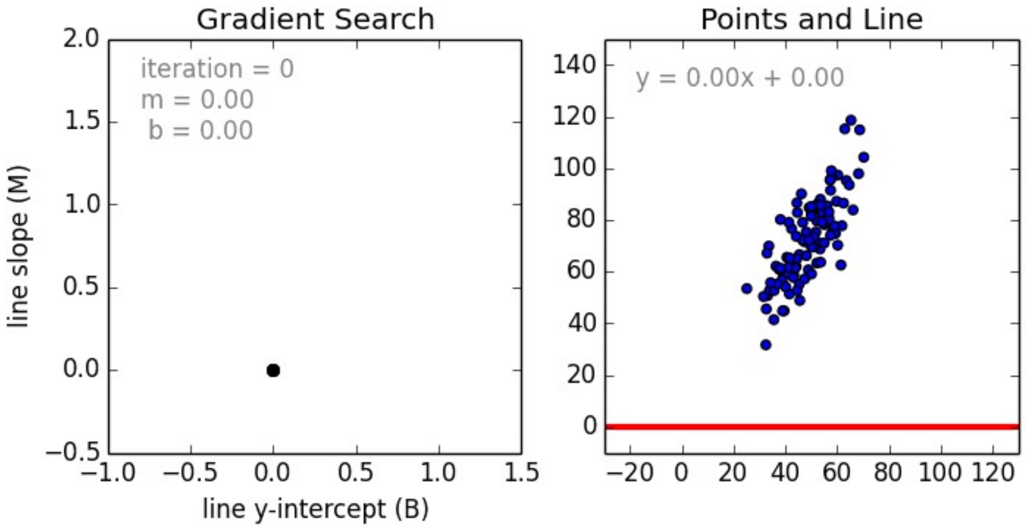
Linear Regression



Simple Linear Regression

- Simple Linear regression
- Least Squares Method
- Coefficient of determination
- Model Assumptions
- Testing for Significance
- Using the Estimated regression Equation for Estimation an Prediction

USES OF REGRESSION

Major uses of regression analysis are:-

- Determining the strength of predictors
- Forecasting an effect, and
- Trend Forecasting
- Evaluating Trends and Sales estimates
- Analyzing the Impact of price Changes
- Assessment of risk in financial services and insurance domain

EDU FABRICA

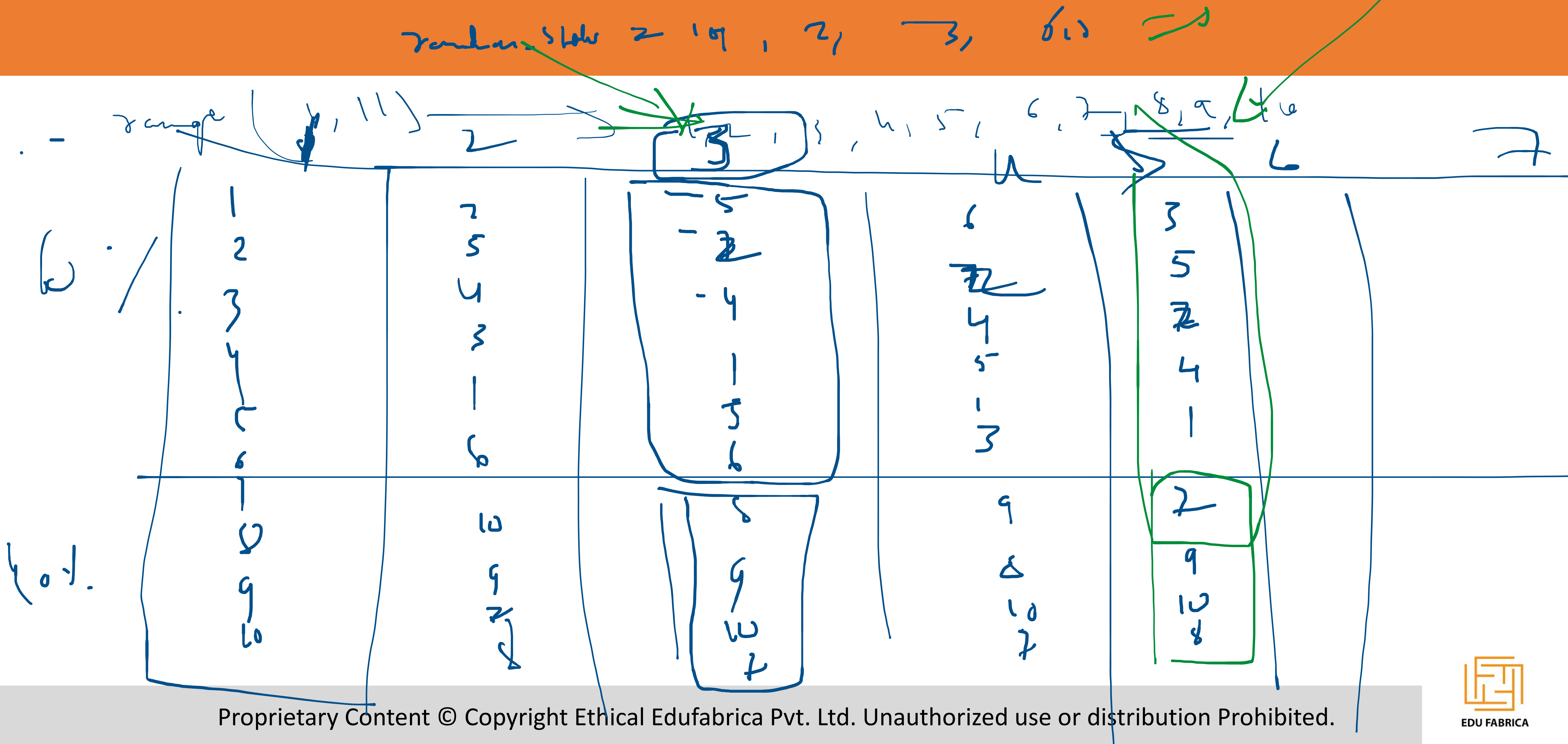
Linear Regression

- The key point in the linear regression is that our dependent value should be continuous
- and cannot be a discreet value. However, the independent variable(s) can be
- measured on either a categorical or continuous measurement scale.
- There are two types of linear regression models.
- They are: simple regression and multiple regression.

SELECTION CRITERIA

- Classification and Regression Capabilities
- Data Quality
- Computational Complexity
- Comprehensible and Transparent

Random State



EDU FABRICA

What is Regression ?

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

EDU FABRICA

What is Regression ?

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

EDU FABRICA

What is Regression ?

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

What is Regression ?

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a Continuous value.

What is Regression ?

X : Independent Variable

Y : Dependent Variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a Continuous value.

What is Regression ?

X : Independent Variable

Y : Dependent Variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a Continuous value.

What is Regression ?

X : Independent Variable

Y : Dependent Variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Regression is the process of predicting a Continuous value.

EDU FABRICA

What is Regression ?

X : Independent Variable

Y : Dependent Variable

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION_COMB	CO2EMISSIONS
0	2.0	4	8.5	196
1	2.4	4	9.6	221
2	1.5	4	5.9	136
3	3.5	6	11.1	255
4	3.5	6	10.6	244
5	3.5	6	10.0	230
6	3.5	6	10.1	232
7	3.7	6	11.1	255
8	3.7	6	11.6	267
9	2.4	4	9.2	?

Continuous Variable

Regression is the process of predicting a Continuous value.

EDU FABRICA

Types of Linear Regression

- **Simple Linear Regression**
  - Simple Linear Regression
  - Simple Non – Linear Regression

Predict co2emission vs EngineSize of all cars

- **Multiple Regression**
  - Multiple Linear Regression
  - Multiple Non – Linear Regression

Predict co2emission vs EngineSize and Cylinders of all cars

Empirical Model

- As an illustration , consider the data in the table
- In this table y is the purity of oxygen produced in a chemical distillation process, and x is the percentage of hydro carbons that are present tin the main condenser of the distillation unit.

Hydrocarbon Level (X)	Purity (Y)
0.99	90.01
1.02	89.05
1.15	91.43
1.29	93.74
1.46	96.73
1.36	94.45
0.87	87.59
1.23	91.77
1.55	99.42
1.4	93.65

Linear vs Logistic

Basis	Linear Regression	Logistic Regression
Core Concept	The data is modelled using a straight line	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
Used with	Continuous Variables	Categorical Variable
Output / prediction	Values of the Variable	Probability of occurrence of the event.
Accuracy and Goodness of fit	Measured by Loss, R squared, Adjusted R squared etc.	Accuracy, Precision, Recall, F1 Score , ROC curve, Confusion Matrix , etc.

Simple Linear Regression Model

- Regression analysis. is a form of predictive modelling technique which investigates the relationship the relationship between a dependent and independent variable.

- The equation that describes how y is related to x and an error term is called the regression model.

- The Simple Linear Regression Model is :

$$\boxed{y = \beta_0 + \beta_1 x + \epsilon}$$

Where:

β_0 and β_1 are called parameters of the model,
ϵ is a random variable called the error term.

Simple Linear Regression Model

- The Simple Linear Regression Model is :

