



Multivariate Regression Diamond

Kelompok 6

Anggota **Kel 6**

Abdul Hakim

G64190078

**Farhan
Fathurrahman**

G64190088



**Yosar Awandi
Porseda**

G64190076

**Mahara Ihsan
Kahfi**

G64190084

01

Multivariate Regression

Penjelasan rumus singkat

02

Data & Information

Melihat data dan variabel

03

Microsoft Excel

Pengerjaan manual dan
Data Analysis

Table of Contents

R Program

Pengerjaan dengan R

Comparison

Membandingkan hasil
kedua metode

QnA

Tanya aku 🐦

04

05

06



01

Multivariate Regression

Rumus yang digunakan

Multivariate Regression Model

Model regresi linier berganda merupakan model regresi dengan peubah penjelas lebih dari satu.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Dalam bentuk matriks

$$y_{n \times 1} = X_{n \times (k+1)} \beta_{(k+1) \times 1} + \epsilon_{n \times 1}$$

Penduga parameter regresi yang digunakan adalah metode kuadrat terkecil.

$$\hat{\beta} = (X'X)^{-1}X'y$$



02

Data & Information

Lihat data dan variabel

Data

Context

- This classic dataset contains the prices and other attributes of almost 54,000 diamonds. It's a great dataset for beginners learning to work with data analysis and visualization.

Content

- **price** price in US dollars (\\$326--\\$18,823)
- **carat** weight of the diamond (0.2--5.01)
- **cut** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- **color** diamond colour, from J (worst) to D (best)
- **clarity** a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
- **x** length in mm (0--10.74)
- **y** width in mm (0--58.9)
- **z** depth in mm (0--31.8)
- **depth** total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43--79)
- **table** width of top of diamond relative to widest point (43--95)

Data

	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
2	0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
3	0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
4	0.29	Premium	I	VS2	62.4	58	334	4.2	4.23	2.63
5	0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
...
53940	0.75	Ideal	D	SI2	62.2	55	2757	5.83	5.87	3.64

Information

Kami ingin menggunakan **price** sebagai variabel respon (y) dan **carat, depth, dan table** sebagai variabel penentu (x)

Maka,

Y = Price

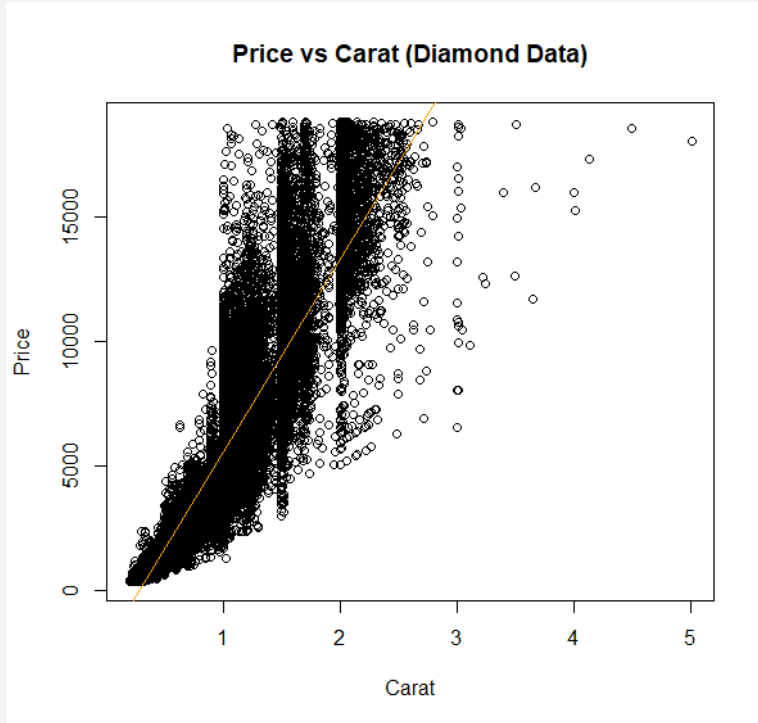
X_1 = carat

X_2 = depth

X_3 = table

	Min	Median	Mean	Max
Price	326	2401	3933	18823
Carat	0.20	0.70	0.7979	5.01
Depth	43	61.80	61.75	79.00
Table	43	57	57.46	95

