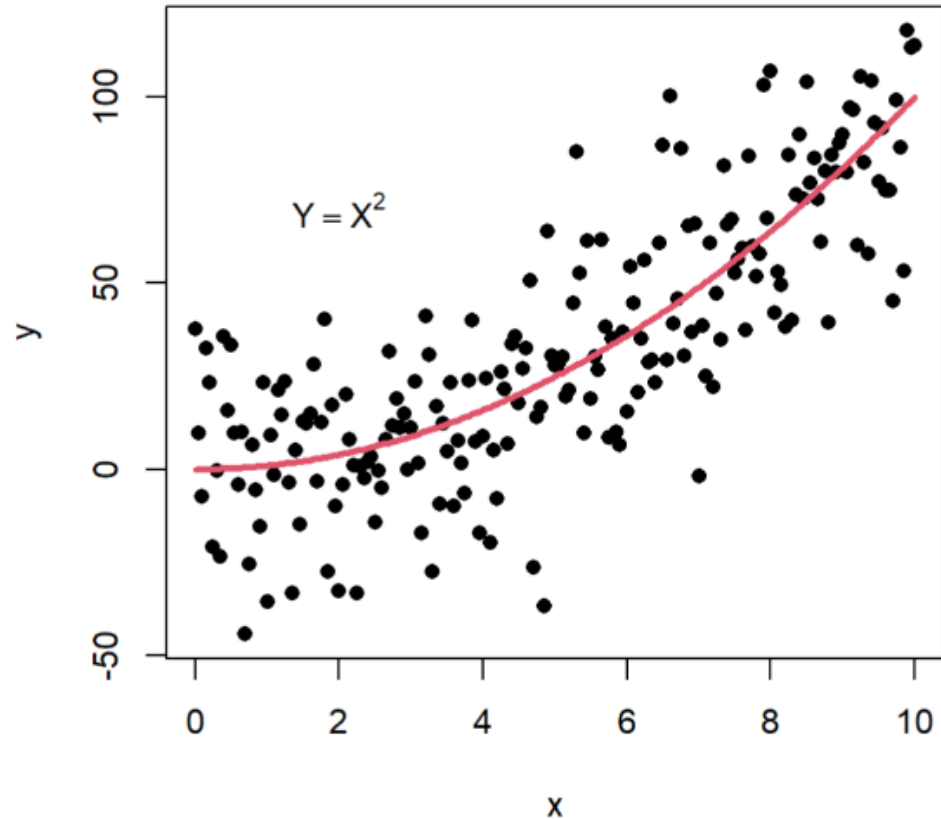


Elok Faiqotul Himmah
STMIK Palangkaraya

Analisis Regresi Linier Sederhana

Apa itu Regresi?



- Metode statistik untuk mempelajari hubungan antara dua variabel atau lebih
- Memprediksi nilai variabel dependen (Y) berdasarkan nilai variabel independen (X)

Perbedaan antara Korelasi dan Regresi Linear



Korelasi: mengukur kekuatan dan arah hubungan antara variabel, menunjukkan sejauh mana variabel berubah bersamaan.



Regresi linier: berfokus pada hubungan sebab akibat antara variabel, memberikan wawasan lebih lanjut tentang bagaimana variabel berinteraksi dan mempengaruhi satu sama lain.

Jenis regresi

01

Regresi Linier
Sederhana (satu
variabel independen)

02

Regresi Linier
Berganda (lebih dari
satu variabel
independen)

03

Regresi Non Linier

Regresi Linear

Regresi linier adalah teknik statistik yang digunakan untuk membuat model dan menyelidiki pengaruh satu atau beberapa variabel independen terhadap satu variabel dependen.

