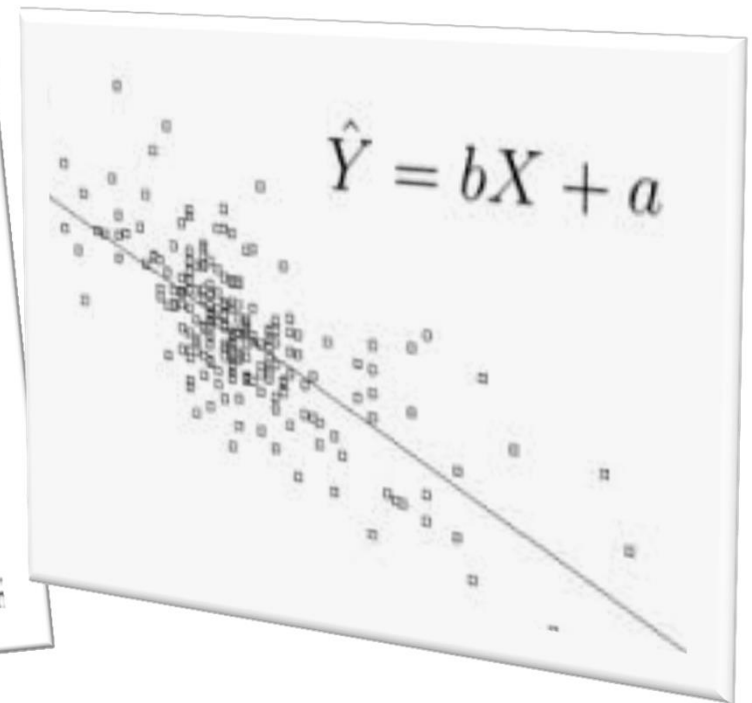
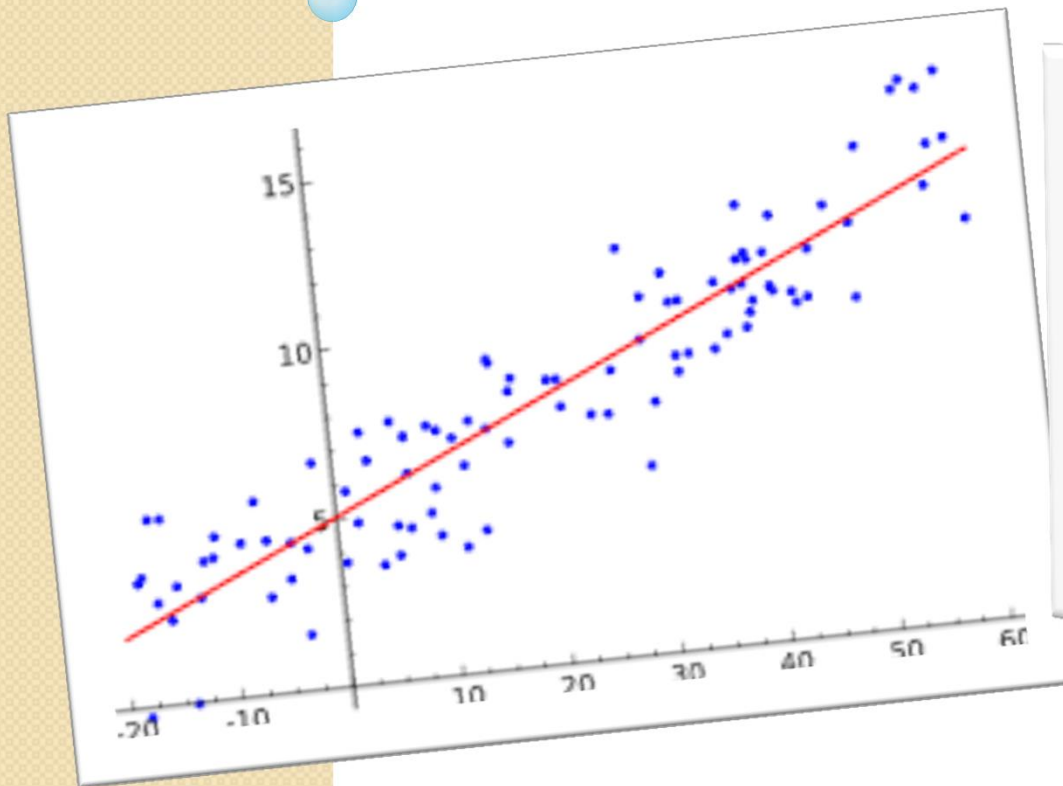