Plot Price vs Carat

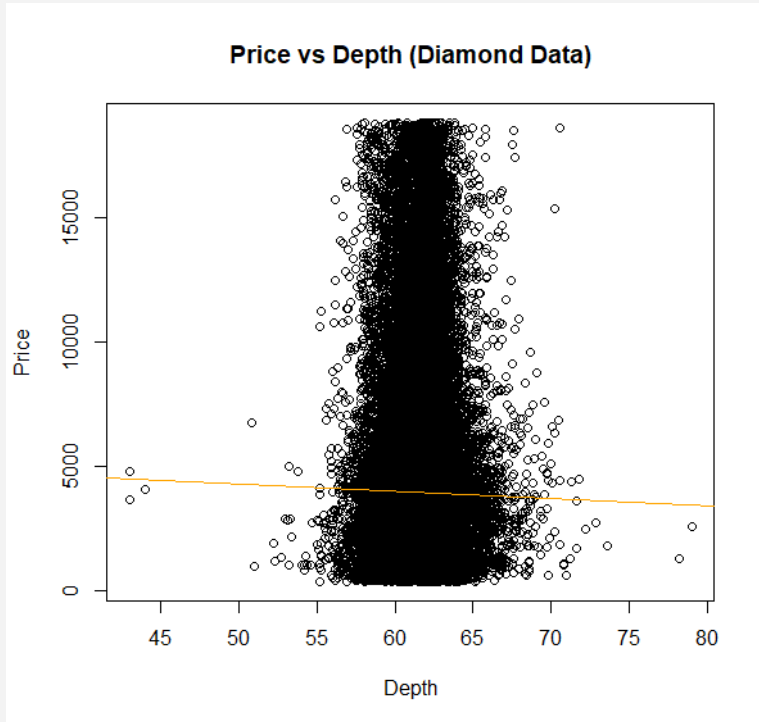


Korelasi
0.9215913

Model Regresi
 $\hat{y} = -2256.35 + 7756.43x$

R²
0.8493305

Plot Price vs Depth



Korelasi

-0.0106474

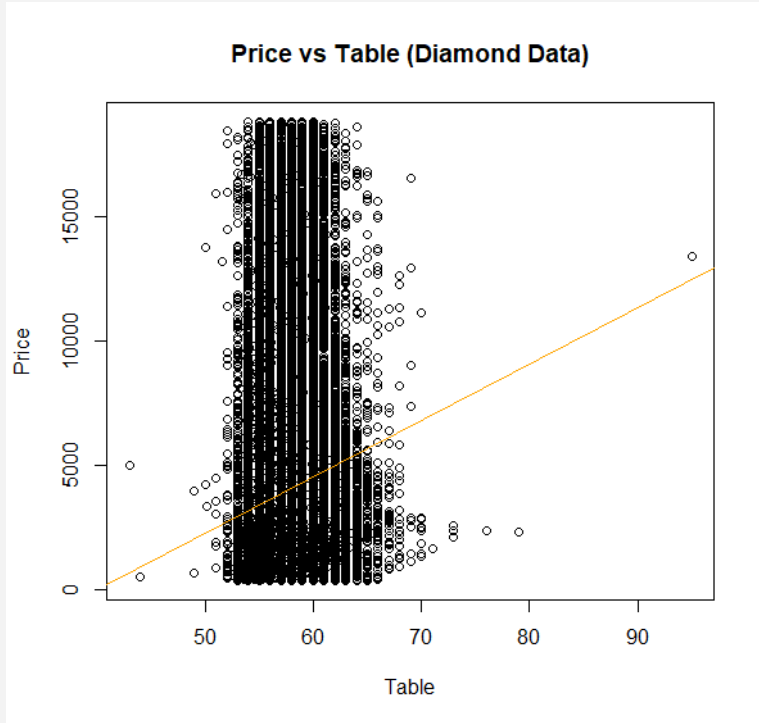
Model Regresi

$$\hat{y} = 5763.67 - 29.65x$$

R²

0.0001133672

Plot Price vs Table



Korelasi

0.1271339

Model Regresi

$$\hat{y} = -9109.047 + 226.984x$$

R²

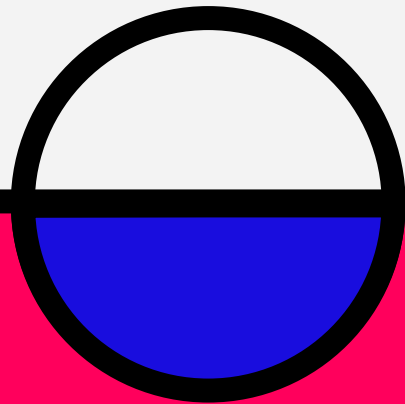
0.01616303



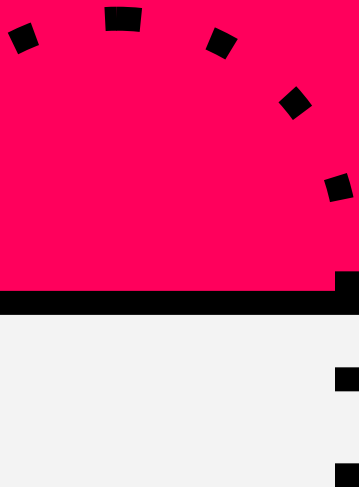
03

Microsoft Excel

Manual dan Data Analysis



Manual



Manual

Penduga parameter regresi yang digunakan adalah metode kuadrat terkecil.

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

Manual

Untuk mengisi matriks, maka harus dicari nilai tersebut

$$X'X = \begin{pmatrix} \text{sum}(nn) & \text{sum}(nx_1) & \text{sum}(nx_2) & \text{sum}(nx_3) \\ \text{sum}(x_1n) & \text{sum}(x_1x_1) & \text{sum}(x_1x_2) & \text{sum}(x_1x_3) \\ \text{sum}(x_2n) & \text{sum}(x_2x_1) & \text{sum}(x_2x_2) & \text{sum}(x_2x_3) \\ \text{sum}(x_3n) & \text{sum}(x_3x_1) & \text{sum}(x_3x_2) & \text{sum}(x_3x_3) \end{pmatrix}$$

$$X'y = \begin{pmatrix} \text{sum}(yn) \\ \text{sum}(yx_1) \\ \text{sum}(yx_2) \\ \text{sum}(yx_3) \end{pmatrix}$$

