


Nama: AUREL REGINA NIM: 065002300023	 UNIVERSITAS TRISAKTI PRAKTIKUM STATISTIKA	MODUL 12 STATISTIKA Nama Dosen: Dedy Sugiarto
Hari/Tanggal: Rabu, 29 Mei 2024		Nama Aslab: 1. Tarum Widyasti (064002200027) 2. Kharisma Maulida (064002200024)

Modul 9

Regresi Linier Sederhana dan Berganda

Teori Singkat

Pada regresi linier akan dibicarakan masalah pendugaan atau peramalan sebuah variabel dependen Y dengan sebuah variabel independen X yang telah diketahui nilainya.

Model persamaan linier yang digunakan di sini adalah : $\hat{y} = a + bx$.

Regresi linier berganda Jika variabel dependen-nya dihubungkan dengan lebih dari satu variabel independen, maka persamaan yang dihasilkan adalah persamaan regresi linier berganda (*multiple linier regression*). Dalam hal ini kita membatasi pada kasus dua peubah bebas X_1 dan X_2 saja. Dengan hanya dua peubah bebas, persamaan regresi contohnya menjadi :

$$\hat{y} = b_0 + b_1x_1 + b_2x_2$$

Salah satu ukuran kebaikan model adalah dengan melihat koefisien determinasi R^2 yang menyatakan proporsi keragaman variabel Y yang dapat dijelaskan oleh variabel X. Namun penggunaan yang lebih baik adalah dengan menggunakan nilai **R-Sq(adj)**, yang merupakan nilai estimasi yang tidak bias (*unbiased estimate*) dari populasi.

ELEMEN KOMPETENSI I

No	X	Y
1	40	385
2	20	400
3	25	395
4	20	365
5	30	475
6	50	440

Misalkan ingin dilakukan pendugaan terhadap nilai penjualan dalam USD (variabel Y) berdasarkan nilai biaya iklan yang dikeluarkan dalam USD (variabel X) di suatu perusahaan. Data sampel dalam 12 bulan terakhir adalah sebagai berikut :

- buatlah persamaan regresi untuk menduga penjualan mingguan (Y) berdasarkan pengeluaran iklan (X).
- Hitunglah R-square (Koefisien determinasi) dan korelasinya.

R Studio:

```
df_nama=read.delim("clipboard")
View(df_nama)
model_reg=lm(df_nama$Y~df_nama$X)
summary(model_reg)

> df_aurel=read.delim("clipboard")
> View(df_aurel)
> model_reg=lm(df_aurel$Y~df_aurel$X)
> summary(model_reg)

Call:
lm(formula = df_aurel$Y ~ df_aurel$X)

Residuals:
    1      2      3      4      5      6 
-37.399   4.653  -7.110 -30.347  66.127   4.075 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   368.295     50.184   7.339  0.00184 **
df_aurel$X     1.353      1.534   0.882  0.42760
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.17 on 4 degrees of freedom
Multiple R-squared:  0.1628,    Adjusted R-squared:  -0.04648 
F-statistic: 0.7779 on 1 and 4 DF,  p-value: 0.4276
```

Nilai R Squarenya adalah 0.1628

Korelasi nya adalah 0.40350105

Deskripsi (minimal 4 baris)

R menunjukkan hubungan antara variabel independen (X) dan variabel dependen (Y) teridentifikasi sebagai lemah, dengan koefisien determinasi (R^2) sebesar 0.1628. P-value untuk slope menunjukkan bahwa hubungan antara X dan Y tidak signifikan secara statistik. Intercept dari model adalah 368.295.

Python:

```
import numpy as np
import pandas as pd

# Data
data = {
    "X": [40, 20, 25, 20, 30, 50],
    "Y": [385, 400, 395, 365, 475, 440]
}
df = pd.DataFrame(data)

# Menghitung rata-rata
mean_X = np.mean(df["X"])
mean_Y = np.mean(df["Y"])

# Menghitung b1 (slope)
df['XY'] = df['X'] * df['Y']
df['X^2'] = df['X'] ** 2
b1 = (sum(df['XY']) - len(df) * mean_X * mean_Y) / (sum(df['X^2']) - len(df) * mean_X ** 2)

# Menghitung b0 (intersep)
b0 = mean_Y - b1 * mean_X

# Persamaan regresi
regression_eq = f"Y = {b0:.2f} + {b1:.2f}X"

# Menghitung koefisien korelasi (r)
correlation_matrix = np.corrcoef(df["X"], df["Y"])
r = correlation_matrix[0, 1]

# Menghitung koefisien determinasi ( $R^2$ )
R2 = r ** 2

# Menampilkan hasil dengan deskripsi
print(f"Rata-rata X: {mean_X:.2f}")
print(f"Rata-rata Y: {mean_Y:.2f}")
print(f"Nilai intersep (b0): {b0:.2f}")
print(f"Nilai kemiringan (b1): {b1:.2f}")
print(f"Koefisien korelasi (r): {r:.2f}")
print(f"Koefisien determinasi ( $R^2$ ): {R2:.2f}")
print(f"Persamaan regresi: {regression_eq}")
```

```
# Menampilkan DataFrame
print("\nDataFrame:")
print(df)
```

Output:

```
print(f"Rata-rata X: {mean_X:.2f}")
print(f"Rata-rata Y: {mean_Y:.2f}")
print(f"Nilai intersep (b0): {b0:.2f}")
print(f"Nilai kemiringan (b1): {b1:.2f}")
print(f"Koefisien korelasi (r): {r:.2f}")
print(f"Koefisien determinasi (R^2): {R2:.2f}")
print(f"Persamaan regresi: {regression_eq}")

# Menampilkan DataFrame
print("\nDataFrame:")
print(df)
```

Rata-rata X: 30.83
Rata-rata Y: 410.00
Nilai intersep (b0): 368.29
Nilai kemiringan (b1): 1.35
Koefisien korelasi (r): 0.40
Koefisien determinasi (R^2): 0.16
Persamaan regresi: Y = 368.29 + 1.35X

```
DataFrame:
   X  Y  XY  X^2
0  40 385 15400 1600
1  20 400  8000  400
2  25 395  9875  625
3  20 365  7300  400
4  30 475 14250  900
5  50 440 22000 2500
```

Deskripsi (minimal 4 baris)

menemukan rata2 variabel X dan Y, nilai kemiringan, intersep, dan menemukan persamaan garis regresi, menghitung statistik seperti koefisien korelasi dan determinasi, serta menyajikan hubungan antara kedua variabel dalam bentuk persamaan regresi dan statistik terkait.

Excel:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	X	Y	x^2	xy	cara lain		(x-xbar)	(y-ybar)	(x-xbar)(y-ybar)	(x-xbar)^2		y_duga	(y_duga - ybar)^2	(y-ybar)^2
	40	385	1600	15400	*buku anderson		9.16666667	-25	-229.166667	84.0277778		422.3988439	153.7313308	625
	20	400	400	8000			-10.83333333	-10	108.3333333	117.3611111		395.3468208	214.7156604	100
	25	395	625	9875			-5.833333333	-15	87.5	34.02777778		402.1098266	62.25483645	225
	20	365	400	7300			-10.83333333	-45	487.5	117.3611111		395.3468208	214.7156604	2025
	30	475	900	14250			-0.833333333	65	-54.1666667	0.694444444		408.8728324	1.270506866	4225
	50	440	2500	22000			19.16666667	30	575	367.3611111		435.9248555	672.0981322	900
sum	185	2460	6425	76825					975	720.8333333			1318.786127	8100
y_bar		410											SSR	SST
x_bar	30.83333333													
n	6				nama: Aurel Regina									
b1	1.352601156						b1	1.352601156				rsquare	0.1628131021	
b0	368.2947977						b0	368.2947977				correlation	0.4035010559	
y_duga = 368.2948 + 1.352601x														
artinya penambahan satu satuan x akan meningkatkan 1.352601 satuan y														

Nilai R Squarenya adalah 0.1628

Korelasi nya adalah 0.40350105

Deskripsi (minimal 4 baris)

variabel x dan y, menghasilkan persamaan regresi $(y_{\text{duga}} = -35 + 1.3x)$. Ini mengindikasikan bahwa setiap kenaikan satu satuan pada x akan meningkatkan y

sebesar 1.3 satuan. Koefisien determinasi (R^2) sebesar 0.1628, bahwa 16.28% variabilitas y dapat dijelaskan oleh x. koefisien korelasi sebesar 0.4035, menunjukkan adanya hubungan positif moderat antara x dan y.