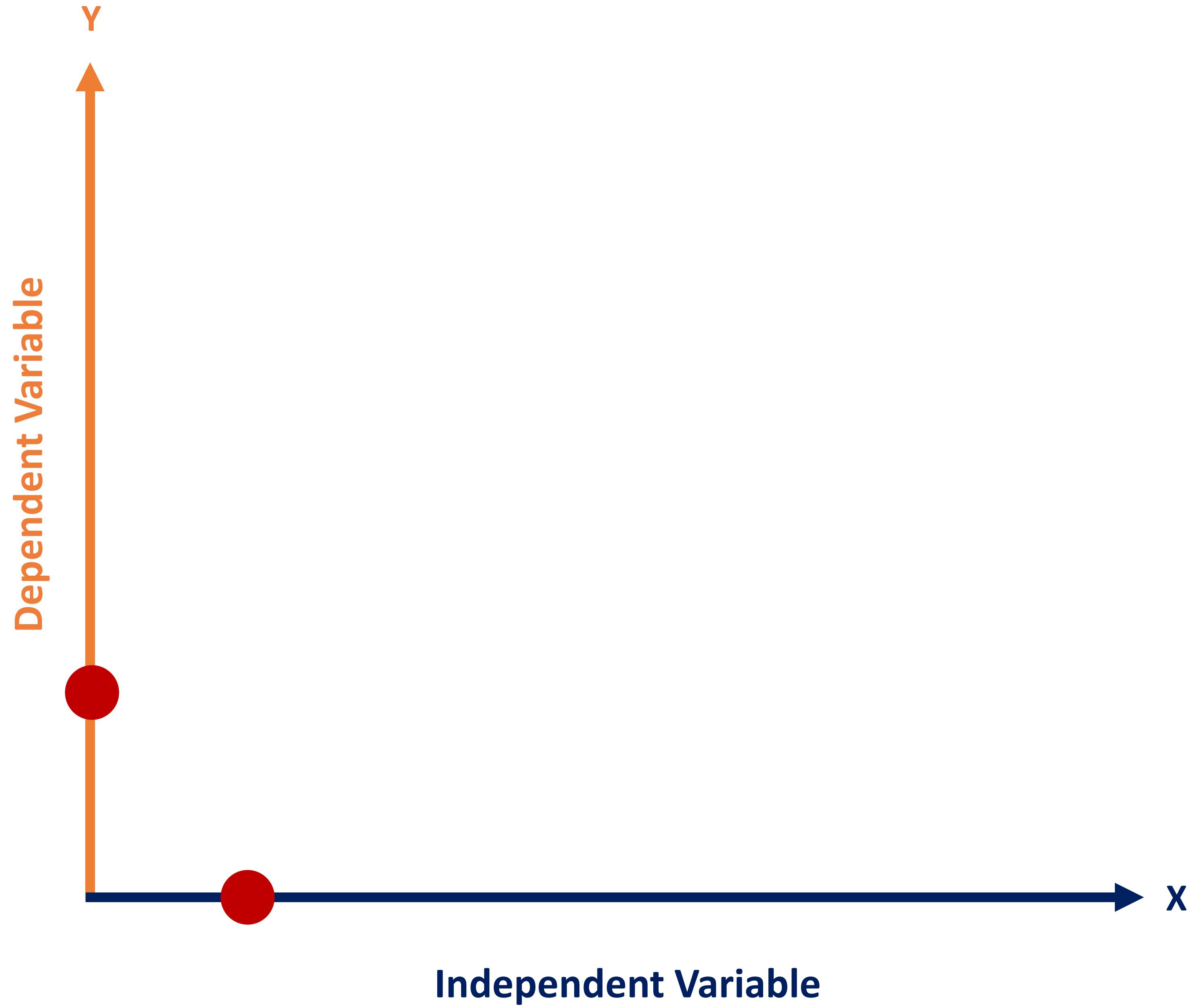
$$E(y) = \beta_0 + \beta_1 x$$

→ mean value of y not actual value

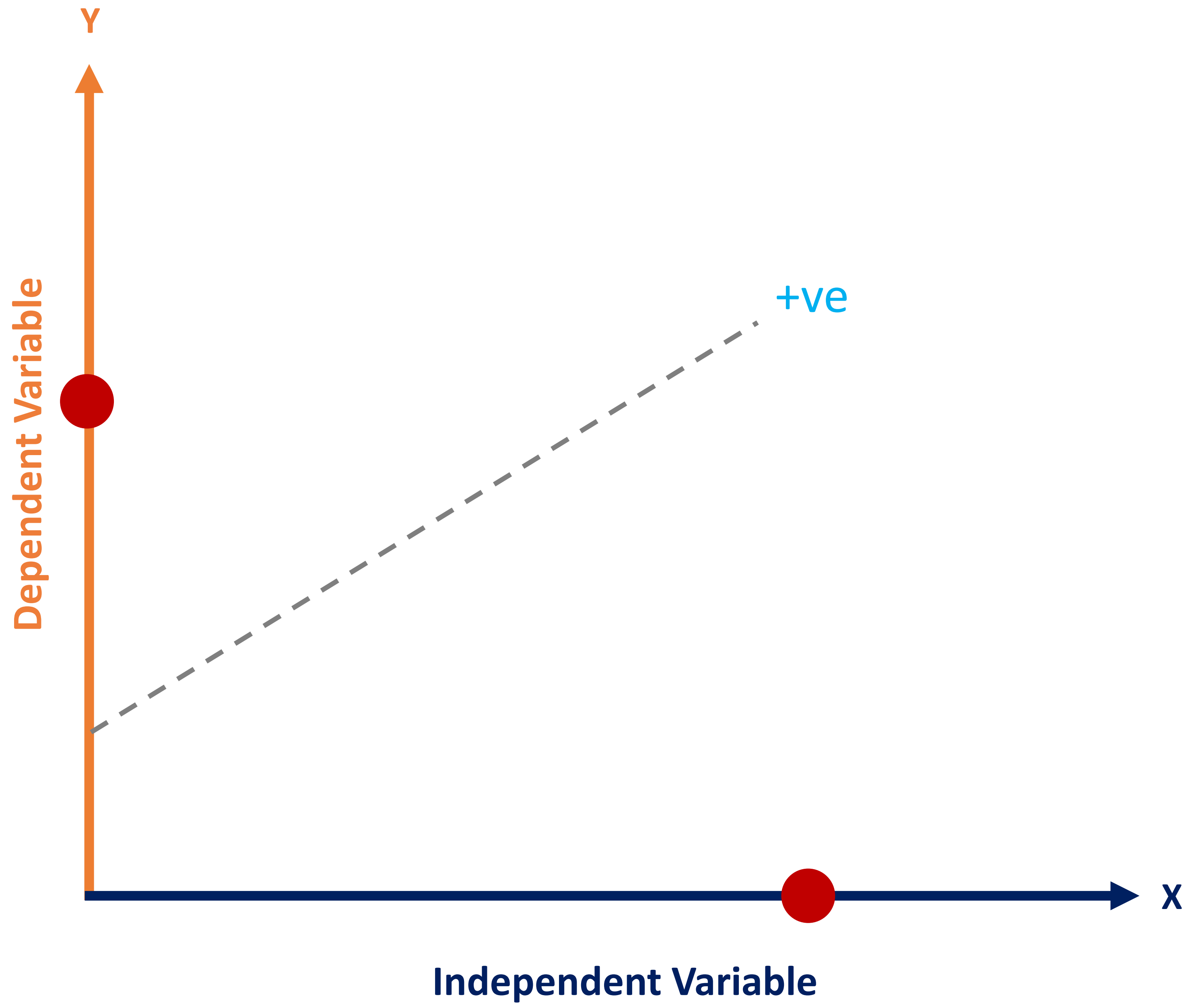
- Graph of the regression equation is a straight line
- β_0 is the intercept of the regression line
- B_1 is the slope of the regression line
- $E(y)$ is the expected value of y for a given x value

