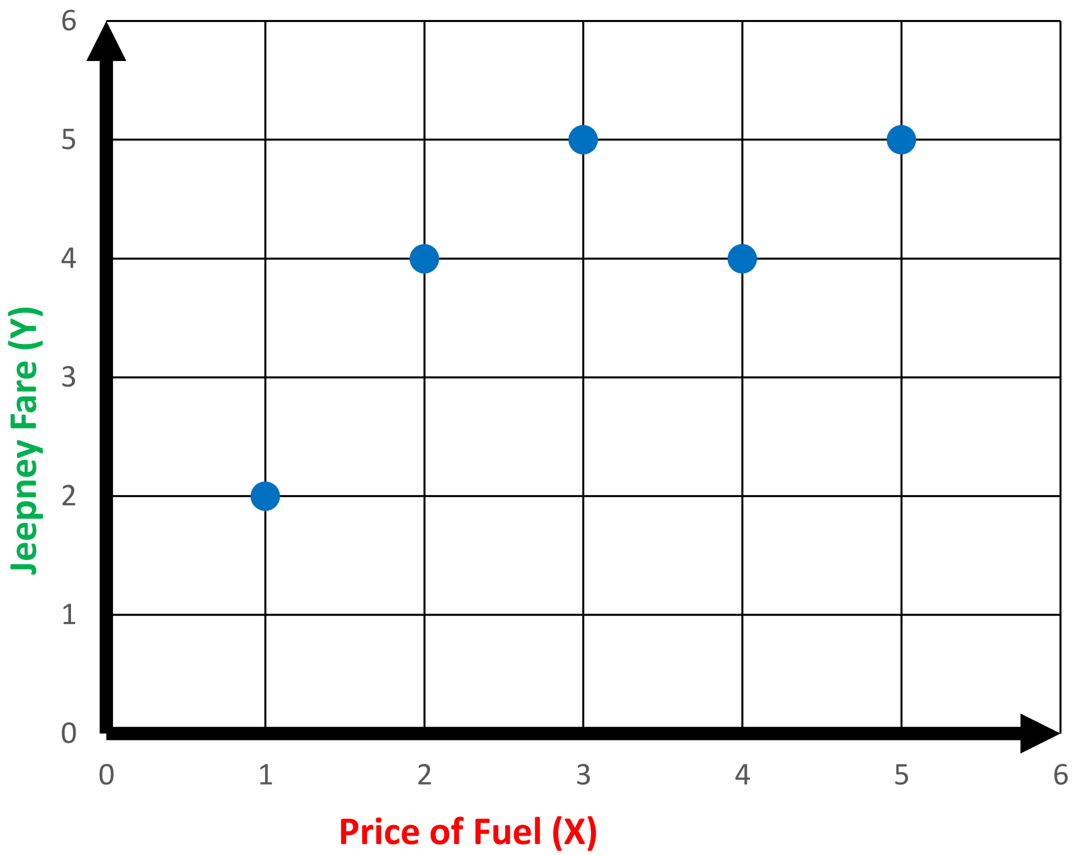
Intuition Behind the Regression Line

Price of Fuel (X)	Jeepney Fare (Y)
1	2
2	4
3	5
4	4
5	5



Intuition Behind the Regression Line

Price of Fuel (X)	Jeepney Fare (Y)	(X * Y)	X^2
1	2	2	1
2	4	8	4
3	5	15	9
4	4	16	16
5	5	25	25

$\Sigma x = 15$ $\Sigma y = 20$ $\Sigma xy = 66$ $\Sigma x^2 = 55$

Step 1: Get the sum of **X, Y, (X * Y)** and **X^2**

Regression Equation

The simplest form of a simple linear regression equation with one dependent and one independent variable is represented by:

$$\boxed{y = m(x) + b}$$

Where:

y is the value of the **dependent variable**

X is the value of the **independent variable**

m is the **slope** of the line

b is the **y-intercept**

Regression Equation

Calculating the **slope** is given by this formula:

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

Where:

m is the **slope** of the line

n is the total number of data points

x is the value of the **independent variable**

y is the value of the **dependent variable**

Regression Equation

Calculating the **intercept** is given by this formula:

$$\boxed{b = \frac{\Sigma y - m(\Sigma x)}{n}}$$

Where:

b is the **y-intercept**

m is the **slope** of the line

n is the total number of data points

x is the value of the **independent variable**

y is the value of the **dependent variable**

Regression Equation

Price of Fuel (X)	Jeepney Fare (Y)
1	1.5
2	3.8
3	6.7
4	9
5	11.2
6	13.6
7	16

$$\boxed{y = m(x) + b}$$

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

$$\boxed{b = \frac{\Sigma y - m(\Sigma x)}{n}}$$

Calculate the Slope

Σx	Σy	Σxy	Σx^2
15	20	66	55

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

$$m = \frac{5\,(66) - (15)\,(20)}{5\,(55) - (15)^2}$$

$$\boxed{m = 0.6}$$

Calculate the Intercept

Σx	Σy	Σxy	Σx^2
15	20	66	55

$$\boxed{b = \frac{\Sigma y - m(\Sigma x)}{n}}$$

$$b = \frac{20 - 0.6\,(15)}{5}$$

$$\boxed{b = 2.2}$$

Predicting the Jeepney Fare

Price of Fuel (X)	Jeepney Fare (Y)	Predicted Jeepney Fare ($Y_{predict}$)
1	2	2.8
2	4	3.4
3	5	4
4	4	4.6
5	5	5.2

m	b
0.6	2.2

$$\boxed{y = m(x) + b}$$

$$y = 0.6(2) + (2.2)$$

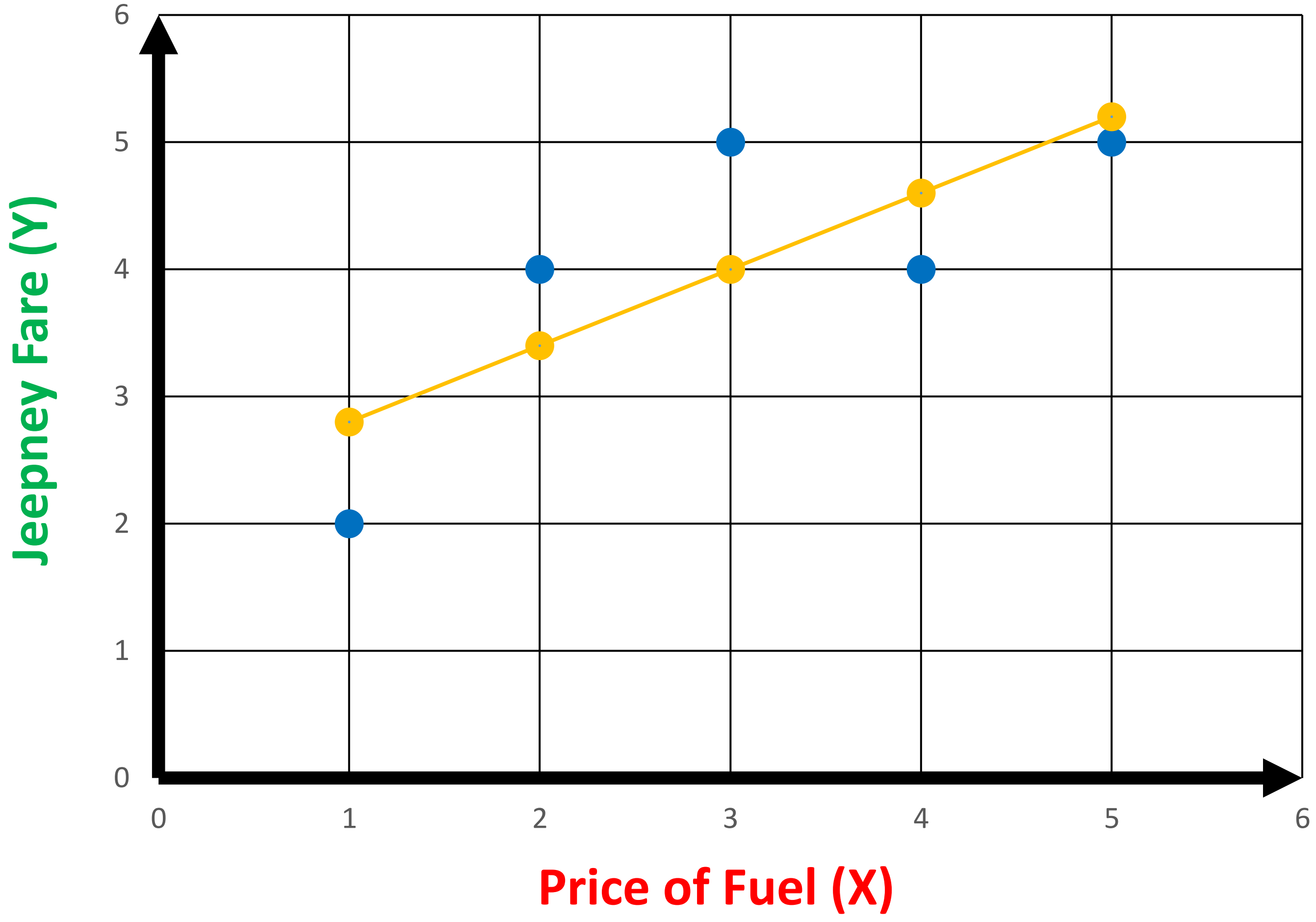
$$\boxed{y_{predict} = 3.4}$$

$$y = 0.6(5) + (2.2)$$

$$\boxed{y_{predict} = 5.2}$$

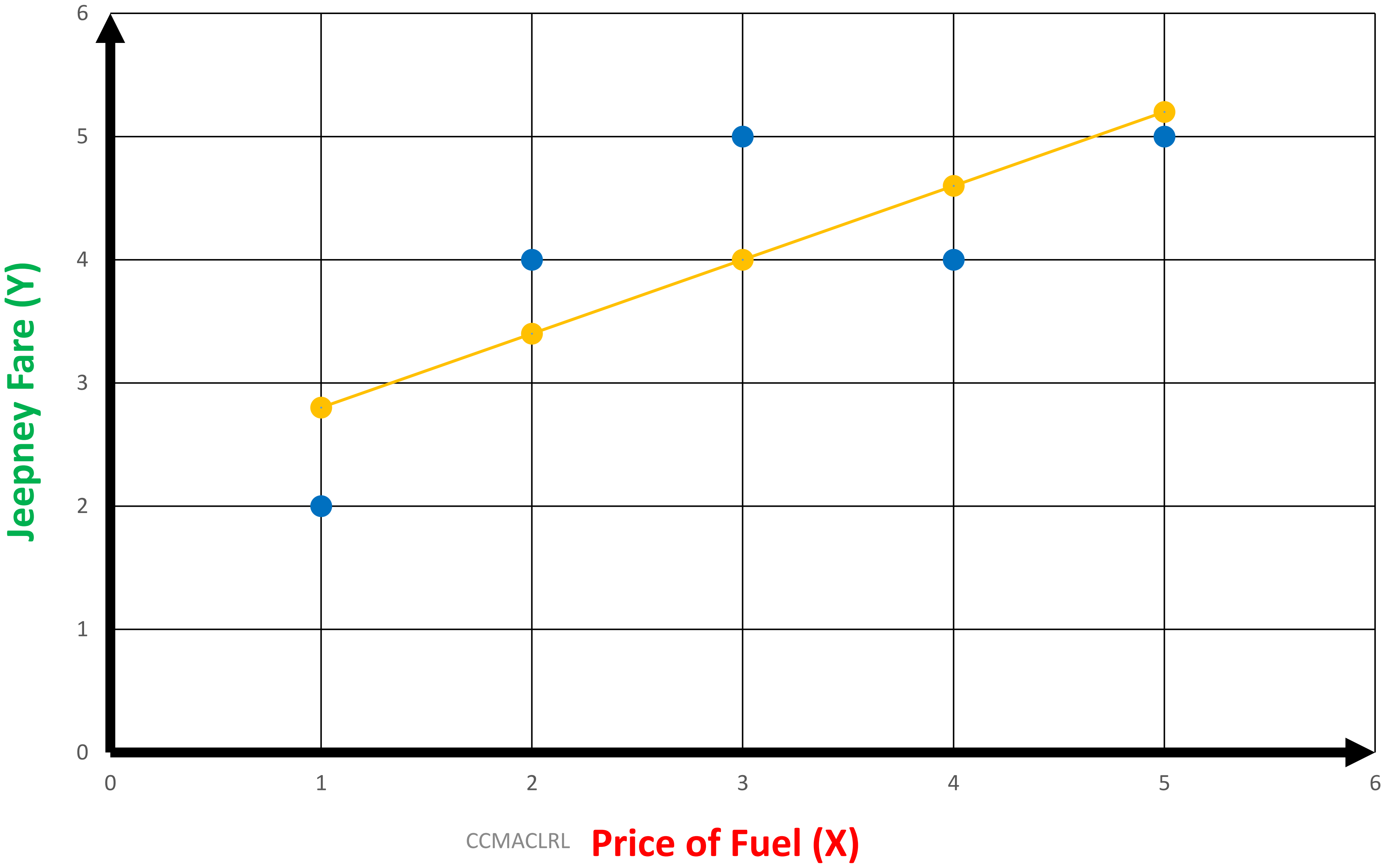
Drawing the Regression Line

Price of Fuel (X)	Jeepney Fare (Y)	Predicted Jeepney Fare ($Y_{predict}$)
1	2	2.8
2	4	3.4
3	5	4
4	4	4.6
5	5	5.2



Drawing the Regression Line

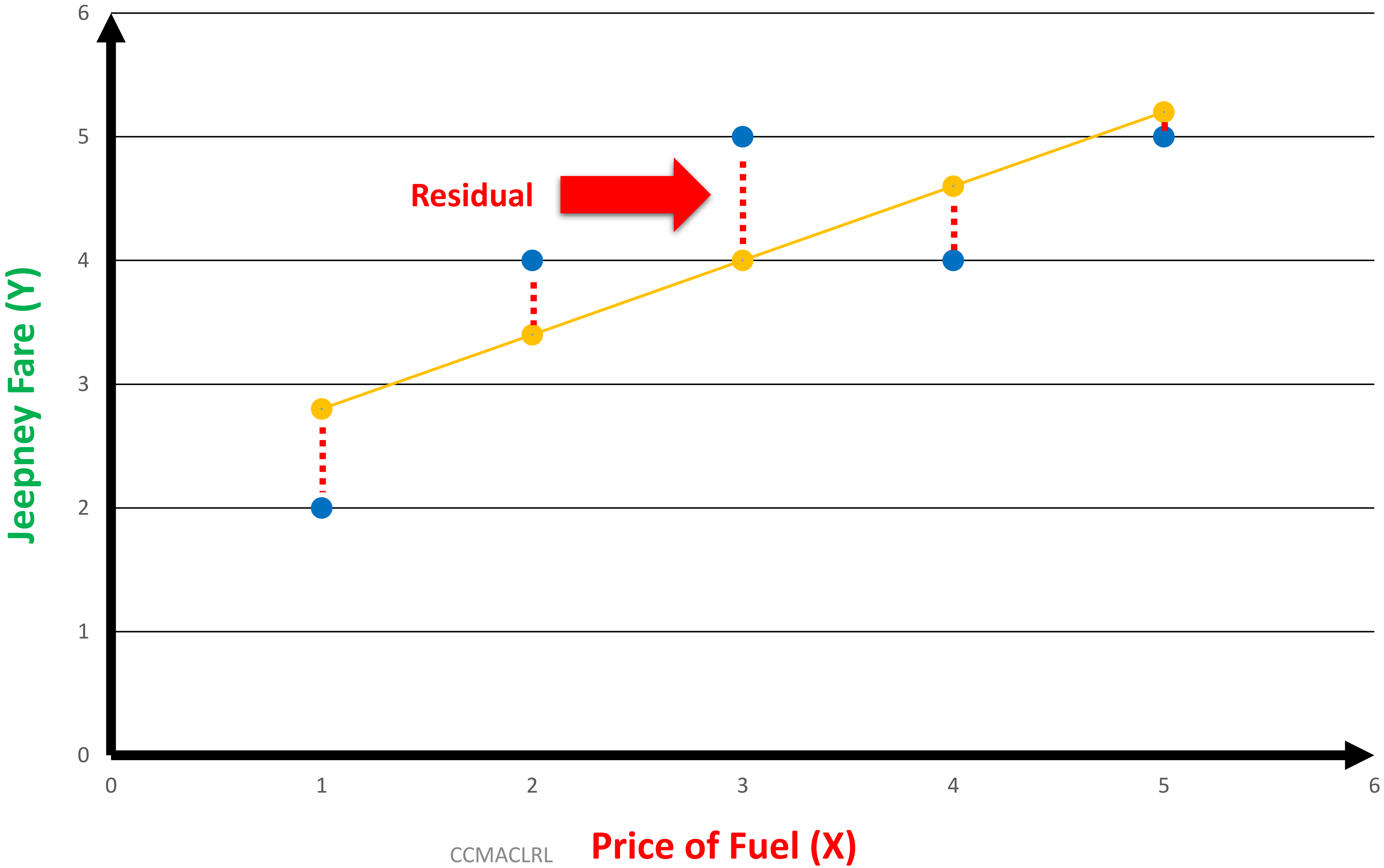
The **blue points** represent **the actual Y values** and **the orange points** represent the **predicted Y values**