TUGAS

Delapan orang lulusan dipilih secara random dan ditanyakan berapa nilai IPK kelulusan (X) serta total gaji pertama kali (Y) dalam jutaan rupiah. Data yang diperoleh sebagai berikut:

Nama	X	Y
Amir	2,8	5,4
Agus	2,5	5,1
Charlie	3,5	7,2
Debi	3,1	6,2
Faishal	3,0	6,0
Jojo	3,8	7,5
Kamal	3,3	6,8
Caca	3,5	8,9

Pertanyaan:

- Hitung nilai intersep (b_0) dan slope/koefisien regresi (b_1). Jelaskan perhitungan manualnya
- Tuliskan persamaan regresi linear sederhana dan interpretasikan nilai b_1 dalam regresi ini.
- Hitung nilai koefisien determinasi dan koefisien korelasi serta interpretasikan nilainya. Jelaskan perhitungan manualnya

R Studio:

```
>df_aurel=read.delim("clipboard")
> View(df_aurel)
> model_reg=lm(df_aurel$Y~df_aurel$X)
> summary(model_reg)
```

```

> df_aurel=read.delim("clipboard")
> View(df_aurel)
> model_reg=lm(df_aurel$Y~df_aurel$X)
> summary(model_reg)

Call:
lm(formula = df_aurel$Y ~ df_aurel$X)

Residuals:
    Min       1Q   Median       3Q      Max
-0.66967 -0.23146 -0.19354 -0.04362  1.48078

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.3360     1.9930  -0.670  0.52756
df_aurel$X     2.5015     0.6205   4.031  0.00687 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6934 on 6 degrees of freedom
Multiple R-squared:  0.7304,    Adjusted R-squared:  0.6854
F-statistic: 16.25 on 1 and 6 DF,  p-value: 0.006869

```

Deskripsi (minimal 4 baris)

variabel $Y \sim X$, dengan koefisien intersep -1.3360 tidak signifikan, dan koefisien X 2.5015 signifikan pada tingkat 0.01. R-squared sebesar 0.7304, model ini mampu menjelaskan sekitar 73% dari variasi data, dengan nilai adjusted R-squared sebesar 0.6854. Penilaian p untuk model ini adalah 0.006869, menunjukkan signifikansi statistik. Selain itu, residual standard error dari model adalah 0.6934.

Python:

```

import numpy as np
import pandas as pd

# Data
data={
    "X": [2.8,2.5,3.5,3.1,3,3.8,3.3,3.5],
    "Y": [5.4, 5.1, 7.2, 6.2, 6, 7.5, 6.8, 8.9]
}
df = pd.DataFrame(data)

# Menghitung rata-rata
mean_X = np.mean(df["X"])
mean_Y = np.mean(df["Y"])

# Menghitung b1 (slope)

```

```

df['XY'] = df['X'] * df['Y']
df['X^2'] = df['X'] ** 2
b1 = (sum(df['XY'])- len(df) * mean_X * mean_Y) / (sum(df['X^2'])- len(df) *
mean_X ** 2)

# Menghitung b0 (intersep)
b0 = mean_Y- b1 * mean_X

# Persamaan regresi
regression_eq = f"Y = {b0:.2f} + {b1:.2f}X"

# Menghitung koefisien korelasi (r)
correlation_matrix = np.corrcoef(df["X"], df["Y"])
r = correlation_matrix[0, 1]

# Menghitung koefisien determinasi (R^2)
R2 =r ** 2

# Menampilkan hasil dengan deskripsi
print(f'Rata-rata X: {mean_X:.2f}')
print(f'Rata-rata Y: {mean_Y:.2f}')
print(f'Nilai intersep (b0): {b0:.2f}')
print(f'Nilai kemiringan (b1): {b1:.2f}')
print(f'Koefisien korelasi (r): {r:.2f}')
print(f'Koefisien determinasi (R^2): {R2:.2f}')
print(f'Persamaan regresi: {regression_eq}')

# Menampilkan DataFrame
print("\nDataFrame:")
print(df)

```

HASIL :

```

Rata-rata X: 3.19
Rata-rata Y: 6.64
Nilai intersep (b0): -1.34
Nilai kemiringan (b1): 2.50
Koefisien korelasi (r): 0.85
Koefisien determinasi (R^2): 0.73
Persamaan regresi: Y = -1.34 + 2.50X

