

## BAB 10. Regresi dan Korelasi

### A. Pengantar Regresi dan Korelasi

Dalam kehidupan sehari-hari sering ditemukan masalah/kejadian yang saling berkaitan satu sama lain. Manusia memerlukan analisis hubungan antara kejadian tersebut. Dalam bab ini akan membahas dua kejadian yg saling berhubungan, khususnya 2 kejadian yg dapat diukur secara matematis. Dua hal yang perlu dianalisis yaitu hubungan fungsional (persamaan matematis) dan hubungan kekuatan.

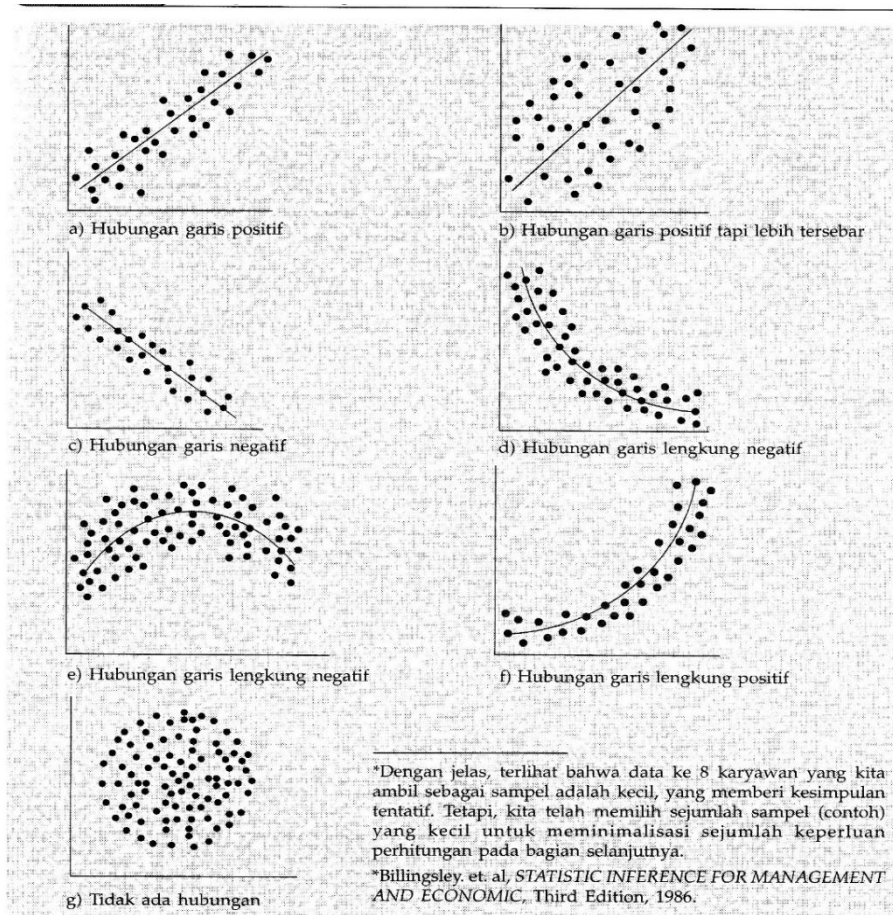
Analisis regresi merupakan suatu analisis yang digunakan untuk mempelajari dan mengukur hubungan fungsional (statistik atau persamaan matematis) yang terjadi antara dua variabel atau lebih variabel. Variabel tersebut adalah variabel X (variabel independent / variabel yang mempengaruhi / variabel yang diketahui), dan variabel Y (variabel dependent / variabel yang dipengaruhi/ variabel yang tidak diketahui). Sementara itu, analisis korelasi merupakan suatu analisis yang bertujuan untuk mengukur “seberapa kuat” atau “derajat kedekatan”, suatu relasi yang terjadi antar variabel.

