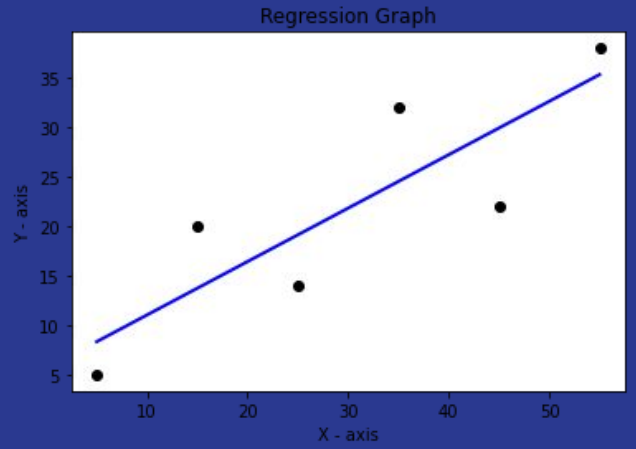
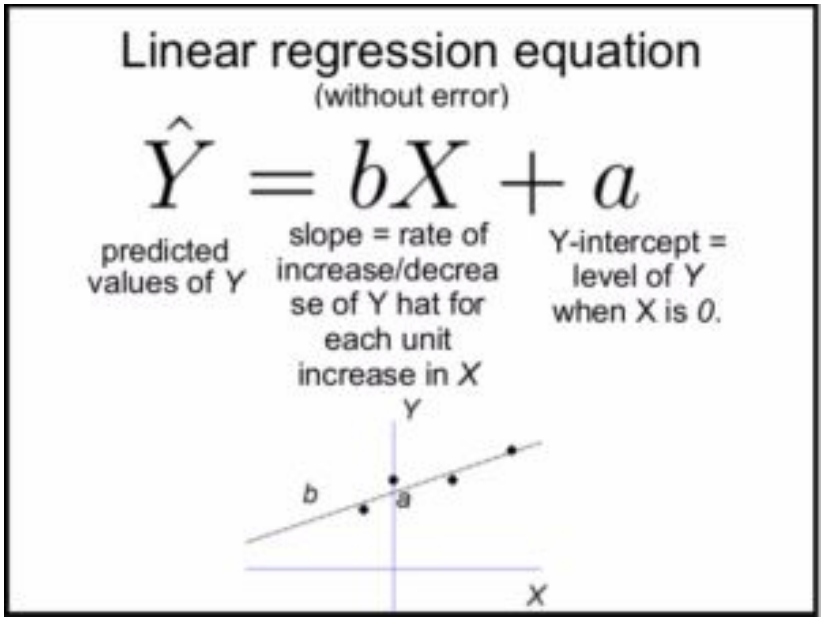
Linear Regression

Least squares method



Equation


Linear regression equation
(without error)
$\hat{Y} = bX + a$
predicted values of Y
slope = rate of increase/decrease of Y hat for each unit increase in X
Y-intercept = level of Y when X is 0.
Y
b
a
X

Step-by-step

Step 1: For each (x,y) point calculate x^2 and xy

Step 2: Sum all x, y, x^2, xy

Step 3: Calculate slope **b**: $\frac{N\,\Sigma(xy) - \Sigma x\,\Sigma y}{N\,\Sigma(x^2) - (\Sigma x)^2}$

Step 4: Calculate intercept **a**: $\frac{\Sigma y - b\,\Sigma x}{N}$

Step 5: Assemble the equation of a line: **y = bx + a**

Data

"x" Hours of Sunshine	"y" Ice Creams Sold
2	4
3	5
5	7
7	10
9	15

Calculation

Step 1: For each (x,y) calculate x^2 and xy:

x	y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135

Calculation

Step 2: Sum x, y, x^2 and xy (gives us Σx, Σy, $Σx^2$ and Σxy):

x	y	x^2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135
Σx: 26	**Σy: 41**	**$Σx^2$: 168**	**Σxy: 263**

Also **N** (number of data values) = **5**

Calculation

Step 3: Calculate Slope **b**

$$b = \frac{N\,\Sigma(xy) - \Sigma x\,\Sigma y}{N\,\Sigma(x^2) - (\Sigma x)^2}$$

$$= \frac{5 \times 263 - 26 \times 41}{5 \times 168 - 26^2}$$

$$= \frac{1315 - 1066}{840 - 676}$$

$$= \frac{249}{164} = 1.5183...$$

Calculation

Step 4: Calculate Intercept **a**

$$a = \frac{\Sigma y - b \Sigma x}{N}$$

$$= \frac{41 - 1.5183 \times 26}{5}$$

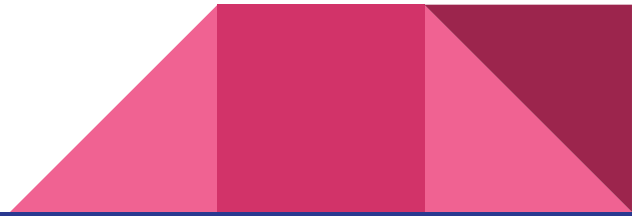
$$= 0.3049...$$

Calculation

Step 5: Assemble the equation of a line:

$y = bx + a$

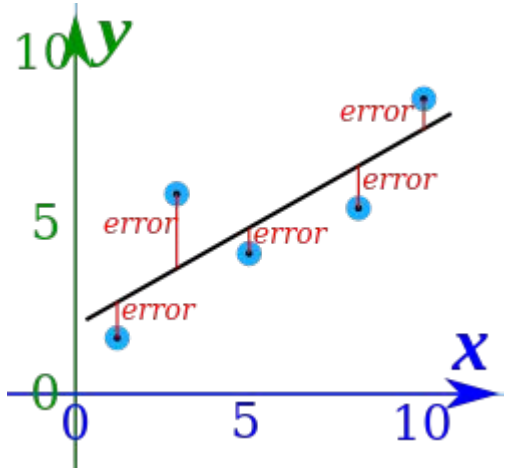
$y = 1.518x + 0.305$



Evaluation - Square of the errors

error = y’ - y

x	y	y = 1.518x + 0.305	error
2	4	3.34	−0.66
3	5	4.86	−0.14
5	7	7.89	0.89
7	10	10.93	0.93
9	15	13.97	−1.03



Mean square error - MSE

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

where N is the number of data points,
f_i the value returned by the model and
y_i the actual value for data point i.

MSE

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

x	y	y = 1.518x + 0.305	error	(y' - y) ^2
2	4	3.34	−0.66	0.4356
3	5	4.86	−0.14	0.0196
5	7	7.89	0.89	0.7921
7	10	10.93	0.93	0.8649
9	15	13.97	−1.03	1.0609

Sum = 3.1731

MSE = 3.1731/5
= 0.63462

Reference

- https://www.statisticshowto.com/probability-and-statistics/regression-analysis/find-a-linear-regression-equation/
- https://www.slideserve.com/bern/chapter-12-simple-linear-regression
- https://www.mathsisfun.com/data/least-squares-regression.html
-