Steps

Var	y	n	x1	x2	x3	x1x1	x1x2	x1x3	x2x2	x2x3	x3x3	ny	x1y	x2y	x3y
#	price		carat	depth	table										
1	326	1	0.23	61.5	55	0.0529	14.145	12.65	3782.2 5	3382.5	3025	326	74.98	20049	17930
2	326	1	0.21	59.8	61	0.0441	12.558	12.81	3576.0 4	3647.8	3721	326	68.46	19494.8	19886
3	327	1	0.23	56.9	65	0.0529	13.087	14.95	3237.61	3698.5	4225	327	75.21	18606.3	21255
...
53940	2757	1	0.75	62.2	55	0.5625	46.65	41.25	3868.8 4	3421	3025	2757	2067.7 5	171485. 4	151635
Total						46463. 39	265878 2	248338 3	2.06E+ 08	1.91E+0 8	1.78E+0 8	2.12E+0 8	2.63E+ 08	1.31E+10	1.22E+1 0

Steps



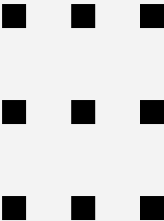
Dari hasil perhitungan sebelumnya didapatkan:

$$X'X = \begin{pmatrix} 53940 & 43040.87 & 3330762.9 & 3099240.5 \\ 43040.87 & 46463.3947 & 2658781.93 & 2483383.133 \\ 3330762.9 & 2658781.93 & 205783331.5 & 191325184.9 \\ 3099240.5 & 2483383.133 & 191325184.9 & 178342946 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 212135217 \\ 263274142.6 \\ 13095941018 \\ 12249822220 \end{pmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 13003.44052 \\ 7858.77051 \\ -151.2363469 \\ -104.4727802 \end{pmatrix}$$