Pada dasarnya hubungan antar 2 variabel dapat dibedakan atas menjadi 3 yaitu hubungan searah/positif, hubungan bersifat kebalikan/negative serta tidak ada hubungan. Hubungan yang searah diartikan apabila perubahan variabel x (independent) akan mempengaruhi variabel y (dependent) yang searah atau jika variabel x bertambah, maka variabel y bertambah pula, dan sebaliknya. Contohnya adalah hubungan antara pengeluaran iklan (x) dan jumlah penjualan (y) atau hubungan antara penghasilan (X) dan pengeluaran konsumsi (Y). Selanjutnya, dua variabel dikatakan mempunyai hubungan yang bersifat kebalikan atau negatif, apabila perubahan variabel independent (x) akan mempengaruhi variabel dependent (Y) pada arah yang berlawanan. Artinya apabila variabel x bertambah, maka variabel y berkurang atau sebaliknya, jika variabel x berkurang maka variabel y bertambah. Contohnya yaitu hubungan antara usia kendaraan (X) dengan tingkat harga (Y) atau hubungan antara harga barang (x) dengan jumlah yang diminta (Y). Terakhir, dua variabel dikatakan tidak punya hubungan apabila perubahan pada variabel independent (x) **tidak** mempengaruhi perubahan pada variabel dependent (y). Sebagai contoh adalah hubungan antara konsumsi pangan (x) dengan tingginya gedung (y).

### B. Regresi Linier Sederhana

Pada regresi linier sederhana, penggambaran garis regresi dapat dilakukan salah satunya adalah metode diagram berserak (*the scatter diagram*). Setelah ditetapkan bahwa terdapat

hubungan logis di antara variabel, maka untuk mendukung analisis lebih jauh, tahap selanjutnya adalah membuat diagram pencar, yang menunjukkan titik-titik tertentu. Setiap titik memperlihatkan suatu hasil yang kita nilai sebagai variabel bebas maupun variabel tak bebas. Diagram pencar ini memiliki 2 manfaat, yaitu dapat membantu menunjukkan apakah terdapat hubungan yang bermanfaat antara dua variabel, serta membantu menetapkan tipe persamaan yang menunjukkan hubungan antara kedua variabel tersebut. Contoh gambaran umum dari diagram berserak atau diagram pencar dapat dilihat sebagaimana berikut.



Gambar 10.1 Diagram berserak (*The scatter diagram*)

Persamaan yang digunakan untuk mendapatkan garis regresi pada data diagram pencar disebut persamaan regresi. Untuk menempatkan garis regresi pada data yang diperoleh maka digunakan metode kuadrat terkecil, sehingga bentuk persamaan regresi adalah sebagai berikut:

$$Y' = a + b X \dots\dots\dots (10.1)$$

Dimana:

$Y'$ : nilai estimasi/taksiran untuk variabel terikat (tak bebas  $Y$ )

$a$ : titik potong garis regresi pd sumbu  $y$  (nilai estimate  $Y'$  bila  $x=0$ )

b: gradien garis regresi (perub nilai estimasi  $Y'$  per satuan perubahan nilai  $x$ ) atau koefidien arah dari garis regresi

X: nilai variabel bebas

Kesamaan diantara garis regresi dan garis trend tidak dapat berakhir dengan persamaan garis lurus. Dalam hal ini dicari persamaan regresi yg *paling baik* untuk mewakili sebaran titik data tersebut. Suatu kriteria bahwa persamaan regresi yg paling baik adalah regresi yg mempunyai total kuadrat selisih yg paling minimum. Garis regresi) memiliki dua sifat matematis berikut  $\Sigma(Y - Y') = 0$  dan  $\Sigma(Y - Y')^2 =$  nilai terkecil atau terendah. Dengan perkataan lain, garis regresi akan ditempatkan pada data dalam diagram sedemikian rupa sehingga penyimpangan (perbedaan) positif titik-titik terhadap titik-titik pencar di atas garis akan mengimbangi penyimpangan negatif titik-titik pencar yang terletak di bawah garis, sehingga hasil penyimpangan keseluruhan titik-titik terhadap garis lurus adalah nol. Untuk memperoleh total kuadrat error paling minimum, dipakailah metode kuadrat minimum. Dari persamaan regresi linear sebelumnya akan memiliki total kuadrat error minimum bila koefisien regresi a dan b dihitung dengan rumus berikut.

