

Program Fresh Graduate Academy Digital Talent Scholarship 2019 | Machine Learning

# Simple Linear Regression

M. Ramli & M. Soleh





Bagian 1

# Pendahuluan



# Pendahuluan

- Penggunaan statistika dalam mengolah data penelitian berpengaruh terhadap tingkat analisis hasil penelitian.
- Penelitian-penelitian dalam bidang ilmu pengetahuan alam (science) yang menggunakan perhitungan-perhitungan statistika, akan menghasilkan data yang mendekati benar, jika memperhatikan tata cara analisis data yang digunakan.
- Dalam memprediksi dan mengukur nilai dari pengaruh satu variabel (bebas/independent/ predictor) terhadap variabel lain (tak bebas/ dependent/ response) dapat digunakan uji regresi

# Apa itu Regresi?

- **Analisis/uji regresi** merupakan suatu kajian dari hubungan antara satu variabel, yaitu **variabel yang diterangkan (the explained variabel)** dengan satu atau lebih variabel, yaitu **variabel yang menerangkan (the explanatory)**.
- Apabila **variabel bebasnya** hanya **satu**, maka analisis regresinya disebut dengan **regresi linear sederhana**.
- Apabila **variabel bebasnya lebih dari satu**, maka analisis regresinya dikenal dengan **regresi linear berganda**.
  - Dikatakan berganda karena terdapat beberapa variabel bebas yang mempengaruhi variabel tak bebas

# Apa itu Regresi? (2)

- Analisis/uji regresi banyak digunakan dalam perhitungan hasil akhir untuk penulisan karya ilmiah/penelitian.
- Hasil perhitungan analisis/uji regresi akan dimuat dalam kesimpulan penelitian dan akan menentukan apakah penelitian yang sedang dilakukan berhasil atau tidak.
- Analisis perhitungan pada uji regresi menyangkut beberapa perhitungan statistika seperti uji signifikansi (uji-t, uji-F), anova dan penentuan hipotesis.
- Hasil dari analisis/ uji regresi berupa suatu persamaan regresi.
  - Persamaan regresi ini merupakan suatu fungsi prediksi variable yang mempengaruhi variabel lain

# Regresi Linear Sederhana

- **Regresi Linear Sederhana** adalah Metode Statistik yang berfungsi untuk menguji sejauh mana hubungan sebab akibat antara Variabel Faktor Penyebab (X) terhadap Variabel Akibatnya.
- **Faktor Penyebab** pada umumnya dilambangkan dengan **X** atau disebut juga dengan **Predictor** sedangkan Variabel Akibat dilambangkan dengan Y atau disebut juga dengan Response.
- Regresi Linear Sederhana atau sering disingkat dengan SLR (Simple Linear Regression) juga merupakan salah satu Metode Statistik yang dipergunakan dalam produksi untuk melakukan peramalan ataupun prediksi tentang karakteristik kualitas maupun kuantitas.



# Contoh Penggunaan

- Contoh penggunaan analisis Regresi Linear Sederhana dalam kegiatan produksi, antara lain:
  - Hubungan antara lamanya kerusakan mesin dengan kualitas produk yang dihasilkan
  - Hubungan jumlah pekerja dengan output yang diproduksi
  - Hubungan antara suhu ruangan dengan cacat produksi yang dihasilkan.





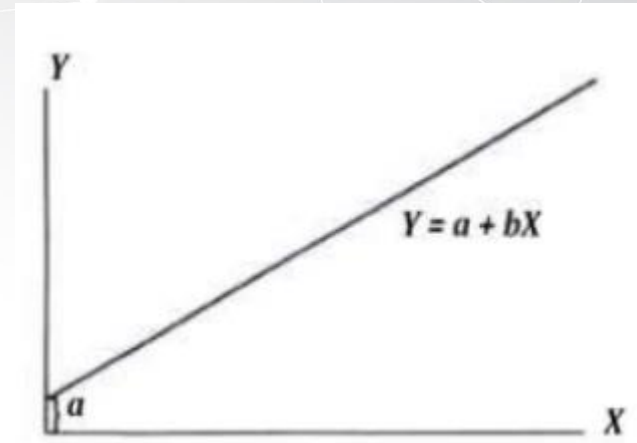
Bagian 2

# Metode Regresi Linear Sederhana



# Model Persamaan Regresi Linear Sederhana

- Model Persamaan Regresi Linear Sederhana adalah seperti berikut ini :
  - $Y = a + bX$
- Dimana :
  - $Y$  = Variabel Response atau Variabel Akibat (Dependent)
  - $X$  = Variabel Predictor atau Variabel Faktor Penyebab (Independent)
  - $a$  = konstanta
  - $b$  = koefisien regresi (kemiringan/slope); besaran Response yang ditimbulkan oleh Predictor .