Parameter	Nilai
$\hat{\beta}_0$	13003.441
$\hat{\beta}_2$	7858.771
$\hat{\beta}_3$	-151.236
$\hat{\beta}_4$	-104.473



Model Formula

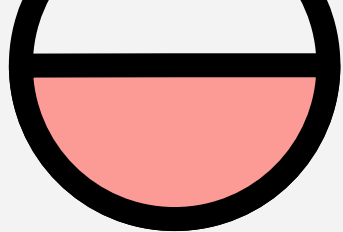
Dari hasil perhitungan parameter, model regresi yang didapatkan adalah

$$\hat{Y} = 13003.4 + 7858.8X_1 - 151.2X_2 - 104.5X_3$$

Koefisien X_1 dapat diinterpretasikan rata-rata Y akan naik sebesar 7858.8 jika X_1 naik satu satuan dan X_2 dan X_3 tetap.

Koefisien X_2 dapat diinterpretasikan rata-rata Y akan turun sebesar 151.2 jika X_2 naik satu satuan dan X_1 dan X_3 tetap.

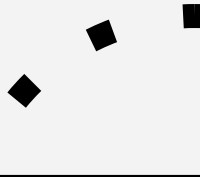
Koefisien X_3 dapat diinterpretasikan rata-rata Y akan turun sebesar 104.5 jika X_3 naik satu satuan dan X_1 dan X_2 tetap.

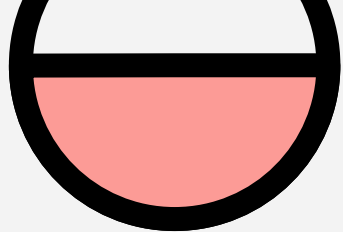


Manual Matrices

Juga bisa menghitung parameter regresi menggunakan fungsi excel yang mendukung perhitungan matriks.

Perkalian matriks : `=MMULT(A, B)`
Transpos matriks : `=TRANSPOSE(A)`
Invers matriks : `=MINVERSE(A)`





Steps

Data Y = B3:B53942

Data X = C3:F53942

Mencari $X'X$ (J3:M6)

$\{=MMULT(TRANSPOSE(C3:F53942),C3:F53942)\}$

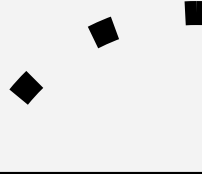
Mencari $X'Y$ (J8:J11)

$\{=MMULT(TRANSPOSE(C3:F53942),B3:B53942)\}$

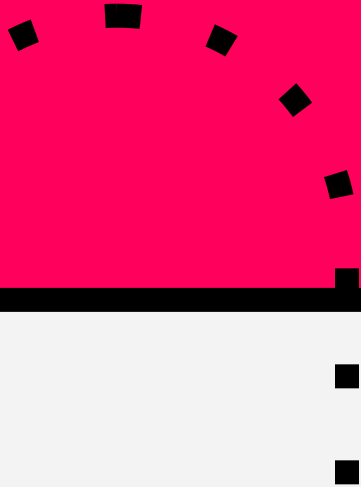
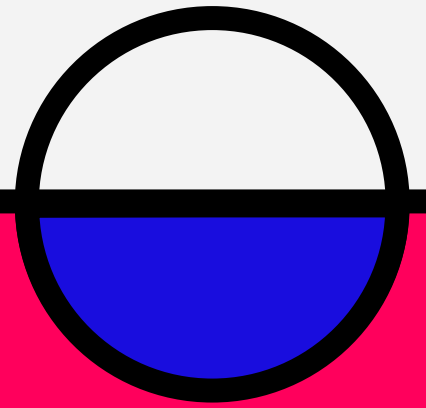
Mencari B penduga

$\{=MMULT(MINVERSE(J3:M6),J8:J11)\}$

Metode ini akan menghasilkan hasil yang sama.

