


Nama: Francisco NIM: 064002300044	 Praktikum Probabilitas dan Statistika	MODUL 12 Nama Dosen: Drs. Joko Riyono, M.Si
Hari/Tanggal: Sabtu, 8 Juni 2024		Nama Asisten Labratorium 1. Adzriel Yusak Noah (064.20.11) 2. Mohammad Akhmal Firdaus (064.20.34)

Regresi Linear Sederhana dan Berganda

1. 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 $R\text{-Sq}(\text{adj})$, yang merupakan nilai estimasi yang tidak bias (unbiased estimate) dari populasi.

2. Alat dan Bahan

Hardware : Laptop/PC

Software : R Studio

3. Elemen Kompetensi

a. Latihan pertama – Regresi

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:

X	40	20	25	20	30	50	40	20	50	40	25	50
Y	385	400	395	365	475	440	490	420	560	525	480	510

1. buatlah persamaan regresi untuk menduga penjualan mingguan (Y) berdasarkan pengeluaran iklan (X).

```
> df_francisco = read.delim("clipboard")
> head(df_francisco)

> model_reg=lm
  (df_francisco$Y~df_francisco$X)

> summary(model_reg)
```

Output:

```
> df_francisco = read.delim("clipboard")
> head(df_francisco)
   X   Y
1 40 385
2 20 400
3 25 395
4 20 365
5 30 475
6 50 440
> model_reg=lm
(df_francisco$Y~df_francisco$X)
> summary(model_reg)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-87.538	-32.700	8.566	39.118	55.774

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	343.706	44.766	7.678	1.68e-05	***
df_ano\$X	3.221	1.240	2.598	0.0266	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.23 on 10 degrees of freedom

Multiple R-squared: 0.403, Adjusted R-squared: 0.343

F-statistic: 6.751 on 1 and 10 DF, p-value: 0.02657

Penjelasan: Menampilkan data dari nilai penjualan dalam USD (variabel Y) berdasarkan nilai biaya iklan yang dikeluarkan dalam USD (variabel X) di suatu perusahaan lalu menunjukan hasil dari R-squarednya.

2. Hitunglah R-square (Koefisien determinasi)

Nilai R Squarenya adalah 0.403

Output:

```
> model_reg=lm  
  (df_francisco$Y~df_francisco$X)  
  