NATIONAL UNIVERSITY

Residuals/Errors

The **distance** between the **actual values** and the **predicted values** are known as **residuals or errors**



CCMACLRL

NATIONAL UNIVERSITY

Sum of Squared Error

Price of Fuel (X)	Jeepney Fare (Y)	Predicted Jeepney Fare ($Y_{predict}$)	$Y - Y_{predict}$	$(Y - Y_{predict})^2$
1	2	2.8	-0.8	0.64
2	4	3.4	0.6	0.36
3	5	4	1	1
4	4	4.6	-0.6	0.36
5	5	5.2	-0.2	0.04

$$Sum\ of\ Squared\ Errors\ (SSE) = \sum_{i=1}^{n} (y_i - y_{predict})^2$$

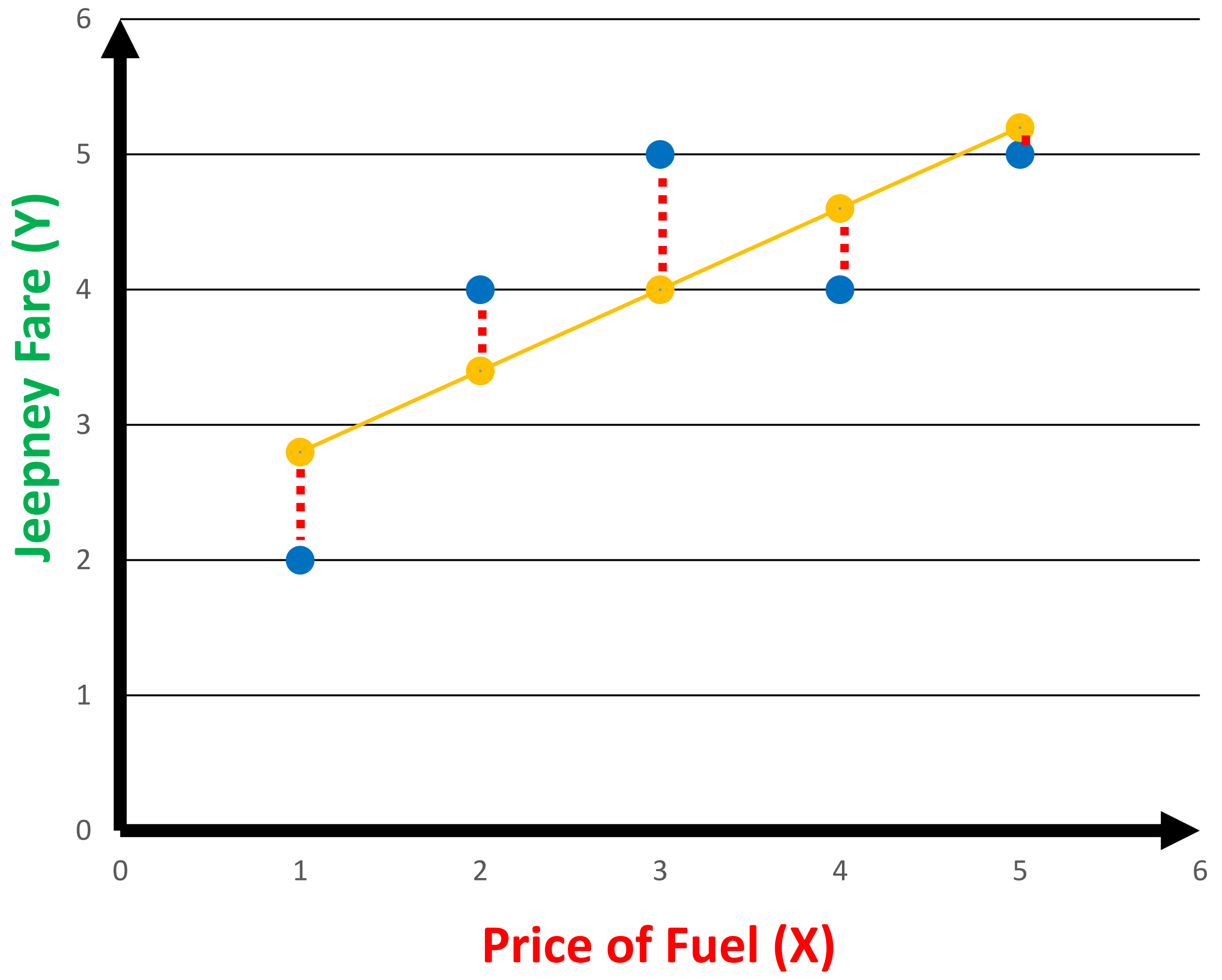
NATIONAL UNIVERSITY

Sum of Squared Error

Price of Fuel (X)	Jeepney Fare (Y)	Predicted Jeepney Fare ($Y_{predict}$)	$Y - Y_{predict}$	$(Y - Y_{predict})^2$
1	2	2.8	-0.8	0.64
2	4	3.4	0.6	0.36
3	5	4	1	1
4	4	4.6	-0.6	0.36
5	5	5.2	-0.2	0.04

SSE = 2.4

The sum of squared errors (SSE) for this regression line is **2.4**. This tells you how **good a line is fitted to the data**. **The best fit line will have the least amount of this value.**



Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2			
2	4			
3	5			
4	4			
5	5			

$$\boxed{\bar{Y} = \frac{\sum Y_i}{n}}$$

Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2	4		
2	4	4		
3	5	4		
4	4	4		
5	5	4		

$$\boxed{\bar{Y} = \frac{20}{5} = 4}$$

Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2	4	-2	
2	4	4	0	
3	5	4	1	
4	4	4	0	
5	5	4	1	

$$\bar{Y} = \frac{20}{5} = 4$$

NATIONAL UNIVERSITY 1900

Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2	4	-2	4
2	4	4	0	0
3	5	4	1	1
4	4	4	0	0
5	5	4	1	1

$$\bar{Y} = \frac{20}{5} = 4$$

NATIONAL UNIVERSITY 1900

Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2	4	-2	4
2	4	4	0	0
3	5	4	1	1
4	4	4	0	0
5	5	4	1	1

$$\textit{Sum of Squared Total (SST)} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Sum of Squared Total

Price of Fuel (X_i)	Jeepney Fare (Y_i)	$\bar{Y}$	$Y_i - \bar{Y}$	$(Y_i - \bar{Y})^2$
1	2	4	-2	4
2	4	4	0	0
3	5	4	1	1
4	4	4	0	0
5	5	4	1	1

$\boxed{\text{SST} = 6}$

$$\boxed{\boldsymbol{Sum\ of\ Squared\ Total\ (SST)} = 4 + 0 + 1 + 0 + 1 = 6}$$

NATIONAL UNIVERSITY

Computing R^2

$$\boldsymbol{Sum\ of\ Squared\ Errors\ (SSE)} = \sum_{i=1}^{n} (y_i - y_{predict})^2$$

$$\boldsymbol{Sum\ of\ Squared\ Total\ (SST)} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\boldsymbol{R^2 = 1 - \frac{SSE}{SST}}$$

NATIONAL UNIVERSITY

Computing $\boldsymbol{R^2}$

$$\boxed{\boldsymbol{Sum\ of\ Squared\ Errors\ (SSE)} = 2.4}$$

$$\boxed{\boldsymbol{Sum\ of\ Squared\ Total\ (SST)} = 6}$$

$$\boxed{\boldsymbol{R^2 = 1 - \frac{2.4}{6} = 0.6}}$$