# Model Persamaan Regresi Linear Sederhana (2)

- Besarnya konstanta **a** dan **b** dapat ditentukan menggunakan persamaan:

$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$b = \frac{n (\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

yang mana  $n$  = jumlah data

# Langkah-langkah Analisis dan Uji Regresi Linier Sederhana

- Adapun langkah-langkah yang perlu dilakukan untuk melakukan analisis dan uji regresi linier sederhana adalah sebagai berikut:
  1. Menentukan tujuan dari Analisis Regresi Linear Sederhana
  2. Mengidentifikasi variabel predictor dan variabel response
  3. Melakukan pengumpulan data dalam bentuk table
  4. Menghitung  $X^2$ ,  $Y^2$ ,  $XY$  dan total dari masing-masingnya
  5. Menghitung a dan b menggunakan rumus yang telah ditentukan
  6. Membuat model Persamaan Garis Regresi
  7. Melakukan prediksi terhadap variabel predictor atau response
  8. Uji korelasi



Bagian 3

# Contoh Kasus 1



# Berat Badan Mahasiswa

- “ Terdapat suatu data penelitian tentang berat badan 10 mahasiswa yang diprediksi dipengaruhi oleh konsumsi jumlah kalori/hari”.
- Bagaimana menganalisisnya?
  1. Menentukan Tujuan : “Apakah konsumsi jumlah kalori/hari mempengaruhi berat badan mahasiswa?”
  2. Mengidentifikasi variabel predictor dan variabel response
    - X (variable bebas/ predictor) = jumlah kalori/hari
    - Y (variable tak bebas/ response) = berat badan

# Tabel Data

## 3. Melakukan pengumpulan data dalam bentuk tabel

Nama Mahasiswa	Kalori/ hari (X)	Berat Badan (Y)
Adi	530	89
Rudi	300	48
Didi	358	56
Budi	510	72
Intan	302	54
Putu	300	42
Partadiyasa	387	60
Esti	527	85
Ike	415	63
Rojali	512	74

# Tabel Data Bantu

4. Menghitung  $X^2$ ,  $Y^2$ ,  $XY$  dan total dari masing-masingnya

	X	$X^2$	Y	$Y^2$	XY
	530	280900	89	7921	47170
	300	90000	48	2304	14400
	358	128164	56	3136	20048
	510	260100	72	5184	36720
	302	91204	54	2916	16308
	300	90000	42	1764	12600
	387	149769	60	3600	23220
	527	277729	85	7225	44795
	415	172225	63	3969	26145
	512	262144	74	5476	37888
$\Sigma$	4141	1802235	643	43495	279294



# Mencari Koefisien “a”

5. Menghitung a dan b menggunakan rumus yang telah ditentukan

$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{(643)(1802235) - (4141)(279294)}{10(1802235) - (4141)^2} = \frac{2280651}{874469} \cong 2,608$$