> summary(model_reg)
```

Residuals:

Min	1Q	Median	3Q	Max
-87.538	-32.700	8.566	39.118	55.774

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	343.706	44.766	7.678	1.68e-05	***
df_ano\$X	3.221	1.240	2.598	0.0266	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50.23 on 10 degrees of freedom

Multiple R-squared: 0.403, Adjusted R-squared: 0.3433

F-statistic: 6.751 on 1 and 10 DF, p-value: 0.02657

b. Latihan Kedua – Tugas

Lakukan analisis regresi untuk memprediksi variabel Price (harga rumah) berdasarkan variabel SqFt (luas tanah), bedrooms dan bathrooms. . Lakukan evaluasi terhadap model regresi yang digunakan.

Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
114300	1790	2	2	2	No	East
114200	2030	4	2	3	No	East
114800	1740	3	2	1	No	East
94700	1980	3	2	3	No	East
119800	2130	3	3	3	No	East
114600	1780	3	2	2	No	North
151600	1830	3	3	3	Yes	West
150700	2160	4	2	2	No	West
119200	2110	4	2	3	No	East
104000	1730	3	3	3	No	East
132500	2030	3	2	3	Yes	East
123000	1870	2	2	2	Yes	East
102600	1910	3	2	4	No	North
126300	2150	3	3	5	Yes	North
176800	2590	4	3	4	No	West

145800	1780	4	2	1	No	West
147100	2190	3	3	4	Yes	East
83600	1990	3	3	4	No	North
111400	1700	2	2	1	Yes	East
167200	1920	3	3	2	Yes	West
116200	1790	3	2	3	No	East
113800	2000	3	2	4	No	North
91700	1690	3	2	3	No	North
106100	1820	3	2	3	Yes	North
156400	2210	4	3	2	Yes	East
149300	2290	4	3	3	No	North
137000	2000	4	2	3	No	West
99300	1700	3	2	2	No	East
69100	1600	2	2	3	No	North
188000	2040	4	3	1	Yes	West
182000	2250	4	3	3	Yes	West
112300	1930	2	2	2	Yes	North
135000	2250	3	3	3	Yes	East
139600	2280	5	3	4	Yes	East
117800	2000	2	2	3	No	North
117100	2080	3	3	3	No	North
117500	1880	2	2	2	No	North
147000	2420	4	3	4	No	West
131300	1720	3	2	1	No	West
108200	1740	3	2	2	No	North
106600	1560	2	2	1	No	East
133600	1840	4	3	2	No	West
105600	1990	2	2	3	No	East
154000	1920	3	2	1	Yes	East
166500	1940	3	3	2	Yes	West
103200	1810	3	2	3	No	East
129800	1990	2	3	2	No	North
90300	2050	3	2	6	No	North
115900	1980	2	2	2	No	East
107500	1700	3	2	3	Yes	North
151100	2100	3	2	3	Yes	East
91100	1860	2	2	3	No	North

117400	2150	2	3	4	No	North
130800	2100	3	2	3	No	North
81300	1650	3	2	3	No	North
125700	1720	2	2	2	Yes	East
140900	2190	3	2	3	Yes	East
152300	2240	4	3	3	No	West
138100	1840	3	3	1	No	West
155400	2090	4	2	1	No	West
180900	2200	3	3	1	No	West
100900	1610	2	2	2	No	North
161300	2220	4	3	2	No	West
120500	1910	2	3	2	No	East
130300	1860	3	2	2	No	West
111100	1450	2	2	1	Yes	North
126200	2210	3	3	4	No	North
151900	2040	4	3	3	No	East
93600	2140	3	2	4	No	North
165600	2080	4	3	3	No	West
166700	1950	3	3	3	Yes	West
157600	2160	4	2	1	No	West
107300	1650	3	2	3	No	North
125700	2040	3	3	2	No	East
144200	2140	3	3	3	No	West
106900	1900	2	2	2	No	North
129800	1930	3	2	2	No	West
176500	2280	4	3	3	Yes	West
121300	2130	3	2	3	No	North
143600	1780	4	2	1	No	West
143400	2190	3	3	4	Yes	East
184300	2140	4	3	2	Yes	West
164800	2050	2	2	1	Yes	West
147700	2410	3	3	2	No	East
90500	1520	2	2	3	No	North
188300	2250	4	3	2	Yes	West
102700	1900	4	2	4	No	North
172500	1880	3	3	1	Yes	West
127700	1930	3	3	2	No	North

97800	2010	2	2	4	No	North
143100	1920	4	2	2	No	West
116500	2150	3	2	2	No	East
142600	2110	3	2	2	No	West
157100	2080	3	3	2	No	East
160600	2150	4	3	3	Yes	West
152500	1970	2	2	1	Yes	West
133300	2440	3	3	3	No	East
126800	2000	2	2	1	Yes	East
145500	2060	3	2	1	No	West
171000	2080	3	3	2	Yes	West
103200	2010	3	2	5	No	North
123100	2260	3	3	5	No	East
136800	2410	3	3	4	No	East
211200	2440	4	3	3	Yes	West
82300	1910	3	2	4	No	East
146900	2530	4	3	4	No	West
108500	2130	3	2	4	No	North
134000	1890	3	2	1	Yes	East
117000	1990	3	3	3	Yes	East
108700	2110	3	2	3	No	East

1. Regresi Berganda Script