EDU FABRICA

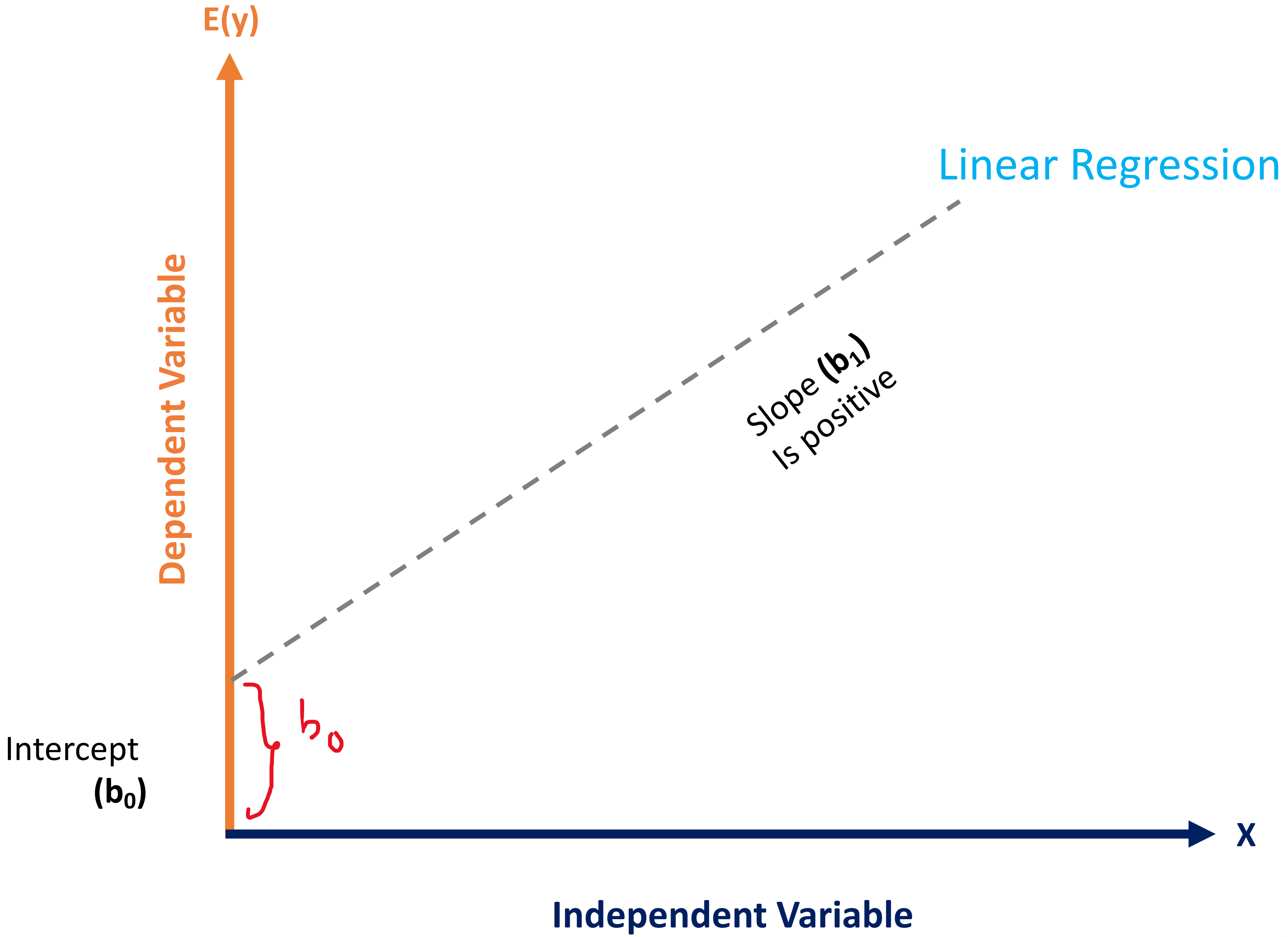
Simple Linear Regression Model



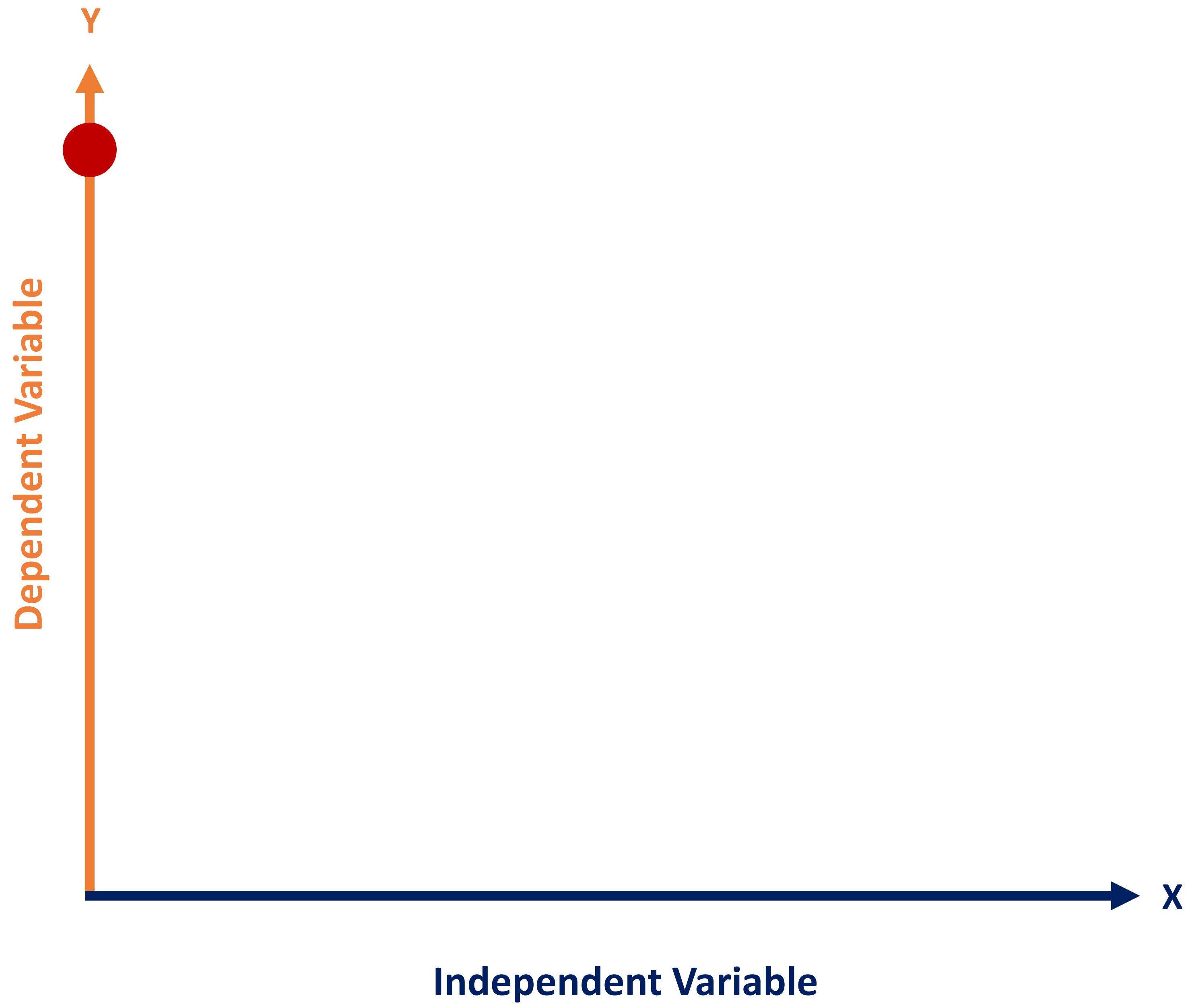
Understanding Linear Regression



Understanding Linear Regression

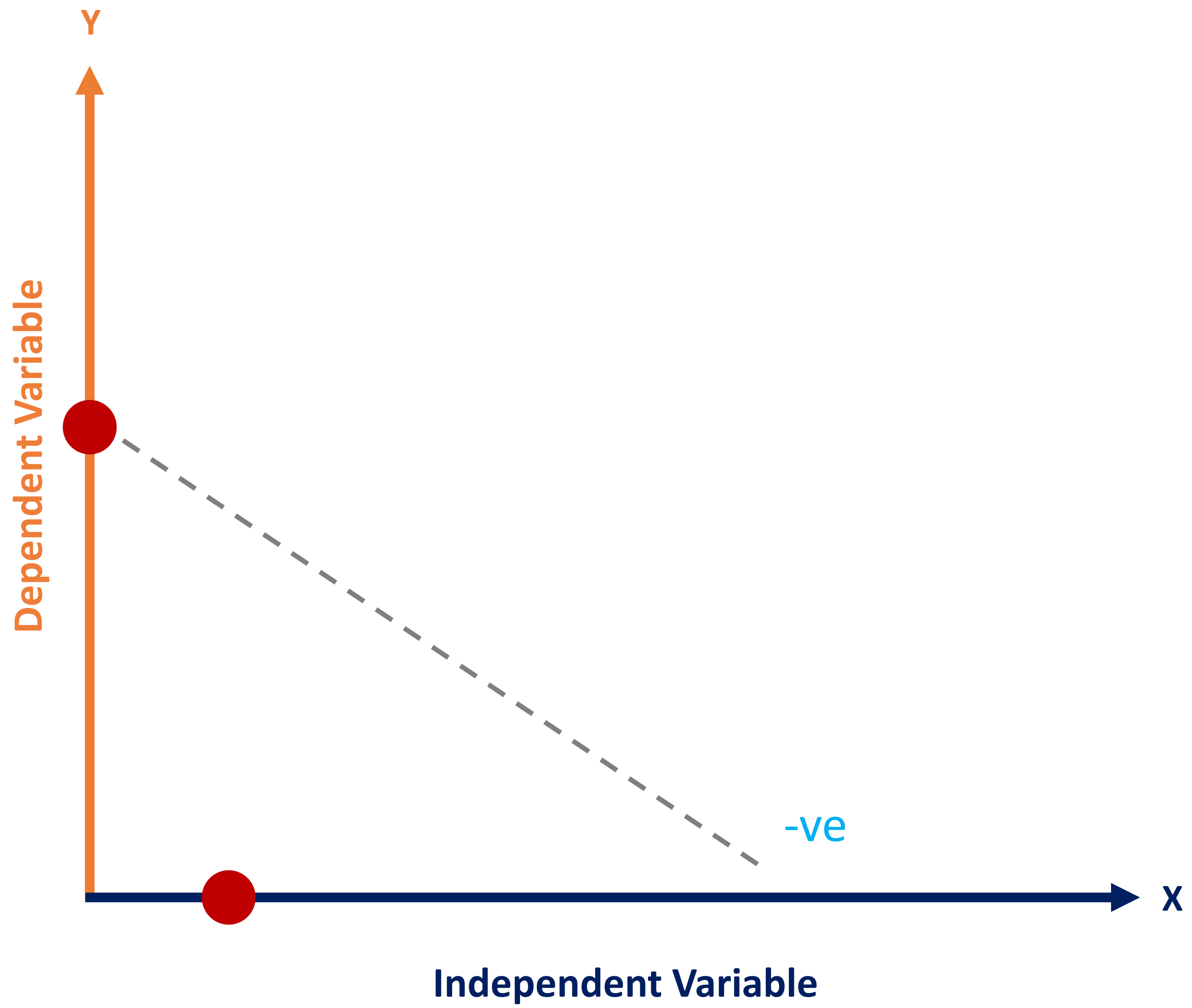


Understanding Linear Regression

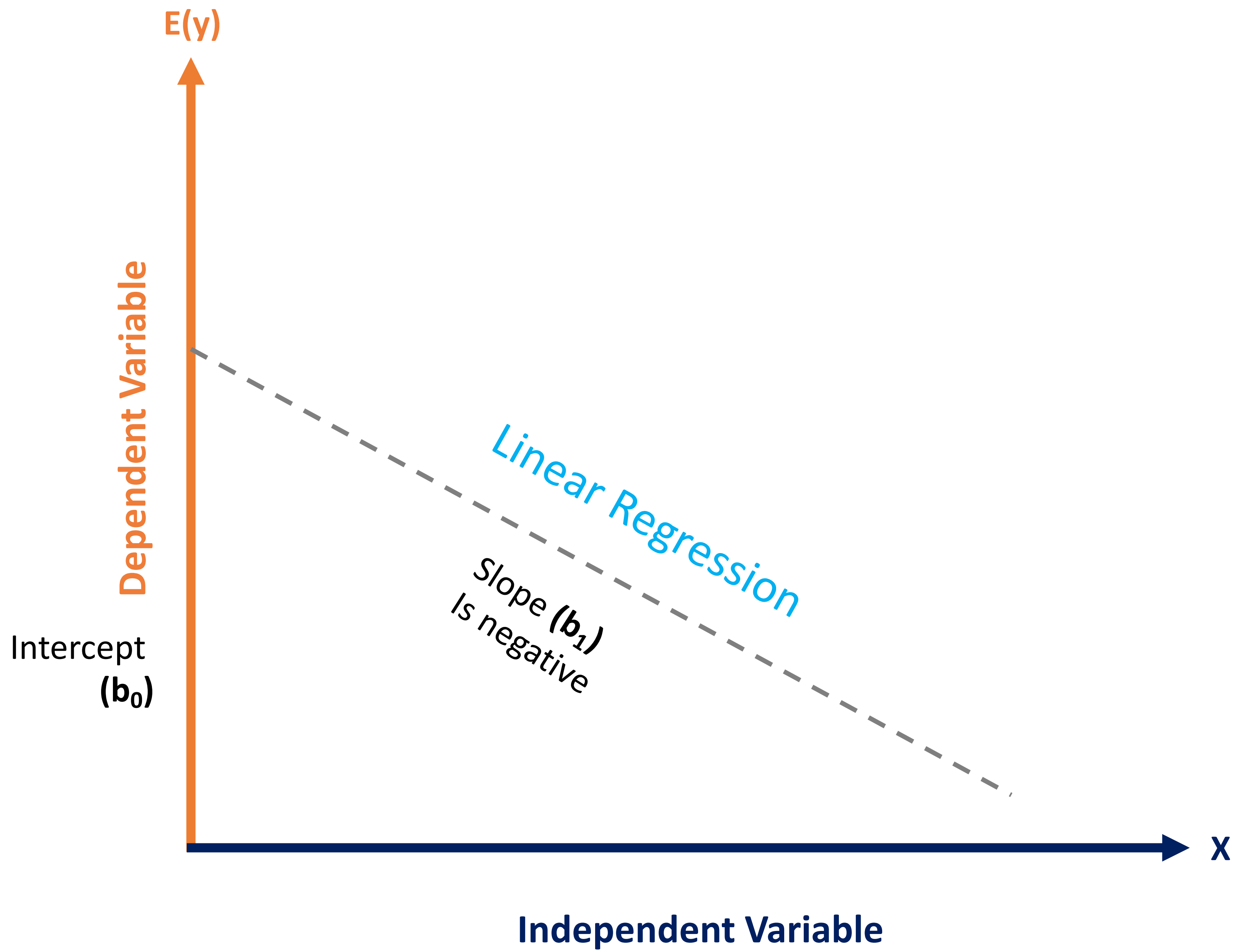


EDU FABRICA

Understanding Linear Regression



Understanding Linear Regression



EDU FABRICA

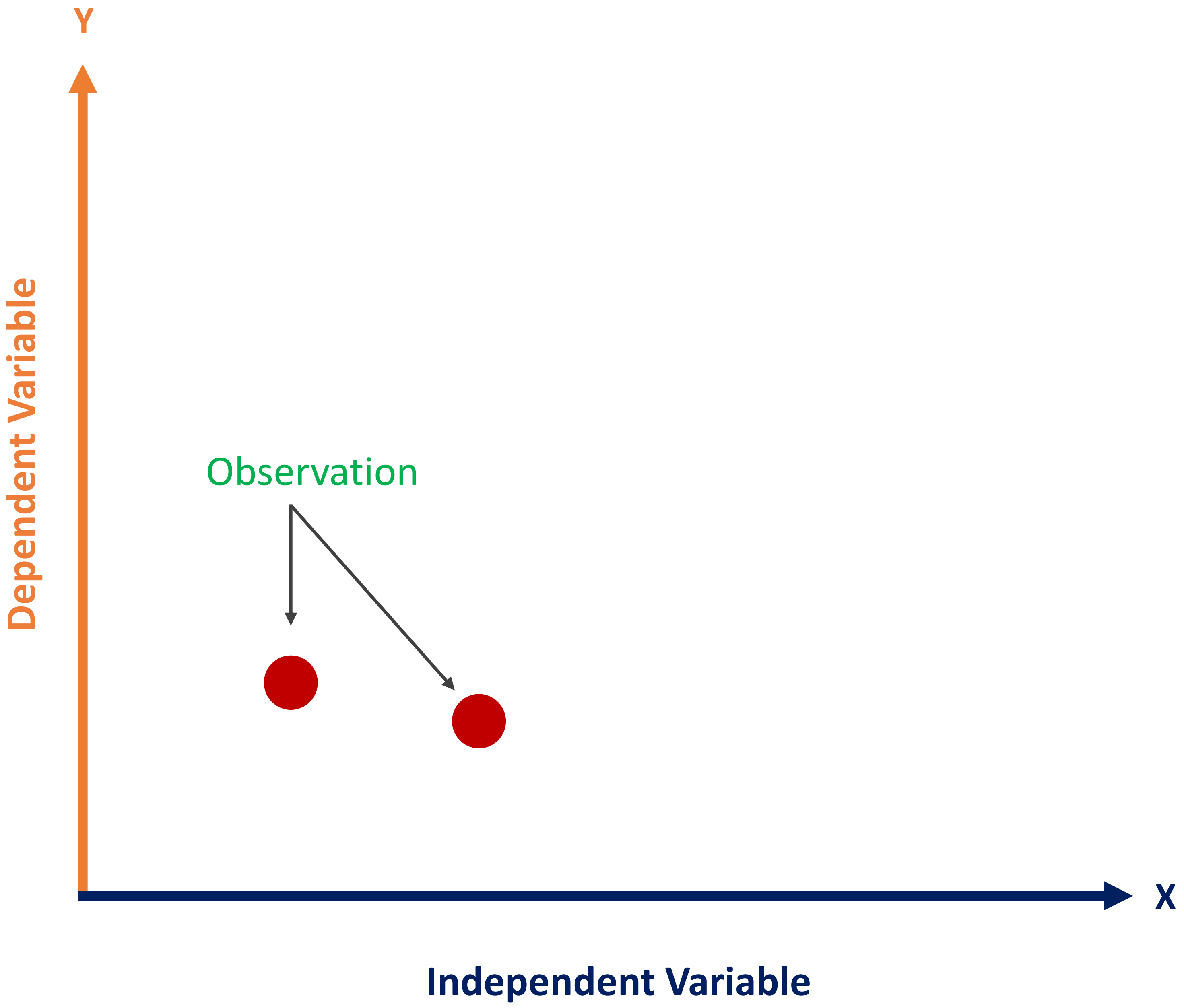
Estimated Simple Linear Regression Equation

- The estimated Simple Linear Regression equation is :

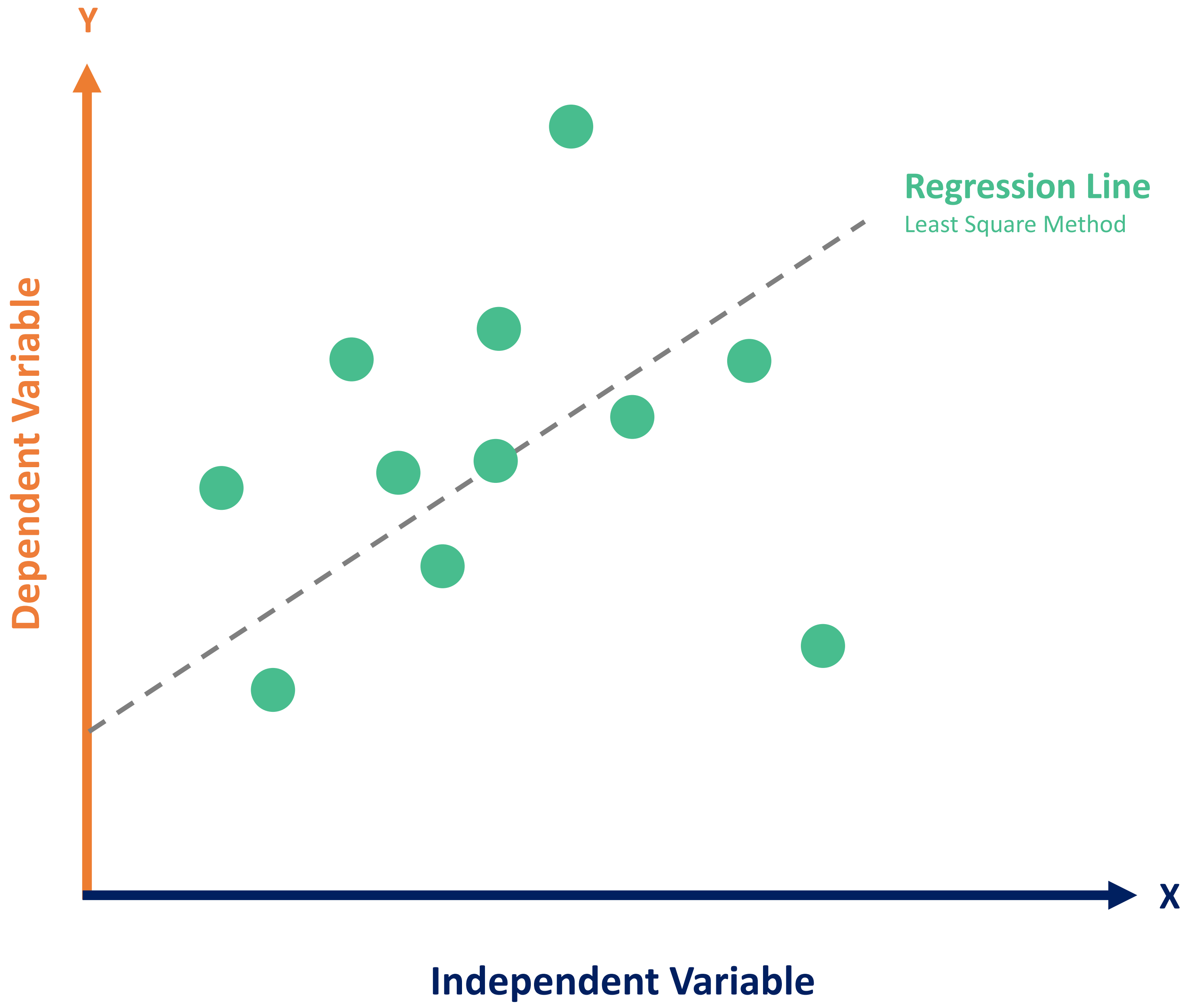
$$\hat{y} = b_0 + b_1 x$$

- The graph is called the estimated regression line
- b_0 is the intercept of the line
- b_1 is the slope of the line
- $\hat{y}$ is the estimated value of y for a given x value