Syarat untuk dilakukan analisis regresi linear

- Kedua variabel memiliki hubungan kausalitas/ sebab-akibat
- Variabel diukur dalam skala interval atau rasio
- Residu (galat acak) dari variabel dependen ($Y - \hat{Y}$) berdistribusi normal (uji normalitas)
- Adanya hubungan (korelasi) yang kuat antara kedua variabel (analisis korelasi)
- Adanya hubungan yang linear antara variabel independen dan variabel dependen (uji linearitas)
- Untuk setiap kelompok variabel bebas, varians galat taksirannya homogen.
- Bersifat homoskedastisitas (residual terdistribusi secara merata)

Rumus Umum Persamaan Regresi Linear

The diagram shows the general linear regression equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon$. Arrows point from labels to the corresponding terms in the equation: 'Dependent Variable (Response Variable)' points to Y ; 'Independent Variables (Predictors)' points to the X terms; 'Y intercept' points to β_0 ; 'Slope Coefficient' points to β_1 and β_2 ; and 'Error Term' points to ε . A large orange decorative shape is in the bottom right corner.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon$$

Labels and arrows in the diagram:

- Dependent Variable (Response Variable) points to Y
- Independent Variables (Predictors) points to X_1, X_2, \dots
- Y intercept points to β_0
- Slope Coefficient points to β_1, β_2
- Error Term points to ε

Prasyarat Analisis Regresi Linear

Jenis Uji Asumsi Klasik (Prasyarat)	Analisis regresi Linear Sederhana	Analisis Regresli Linear Berganda
Uji Normalitas	√	√
Uji Linearitas	√	√
Uji Multikolinearitas	-	√
Uji Heteroskedastisitas	√	√
Uji Autokorelasi	-	√

Regresi Linier Sederhana

- Regresi linier sederhana adalah teknik statistik yang digunakan untuk membuat model dan menyelidiki pengaruh satu variabel independen terhadap satu variabel dependen.
- Regresi linier sederhana → metode yang sederhana namun kuat untuk memodelkan dan menginterpretasikan hubungan.
- Regresi ini melibatkan 2 variabel yaitu satu variabel independen (X) dan satu variabel dependen (Y).
- Variabel independen berarti variabel yang mungkin akan mempengaruhi nilai variabel dependen
- Variabel dependen berarti variabel yang mungkin dipengaruhi oleh variabel independen, nilainya tergantung pada nilai variabel dependennya. Jika korelasi tinggi maka dependensi juga tinggi.
- Tujuan regresi linier sederhana: menemukan hubungan linier antara variabel-variabel yang diobservasi, direpresentasikan oleh suatu persamaan garis lurus.



Aplikasi Regresi Linier Sederhana

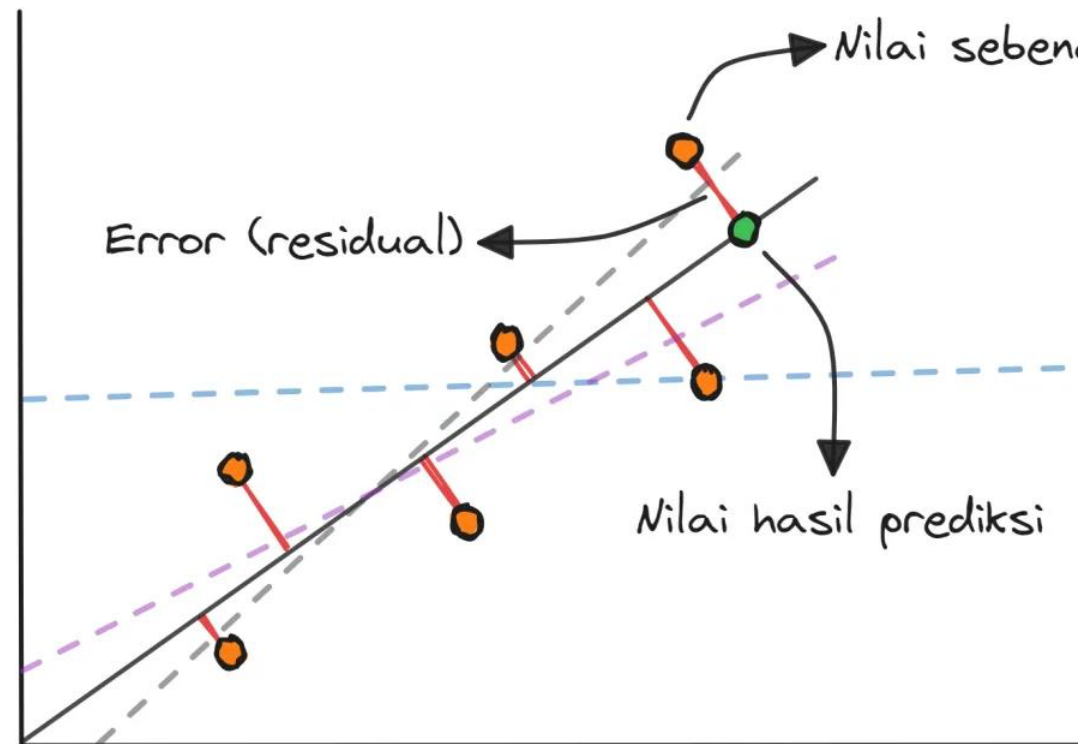
- Bidang ekonomi dan keuangan: menganalisis hubungan antara tingkat bunga dan harga saham, memprediksi harga rumah berdasarkan faktor-faktor seperti luas bangunan atau lokasi.
- Bidang pemasaran: memahami dampak pengeluaran iklan terhadap penjualan, memprediksi nilai seumur hidup pelanggan berdasarkan data historis.
- Bidang kesehatan: mempelajari hubungan antara dosis dan efikasi obat, memprediksi hasil pasien berdasarkan variabel demografis.
- Bidang ilmu lingkungan: menganalisis korelasi antara suhu dan emisi karbon, memprediksi tren iklim di masa depan.
- Bidang pendidikan: menilai hubungan antara waktu belajar dan skor ujian, memprediksi tingkat kelulusan berdasarkan faktor-faktor sosio-ekonomi.
- Bidang ilmu komputer: melakukan prediksi dengan cara supervised learning dalam Machine Learning.



Hasil Analisis

Hasil analisis regresi linier sederhana meliputi:

- Persamaan regresi
- Signifikansi persamaan regresi, artinya kepercayaan dari persamaan regresi benar-benar dapat memprediksi secara nyata atau tidak
- Garis regresi, yaitu membuat grafik regresi
- koefisien korelasi dan koefisien determinasi



Regresi Linier Sederhana

- Slope (kemiringan garis), menunjukkan besarnya kontribusi variabel X terhadap variabel Y.
- Intercept (titik perpotongan antara garis regresi dan sumbu Y saat nilai X adalah 0). Secara statistik, ini berarti nilai rata-rata pada variabel Y saat variabel X bernilai 0.
- Persamaan : $Y = a + bX + e$ (1)

dengan:

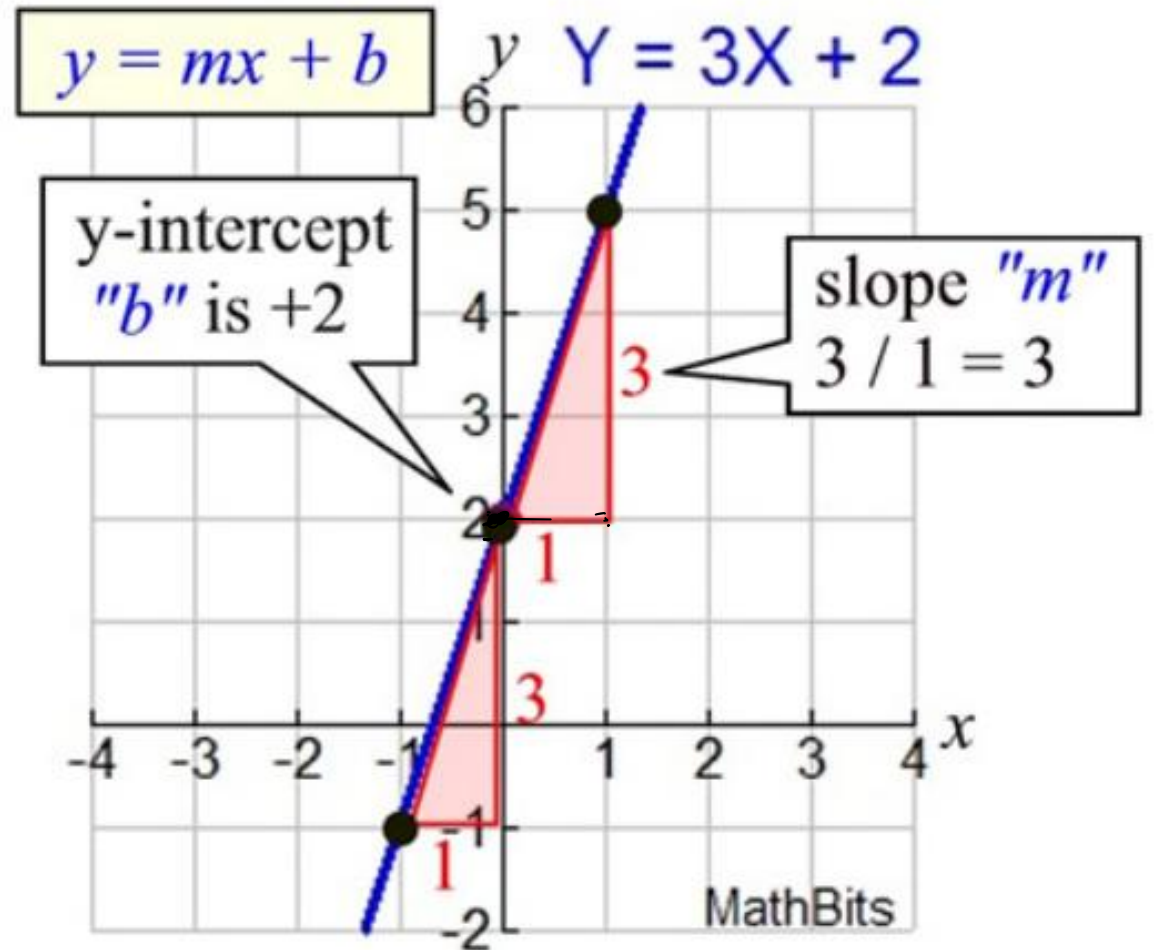
Y : variabel dependen,

X : variabel independen,

a : konstanta (intercept),

b : koefisien regresi (slope),

e : standar error.



Langkah-langkah analisis regresi linier sederhana

- Uji asumsi klasik (uji prasyarat) : uji normalitas, uji linearitas, dan uji heteroskedastisitas.
- Menentukan persamaan regresi
- Uji nyata koefisien regresi menggunakan uji-F, digunakan untuk menguji apakah model fit (tepat) atau tidak untuk memprediksi variabel Y. Nilai F ini disebut ukuran *goodness of fit*
- Uji parsial (uji-t), digunakan untuk mengetahui apakah ada pengaruh variabel X terhadap variabel Y. (
- Menentukan koefisien determinasi, digunakan untuk mengetahui besarnya kontribusi variabel X dalam mempengaruhi variabel Y.
- Pengambilan kesimpulan

Menentukan Persamaan Regresi

Mencari garis regresi terbaik dengan meminimalkan jumlah kuadrat selisih antara nilai Y observasi dengan nilai Y prediksi.

Rumus:

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

dan

$$a = \bar{y} - b\bar{x}$$

Menentukan Persamaan Regresi: Metode Kuadrat Terkecil

Prinsip: mencari garis regresi terbaik dengan meminimalkan jumlah kuadrat selisih antara nilai Y observasi dengan nilai Y prediksi.

Rumus:

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

dan

$$a = \bar{y} - b\bar{x}$$

Sehingga diperoleh persamaan regresi seperti pada Persamaan (1).

Uji Nyata Regresi (Uji-F)

Langkah 1. Formulasi Hipotesis:

H_0 : $b = 0$ atau model tidak dapat digunakan untuk memprediksi variabel Y
(model tidak sesuai)

H_1 : $b \neq 0$ atau model dapat digunakan untuk memprediksi variabel Y (model sesuai)

Langkah 2. Taraf Signifikansi (α)

α digunakan untuk menentukan nilai $F_{tabel} = F_{\alpha;1;n-2}$.