DATA MINING

# REGRESI LINIER SEDERHANA



# Pendahuluan

- Regresi merupakan alat ukur yang digunakan untuk mengetahui ada tidaknya korelasi antar variabel.
- Analisis regresi lebih akurat dalam analisis korelasi karena tingkat perubahan suatu variabel terhadap variabel lainnya dapat ditentukan.
- Jadi pada regresi, peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas akan lebih akurat.

# Pendahuluan

- Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen;  $Y$ ) dengan satu atau lebih variabel bebas (independent;  $X$ ).
- Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai **regresi linier sederhana**, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai **regresi linier berganda**.

# Fungsi Regresi Linier

- Menghitung nilai estimasi rata-rata dan nilai variabel terikat berdasarkan pada nilai variabel bebas.
- Menguji hipotesis karakteristik dependensi.
- Meramalkan nilai rata-rata variabel bebas dengan didasarkan pada nilai variabel bebas diluar jangkauan sampel.

# Analisis Data

1. Persamaan regresi (model)
2. Nilai Prediksi
3. Koefisien determinasi
4. Kesalahan baku estimasi
5. Kesalahan baku koefisien regresinya
6. Nilai F hitung
7. Nilai T hitung
8. Kesimpulan

# Kriteria Penerimaan Hipotesis

## Hipotesis Satu Arah (One Tailed)

- $H_o$  : Tidak terdapat pengaruh positif / negatif variabel X terhadap variabel Y.
- $H_a$  : Terdapat pengaruh positif / negatif variabel X terhadap variabel Y.

## Hipotesis Dua Arah (Two Tailed)

- $H_o$  : Tidak terdapat pengaruh variabel X terhadap variabel Y.
- $H_a$  : Terdapat pengaruh variabel X terhadap variabel Y.
- $H_o$  diterima jika  $b \leq 0$ ,  $t \text{ hitung} \leq t \text{ tabel}$
- $H_a$  diterima jika  $b > 0$ ,  $t \text{ hitung} > t \text{ tabel}$



# Model Regresi Linier Sederhana

Persamaan Regresi Linier dari Y terhadap X

$$Y = a + bX$$

Keterangan :

Y = variabel terikat

X = variabel bebas

a = intercept / konstanta

b = koefisien regresi / slope

# Perhitungan Nilai Koefisien a dan b

- Pendekatan Matriks:

$$\begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma XY \end{pmatrix}$$

$$a = \frac{\det A_1}{\det A} \quad b = \frac{\det A_2}{\det A}$$

$$A = \begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} \quad A_1 = \begin{pmatrix} \Sigma Y & \Sigma X \\ \Sigma XY & \Sigma X^2 \end{pmatrix} \quad A_2 = \begin{pmatrix} \Sigma n & \Sigma Y \\ \Sigma X & \Sigma XY \end{pmatrix}$$

$$\det A = (n)(\Sigma X^2) - (\Sigma X)(\Sigma X)$$

$$\det A_1 = (\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)$$

$$\det A_2 = (n)(\Sigma XY) - (\Sigma Y)(\Sigma X)$$

- Sehingga diperoleh persamaan:

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$\text{atau} \\ a = \bar{Y} - b\bar{X}, \quad a = \frac{\Sigma Y - b(\Sigma X)}{n}$$



# Koefisien Determinasi

- Koefisien determinasi adalah besarnya nilai pengaruh suatu variabel bebas terhadap variabel terikat. Sisanya disebabkan oleh faktor lain yang tidak dimasukkan dalam model.
- Nilai koefisien determinasi ( $R^2$ ) dapat dihitung dengan:

$$R^2 = 1 - \frac{\sum (Y - Y_{pred})^2}{\sum (Y - Y_{rerata})^2}$$

# Kesalahan Baku Estimasi

- Kesalahan Baku Estimasi (*Standard Error*)  
Digunakan untuk mengukur tingkat kesalahan dari model regresi yang dibentuk.
- Persamaan kesalahan baku estimasi (Se):

$$Se = \sqrt{\frac{\sum (Y - Y_{pred})^2}{n - k}}$$

- Dimana:
  - n = jumlah sampel
  - k = jumlah variabel bebas dan variabel terikat

# Standard Error Koefisien Regresi

- *Standard Error* Koefisien Regresi (*Sb*)  
Digunakan untuk mengukur besarnya tingkat kesalahan dari koefisien regresi.
- Nilai *Sb* dapat dihitung menggunakan persamaan:

$$Sb = \frac{Se}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}}$$

# Uji F (*Fisher Distribution*)

- Uji F digunakan untuk uji ketepatan model, apakah nilai prediksi mampu menggambarkan kondisi sesungguhnya:
  - $H_0$ : Diterima jika  $F_{hitung} \leq F_{tabel}$
  - $H_a$ : Diterima jika  $F_{hitung} > F_{tabel}$
- Nilai  $F_{hitung}$  dapat dicari dengan persamaan:

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

# Uji T (*T Distribution*)

- Uji t digunakan untuk mengetahui pengaruh variabel bebas terhadap variabel terikat.
  - $H_o$ : Diterima jika  $t_{hitung} \leq \underline{t_{tabel}}$
  - $H_a$ : Diterima jika  $t_{hitung} > \underline{t_{tabel}}$
- Nilai t hitung bisa dicari dengan persamaan:

$$T_{hitung} = \frac{bj}{Sbj}$$

- Dimana  $bj$  = koefisien regresi j, dan  $Sbj$  = *standard error* koefisien regresi j

# Contoh Studi Kasus

- Seorang manajer pemasaran akan meneliti apakah terdapat pengaruh biaya promosi terhadap penjualan pada perusahaan-perusahaan di wilayah *WaterGold*, untuk kepentingan penelitian tersebut diambil data sampel dari 8 perusahaan sejenis yang telah melakukan promosi. Tingkat/taraf signifikansi pengujian ditentukan sebesar  $\alpha = 5\%$ .



# Pemecahan Masalah

## 1. Judul Penelitian

- Pengaruh biaya promosi terhadap penjualan perusahaan.

## 2. Pertanyaan Penelitian

- Apakah terdapat pengaruh positif biaya promosi terhadap penjualan di perusahaan?

## 3. Hipotesis

- Terdapat pengaruh positif biaya promosi terhadap penjualan perusahaan.

# Kriteria Penerimaan Hipotesis

- $H_o$  : Tidak terdapat pengaruh positif biaya promosi terhadap penjualan di perusahaan.
- $H_a$  : Terdapat pengaruh positif biaya promosi terhadap penjualan di perusahaan.

# Data Sampel (Data Pelatihan)

- Data sampel perusahaan

Nilai Promosi (X)	Nilai Penjualan (Y)
20	64
16	61
34	84
23	70
27	88
32	92
18	72
22	77

# Perhitungan

No	Y	X	XY	X <sup>2</sup>	Y <sup>2</sup>
1	64	20	1280	400	4096
2	61	16	976	256	3721
3	84	34	2856	1156	7056
4	70	23	1610	529	4900
5	88	27	2376	729	7744
6	92	32	2944	1024	8464
7	72	18	1296	324	5184
8	77	22	1694	484	5929
Jumlah	608	192	15032	4902	47094
Rerata	76	24			

# Persamaan / Model Regresi

- Nilai koefisien regresi ( $b$ )

$$b = \frac{[n \sum XY] - [(\sum X)(\sum Y)]}{[n \sum X^2 - (\sum X)^2]}$$

$$b = 1,497$$

- Nilai konstanta ( $a$ )

$$a = \bar{Y} - b\bar{X}$$

$$a = 76 - (1,497 \times 24) = 40,1$$

- Model / Persamaan Regresi:

$$Y = 40,1 + 1,497 X$$

# Perhitungan Lanjutan

No	Y	X	XY	X <sup>2</sup>	Y <sup>2</sup>	Y <sub>pred</sub>	(Y-Y <sub>pred</sub> ) <sup>2</sup>	(Y-Y <sub>rerata</sub> ) <sup>2</sup>
1	64	20	1280	400	4096	70.014	36.163	144
2	61	16	976	256	3721	64.027	9.164	225
3	84	34	2856	1156	7056	90.966	48.525	64
4	70	23	1610	529	4900	74.503	20.281	36
5	88	27	2376	729	7744	80.490	56.403	144
6	92	32	2944	1024	8464	87.973	16.218	256
7	72	18	1296	324	5184	67.020	24.796	16
8	77	22	1694	484	5929	73.007	15.946	1
Jumlah	608	192	15032	4902	47094	608	227.497	886
Rerata	76	24						



# Koefisien Determinasi

- Nilai koefisien determinasi ( $R^2$ ):

$$R^2 = 1 - \frac{\sum (Y - Y_{pred})^2}{\sum (Y - Y_{rerata})^2}$$

- Diperoleh nilai:

$$R^2 = 1 - \frac{(227,497)}{(886)} = 0,743$$

- Nilai determinasi ( $R^2$ ) sebesar 0,743, artinya pengaruh biaya promosi terhadap penjualan adalah sebesar 74,3%. Sisanya 25,7% disebabkan oleh faktor lain yang tidak dimasukkan dalam model.

# Kesalahan Baku Estimasi

- Persamaan kesalahan baku estimasi (Se):

$$Se = \sqrt{\frac{\sum (Y - Y_{pred})^2}{n - k}}$$

- Diperoleh nilai Se:

$$Se = \sqrt{\frac{(227,467)}{8 - 2}} = 6,1576$$

- Se = 6,1576 berarti bahwa batasan seberapa jauh melesetnya perkiraan dalam meramal data memiliki selisih taksir standar sebesar 6,1576

# Standard Error Koefisien Regresi

- Nilai  $Sb$  dihitung dengan persamaan:

$$Sb = \frac{Se}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}}$$

- Diperoleh nilai  $Sb$ :

$$Sb_1 = \frac{6,1576}{\sqrt{(4902) - \frac{(192)^2}{8}}} = 0,359$$

# Uji F (*Fisher Distribution*)

- Nilai  $F_{hitung}$  dicari dengan persamaan:

$$F_{hitung} = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

- Diperoleh nilai F hitung:

$$F = \frac{0,743 / (2 - 1)}{1 - 0,743 / (8 - 2)} = 17,367$$

- Karena F hitung (17,367) > dari F tabel (5,99) maka persamaan regresi dinyatakan **Baik** (*good of fit*)

# Uji T (*T Distribution*)

- Nilai  $t_{hitung}$  dicari dengan persamaan:

$$T_{hitung} = \frac{bj}{Sbj}$$

- Diperoleh nilai  $t_{hitung}$  :

$$t_{hitung} = \frac{1,497}{0,359} = 4,167$$

- Karena  $t_{hitung} (4,167) >$  dari  $t$  tabel (1,943) maka  $H_a$  diterima ada pengaruh positif biaya promosi terhadap penjualan.

# Kesimpulan dan Implikasi

- KESIMPULAN

Terdapat pengaruh positif biaya promosi terhadap nilai penjualan.

- IMPLIKASI

Sebaiknya perusahaan terus meningkatkan promosi agar penjualan meningkat.



Cukup mudahkan?



# Latihan

- Carilah model / persamaan regresi linier sederhana dari data berikut. Tentukan apakah diameter pohon berpengaruh positif terhadap tinggi pohon?

No	Diameter (X)	Tinggi Pohon (Y)
1	8	35
2	9	49
3	7	27
4	6	33
5	13	60
6	7	21
7	11	45
8	12	51