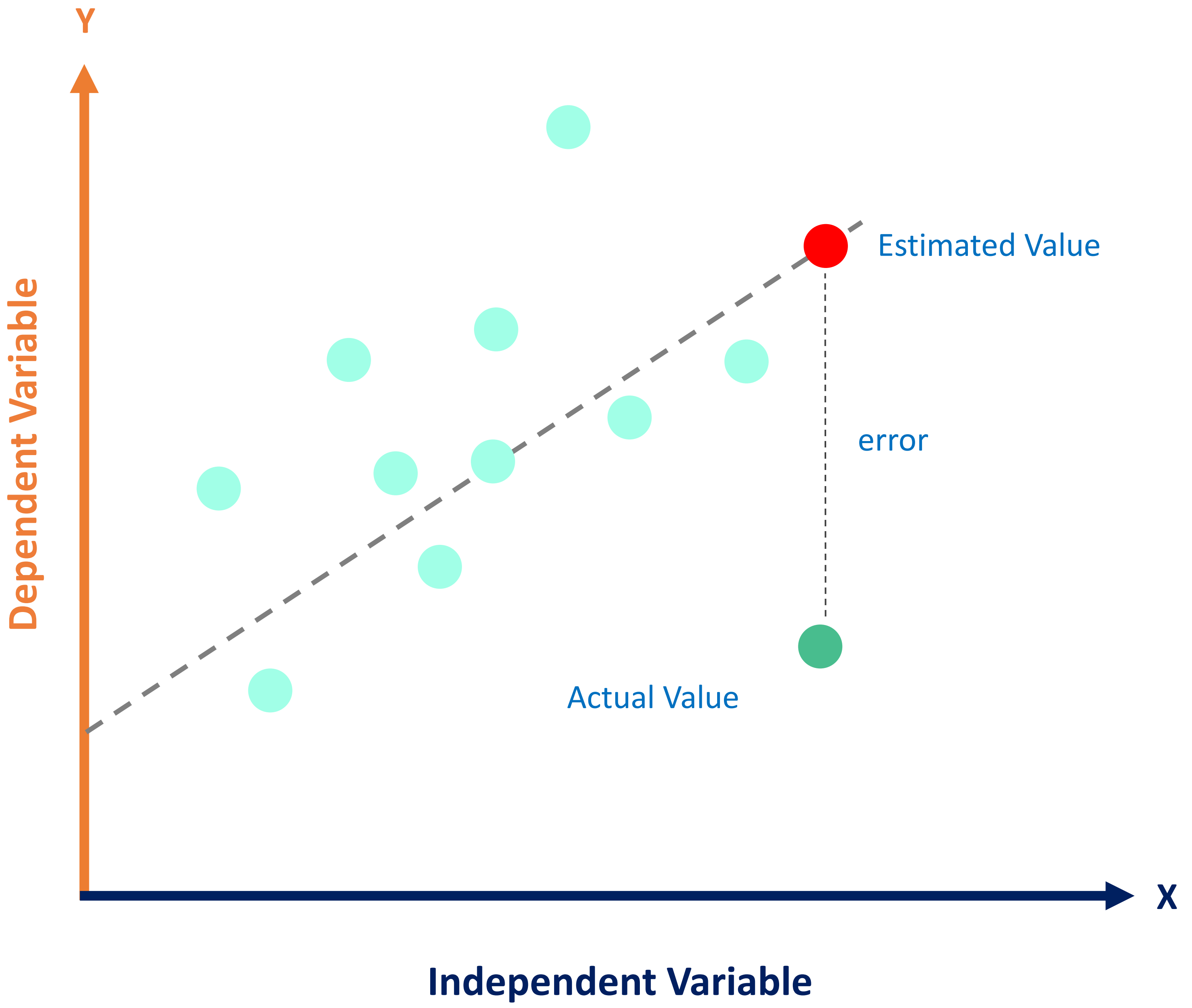
Understanding Linear Regression



Understanding Linear Regression

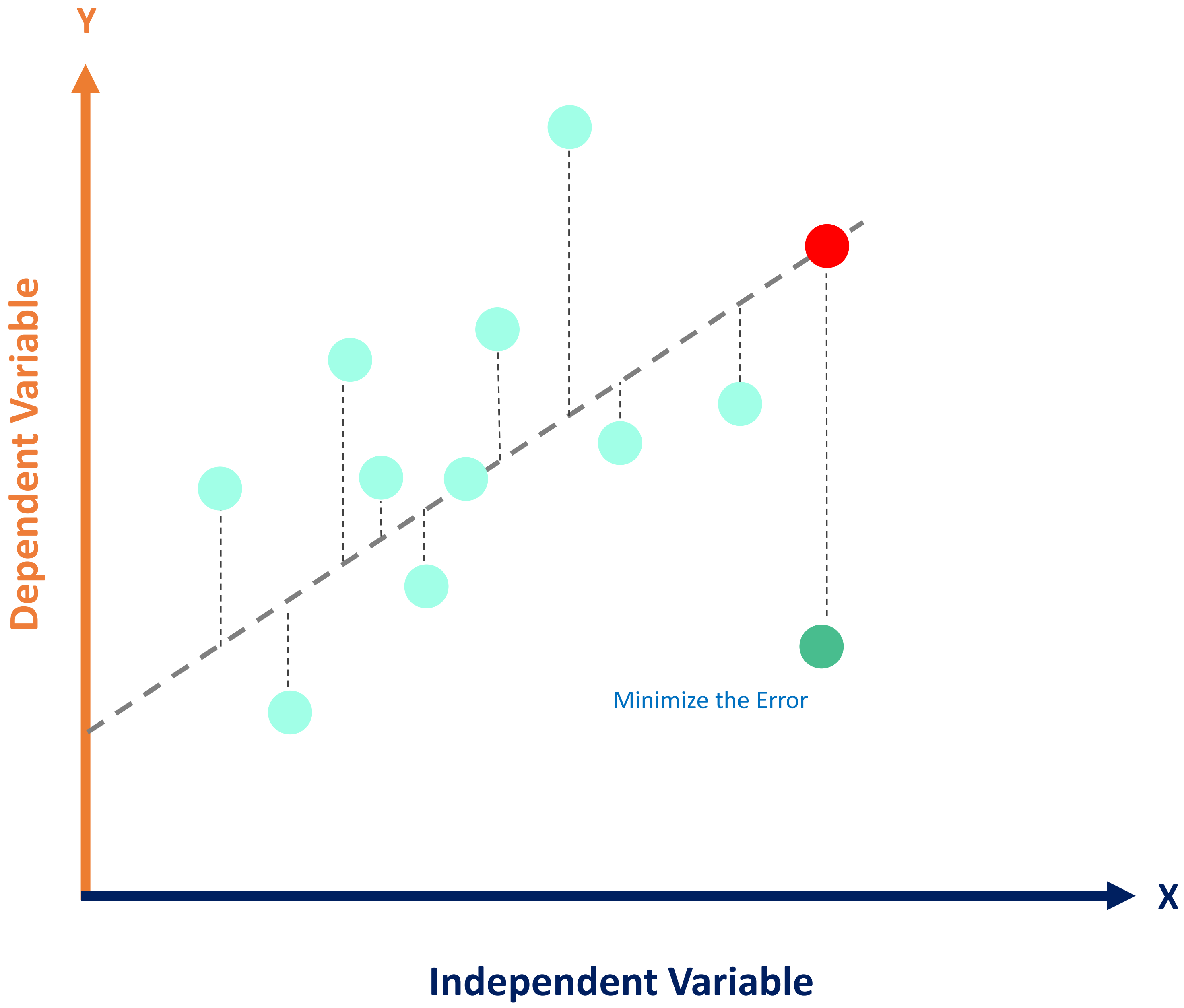


Understanding Linear Regression

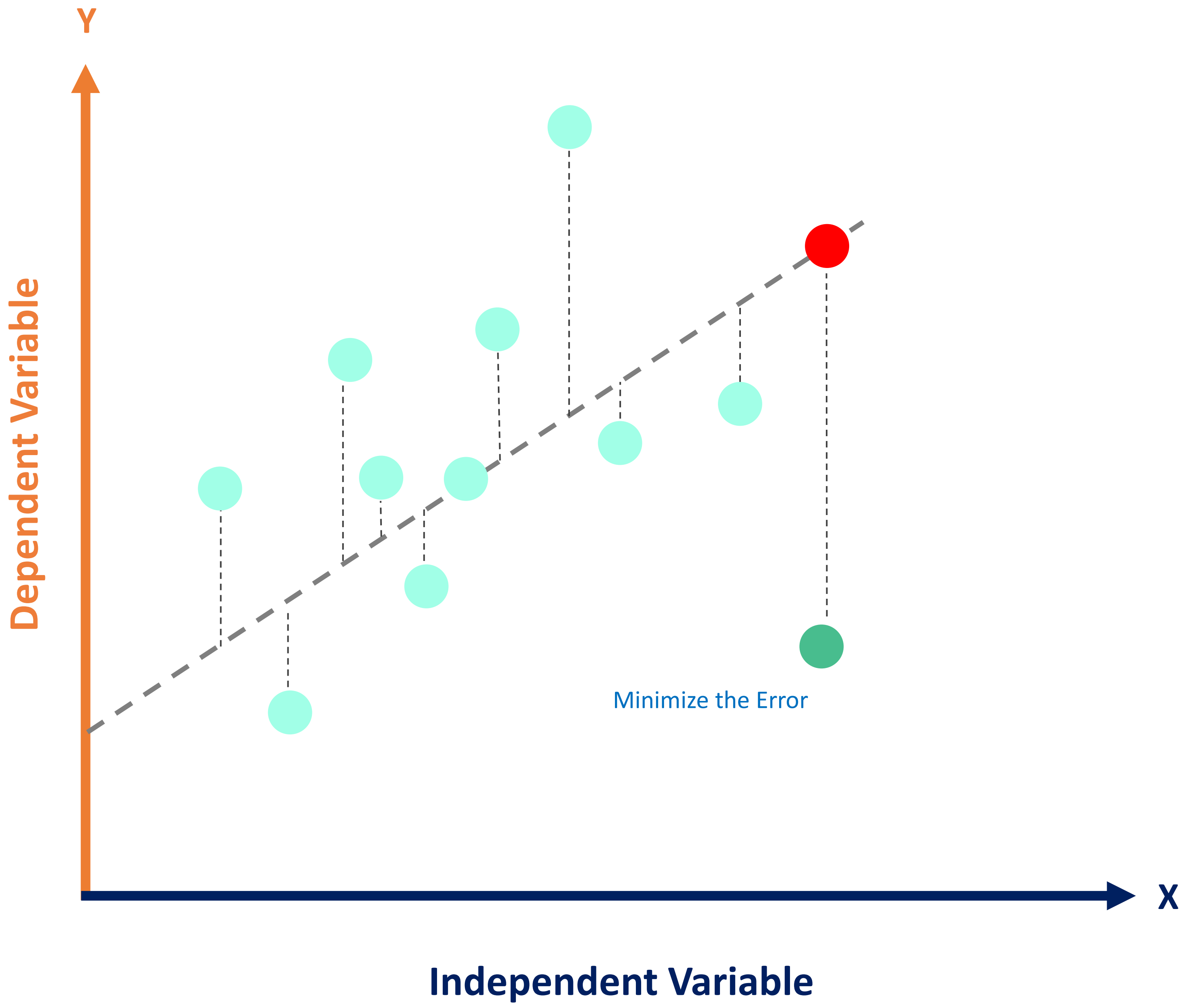


EDU FABRICA

Understanding Linear Regression



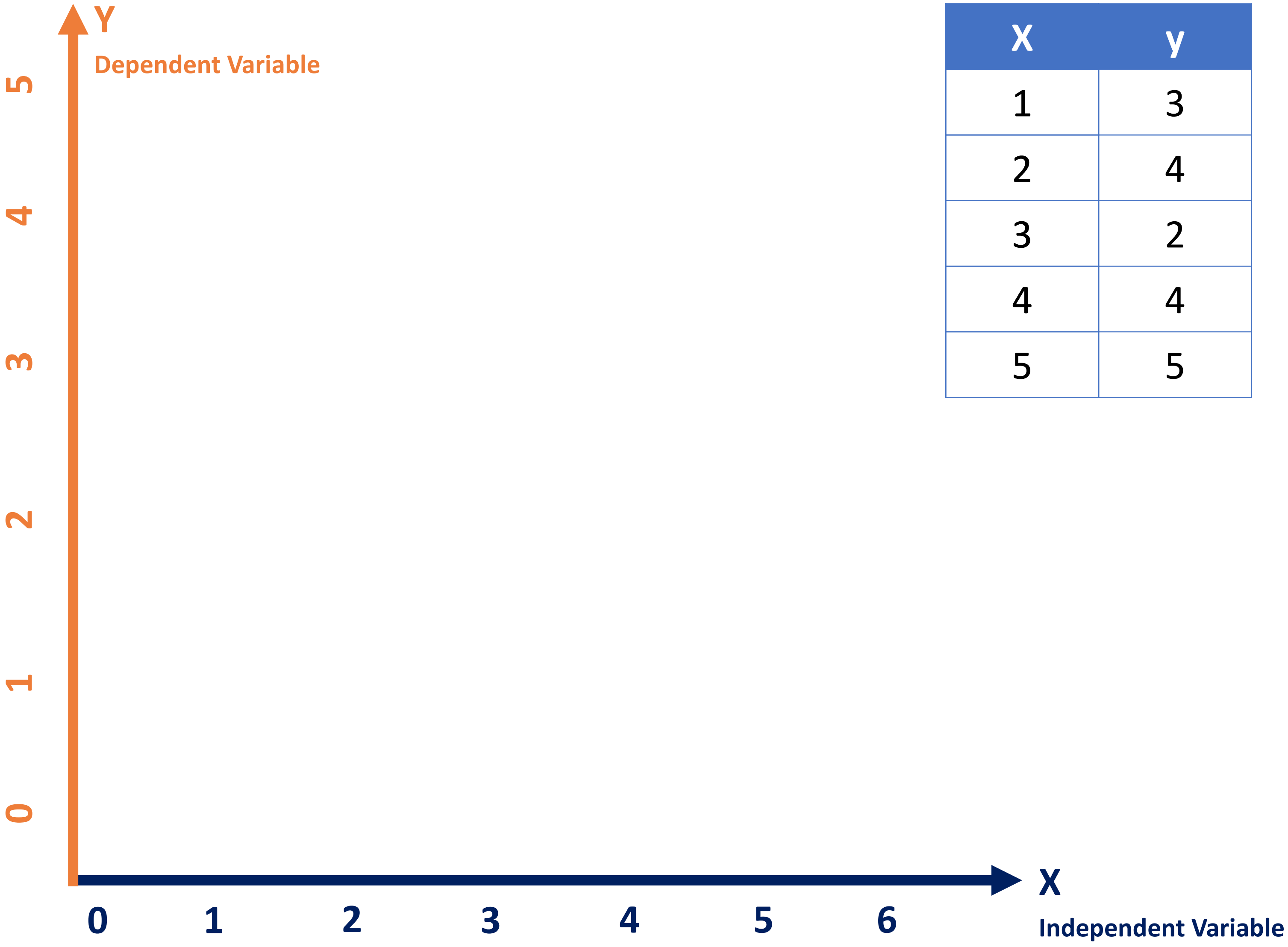
Understanding Linear Regression



Simple Linear Regression

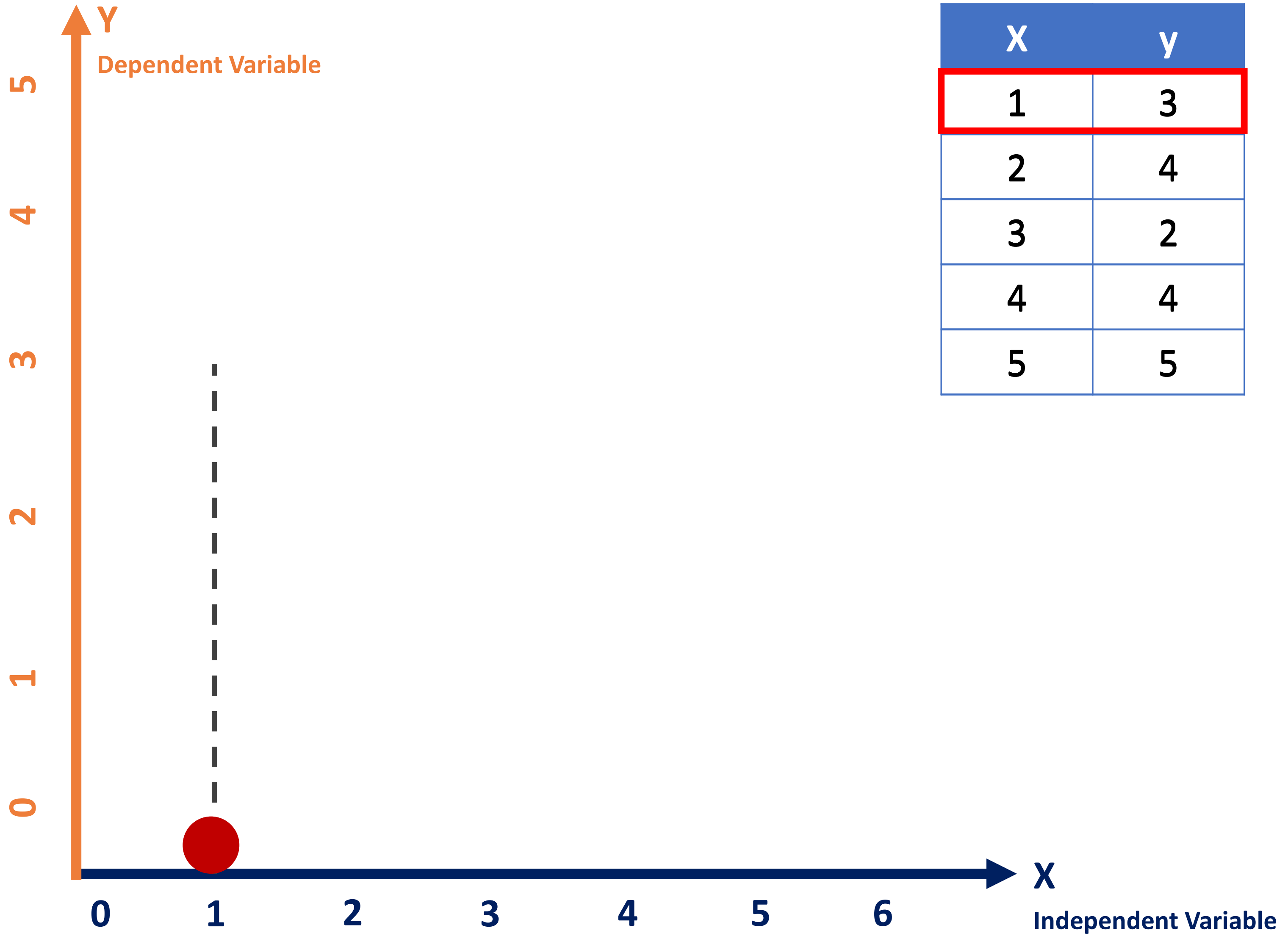
X
1
2
3
4
5

Simple Linear Regression