$$b = \frac{n (\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

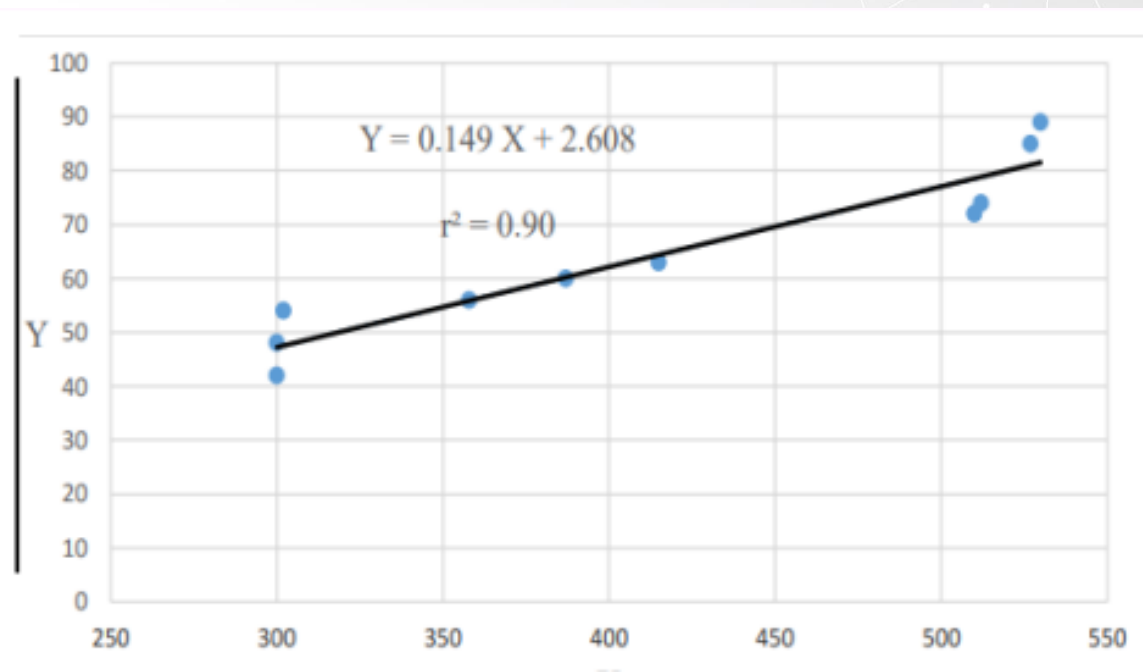
$$= \frac{10(279294) - (4141)(643)}{10(1802235) - (4141)^2} = \frac{130227}{874469} \cong 0,14892 \approx 0,149$$



# Persamaan Garis Regresi

- Membuat model Persamaan Garis Regresi
- Setelah didapat koefisien a dan b, maka persamaan garisnya adalah:

$$Y = 2,608 + 0,149 X$$



# Prediksi atau Peramalan terhadap Variabel Faktor Penyebab

7. Melakukan prediksi terhadap variabel predictor atau response

- Prediksikan berat badan mahasiswa jika asupannya adalah 600 kalori/hari:

$$Y = 2,608 + 0,149 X$$

- Prediksi  $Y = 2,608 + (0,149 * 600) = 92$  kilo gram

- Prediksikan asupan mahasiswa, jika berat badan mahasiswa adalah 40 kilo gram:

$$40 = 2,608 + 0,149 X$$

$$37,392 = 0,149X$$

- Prediksi  $X = 250.59$  kalori/ hari

# Uji Korelasi

- Untuk mengukur kekuatan hubungan antar variable predictor X dan response Y, dilakukan analisis korelasi yang hasilnya dinyatakan oleh suatu bilangan yang dikenal dengan koefisien korelasi.
- Biasanya analisis regresi sering dilakukan bersama-sama dengan analisis korelasi. Persamaan koefisien korelasi ( $r$ ) diekspresikan oleh:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \left( \sum_{i=1}^n X_i \right) \left( \sum_{i=1}^n Y_i \right)}{\sqrt{\left[ n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2 \right] \left[ n \sum_{i=1}^n Y_i^2 - \left( \sum_{i=1}^n Y_i \right)^2 \right]}}$$

# Uji Korelasi (2)

- Berdasarkan data tabel, maka koefisien korelasinya adalah:

$$\begin{aligned}
 r &= \frac{n \sum_{i=1}^n X_i Y_i - \left( \sum_{i=1}^n X_i \right) \left( \sum_{i=1}^n Y_i \right)}{\sqrt{\left[ n \sum_{i=1}^n X_i^2 - \left( \sum_{i=1}^n X_i \right)^2 \right] \left[ n \sum_{i=1}^n Y_i^2 - \left( \sum_{i=1}^n Y_i \right)^2 \right]}} \\
 &= \frac{10(279294) - (4141)(643)}{\sqrt{[10(1802235) - (4141)^2][10(43495) - (643)^2]}} \\
 &= \frac{130277}{137120,2318} = 0,95
 \end{aligned}$$

- Nilai ini memberi arti bahwa, hubungan variable bebas/ predictor X dengan variabel terikat/ response Y adalah sangat kuat, persentasenya 95%.
- Jadi, berat badan memang sangat dipengaruhi oleh konsumsi jumlah kalori/hari

# Koefisien Determinasi ( $r^2$ )

- Koefisien determinasi dapat ditentukan dengan mengkuadratkan koefisien korelasi.
- Dari contoh kasus di atas, maka koefisien determinasinya adalah  $r^2 = 0,90$
- Nilai ini berarti bahwa, 90% variabel bebas/ predictor X dapat menerangkan/ menjelaskan variabel tak bebas/ response Y dan 10% tidak yakin bahwa variable x menerangkan variable y.





