


Nama: Nadhira Anindita Ralena NIM: 065002300021	 UNIVERSITAS TRISAKTI	MODUL 12 STATISTIKA Nama Dosen: Dedy Sugiarto
Hari/Tanggal: Rabu, 29 Mei 2024	PRAKTIKUM STATISTIKA	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

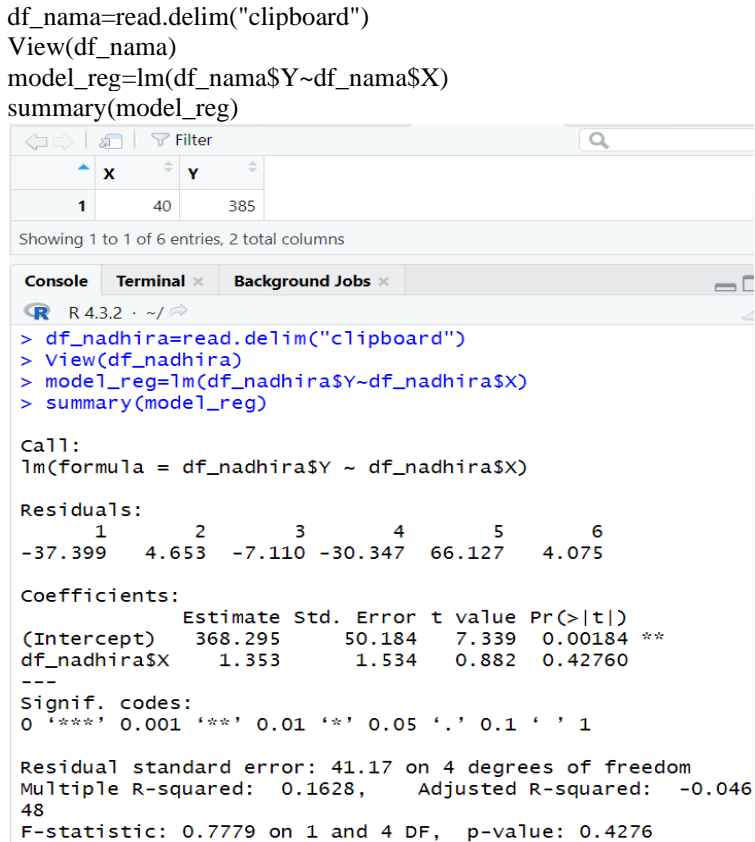
No	X	Y
1	40	385
2	20	400
3	25	395
4	20	365
5	30	475
6	50	440

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)
```



```
Call:
lm(formula = df_nadhira$Y ~ df_nadhira$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_nadhira$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)

Ringkasan output analisis regresi linear sederhana menggunakan R menunjukkan

bahwa 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

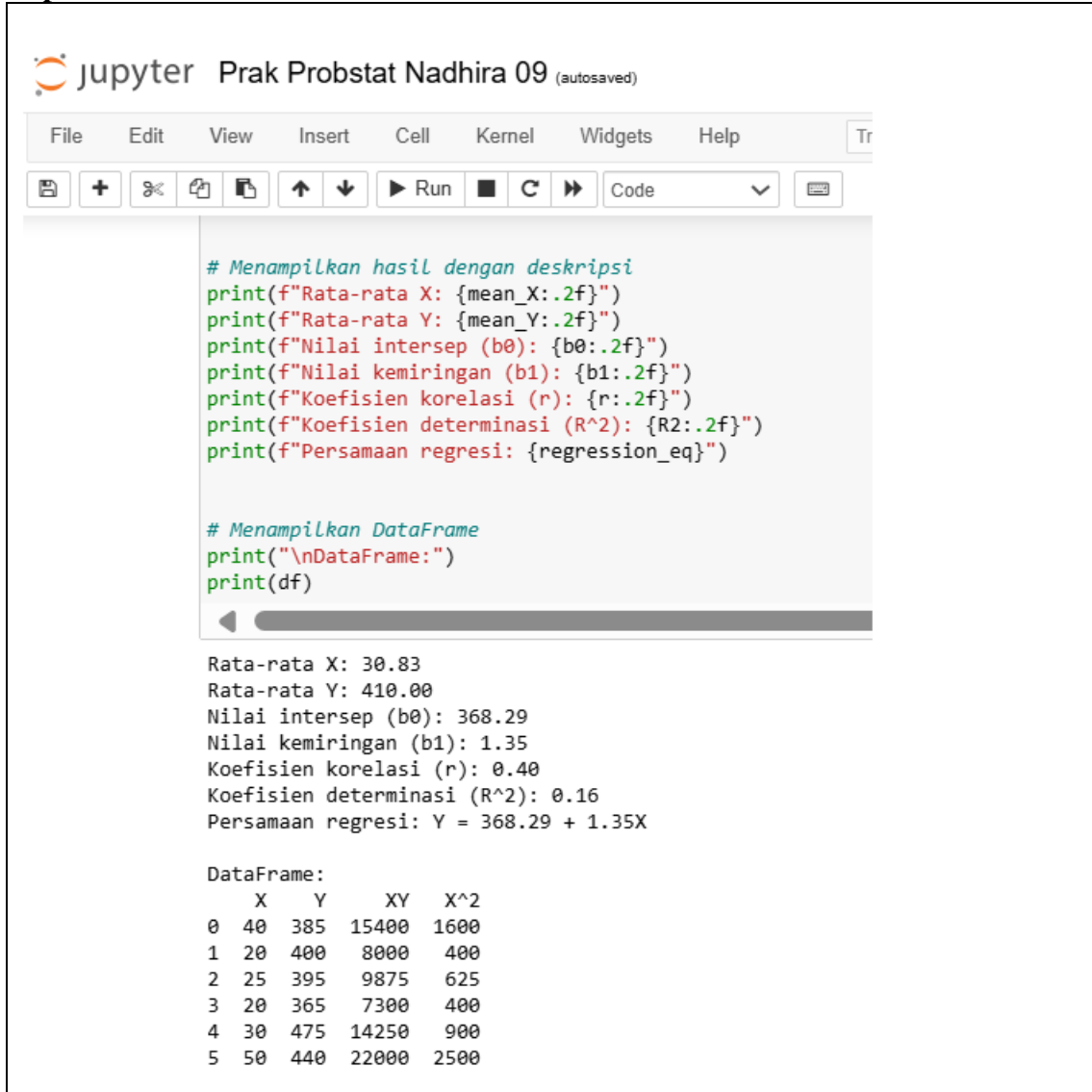
# Persamaan regresi
regression_eq = f"Y = {b0:.2f} + {b1:.2f}X"