$$a = \frac{\Sigma Y \Sigma X^2 - \Sigma X \Sigma XY}{n \Sigma X^2 - (\Sigma X)^2} \text{ dan } b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} \dots\dots\dots (10.2)$$

atau

$$b = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} \text{ dan } a = \frac{\Sigma Y}{n} - b \left( \frac{\Sigma X}{n} \right) \text{ atau } a = \bar{Y} - b\bar{X} \dots\dots\dots (10.3)$$

Koefisien Regresi adalah lereng garis regresi (nilai b). Nilai b positif , menunjukkan hubungan antara variabel x dan y searah atau hubungannya positif. Nilai b negatif, menunjukkan hubungan antara variabel x dan y berlawanan arah atau hubungannya negatif. Besar kecilnya perubahan variabel x terhadap variabel y ditentukan besar kecilnya koefisien regresi. Dalam praktknya ada istilah yang disebut sebagai “kesalahan baku dari penaksiran” atau *standard error of estimation* oleh  $Y'=a+bX$  yaitu sebagai berikut.

$$S_{y'.x} = \sqrt{\frac{\Sigma(Y-Y')^2}{n}} \text{ dijabarkan menjadi } S_{y'.x} = \sqrt{\frac{\Sigma Y^2 - a \Sigma Y - b \Sigma XY}{n}} \dots\dots\dots (10.4)$$

Perbedaan mendasar antara regresi dan korelasi adalah bahwa regresi menunjukkan hubungan antara variabel satu dengan variabel lainnya. Sifat hubungan dapat dijelaskan: variabel yang satu sebagai penyebab, variabel yang lain sebagai akibat. Sedangkan korelasi

tidak menunjukkan hubungan sebab akibat, akan tetapi menunjukkan hubungan antara variabel satu dengan yang lain.

### C. Korelasi Linier Sederhana

Secara matematis bahwasanya koefisien korelasi ( $r$ ) diartikan sebagai ukuran hubungan linier peubah  $X$  dan  $Y$  dimana nilai  $r$  berkisar antara  $(+1)$  sampai  $(-1)$ . Nilai  $r$  yang  $(+)$  ditandai oleh nilai  $b$  yang  $(+)$ , demikian pula nilai  $r$  yang  $(-)$  ditandai oleh nilai  $b$  yang  $(-)$ . Jika nilai  $r$  mendekati  $+1$  atau  $r$  mendekati  $-1$  maka  $X$  dan  $Y$  memiliki korelasi linier yang tinggi. Jika nilai  $r = +1$  atau  $r = -1$  maka  $X$  dan  $Y$  memiliki korelasi linier sempurna. Kemudian, jika nilai  $r = 0$  maka  $X$  dan  $Y$  tidak memiliki relasi (hubungan) linier (dalam kasus  $r$  mendekati  $0$ , anda dapat melanjutkan analisis ke regresi eksponensial).

Koefisien determinasi sampel disimbulkan dengan  $R$  yang nilainya merupakan kuadrae dari koefisien korelasi sehingga  $R = r^2$ . Ukuran proporsi keragaman total nilai peubah  $Y$  yang dapat dijelaskan oleh nilai peubah  $X$  melalui hubungan linier. Penetapan dan interpretasi koefisien krelasi dan koefisien determinasi adalah esbagai berikut.

$$r = \frac{n \sum XY - \left( \sum X \right) \left( \sum Y \right)}{\sqrt{\left[ n \sum X^2 - \left( \sum X \right)^2 \right] \left[ n \sum Y^2 - \left( \sum Y \right)^2 \right]}} \dots \dots \dots (10.5)$$

$$R = r^2 \dots \dots \dots (10.6)$$

Simbol  $r$  adalah Koefisien Korelasi dan koefisien determinasi sampel adalah  $R = r^2$

Contoh soal regresi. Berikut adalah data Biaya Promosi dan Volume Penjualan PT BIMOIL perusahaan Minyak Goreng. Buatlah persmaan regresi linear sederhana dengan minimum kuadrat terkecil. Perkirakan Volume penjualan jika, dikeluarkan biaya promosi Rp. 10 juta.

<u>Tahun</u>	<u>X</u> <u>Biaya</u> <u>Promosi</u> <u>(Juta Rupiah)</u>	<u>Y</u> <u>Volume</u> <u>Penjualan</u> <u>(Ratusan Juta</u> <u>Liter)</u>	<u>XY</u>	<u>X<sup>2</sup></u>	<u>Y<sup>2</sup></u>
1992	2	5	10	4	25
1993	4	6	24	16	36
1994	5	8	40	25	64
1995	7	10	70	49	100
1996	8	11	88	64	121
$\Sigma$	$\Sigma x = 26$	$\Sigma y = 40$	$\Sigma xy = 232$	$\Sigma x^2 = 158$	$\Sigma y^2 = 346$

$n = 5$

Jawab. Bentuk umum persamaan regresi linier sederhana:  $Y = a + bX$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

$$b = \frac{(5 \times 232) - (26 \times 40)}{(5 \times 158) - (26^2)} = \frac{1160 - 1040}{790 - 676} = \frac{120}{114} = 1.05263\dots$$

$$b = 1,053$$

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}$$

$$a = \frac{40}{5} - \left( 1.05263\dots \times \frac{26}{5} \right) = 8 - (1.05263\dots \times 5.2)$$

$$a = 8 - 5.4736\dots = 2.5263\dots \Rightarrow 2.53$$

**Sehingga  $Y = a + b X \rightarrow Y = 2,530 + 1,053X$**

$$Y = 2,530 + 1,053 X$$

Jika  $X = 10$

$$Y = 2,53 + 1,053 (10)$$

$$Y = 2,53 + 10,53 = 13,06 \text{ (ratusan juta liter)}$$

Maka volume penjualan =  $13.06 \times 100\,000\,000$  liter

Contoh Korelasi (Lihat soal regresi). Setelah mendapatkan persamaan Regresi  $Y = 2.52 + 1.053 X$ , hitung koefisien korelasi ( $r$ ) dan koefisien determinasi ( $R$ ). Gunakan data berikut.

- $\Sigma x = 26$
- $\Sigma y = 40$
- $\Sigma xy = 232$
- $\Sigma x^2 = 158$
- $\Sigma y^2 = 346$

Jawab

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

$$r = \frac{(5 \times 232) - (26 \times 40)}{\sqrt{[(5 \times 158) - (26^2)] \times [(5 \times 346) - (40^2)]}}$$

$$r = \frac{1160 - 1040}{\sqrt{[790 - 676] \times [1730 - 1600]}} = \frac{120}{\sqrt{114 \times 130}}$$

$$r = \frac{120}{\sqrt{14820}} = \frac{120}{121.73...} = 0.9857...$$

Nilai  $r = 0,9857$  menunjukkan bahwa peubah X (biaya promosi) dan Y (volume penjualan) berkorelasi linier yang positif dan tinggi.

$$R = r^2 = 0.9857...^2 = 0,97165.... = 97 \%$$

Nilai  $R = 97\%$  menunjukkan bahwa 97% proporsi keragaman nilai peubah Y (volume penjualan) dapat dijelaskan oleh nilai peubah X (biaya promosi) melalui hubungan linier. Sisanya, yaitu 3 % dijelaskan oleh hal-hal lain.