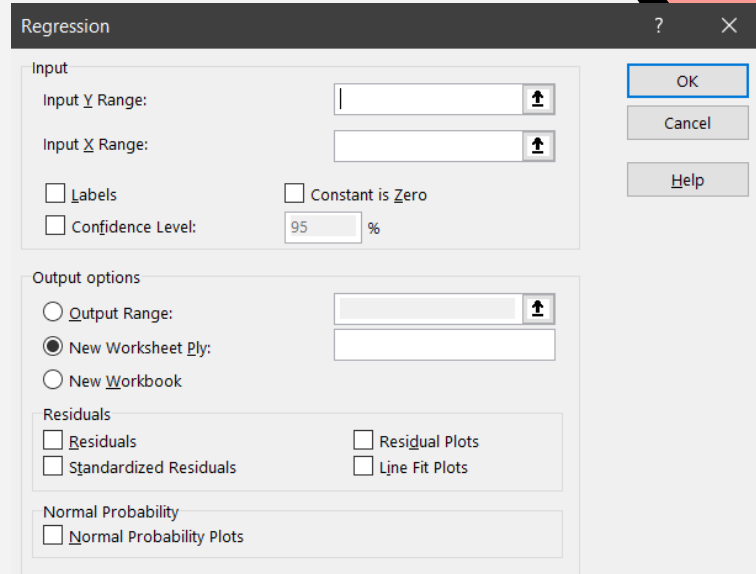
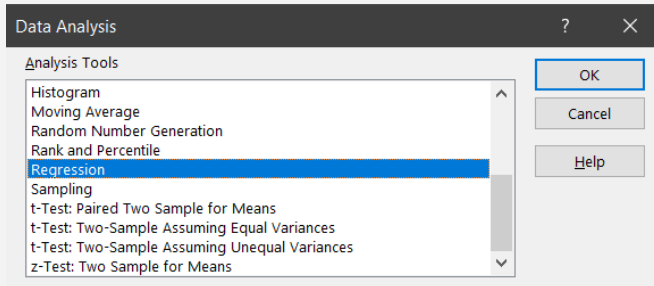


Data Analysis



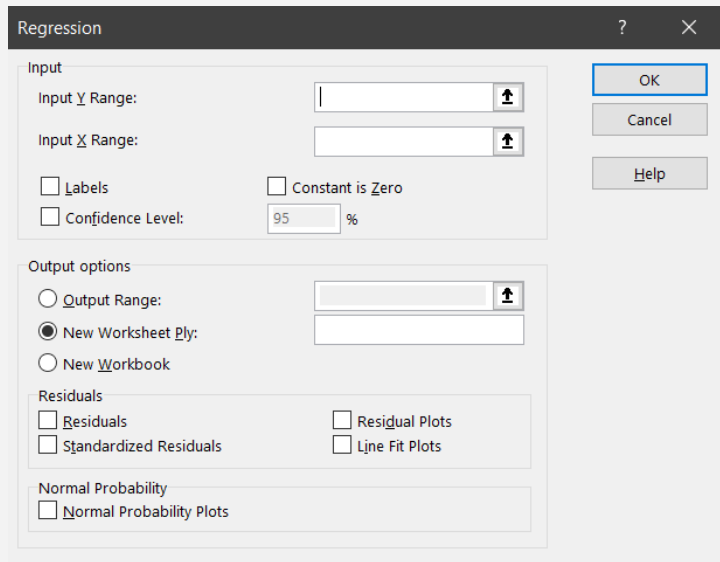
Data Analysis

Excel menyediakan Add-on bernama Data Analysis yang mempunyai banyak metode statistika untuk mengolah data secara otomatis. Salah satu dari metode itu adalah Regression yang dapat digunakan untuk menghitung regresi data



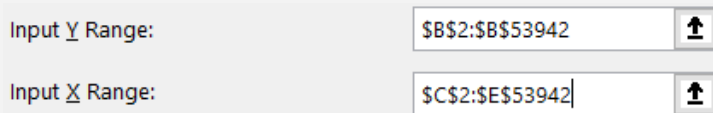
Steps

1. Membuka Data Analysis dan pilih Regression



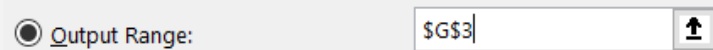
The screenshot shows the 'Regression' dialog box in Excel. It has a title bar with a question mark and a close button. The 'Input' section contains 'Input Y Range' and 'Input X Range' text boxes, each with an upward arrow icon. Below these are checkboxes for 'Labels', 'Constant is Zero', and 'Confidence Level' (set to 95%). The 'Output options' section has three radio buttons: 'Output Range', 'New Worksheet Ply', and 'New Workbook'. The 'Residuals' section has checkboxes for 'Residuals', 'Standardized Residuals', 'Residual Plots', and 'Line Fit Plots'. The 'Normal Probability' section has a checkbox for 'Normal Probability Plots'. On the right side of the dialog are three buttons: 'OK' (highlighted with a blue border), 'Cancel', and 'Help'.

2. Isi Y Range dan X Range dengan cell yang terisi dengan data Y dan X



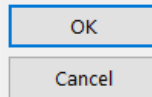
This screenshot shows two input fields. The first is 'Input Y Range:' with the text '\$B\$2:\$B\$53942' and an upward arrow icon. The second is 'Input X Range:' with the text '\$C\$2:\$E\$53942' and an upward arrow icon.

3. Pilih Output Range untuk melihat data yang diolah



This screenshot shows the 'Output Range:' field with the text '\$G\$3' and an upward arrow icon. The radio button next to 'Output Range:' is selected.

4. Pilih OK



This screenshot shows two buttons: 'OK' (highlighted with a blue border) and 'Cancel'.

Result

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.923946								
R Square	0.853676								
Adjusted R Square	0.853668								
Standard Error	1526.094								
Observations	53940								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	3	7.33E+11	2.44E+11	104890.5	0				
Residual	53936	1.26E+11	2328964						
Total	53939	8.58E+11							
Coefficients									
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	13003.44	390.9183	33.26383	3.5E-240	12237.24	13769.64	12237.24	13769.64	
carat	7858.771	14.15088	555.3558	0	7831.035	7886.506	7831.035	7886.506	
depth	-151.236	4.81989	-31.3776	3.5E-214	-160.683	-141.789	-160.683	-141.789	
table	-104.473	3.141234	-33.2585	4.2E-240	-110.63	-98.3159	-110.63	-98.3159	

Didapatkan nilai parameter penduga seperti berikut

Parameter	Nilai
$\hat{\beta}_0$	13003.441
$\hat{\beta}_2$	7858.771
$\hat{\beta}_3$	-151.236
$\hat{\beta}_4$	-104.473

Model Formula

Dari hasil perhitungan parameter, model regresi yang didapatkan adalah

$$\hat{Y} = 13003.4 + 7858.8X_1 - 151.2X_2 - 104.5X_3$$

Koefisien X_1 dapat diinterpretasikan rata-rata Y akan naik sebesar 7858.8 jika X_1 naik satu satuan dan X_2 dan X_3 tetap.

Koefisien X_2 dapat diinterpretasikan rata-rata Y akan turun sebesar 151.2 jika X_2 naik satu satuan dan X_1 dan X_3 tetap.

Koefisien X_3 dapat diinterpretasikan rata-rata Y akan turun sebesar 104.5 jika X_3 naik satu satuan dan X_1 dan X_2 tetap.