```

```

DataFrame:
   X  Y  XY  X^2
0  2.8  5.4  15.12  7.84
1  2.5  5.1  12.75  6.25
2  3.5  7.2  25.20  12.25
3  3.1  6.2  19.22  9.61
4  3.0  6.0  18.00  9.00
5  3.8  7.5  28.50  14.44
6  3.3  6.8  22.44  10.89
7  3.5  8.9  31.15  12.25

```

Deskripsi (minimal 4 baris)

Rata-rata variabel X tercatat sebesar 3.19, rata-rata variabel Y adalah 6.64. Intersep nilai (b_0) dari regresi adalah -1.34, dengan kemiringan (b_1) sebesar 2.50. Koefisien korelasi (r) mencapai 0.85, sementara koefisien determinasi (R^2) mencapai 0.73. Persamaan regresi yang terbentuk adalah $Y = -1.34 + 2.50X$. Data dipresentasikan dalam bentuk dataframe dengan kolom X, Y, XY, dan X^2 untuk 8 baris data.

Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		X	Y	x^2	xy	cara lain		(x-xbar)	(y-ybar)	(x-xbar)(y-ybar)	(x-xbar) ²		y_duga	(y_duga - ybar)	(y-ybar) ²
2			2.8	5.4	7.84	15.12	*buku anderson	-0.39	-1.24	0.4795	0.15015625		5.67	0.939604200	1.53140625
3			2.5	5.1	6.25	12.75		-0.69	-1.54	1.0570	0.47265625		4.92	2.957651098	2.36390625
4			3.5	7.2	12.25	25.2		0.31	0.56	0.1758	0.09765625		7.42	0.611084937	0.31640625
5			3.1	6.2	9.61	19.22		-0.09	-0.44	0.0383	0.00765625		6.42	0.047909059	0.19140625
6			3	6	9	18		-0.19	-0.64	0.1195	0.03515625		6.17	0.219990577	0.40640625
7			3.8	7.5	14.44	28.5		0.61	0.86	0.5283	0.37515625		8.17	2.347543897	0.74390625
8			3.3	6.8	10.89	22.44		0.11	0.16	0.0183	0.01265625		6.92	0.079196607	0.02640625
9			3.5	8.9	12.25	31.15		0.31	2.26	0.7070	0.09765625		7.42	0.611084937	0.31640625
10	sum		25.5	53.1	82.53	172.38				3.1238	1.24875000	SUM		7.814065315	10.69875
11	y_bar			6.6375										SSR	SST
12	x_bar		3.1875												
13	n		8			Aurel Regina							rsquare	0.73037	
14	b1		2.50					b1		2.50			correlation	0.85462	
15	b0		-1.34					b0		-1.34					
16															

Deskripsi (minimal 4 baris)

Data dipresentasikan dalam sebuah dataframe dengan kolom X, Y, XY, dan X^2 untuk 8 baris data. Hasil perhitungan dari dataframe ini akan menjadi dasar untuk membuat prediksi serta menarik kesimpulan yang relevan dari data yang tersedia.

CEK LIST (✓)

1. Melakukan regresi linier sederhana dan berganda.

(✓)

KESIMPULAN

Regresi linier berganda Jika variabel dependen-nya dihubungkan dengan lebih dari satu variabel independen, maka persamaan yang dihasilkan adalah persamaan regresi linier berganda (multiple linier regression). ukuran kebaikan model adalah dengan melihat koefisien determinasi R^2 yang menyatakan proporsi keragaman variabel Y yang dapat dijelaskan oleh variabel X. Namun penggunaan yang lebih baik adalah dengan menggunakan nilai **R-Sq(adj)**, yang merupakan nilai estimasi yang tidak bias (*unbiased estimate*) dari populasi. pentingnya regresi linier sebagai alat analisis yang efektif dalam memahami hubungan antara variabel dalam suatu studi.

LINK GITHUB :

<https://github.com/aurelregina/prak9probabilitas.git>

FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian (menit)
Melakukan regresi linier sederhana dan berganda.	Sulit	Cukup Tertarik	30 Menit

Keterangan Tingkat Kesulitan

- 1: Sangat Mudah
- 2: Mudah
- 3: Biasa
- 4: Sulit
- 5: Sangat Sulit

Keterangan Tingkat Ketertarikan

- 1: Tidak Tertarik
- 2: Cukup Tertarik
- 3: Tertarik
- 4: Sangat Tertarik