Navarro, Genesis F.
COM231
Assignment #2

Linear Regression Activity: Predicting Exam Scores

We want to predict a student's **Exam Score (y)** based on the number of **Hours Studied (x)**.

Data

Student	Hours Studied (x)	Exam Score (y)
1	1	52
2	2	57
3	3	61
4	4	65
5	5	70

Task

We want to fit a **linear regression line** of the form:

$$y = mx + b$$

A new student studied **6 hours**. We want to predict the **Exam Score** using the regression equation.

1. Fill in the table (14 points)

- Compute x^2 for each student.
- Compute xy for each student.
- Find the totals: Σx, Σy, Σx^2, and Σxy.

Student	Hours Studied (x)	Exam Score (y)	xy	x^2
1	1	52	52	1
2	2	57	114	4
3	3	61	183	9
4	4	65	260	16
5	5	70	350	25
	Σx = ? 15	Σx = ? 305	Σxy = 959	Σx^2 = 55

$\sum x = 1 + 2 + 3 + 4 + 5 = 15$

∑y = 52 + 57 + 61 + 65 + 70 = 305

∑xy = 52 + 114 + 183 + 260 + 350 = 959

2. Compute the Slope m (5 points)

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

$m = ?$ 4.4

3: Compute the Intercept b (5 points)

$$b = \frac{\sum y - m\sum x}{n}$$

$b = ?$ 47.8

Solution 2

∑x = 15

∑y = 305

∑xy = 959

∑x2 = 55

n = 5

$$m = \frac{5(959) - (15)(305)}{5(55) - (15)^2}$$

m = 4.4

Solution 3

∑x = 15 n = 5

∑y = 305 m = 4.4

$$b = \frac{305 - (4.4)(15)}{5}$$

$$b = 47.8$$

4. Regression Equation (5 points)

Write the regression line:

$$y = mx + b$$

$$y = ?$$

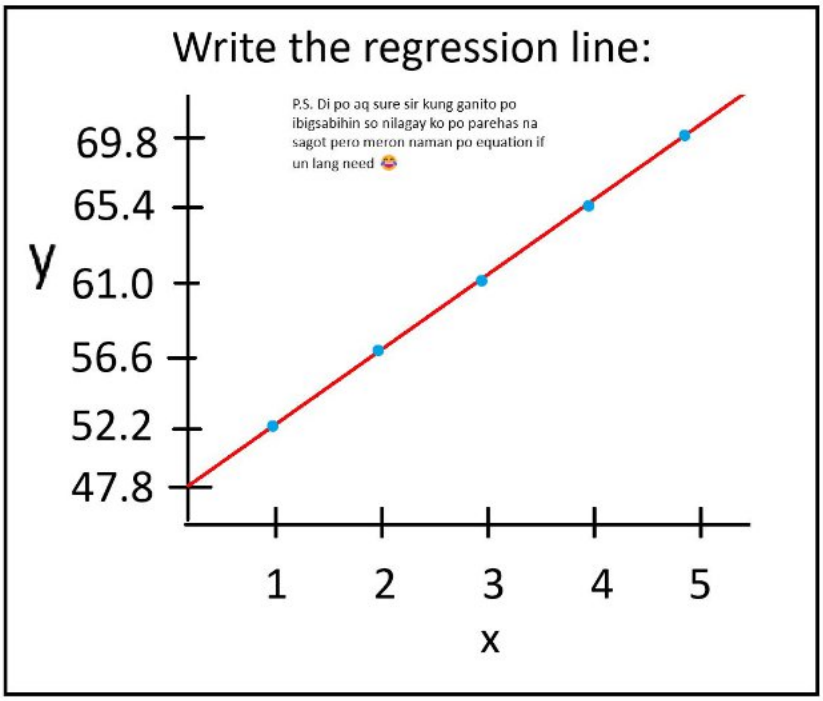
m = 4.4 b = 47.8

$$y = (4.4)(x) + 47.8$$

$$y = 4.4x + 47.8$$

Student	x	$y_{predict}$
1	1	52.2
2	2	56.6
3	3	61.0
4	4	65.4
5	5	69.8

Write the regression line:



5. Draw the regression line using a scatter plot (10 points)

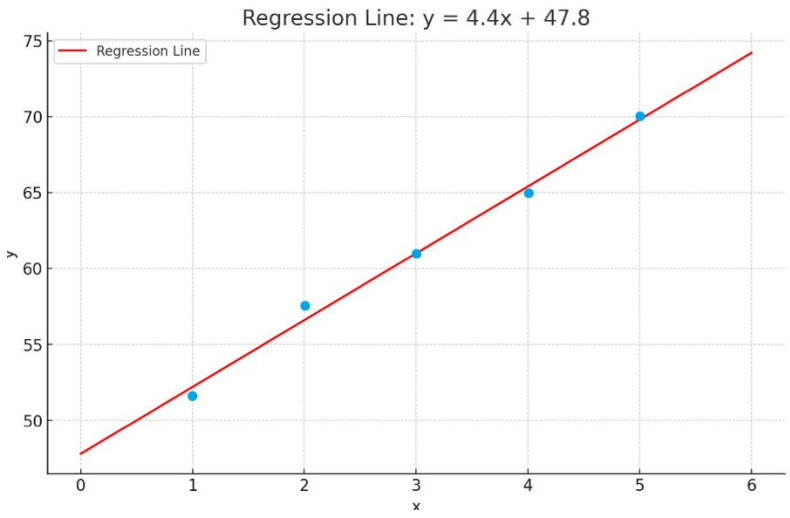
- Calculate $y_{predict}$ for each data points
- Draw a regression line using $y_{predict}$
- Use a circle 🔵 for all data points
- Use a red line for the regression line

Student	Hours Studied (x)	Exam Score (y)	Predicted Exam Score ($y_{predict}$)
1	1	52	?
2	2	57	?
3	3	61	?
4	4	65	?
5	5	70	?

Data

Student	Hours Studied (x)	Exam Score (y)
1	1	52
2	2	57
3	3	61
4	4	65
5	5	70

$y_{predict}$	$y_{predict}$
52.2	$y = 4.4(1) + 47.8 = 52.2$
56.6	$y = 4.4(2) + 47.8 = 56.6$
61.0	$y = 4.4(3) + 47.8 = 61.0$
65.4	$y = 4.4(4) + 47.8 = 65.4$
69.8	$y = 4.4(5) + 47.8 = 69.8$



6. Calculate the Sum of Squared Errors (20 points)

$$ SSE = \sum (yi - y{predict})^2 $$

Student	Hours Studied (x)	Exam Score (y)	Predicted Exam Score ($y_{predict}$)	$y_i - y_{predict}$	$(y_i - y_{predict})^2$
1	1	52	?	?	?
2	2	57	?	?	?
3	3	61	?	?	?
4	4	65	?	?	?
5	5	70	?	?	?
					SSE = ?

Student	Hours Studied (x)	Exam Score (y)	Predicted Exam Score ($y_{predict}$)	$y_i - y_{predict}$	$(y_i - y_{predict})^2$
1	1	52	52.2	0.2	0.04
2	2	57	56.6	0.4	0.16
3	3	61	61.0	0	0
4	4	65	65.4	0.4	0.16
5	5	70	69.8	0.2	0.04
					SSE = 0.4

$y_i - y_{predict}$	$(y_i - y_{predict})^2$
52-52.2 = 0.2	0.2^2 = 0.2 x 0.2 = 0.04
57 - 56.6 = 0.4	0.4^2 = 0.4 x 0.4 = 0.16
61 - 61 = 0	0^2 = 0 x 0 = 0
65 - 65.4 = 0.4	0.4^2 = 0.4 x 0.4 = 0.16
70 - 69.8 = 0.2	0.2^2 = 0.2 x 0.2 = 0.04

7. Calculate the Sum of Squared Total (20 points)

- Get $\bar{y}$ using this formula :

$$\bar{y} = \frac{\sum y_i}{n}$$

- Get SST using this formula :

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

Student	Hours Studied (x)	Exam Score (y)	Mean ($\bar{y}$)	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	1	52	61	9	81
2	2	57	61	4	16
3	3	61	61	0	0
4	4	65	61	4	16
5	5	70	61	9	81
					SST = 194

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

= 81 + 16 + 0 + 16
+ 81 = 194

Exam Score (y)	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
52	52 - 61 = |-9| = 9	9^2 = 9 x 9 = 81
57	57 - 61 = |-4| = 4	4^2 = 4 x 4 = 16
61	61 - 61 = 0	0^2 = 0 x 0 = 0
65	65 - 61 = 4	4^2 = 4 x 4 = 16
70	70 - 61 = 9	9^2 = 9 x 9 = 81

8. Compute R^2 (20 points)

- Get R^2 using this formula :

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

$$R^2 = ?$$

SSE = 0.4

SST = 194

$R^2 = 1 - \frac{0.4}{194} = 0.99$

m = 4.4, b = 47.8

y = m(6)+b

= 4.4(6)+47.8

= 74.2

9. Prediction (1 point)

Use your equation to predict the exam score for a student who studied **6 hours.**