Bagian 4

# Contoh Kasus 2



# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02

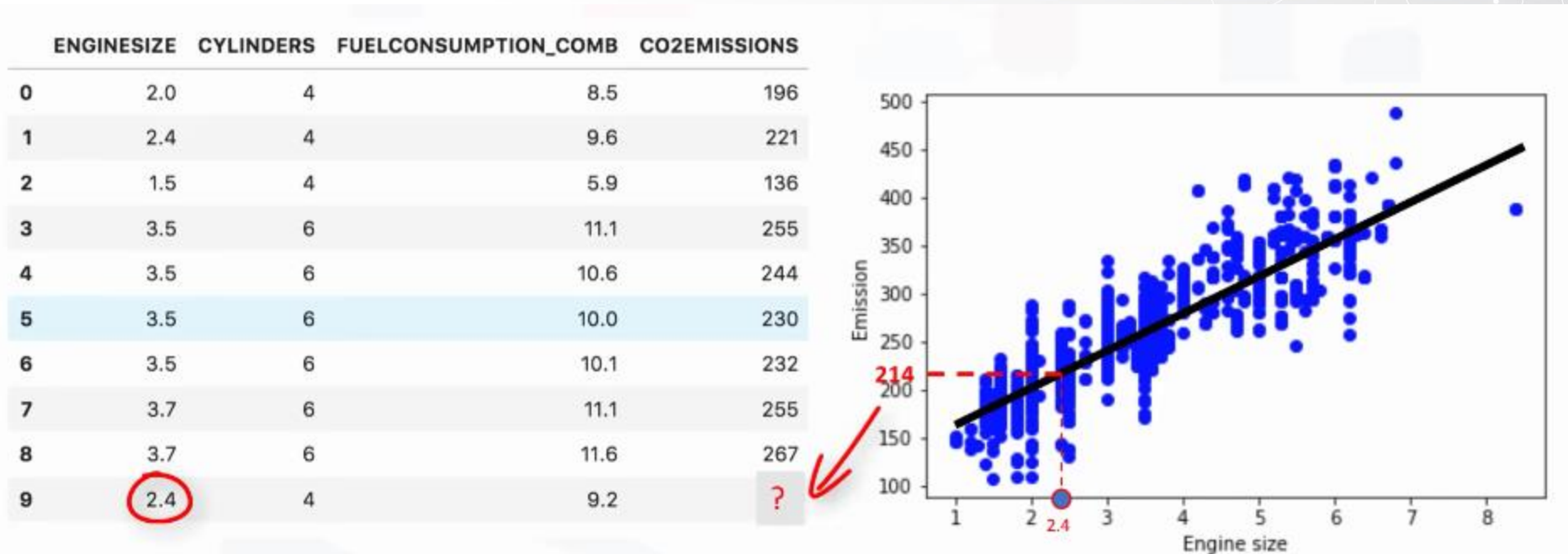
X: Independent variable

Y: Dependent variable

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
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 Values

# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 (2)



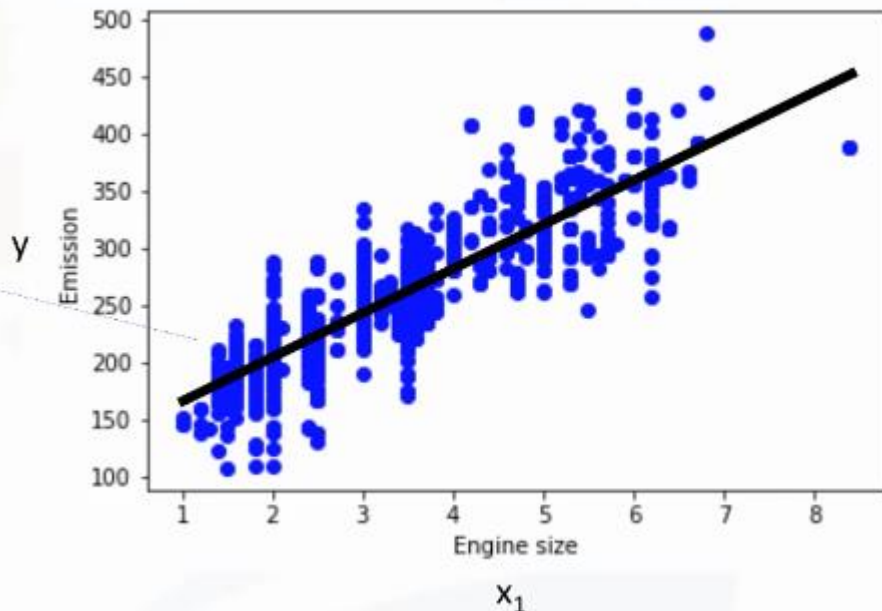
# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 (4)