04

R Program

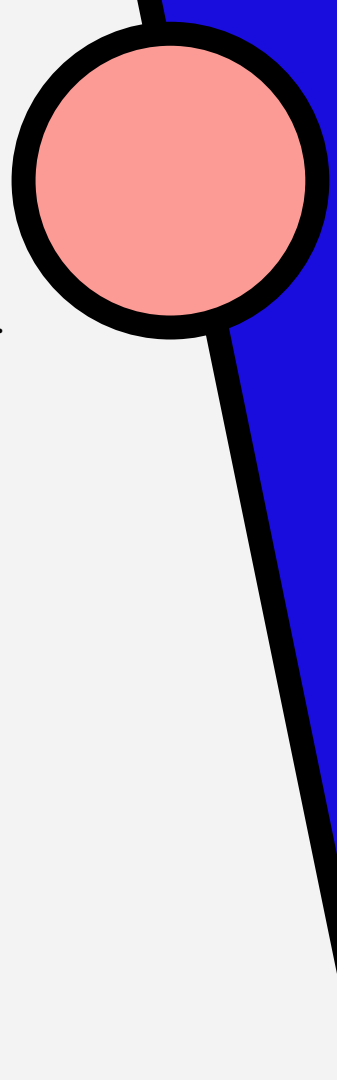
Otomatis :)

R

R merupakan bahasa dan environment yang digunakan untuk perhitungan dan grafik statistika.

R dapat digunakan untuk mencari nilai penduga regresi serta membuat scatterplot secara relatif cepat untuk data yang sangat besar.

Kami akan menggunakan Rstudio untuk mengolah data.



Code

Berikut merupakan code R sederhana untuk mencari model regresi dari data yang disimpan di dalam file diamonds.csv

```
data <-  
  read.csv("C:/Users/akimp/Desktop/anreg  
prak 5/diamonds.csv")  
View(data)  
  
model <-  
  lm(price ~ carat + depth + table, data)  
summary(model)
```

Steps

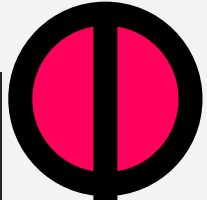
```
data <-  
  read.csv("C:/Users/akimp/Desktop/anreg  
           prak 5/diamonds.csv")  
View(data)
```

Data table dari diamond.csv
dibaca terlebih dahulu

```
model <-  
  lm(price ~ carat + depth + table, data)  
summary(model)
```

Model linear dibuat dengan
memasukkan variabel.

Hasil dari model didapatkan
dengan summary.



Result

```
lm(formula = price ~ carat + depth + table, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-18288.0	-785.9	-33.2	527.2	12486.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13003.441	390.918	33.26	<2e-16 ***
carat	7858.771	14.151	555.36	<2e-16 ***
depth	-151.236	4.820	-31.38	<2e-16 ***
table	-104.473	3.141	-33.26	<2e-16 ***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1526 on 53936 degrees of freedom

Multiple R-squared: 0.8537, Adjusted R-squared: 0.8537

F-statistic: 1.049e+05 on 3 and 53936 DF, p-value: < 2.2e-16

Parameter	Nilai
$\hat{\beta}_0$	13003.441
$\hat{\beta}_2$	7858.771
$\hat{\beta}_3$	-151.236
$\hat{\beta}_4$	-104.473

Model Formula

Dari hasil perhitungan parameter, model regresi yang didapatkan adalah

$$\hat{Y} = 13003.4 + 7858.8X_1 - 151.2X_2 - 104.5X_3$$

Koefisien X_1 dapat diinterpretasikan rata-rata Y akan naik sebesar 7858.8 jika X_1 naik satu satuan dan X_2 dan X_3 tetap.

Koefisien X_2 dapat diinterpretasikan rata-rata Y akan turun sebesar 151.2 jika X_2 naik satu satuan dan X_1 dan X_3 tetap.

Koefisien X_3 dapat diinterpretasikan rata-rata Y akan turun sebesar 104.5 jika X_3 naik satu satuan dan X_1 dan X_2 tetap.



05

Result Comparison

Banding hasil metode

Comparison

Variabel	Penduga Parameter Regresi	Excel			R
		Manual	Matriks	Data Analysis	
Intercept	$\hat{\beta}_0$	13003.441	13003.441	13003.441	13003.441
Carat	$\hat{\beta}_2$	7858.771	7858.771	7858.771	7858.771
Depth	$\hat{\beta}_3$	-151.236	-151.236	-151.236	-151.236
Table	$\hat{\beta}_4$	-104.473	-104.473	-104.473	-104.473

Hasil dari kedua metode, Excel dan R Program, adalah **sama** untuk data diamonds.csv

Thank You

Any questions?