Langkah 3. Kriteria Pengujian:

H_0 ditolak jika $F_{hitung} > F_{tabel}$

Uji Nyata Regresi (Uji-F)

Langkah 4. Uji Statistik (Menentukan F_{hitung} atau F_{reg})

1). Menghitung jumlah kuadrat regresi JK_{reg} dan jumlah kuadrat residu JK_{res} .

$$JK_{reg} = \frac{(\sum XY)^2}{\sum X^2} \text{ dan } JK_{res} = \sum Y^2 - \frac{(\sum XY)^2}{\sum X^2} = \sum Y^2 - JK_{reg}$$

2). Menghitung derajat kebebasan regresi (db_{reg}) dan derajat kebebasan residu (db_{res}).

$$db_{reg} = m \text{ (a prediktor)} = 1, db_{res} = n - 2$$

3). Menghitung rata-rata kuadrat regresi (RK_{reg}) dan rata-rata kuadrat residu (RK_{res}).

$$RK_{reg} = \frac{JK_{reg}}{db_{reg}} \text{ dan } RK_{res} = \frac{JK_{res}}{db_{res}}$$

4). Menentukan F_{hitung} atau F_{reg}

$$F_{reg} = \frac{RK_{reg}}{RK_{res}}$$

Langkah 5. Pengambilan Kesimpulan, apakah model fit (tepat) atau tidak

Uji Parsial (Uji-t)

Uji parsial untuk menentukan apakah ada hubungan nyata variabel X terhadap variabel Y.

Langkah 1. Formulasi Hipotesis:

$H_0 : \rho = 0$ (tidak ada pengaruh variabel X terhadap variabel Y)

$H_1 : \rho \neq 0$ (ada pengaruh variabel X terhadap variabel Y)

Langkah 2. Taraf Signifikansi (α)

α digunakan untuk menentukan nilai $t_{tabel} = t_0 = t_{\frac{\alpha}{2}; n-2}$.

Langkah 3. Kriteria Pengujian:

- H_0 ditolak jika $t_0 < -t_{\frac{\alpha}{2}; n-2}$ atau $t_0 > t_{\frac{\alpha}{2}; n-2}$.

Langkah 4. Uji statistik

Menghitung nilai koefisien korelasi:

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

Menentukan $t_{hitung} = t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

Langkah 5. Pengambilan kesimpulan

Koefisien Determinasi (R^2)

- Koefisien determinasi digunakan untuk menentukan besarnya kontribusi variabel X dalam model regresi mampu menjelaskan variasi dari variabel Y.
 - Menurut Ghozali (2016), nilai koefisien determinasi yang kecil berarti kemampuan variabel independen (X) dalam menjelaskan variabel (Y) sangat terbatas, dan semakin tinggi nilai koefisien determinasi maka semakin baik model regresi yang diperoleh.
 - Kategori nilai koefisien determinasi menurut Chin (1998):
 - $0.19 < R^2 \leq 0.33$: lemah
 - $0.33 < R^2 \leq 0.67$ K : moderat
 - $R^2 > 0,67$: kuat
-

Koefisien Determinasi (R^2)

1) Menghitung jumlah kuadrat xy (JK_{xy})

$$JK_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n},$$

2) Menghitung jumlah kuadrat x (JK_x)

$$JK_x = \sum x^2 - \frac{(\sum x)^2}{n},$$

3) Menghitung jumlah kuadrat y (JK_y)

$$JK_y = \sum y^2 - \frac{(\sum y)^2}{n}.$$

4) Menghitung Koefisien Determinasi (R^2)

$$R^2 = \frac{JK_{xy}}{(JK_x)(JK_y)}$$

dengan:

Nilai koefisien determinasi mempunyai rentang $0 \leq R^2 \leq 1$, biasanya diubah menjadi persentase.

Contoh:

Seorang pengusaha sepatu melakukan penelitian untuk mengetahui pengaruh biaya pemasangan iklan terhadap hasil penjualan produknya. Data diberikan pada tabel.

Untuk itu dilakukan analisis regresi linier sederhana dengan taraf nyata 5% sebagai berikut. Diasumsikan data telah memenuhi prasyarat analisis regresi linier sederhana.

Biaya iklan (jutaan)	Hasil penjualan (jutaan)
0.75	3.55
1.00	4.00
1.50	4.00
0.90	4.25
1.25	3.75
0.65	3.00
0.85	4.00
1.55	6.00
2.00	7.50
2.25	7.25
2.25	7.00

Penyelesaian:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	0.75	3.55	0.5625	12.6025	2.6625
2	1.00	4.00	1.0000	16.0000	4.0000
3	1.50	4.00	2.2500	16.0000	6.0000
4	0.90	4.25	0.8100	18.0625	3.8250
5	1.25	3.75	1.5625	14.0625	4.6875
6	0.65	3.00	0.4225	9.0000	1.9500
7	0.85	4.00	0.7225	16.0000	3.4000
8	1.55	6.00	2.4025	36.0000	9.3000
9	2.00	7.50	4.0000	56.2500	15.0000
10	2.25	7.25	5.0625	52.5625	16.3125
11	2.25	7.00	5.0625	49.0000	15.7500
Jumlah	14.95	54.30	23.8575	295.5400	82.8875

1. Menentukan Persamaan Regresi

Menentukan koefisien regresi (slope):

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$
$$b = \frac{11(82.8875) - (14.95)(54.30)}{11(23.8575) - (14.95)^2} = 2,568$$

Menentukan konstanta (intercept):

$$a = \bar{y} - b\bar{x}$$
$$= 4.93636 - (2.568)(1.35909)$$
$$= 1.446$$

$$\bar{x} = 1.35909, \bar{y} = 4.93636$$

Jadi, persamaan regresinya adalah:

$$y = 1.446 + 2.568x \quad (2)$$

Interpretasi (penafsiran) model (2):

Persamaan tersebut memprediksi bahwa hasil penjualan rata-rata akan berubah sebesar 2.568 jt untuk setiap satuan perubahan yang terjadi pada biaya iklan.

2. Uji Nyata Regresi (Uji-F)

Langkah 1. Formulasi Hipotesis:

H_0 : $b = 0$ atau model tidak dapat digunakan untuk memprediksi variabel Y (model tidak sesuai)

H_1 : $b \neq 0$ atau model dapat digunakan untuk memprediksi variabel Y (model sesuai)

Langkah 2. Taraf Signifikansi (α)

$\alpha = 0.05$, maka diperoleh nilai $F_{tabel} = F_{\alpha;1;n-2} = F_{0.05;1;9} = 5.12$.

Langkah 3. Kriteria Pengujian:

H_0 ditolak jika $F_{hitung} > F_{tabel}$

2. Uji Nyata Regresi (Uji-F)

Langkah 4. Uji Statistik (Menentukan F_{hitung} atau F_{reg})

1). Menghitung jumlah kuadrat regresi JK_{reg} dan jumlah kuadrat residu JK_{res} .

$$JK_{reg} = \frac{(\sum XY)^2}{\sum X^2} = \frac{(82.8875)^2}{23.8575} = 287.973 \text{ dan } JK_{res} = \sum Y^2 - \frac{(\sum XY)^2}{\sum X^2} = 295.54 - 287.973 = 7.567$$

2). menghitung derajat kebebasan regresi (db_{reg}) dan derajat kebebasan residu (db_{res}).

$$db_{reg} = 1 \text{ dan } db_{res} = n - 2 = 11 - 2 = 9.$$

3). Menghitung rata-rata kuadrat regresi (RK_{reg}) dan rata-rata kuadrat residu (RK_{res}).

$$RK_{reg} = \frac{JK_{reg}}{db_{reg}} = \frac{287.973}{1} = 287.973 \text{ dan } RK_{res} = \frac{7.567}{9} = 0.8408$$

4). Menentukan F_{hitung} atau F_{reg}

$$F_{hitung} = F_{reg} = \frac{287.973}{0.8408} = 342.6618.$$

Langkah 5. Mengambil Kesimpulan

Karena $F_{hitung} > F_{tabel}$, maka H_0 ditolak. Jadi, model regresi tepat digunakan untuk memprediksi hasil penjualan.

3. Uji Parsial (Uji-t)

Uji parsial untuk menentukan apakah ada hubungan nyata variabel X terhadap variabel Y.

Langkah 1. Formulasi Hipotesis:

$H_0 : \rho = 0$ (tidak ada pengaruh biaya iklan terhadap hasil penjualan)

$H_1 : \rho \neq 0$ (ada pengaruh biaya iklan terhadap hasil penjualan)

Langkah 2. Taraf Signifikansi (α)

$\alpha = 5\% = 0.05$, maka nilai $t_{tabel} = t_{\frac{\alpha}{2}; n-2} = t_{0.025; 9} = 2.262$.