X	y
1	3
2	4
3	2
4	4
5	5

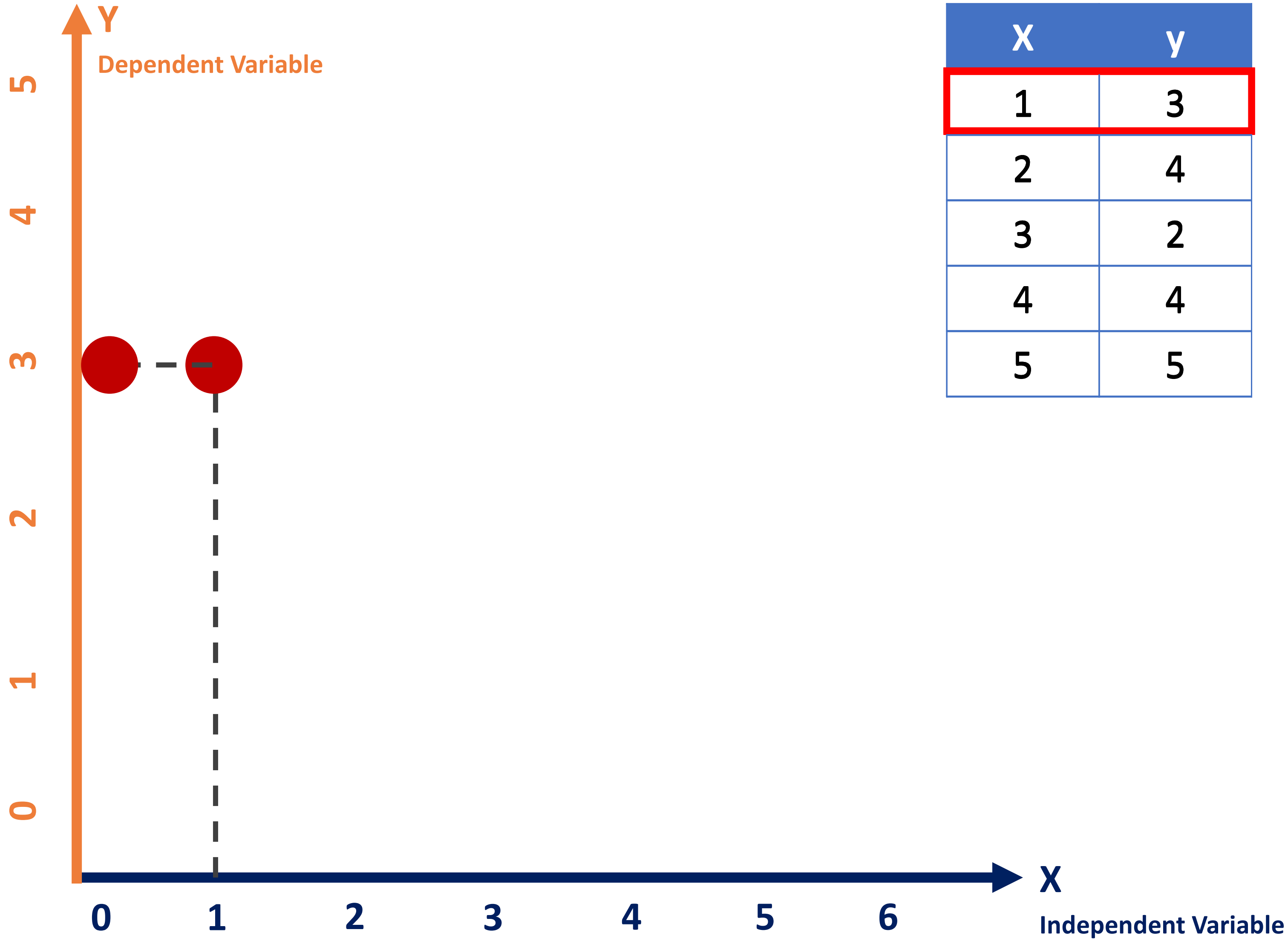
Simple Linear Regression



X	y
1	3
2	4
3	2
4	4
5	5

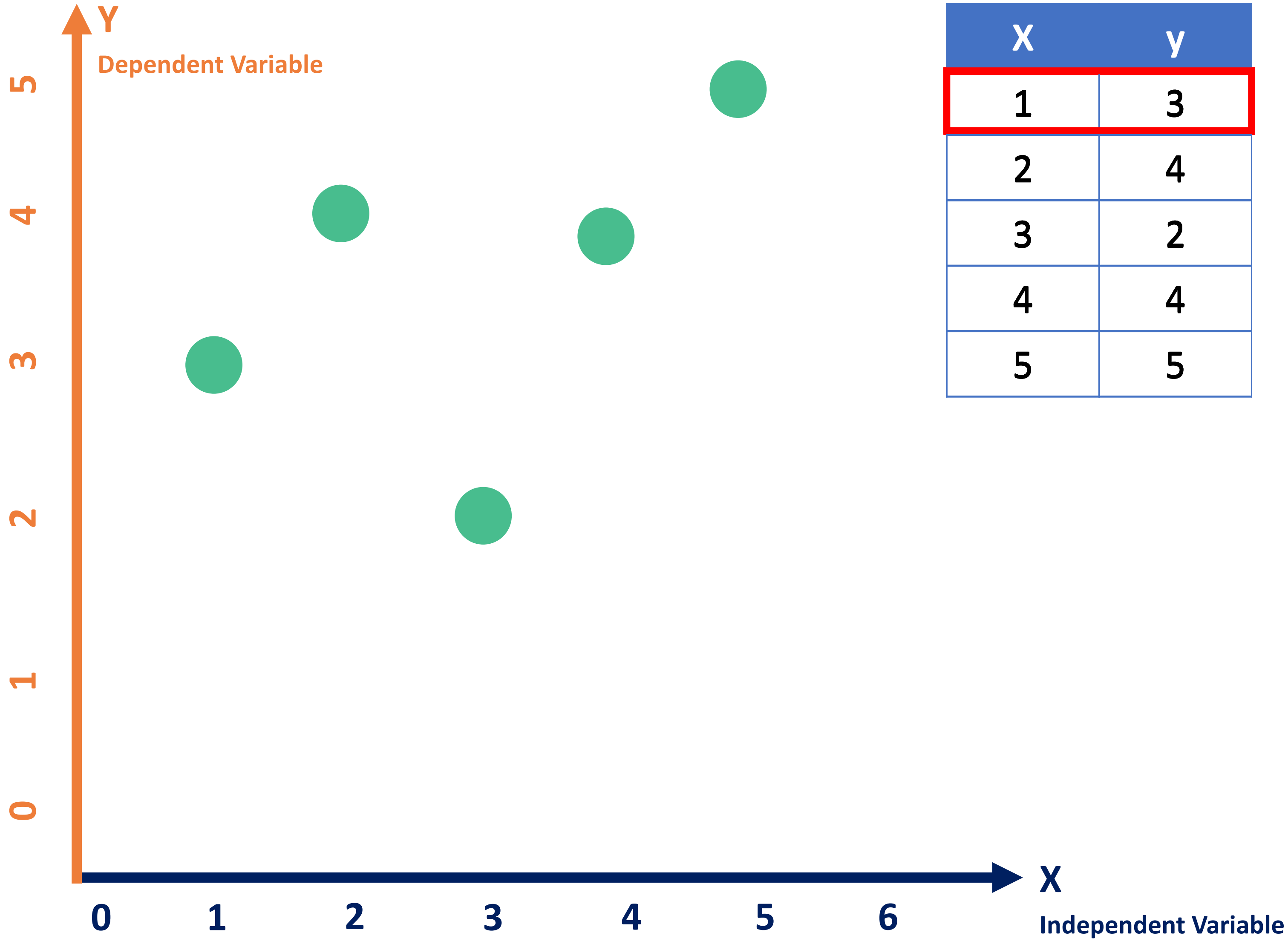
EDU FABRICA

Simple Linear Regression



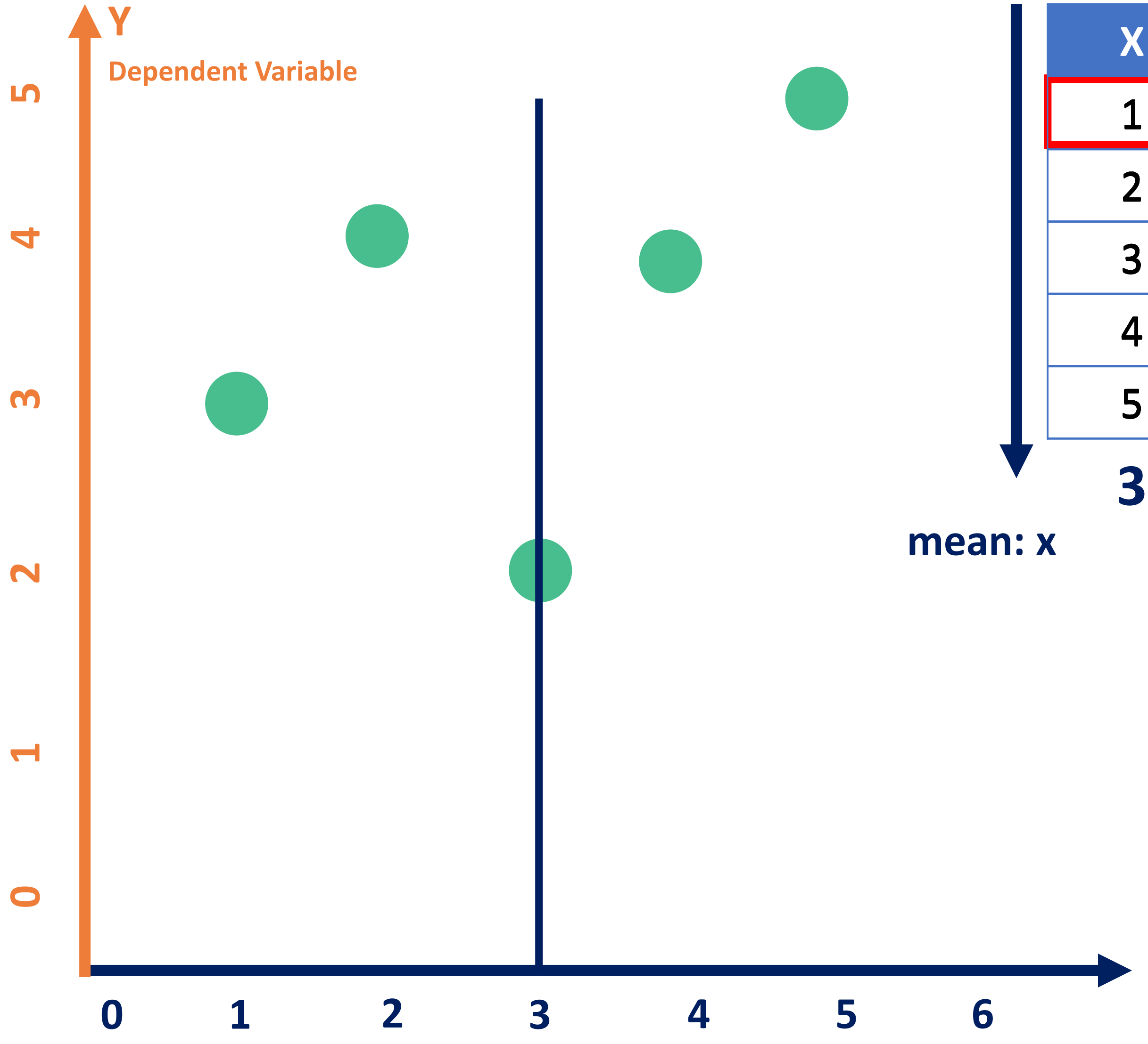
X	y
1	3
2	4
3	2
4	4
5	5

Simple Linear Regression



X	y
1	3
2	4
3	2
4	4
5	5

Simple Linear Regression

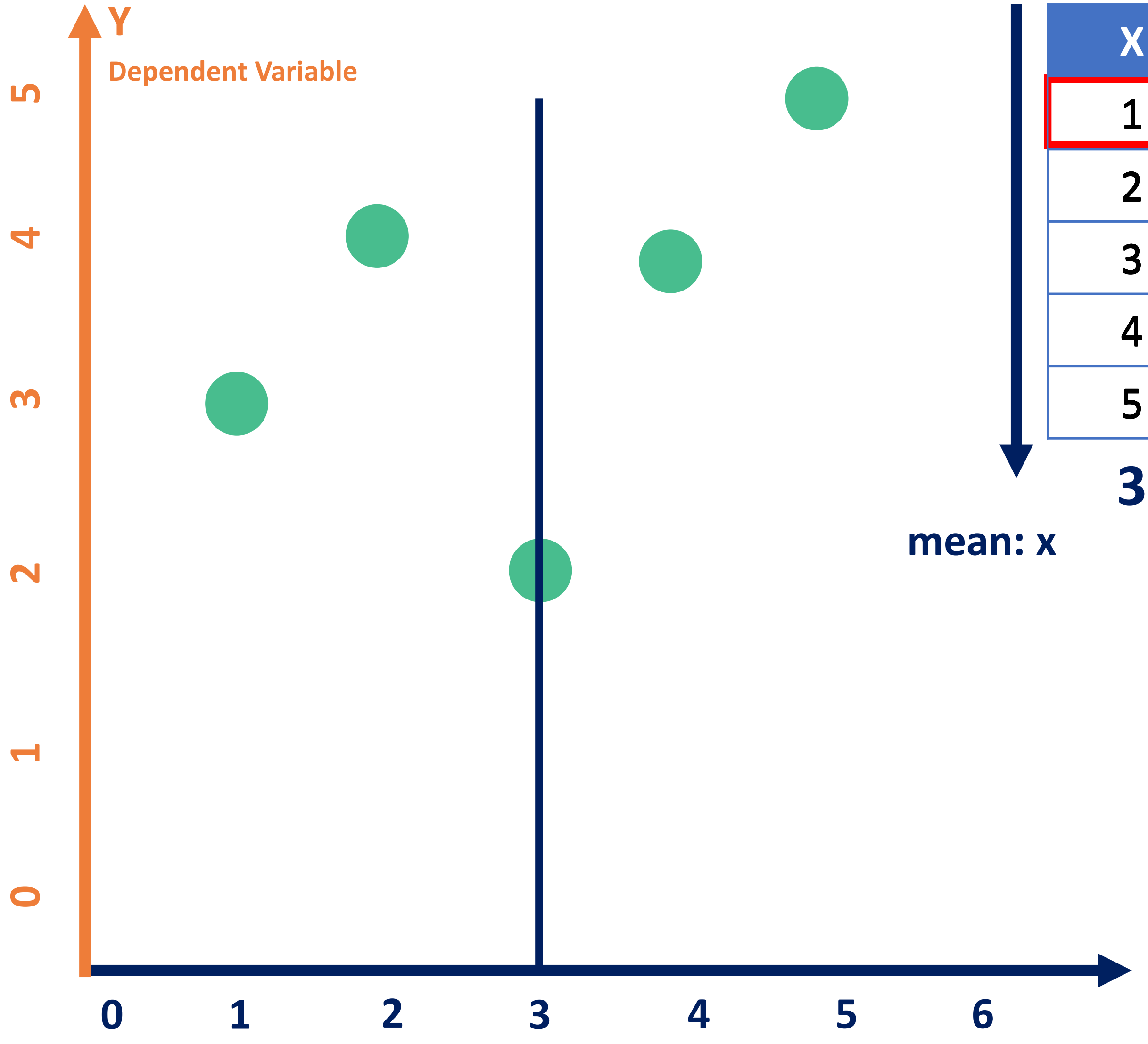


x	y
1	3
2	4
3	2
4	4
5	5

3

mean: x

Simple Linear Regression

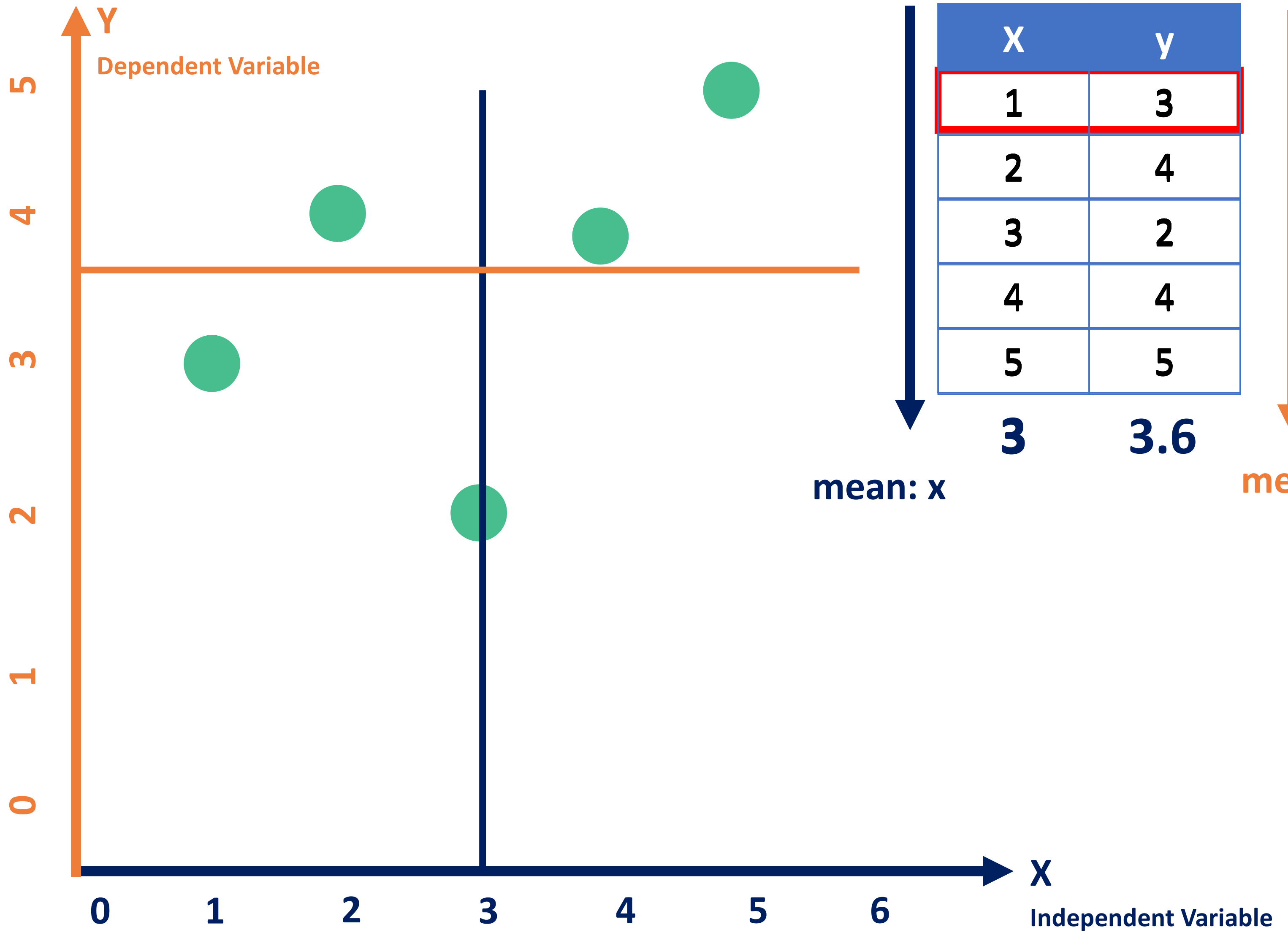


x	y
1	3
2	4
3	2
4	4
5	5

3

mean: x

Simple Linear Regression

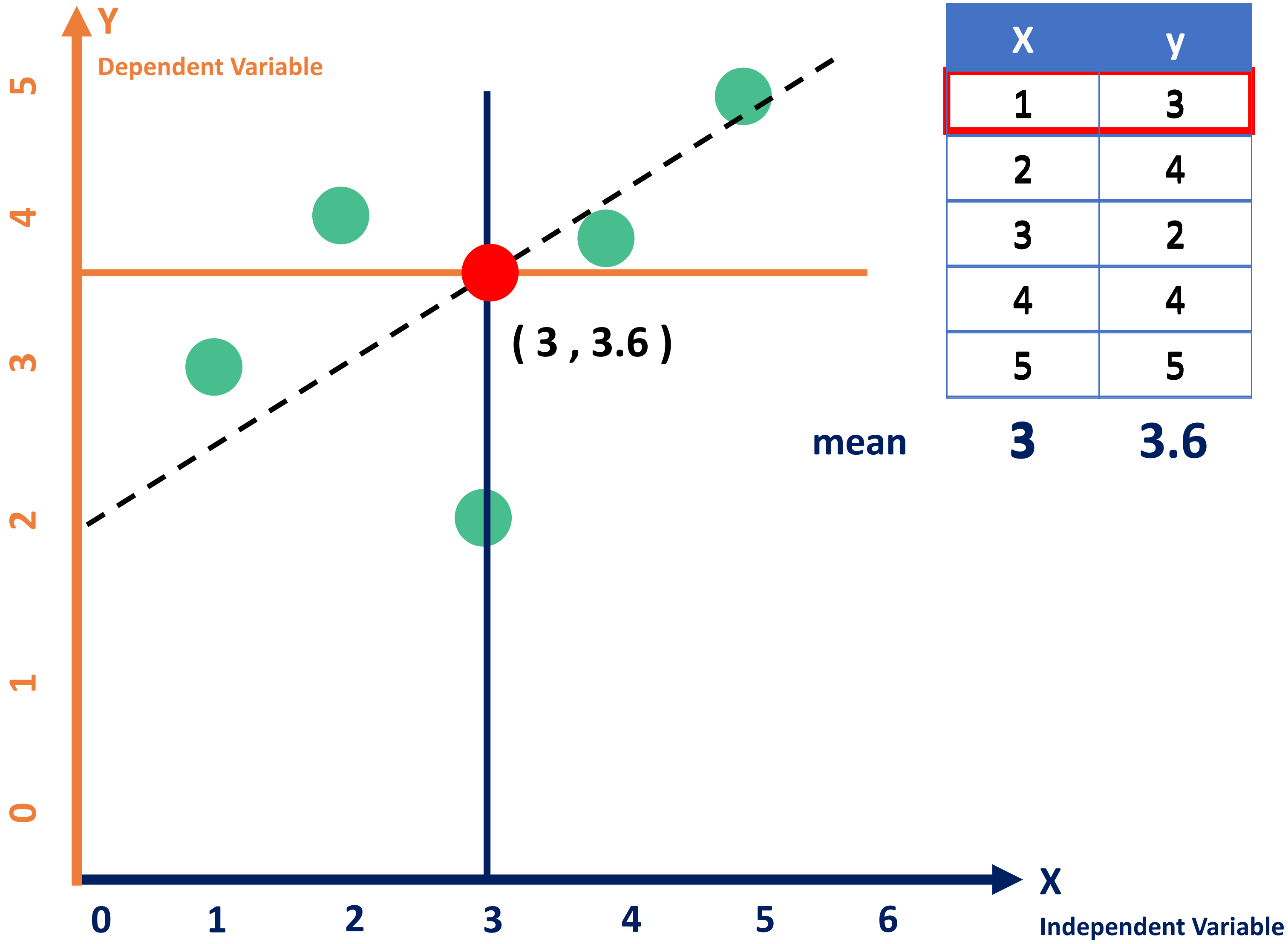


x	y
1	3
2	4
3	2
4	4
5	5
3	**3.6**

mean: x

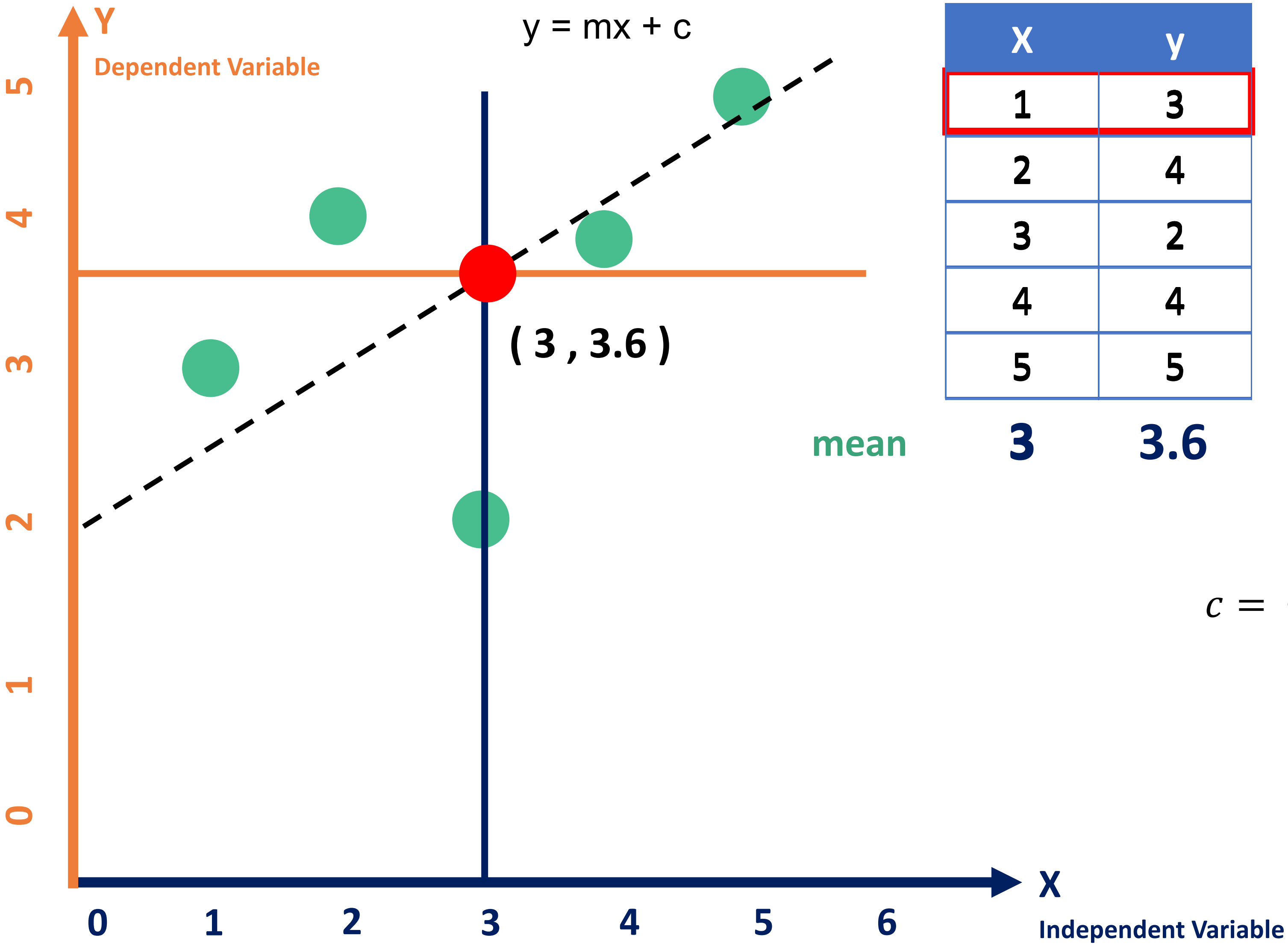
mean: y

Simple Linear Regression



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

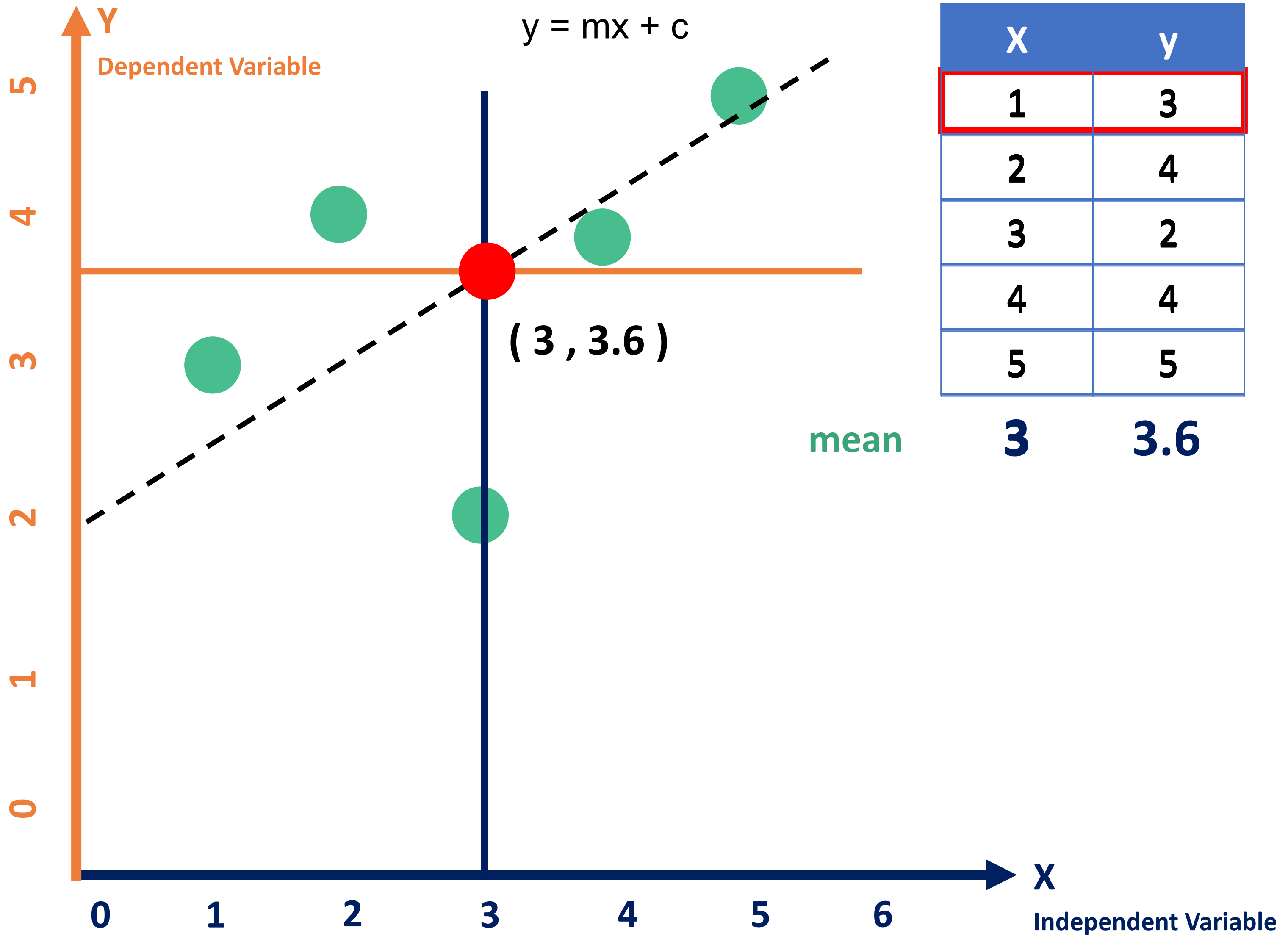
Simple Linear Regression



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

