	$(y - \bar{y})$	$(y_p - \bar{y})^2$	$(y - \bar{y})^2$
3.2	-0.4	0.4	0.16	0.16
3.1	-0.5	-0.6	0.25	0.36
3.0	-0.6	0.4	0.36	0.16
2.9	-0.7	-1.6	0.49	2.56
2.8	-0.8	1.4	0.64	1.96

$\sum$ 1.9 $\sum$ 5.2

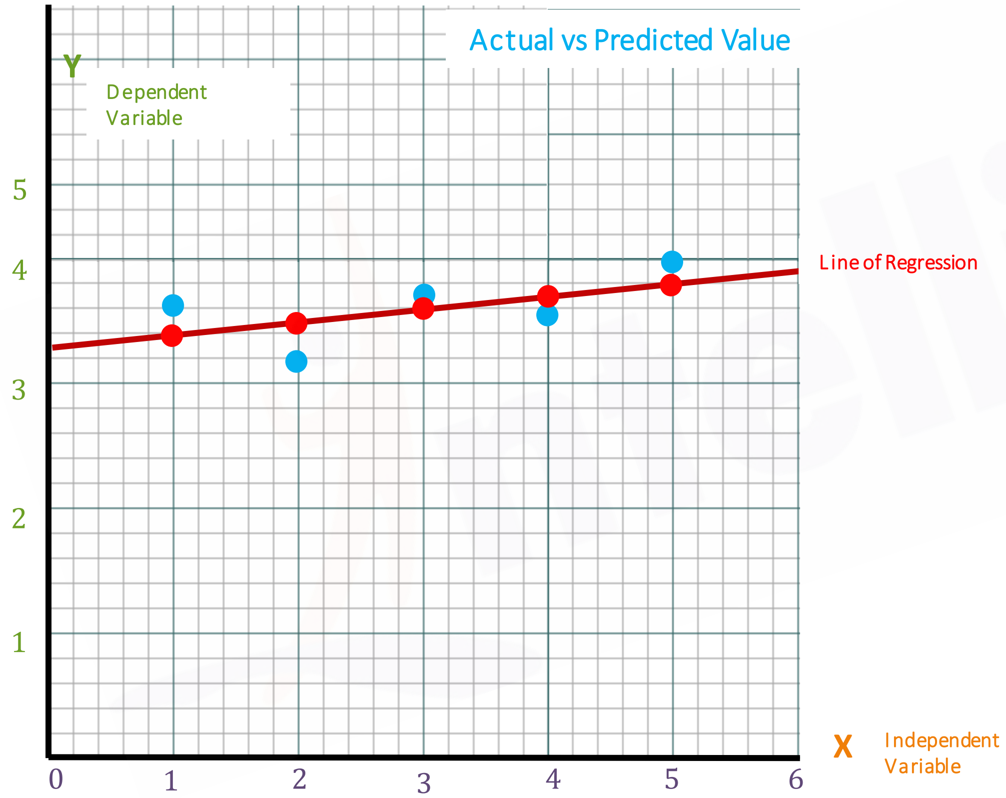
$$R^2 = \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2} = 0.36$$

Goodness of Fit – R^2



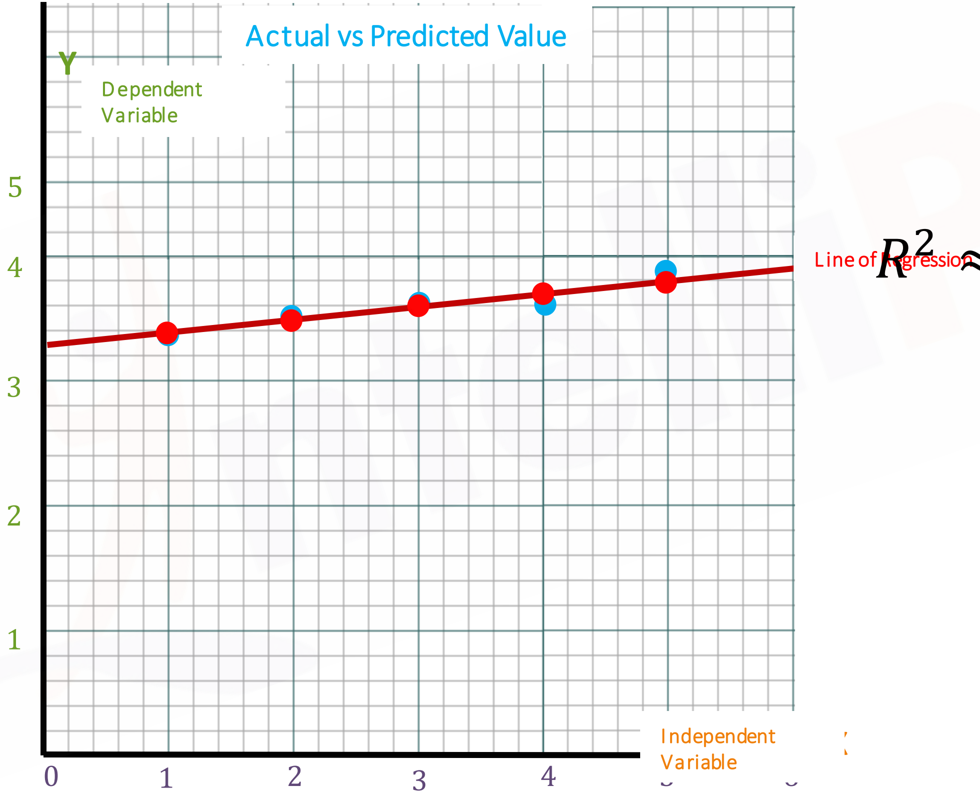
$R^2 \approx 0.6$

Goodness of Fit – R^2



$R^2 \approx 0.9$

Goodness of Fit – R^2



$R^2 \approx 1$

Thank You

www.intellipaat.com

CALL US NOW

India : +91-7847955955

US : 1-800-216-8930 (TOLL FREE)

sales@intellipaat.com