Langkah 3. Kriteria Pengujian:

H_0 ditolak jika $t_0 < -t_{0.025; 9} = -2.262$ atau

$t_0 > t_{0.025; 9} = 2.262$.

Langkah 4. Uji statistik

Menghitung nilai koefisien korelasi:

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

$$r = \frac{11(82.8875) - (14.95)(54.30)}{\sqrt{((11)(23.8575) - (14.95)^2)(11(295.54) - (54.30)^2)}} = 0.921$$

Menentukan $t_{hitung} = t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(0.921)(\sqrt{9})}{\sqrt{1-(0.921)^2}} = 7.093$

Langkah 5. Pengambilan kesimpulan

Diperoleh $t_0 = 7.093$ dan $t_{\frac{\alpha}{2}; n-2} = 2.262$, karena $t_0 > t_{\frac{\alpha}{2}; n-2}$ maka H_0 ditolak, artinya menerima H_1 . Jadi dapat disimpulkan bahwa ada pengaruh biaya iklan terhadap hasil penjualan.

4. Koefisien Determinasi (R^2)

1) Menghitung jumlah kuadrat xy (JK_{xy})

$$JK_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 82.8875 - \frac{(14.95)(54.30)}{11} = 9.089$$

2) Menghitung jumlah kuadrat x (JK_x)

$$JK_x = \sum x^2 - \frac{(\sum x)^2}{n} = 23.8575 - \frac{(14.95)^2}{11} = 3.5391$$

3) Menghitung jumlah kuadrat y (JK_y)

$$JK_y = \sum y^2 - \frac{(\sum y)^2}{n} = 295.54 - \frac{(54.30)^2}{11} = 27.4955$$

4) Menghitung Koefisien Determinasi (R^2)

$$R^2 = \frac{JK_{xy}}{(JK_x)(JK_y)} = \frac{9.089}{(3.5391)(27.4955)} = 0.09342$$

Nilai koefisien determinasi (R^2) sebesar 0.09342 mempunyai arti bahwa variabel biaya iklan berkontribusi sebesar 9.342% dalam menjelaskan variasi dari variabel hasil penjualan. Dengan kata lain, 9.342% hasil penjualan dipengaruhi oleh biaya iklan, sisanya dipengaruhi oleh faktor lain yang tidak diamati.

Latihan soal.

1. Seorang pengembang *website* ingin mengetahui hubungan antara kecepatan *loading* suatu *website* (dalam detik) dengan jumlah pengunjung harian *website* tersebut. Berikut data yang ia kumpulkan selama 7 hari:

Hari	Kecepatan Loading	Jumlah Pengunjung
1	2	1000
2	3	800
3	4	700
4	2.5	900
5	5	600
6	3.5	750
7	4.5	650

Lakukan analisis regresi linear sederhana untuk mengetahui apakah kecepatan loading mempengaruhi jumlah pengunjung. Gunakan tingkat signifikan 5%. Berapakah prediksi jumlah pengunjung jika kecepatan loading website adalah 2 detik?

Latihan soal.

2. Seorang apoteker ingin mengetahui hubungan antara dosis obat (dalam mg) dengan waktu reaksi obat (dalam menit) pada pasien. Berikut data yang ia peroleh dari 7 pasien:

Pasien	Dosis Obat	Waktu Reaksi
1	10	25
2	15	20
3	20	15
4	25	10
5	30	5
6	35	3
7	40	2

Lakukan analisis regresi linear sederhana untuk mengetahui apakah kecepatan loading mempengaruhi jumlah pengunjung. Gunakan tingkat signifikan 5%. Berapakah prediksi waktu reaksi obat jika dosisnya 45 mg?

Referensi:

- Chin, W.W. (1998). The Partial Least Squares approach to Structural Equation Modeling: Modern Methods for Business Research, 295,336.
- Ghozali, I. (2016). Aplikasi Analisis Multivariat dengan Program IBM SPSS 23, Edisi 8. Semarang: Badan Penerbit Universitas Diponegoro.