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Simple Linear Regression



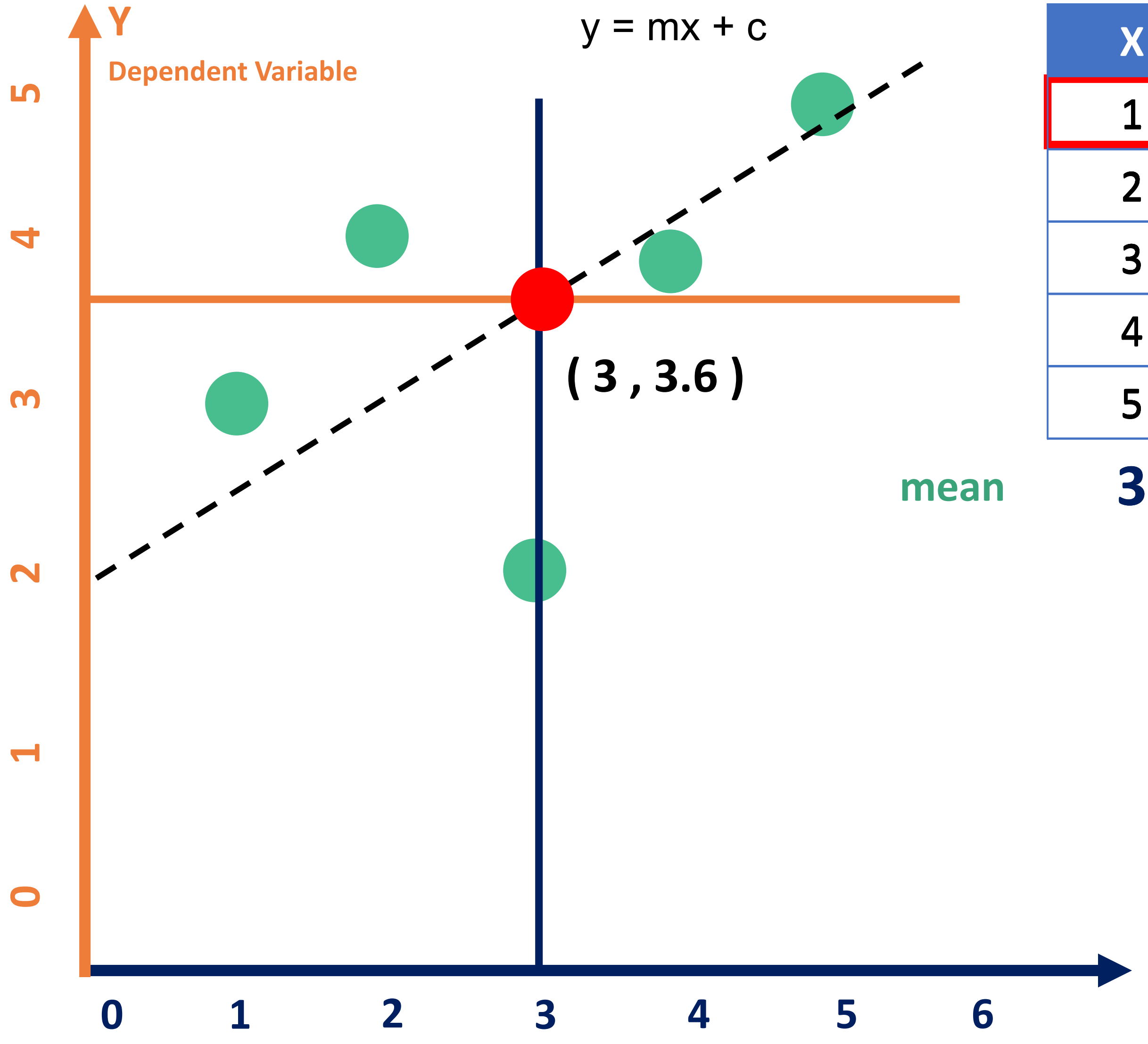
x	y
1	3
2	4
3	2
4	4
5	5
mean **3**	**3.6**

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σy) = 3 + 4 + 2 + 4 + 5 = **18**

EDU FABRICA

Simple Linear Regression



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

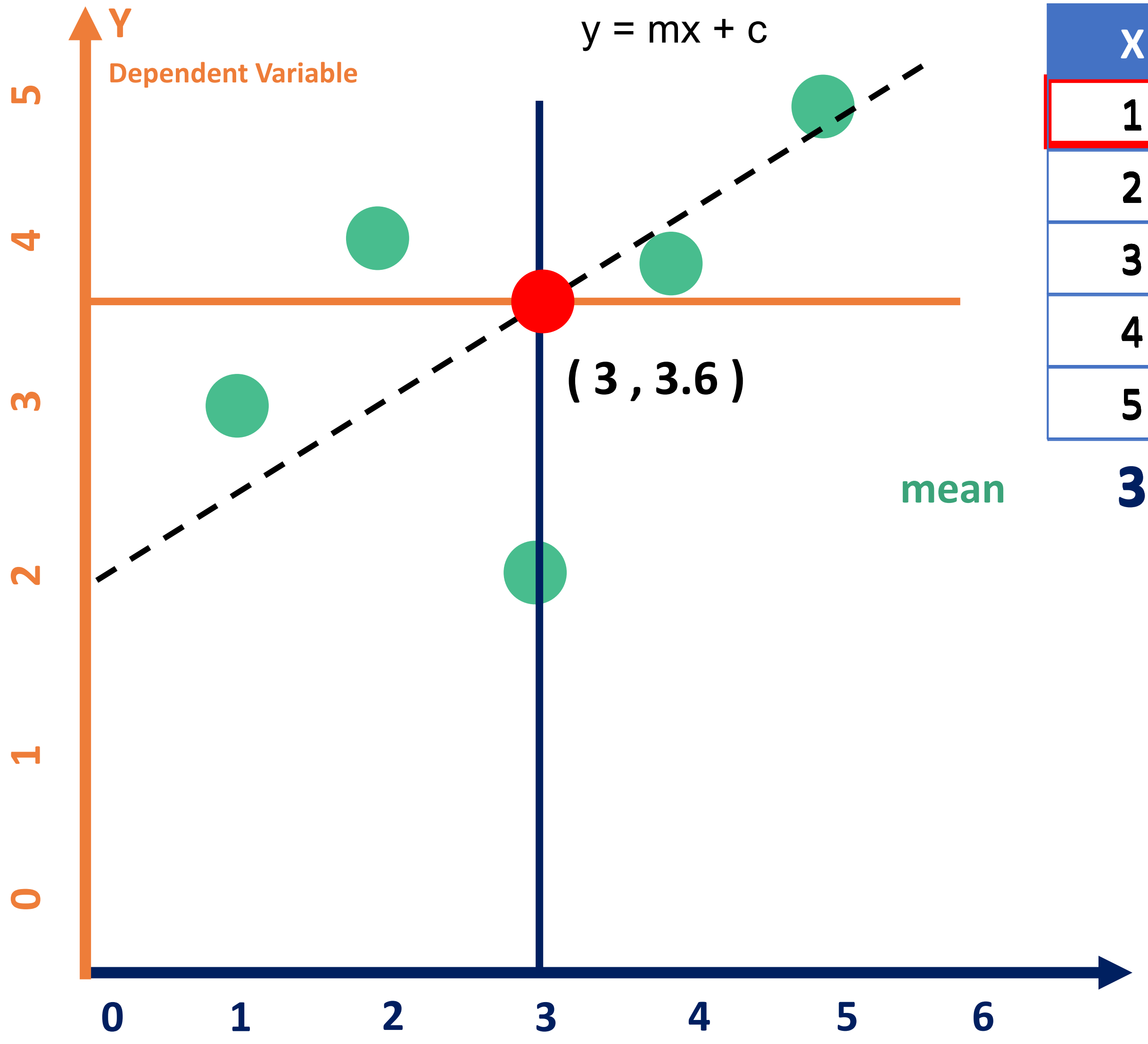
$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σy) = **18**