```

> df_francisco = read.delim("clipboard")
> df_francisco$Brick_dummy <- ifelse(df_francisco$Brick == "weat", 1,0)
> df_francisco$N_dummy1 <- ifelse(df_francisco$Neighborhood == "west", 1.0)
> df_francisco$N_dummy2 <- ifelse(df_francisco$Neighborhood == "North", 1.0)
> model=lm(df_francisco$Price~df_francisco$SqFt+df_francisco$Bedrooms+df_francisco$
ooms+df_francisco$N_dummy1+df_francisco$N_dummy2)
> summary(model)

```


Output:

Residuals:

Min	1Q	Median	3Q	Max
-31771	-7432	-234	7946	31739

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	23476.62	11587.57	2.026	0.04535	*
df_ano\$SqFt	33.38	6.85	4.872	4.01e-06	***
df_ano\$Bedrooms	2247.85	2116.26	1.062	0.29064	
df_ano\$Bathrooms	8572.06	2912.53	2.943	0.00401	**
df_ano\$brick_dummy	17816.32	2650.49	6.722	1.02e-09	***
df_ano\$N_dummy1	28076.12	3059.80	9.176	5.02e-15	***
df_ano\$N_dummy2	-6985.46	3016.65	-2.316	0.02256	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12270 on 103 degrees of freedom
Multiple R-squared: 0.8106, Adjusted R-squared: 0.7996
F-statistic: 73.47 on 6 and 103 DF, p-value: < 2.2e-16

R Square

Residual standard error: 12270 on 103 degrees of freedom
Multiple R-squared: 0.8106, Adjusted R-squared: 0.7996
F-statistic: 73.47 on 6 and 103 DF, p-value: < 2.2e-16

Penjelasan:

Nilai R-squarednya yaitu 0.7996

2. Menampilkan harga
perbagian

Script:

```
> by(df_francisco$Price, df_francisco$Neighborhood, mean)
> by(df_francisco$Price, df_francisco$Brick, mean)
```

Output :

```
> by(df_francisco$Price, df_francisco$Neighborhood, mean)
df_francisco$Neighborhood: East
[1] 125231.1
-----
df_francisco$Neighborhood: North
[1] 110154.5
-----
df_francisco$Neighborhood: West
[1] 159294.9
> by(df_francisco$Price, df_francisco$Brick, mean)
df_francisco$Brick: No
[1] 121958.1
-----
df_francisco$Brick: Yes
[1] 147769
```

Penjelasan:

Nilai Mean Neighborhood dari East yaitu : 124853.8 dan untuk West : 158462.2

Nilai Mean Brick dari No yaitu : 122712 dan untuk Yes yaitu : 148711.4

4. File Praktikum

Github Repository:

https://github.com/franafuk/Praktikum_stat

5. Soal Latihan

Soal:

1. Apa itu Regresi Linear Sederhana dan Regresi Linear Berganda?
2. Sebutkan perbedaan mendasar dari Regresi Linear Sederhana dan Regresi Linear Berganda?

Jawaban:

1. Regresi linear sederhana dan regresi linear berganda merupakan dua teknik analisis regresi yang digunakan dalam statistika untuk memahami hubungan antara satu atau lebih variabel independen (disebut variabel prediktor) dan variabel dependen (disebut variabel respons).
2. Perbedaan mendasar antara regresi linear sederhana dan regresi linear berganda yaitu :
 - Regresi Linear Sederhana: Hanya melibatkan satu variabel independen (X) untuk memprediksi variabel dependen (Y), dan hubungan antara variabel independen dan variabel dependen dimodelkan sebagai garis lurus (linear).
 - Regresi Linear Berganda : Hanya melibatkan dua atau lebih variabel independen (X_1, X_2, \dots, X_n) untuk memprediksi variabel dependen (Y), dan hubungan antara variabel independen dan variabel dependen dimodelkan sebagai kombinasi linear dari variabel independen.

6. Kesimpulan

- a. Dalam pengerjaan praktikum Statistika, Tergolong mudah.
- b. Kita juga dapat mengetahui bahwasannya di modul ini kita dapat memahami perbedaan antara regresi linear sederhana dan regresi linear berganda, serta keunggulan dan kelemahan masing-masing, kita dapat memilih metode yang paling sesuai dengan data dan tujuan analisisnya. Regresi linear sederhana lebih sederhana dan mudah diinterpretasikan, sementara regresi linear berganda memungkinkan untuk memahami dampak variasi beberapa variabel independen terhadap variabel dependen secara bersama-sama.

7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Pertama	✓	
2.	Latihan Kedua	✓	

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	10 Menit	Menarik
2.	Latihan Kedua	10 Menit	Menarik

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang