

REGRESSION

Linear Regression

Linear Regression

- **Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**.
- Regression models a target prediction value based on independent variables.
- It is mostly used for finding out the relationship between variables and forecasting.

Linear Regression

- Linear Regression Function

$$y = a + bx$$

- Y is the output we want. x is the input variable. a = constant and b is the slope of the line
- The output varies linearly based upon the input.
- 'a' is the constant (value of y when x is zero)
- Slope (b) of Regression line

$$b = r \frac{S_y}{S_x}$$

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

- Y-intercept (a) of Regression Line

$$a = \text{mean}(y) - b * \text{mean}(x)$$

Example

Student	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$
1	95	85	17	8
2	85	95	7	18
3	80	70	2	-7
4	70	65	-8	-12
5	60	70	-18	-7
Sum	390	385		
Mean	78	77		

Example.....

Student	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

- And for each student, we also need to compute the squares of the deviation scores

Example.....

Student	x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$
1	95	85	136
2	85	95	126
3	80	70	-14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

- And finally, for each student, we need to compute the product of the deviation scores.

Slope.....

Pearson correlation coefficient

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

$$r = 470 / \sqrt{730 \times 630}$$
$$470 / 678.15$$

$$r = 0.693$$

Compute the standard deviation of x (σ_x / s_x):

$$\sigma_x = \sqrt{\sum (x_i - \bar{x})^2 / N - 1}$$

$$\sigma_x = \sqrt{730 / 4} = \sqrt{182.5} = 13.5$$

Next, we find the standard deviation of y, (σ_y / s_y):

$$\sigma_y = \sqrt{\sum (y_i - \bar{y})^2 / N - 1}$$

$$\sigma_y = \sqrt{630 / 4} = \sqrt{157.5} = 12.5$$

$$\text{Slope } b \text{ of Regression line is } = b = r \frac{s_y}{s_x} = 0.693 \times 12.5 / 13.5 = 0.641$$

Y-intercept

- Y-intercept
- $a = \text{mean}(y) - b \times \text{mean}(x)$
- $77 - 0.64 \times 78$
- $77 - 49.92$
- $a = 27.08$

Linear Regression Function

- Linear Regression Function

$$y = a + bx$$

$$y = 27.08 + 0.641(x)$$

If a student made an 80 in Intro to Programming, the estimated statistics grade (y) would be:

$$y = 27.08 + 0.641(80)$$

$$= 27.08 + 51.28$$

$$= 78.36$$

Linear Regression Graph