(Σx) = 1 + 2 + 3 + 4 + 5 = **15**

EDU FABRICA

Simple Linear Regression



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

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

Simple Linear Regression

X	y
1	3
2	4
3	2
4	4
5	5

3

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

y = mx + c

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15** (Σy) = **18**

$(\boldsymbol{\Sigma xy})$ = 3 + 8 + 6 + 16 + 25 = **58**

EDU FABRICA

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15** (Σy) = **18**

(Σxy) = **58**

(Σx^2) = 1 + 4 + 9 + 16 + 25 = **55**

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

(Σx) = **15**

(Σy) = **18**

(Σxy) = **58**

(Σx^2) = **55**

EDU FABRICA

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$c = \frac{(18)(55) - (15)(58)}{5(55) - (225)}$$

EDU FABRICA

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

$$c = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$c = \frac{(18)(55) - (15)(58)}{5(55) - (225)} = \frac{990 - 870}{275 - (225)}$$

$$c = \frac{120}{50} = 12 / 5 = \mathbf{2.4}$$

y = mx + c

$c = 2.4$

EDU FABRICA

Simple Linear Regression

x	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25

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

$$m = \frac{5(58) - (15)(18)}{5(55) - (15)^2}$$

$$m = \frac{290 - 270}{275 - 225}$$

y = mx + c

$$m = \frac{20}{70} = \frac{2}{7} = 0.28$$

Simple Linear Regression

X	y	xy	x^2
1	3	3	1
2	4	8	4
3	2	6	9
4	4	16	16
5	5	25	25
6	4	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

y = mx + c ;

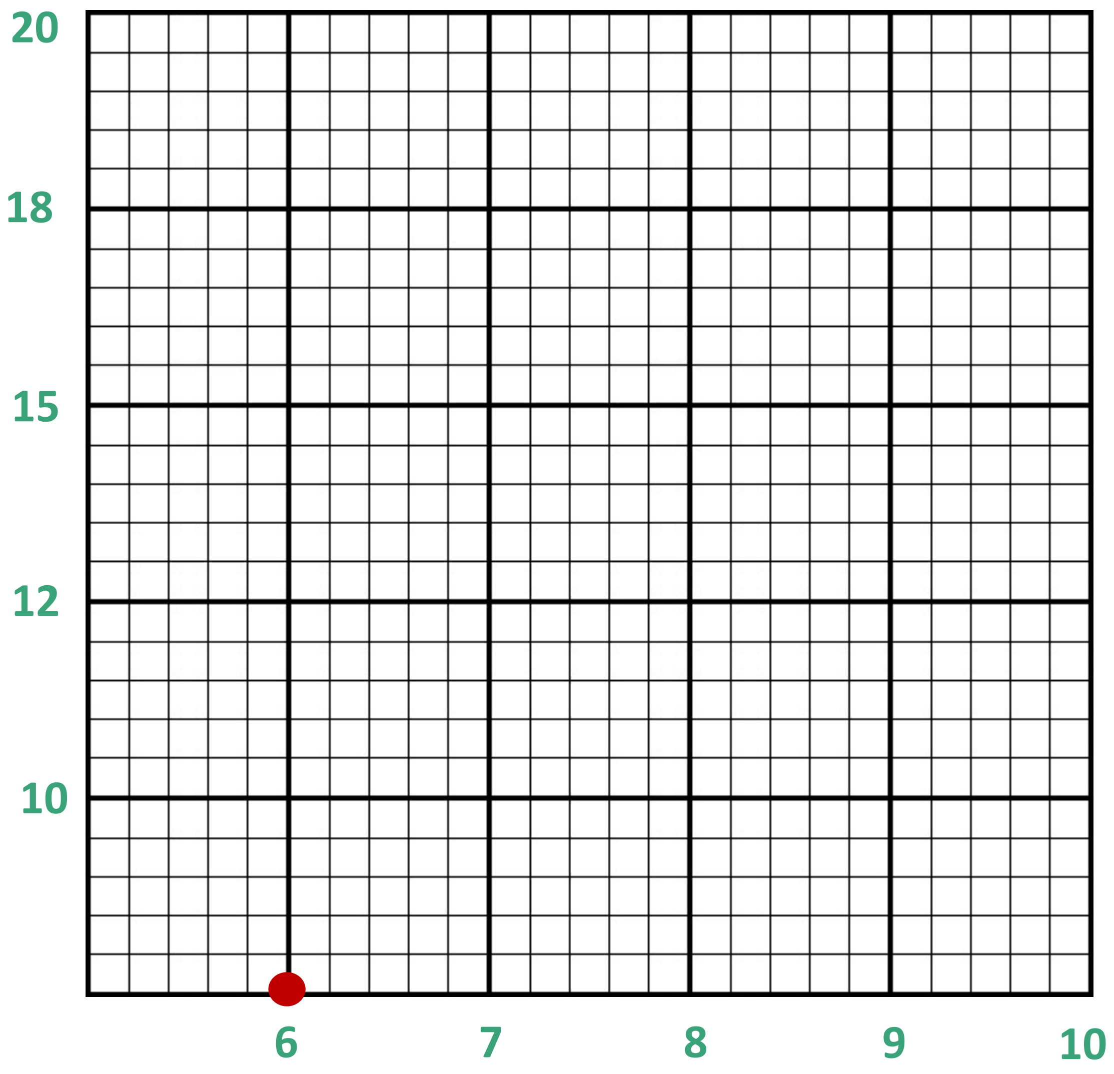
m = 0.28 ; c = 2.4

y = 0.28x + 2.4

Let's calculate y for

X = {6,7,8,9,10}

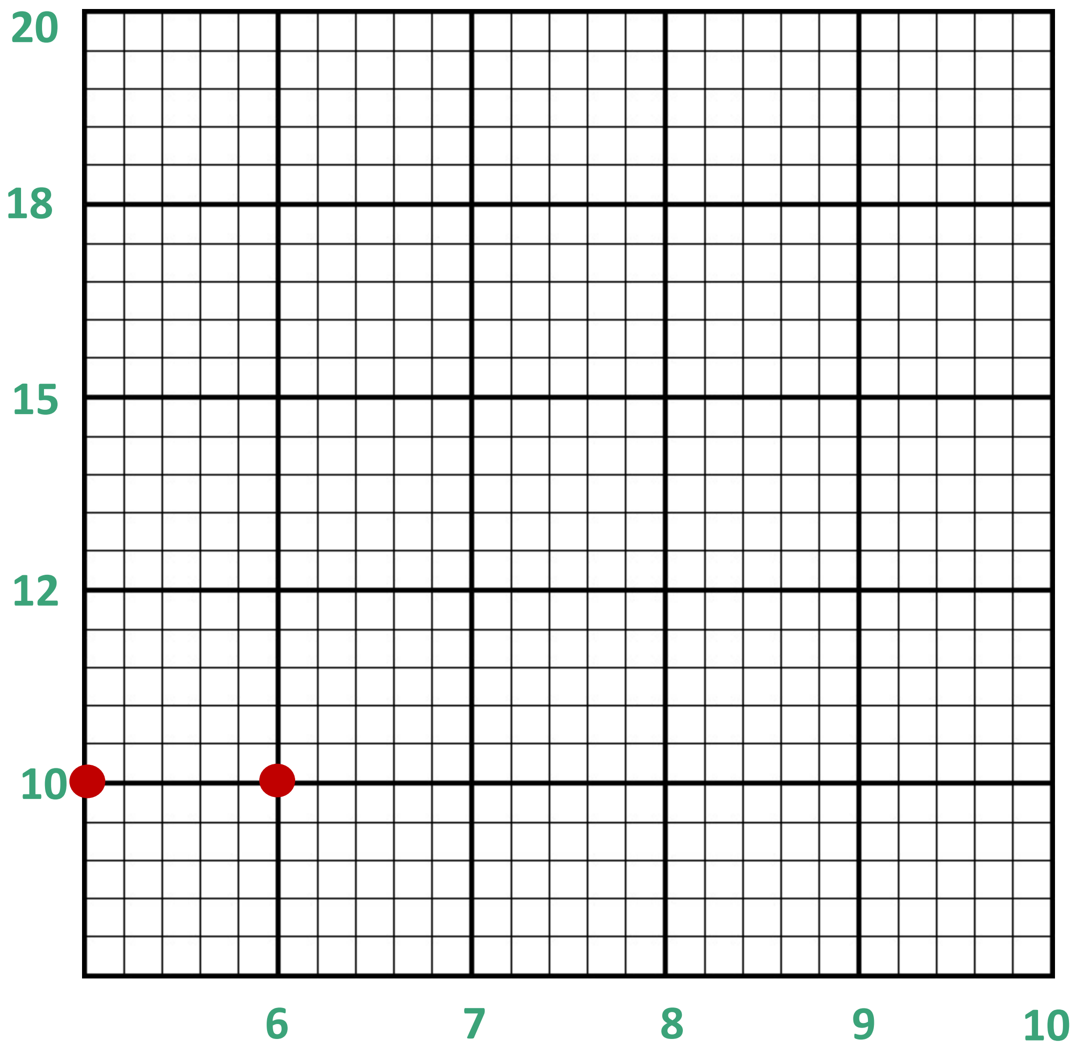
Simple Linear Regression



X	y	xy	x^2
6	10	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

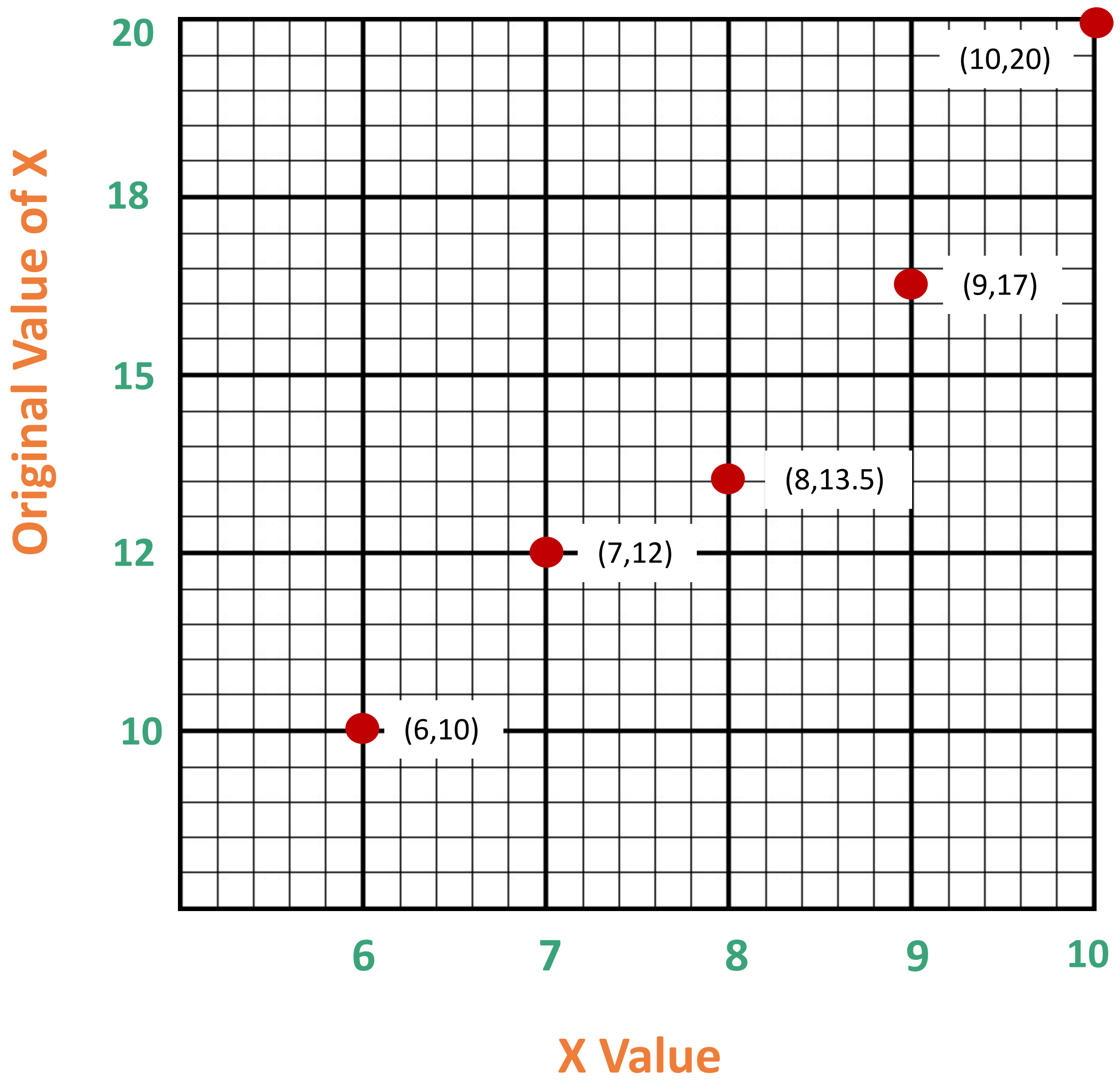
EDU FABRICA

Simple Linear Regression



X	y	xy	x^2
6	10	24	36
7	12	84	49
8	15	120	64
9	18	162	81
10	20	200	100

Simple Linear Regression



X	y	xy	x^2
6	10	24	36
7	12	84	49
8	13.5	120	64
9	17	162	81
10	20	200	100