


<p>Nama: MUHAMMAD RAYYAN NAUFAL</p> <p>NIM: 065002300024</p>	 ,UNIVERSITAS TRISAKTI PRAKTIKUM STATISTIKA	<p>MODUL 12 STATISTIKA</p> <p>Nama Dosen: Dedy Sugiarto</p>
<p>Hari/Tanggal: Rabu, 29 May 2024</p>		<p>Nama Aslab:</p> <ol style="list-style-type: none"> 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_rayyan=read.delim("clipboard")
> View(df_rayyan)
> model_reg=lm(df_rayyan$Y~df_rayyan$X)
> summary(model_reg)
```

```
Call:
lm(formula = df_rayyan$Y ~ df_rayyan$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_rayyan$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

Langkah pertama adalah membuat variabel dataframe untuk membaca data

dari clipboard. Setelah itu, gunakan fungsi view untuk melihat data frame yang telah dibuat. Kemudian, terapkan rumus statistik untuk mencari nilai regresi, yang dilakukan dengan membaca data frame dan mengurangi kolom Y dengan kolom X. Terakhir, gunakan fungsi summary untuk melihat hasil dari perhitungan regresi tersebut.

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:

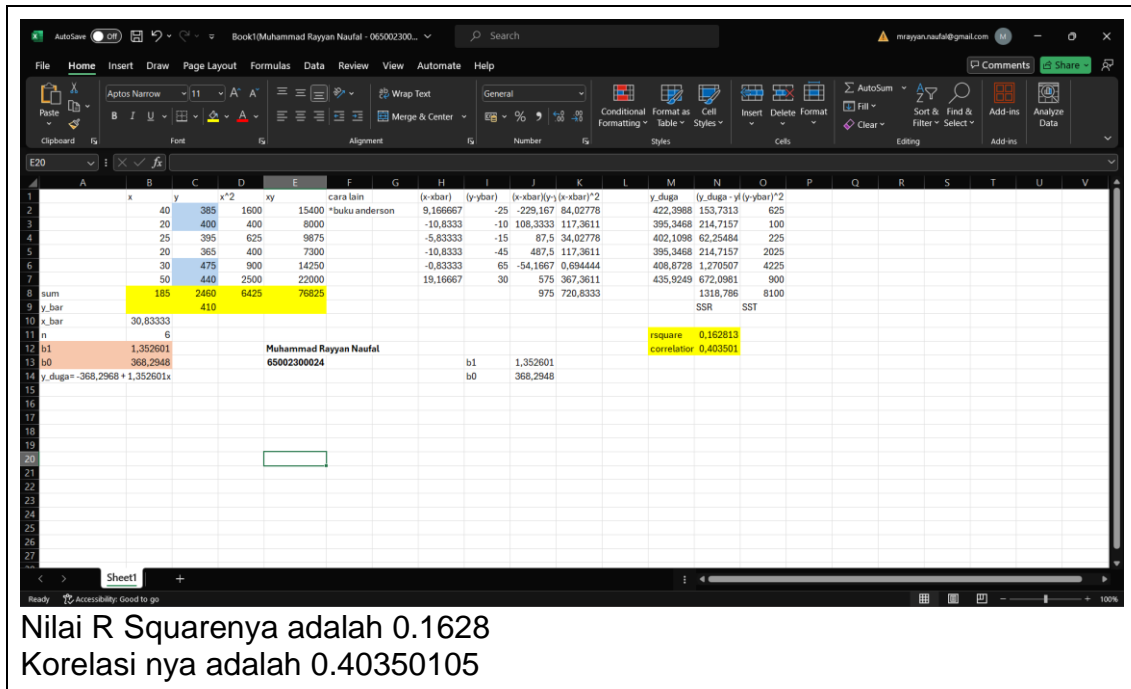
```
... Rata-rata X: 30.83
    Rata-rata Y: 410.00
    Nilai intersep (b0): 368.29
    Nilai kemiringan (b1): 1.35
    Koefisien korelasi (r): 0.40
    Koefisien determinasi (R^2): 0.16
    Persamaan regresi: Y = 368.29 + 1.35X

    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

Pertama, kita membuat DataFrame dari data kolom X dan Y. Lalu, kita Menghitung rata-rata masing-masing kolom. Selanjutnya, kita menghitung Kemiringan (b1) dan intersep (b0) untuk persamaan regresi $(Y = b_0 + b_1X)$. Kemudian, kita menghitung koefisien korelasi (r) dan koefisien determinasi (R^2) Untuk menilai model regresi. Hasilnya, termasuk persamaan regresi, Ditampilkan untuk memberikan gambaran lengkap tentang analisis regresi ini.

Excel:



Deskripsi

X dan Y adalah data frame. X^2 adalah hasil pangkat X, begitu pula dengan Y^2 . $(X - \bar{X})$ adalah nilai data X yang dikurangi rata-rata X (\bar{X}). Referensi absolut ke sel di kolom B dan baris 10 ($\$B\10) memastikan bahwa referensi ini tetap konstan, tidak berubah ketika rumus disalin atau diisi ke sel lain. Begitu pula dengan $(Y - \bar{Y})$.

Selanjutnya adalah mengalikan \bar{X} dengan \bar{Y} . Berikutnya adalah hasil \bar{X} pangkat 2. Y dugaan adalah nilai prediksi yang dihitung dari b_0 ditambah b_1 dikalikan dengan data dari kolom X. Sum adalah hasil penjumlahan masing-masing kolom. \bar{X} dan \bar{Y} pada baris bawah adalah hasil rata-rata dari data X dan Y. n adalah jumlah data.

$(Y \text{ dugaan} - \bar{Y})^2$ adalah hasil perhitungan dari kolom Y dugaan dikurangi \bar{Y} , kemudian dipangkatkan 2. $(Y - \bar{Y})^2$ adalah hasil dari kolom Y dikurangi \bar{Y} , kemudian dipangkatkan 2. Nilai R^2 adalah hasil penjumlahan kolom $(Y \text{ dugaan} - \bar{Y})^2$ dibagi dengan penjumlahan kolom $(Y - \bar{Y})^2$. Korelasi adalah akar atau sqrt dari nilai R^2 .

Jadi, nilai regresi adalah b_0 ditambah b_1 , dan peningkatan satu satuan X akan meningkatkan nilai Y sesuai dengan b_1 .

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/koeffisien 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_rayyan=read.delim("clipboard")
> View(df_rayyan)
> model_reg=lm(df_rayyan$y~df_rayyan$x)
> summary(model_reg)

Call:
lm(formula = df_rayyan$y ~ df_rayyan$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_rayyan$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

Model regresi menunjukkan bahwa intersep signifikan ($p < 0.01$), namun variabel x tidak signifikan ($p > 0.05$). Koefisien untuk variabel x adalah 1.353, dengan nilai R-squared sebesar 0.1628, mengindikasikan bahwa variabel x hanya mampu menjelaskan 16.28% variasi dalam variabel y. Model ini dianggap kurang sesuai untuk tujuan prediksi karena nilai p yang tinggi dan adjusted R-squared yang negatif.

Python:

```

# 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)

```

✓ 1.5s

```

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

Kode tersebut mengimplementasikan regresi linear sederhana menggunakan dataset X dan Y. Hasil analisis menunjukkan bahwa rata-rata X adalah 30.83 dan rata-rata Y

adalah 410.00. Diperoleh nilai intersep (b_0) sebesar 368.29 dan nilai kemiringan (b_1) sebesar 1.35. Koefisien korelasi (r) sebesar 0.40 menunjukkan adanya hubungan sedang antara variabel X dan Y. Selain itu, nilai koefisien determinasi (R^2) sebesar 0.16 mengindikasikan bahwa hanya 16% variasi dalam Y yang dapat dijelaskan oleh X. Persamaan regresi yang dihasilkan adalah $Y=368.29+1.35X$. DataFrame tersebut berisi data awal serta hasil perhitungan regresi tambahan.

Excel:

AutoSave

Book1[Muhammad Rayyan Naufal - 065002300... - Saved]

Search

muhammad.naufal@gmail.com

FileHomeInsertDrawPage LayoutFormulasDataReviewViewAutomateHelp

Clipboard

Font

Alignment

Number

General

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Cells

AutoSum

Fill

Sort & Filter

Find & Select

Add-ins

Analyze Data

CommentsShare

Deskripsi

Dalam gambar tersebut, terdapat hasil perhitungan regresi linear dengan memanfaatkan data x dan y. Tabel tersebut terdiri dari kolom-kolom yang mencakup nilai x, y, x^2 , xy , dan langkah-langkah perhitungan lainnya seperti $(x-\bar{x})(x-\bar{x})$, $(y-\bar{y})(y-\bar{y})$, $(x-\bar{x})(y-\bar{y})$, serta $(x-\bar{x})^2$ dan $(y-\bar{y})^2$. Dari hasil perhitungan, diperoleh nilai rata-rata \bar{x} sebesar 3.1875 dan rata-rata \bar{y} sebesar 6.6375. Selain itu, didapatkan nilai kemiringan ($b_1 = 2.5015$) dan intersep ($b_0 = -1.3360$), persamaan regresi $y=2.5015x-1.3360$, serta koefisien determinasi ($R^2 = 0.7303$) dan korelasi (0.8546). Berdasarkan analisis, dapat disimpulkan bahwa setiap peningkatan satu satuan nilai x akan mengakibatkan kenaikan sebesar 1.3260 satuan pada nilai y.

GITHUB

<https://github.com/rayyan-naufal/Uni/upload/main/S2/ProbabilitasStatistika/Praktikum/Modul9>

CEK LIST (✓)

1. Melakukan regresi linier sederhana dan berganda. ()

KESIMPULAN

Regresi linier sederhana dan regresi linier berganda adalah teknik statistik yang digunakan untuk mengidentifikasi keterkaitan antara satu atau beberapa variabel independen (X) dengan variabel dependen (Y). Pada regresi linier sederhana, terdapat satu variabel independen yang digunakan untuk meramalkan variabel dependen, sementara pada regresi linier berganda, lebih dari satu variabel independen digunakan untuk melakukan prediksi.

FORM UMPAN BALIK

Elemen Kompetensi	Tingkat Kesulitan	Tingkat Ketertarikan	Waktu Penyelesaian (menit)
Melakukan regresi linier sederhana dan berganda.	3	3	30

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