- Model Persamaan Regresi Linear Sederhana adalah seperti berikut ini :
- $\hat{y} = \theta_0 + \theta_1 X_1$

$$\hat{y} = \theta_0 + \theta_1 x_1$$

response variable

a single predictor



# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 – Mencari Best Fit (5)

$x_1 = 2.4$  independent variable

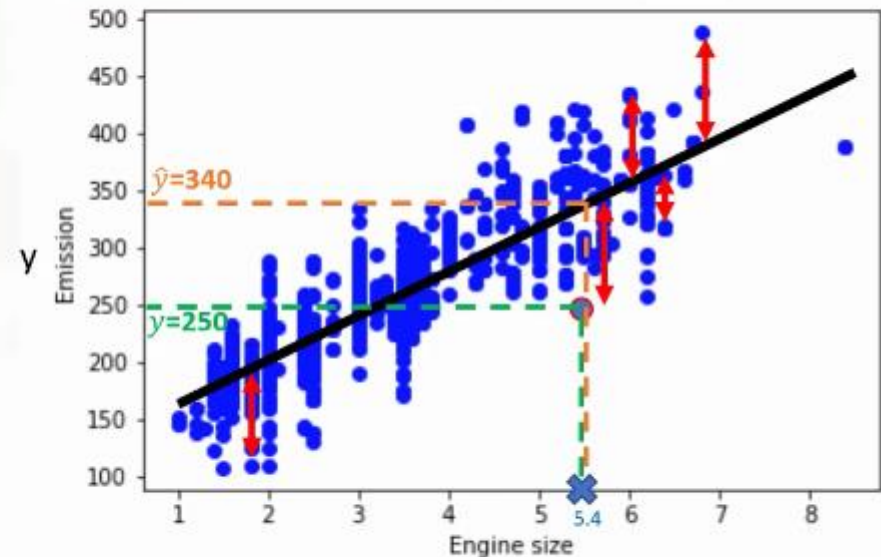
$y = 250$  actual Co2 emission of x1

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$\hat{y} = 340$  the predicted emission of x1

$$\begin{aligned} \text{Error} &= y - \hat{y} \\ &= 250 - 340 \\ &= -90 \end{aligned}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 – Estimasi Parameter (6)

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
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

$X_1$  is indicated by a bracket on the left of the table, and  $y$  is indicated by a bracket on the right of the table.

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_1 = 39$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_0 = 256 - 39 * 3.34$$

$$\theta_0 = 125.74$$



# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 – Estimasi Parameter (7)

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
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

$X_1$  is indicated by a bracket on the left of the table, and  $y$  is indicated by a bracket on the right of the table.

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$\theta_1 = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

$$\bar{x} = (2.0 + 2.4 + 1.5 + \dots) / 9 = 3.34$$

$$\bar{y} = (196 + 221 + 136 + \dots) / 9 = 256$$

$$\theta_1 = \frac{(2.0 - 3.34)(196 - 256) + (2.4 - 3.34)(221 - 256) + \dots}{(2.0 - 3.34)^2 + (2.4 - 3.34)^2 + \dots}$$

$$\theta_1 = 39$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_0 = 256 - 39 * 3.34$$

$$\theta_0 = 125.74$$

$$\hat{y} = 125.74 + 39x_1$$



# Prediksi Pengaruh Volume Mesin Terhadap Emisi C02 – Hasil Regresi (8)

	ENGINE SIZE	CYLINDERS	FUEL CONSUMPTION_COMB	CO2 EMISSIONS
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	?

$$\hat{y} = \theta_0 + \theta_1 x_1$$

$$Co2Emission = \theta_0 + \theta_1 EngineSize$$

$$Co2Emission = 125 + 39 EngineSize$$

$$Co2Emission = 125 + 39 \times 2.4$$

$$Co2Emission = 218.6$$





Bagian 5

# Praktikum Lab

ML0101EN-Reg-Simple-Linear-Regression-Co2-py-v1.ipynb