# 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:


The screenshot shows a Jupyter Notebook titled "Prak Probstat Nadhira 09 (autosaved)". The code cell contains the following Python code:

```
# 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)
```

The output of the code is as follows:

```
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
```

Below the text output, a DataFrame is displayed:

	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)

Kode tersebut menjalankan regresi linear sederhana antara variabel X dan Y untuk 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
1	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
2		40	385	1600	15400	*buku anderson		9,166667	-25	-229,1666667	84,02778		422,3988439	153,7313308	625
3		20	400	400	8000			-10,83333	-10	108,3333333	117,3611		395,3468208	214,7156604	100
4		25	395	625	9875			-5,83333	-15	87,5	34,02778		402,1098266	62,25483645	225
5		20	365	400	7300			-10,83333	-45	487,5	117,3611		395,3468208	214,7156604	2025
6		30	475	900	14250			-0,83333	65	-54,16666667	0,694444		408,8728324	1,270506866	4225
7		50	440	2500	22000			19,16667	30	575	367,3611		435,9248555	672,0981322	900
8	sum	185	2460	6425	76825					975	720,8333			1318,786127	8100
9	y_bar		410											SSR	SST
10	x_bar	30,83333													
11	n	6											rsquare	0,162813102	
12	b1	1,352601						b1	1,352601156				correlation	0,403501056	
13	b0	368,2948						b0	368,2947977						
14															
15	y_duga = 368,2948 + 1.352601x												Nama : Nadhira Anindita Ralena		
16	artinya penambahan satu satu satuan x akan meningkatkan 1.352601 satuan y												NIM : 065002300021		
17															
18															

Nilai R Squarenya adalah 0.162
Korelasi nya adalah 0.40350105

Deskripsi (minimal 4 baris)

Lembar kerja Excel tersebut menampilkan analisis regresi linear sederhana antara 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, menandakan bahwa 16.28% variabilitas y dapat dijelaskan oleh x. Sementara itu, 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 (b0) dan slope/koefisien regresi (b1). Jelaskan perhitungannya manualnya
- Tuliskan persamaan regresi linear sederhana dan interpretasikan nilai b1 dalam regresi ini.

- c. Hitung nilai koefisien determinasi dan koefisien korelasi serta interpretasikan nilainya. Jelaskan perhitungan manualnya

R Studio:

```
>df_nadhira2=read.delim("clipboard")
> View(df_nadhira2)
> model_reg=lm(df_nadhira2$Y~df_nadhira2$X)
> summary(model_reg)
```

```
> df_nadhira2=read.delim("clipboard")
> View(df_nadhira2)
> model_reg=lm(df_nadhira2$Y~df_nadhira2$X)
> summary(model_reg)
```

Call:
lm(formula = df_nadhira2\$Y ~ df_nadhira2\$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_nadhira2\$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)

Hasil analisis regresi menunjukkan bahwa model persamaan $Y \sim X$, dengan koefisien intersep sebesar -1.3360 (yang tidak signifikan), dan koefisien X sebesar 2.5015 (yang signifikan pada tingkat 0.01). Dengan nilai 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 :

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

Rata-rata X: 3.19
Rata-rata Y: 6.64
Nilai intersep (b_0): -1.34
Nilai kemiringan (b_1): 2.50
Koefisien korelasi (r): 0.85
Koefisien determinasi (R^2): 0.73
Persamaan regresi: $Y = -1.34 + 2.50X$

	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 dari variabel X tercatat sebesar 3.19, sedangkan rata-rata dari 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:

[illegible]**Deskripsi (minimal 4 baris)**

Data dipresentasikan dalam sebuah dataframe dengan kolom X, Y, XY, dan X² 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

Dari praktikum ini, kita bisa menyimpulkan bahwa regresi linier sederhana telah berhasil digunakan untuk menganalisis hubungan antara kedua variabel. Model regresi yang terbentuk menunjukkan kekokohan dan signifikansi, menegaskan bahwa variabel independen memiliki dampak yang signifikan terhadap variabel dependen. Analisis ini memberikan wawasan yang berharga tentang dinamika hubungan antara kedua variabel, memperkuat pemahaman kita tentang faktor-faktor yang memengaruhi variabel dependen dalam konteks spesifik yang diamati. Kesimpulan ini menggarisbawahi pentingnya regresi linier sebagai alat analisis yang efektif dalam memahami hubungan antara variabel dalam suatu studi.

LINK GITHUB :

<https://github.com/NadhiraAninditaRalena/probstat9prak.git>

FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian (menit)
Melakukan regresi linier sederhana dan berganda.	Biasa	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