

REGRESI LINIER

1. Hubungan Fungsional Antara Variabel

Variabel dibedakan dalam dua jenis dalam analisis regresi:

- Variabel bebas atau variabel prediktor -> variabel yang mudah didapat atau tersedia. Dapat dinyatakan dengan $X_1, X_2, \dots, X_k$ Dengan $k \geq 1$
- Variabel tak bebas atau variabel respon -> variabel yang terjadi karena variabel bebas. Dapat dinyatakan dengan Y.

Contoh: fenomena yang meliputi hasil panen padi dengan volume pupuk yang digunakan, sebaiknya diambil variabel bebas X = volume pupuk dan variabel takbebas Y = hasil panen padi.

Persamaan regresi secara umum:

$$\mu_{y,x_1,x_2,\dots,x_k} = f(X_1, X_2, \dots, X_k | \theta_1, \theta_2, \dots, \theta_m)$$

Dengan $\theta_1, \theta_2, \dots, \theta_m$ parameter-parameter yang ada dalam regresi itu.

Regresi linier sederhana adalah regresi dengan satu variabel bebas, regresi dengan variabel bebas X dan variabel takbebasnya Y atau dinamakan juga regresi Y atas X, bentuk persamaannya:

$$\mu_{y.x} = \theta_1 + \theta_2 X$$

Regresi linear sederhana berdasarkan sampel, maka θ_1 ditaksir dengan a dan θ_2 ditaksir dengan b diperoleh:

$$\hat{Y} = a + bX$$

Model regresi populasi pangkat dua atau parabola untuk sebuah variabel bebas dengan parameter θ_1, θ_2 , dan θ_3 adalah

$$\mu_{y.x,x^2} = \theta_1 + \theta_2 X + \theta_3 X^2$$

Regresi dari hasil penelitian yang dipakai untuk menaksir regresi parabola:

$$\hat{Y} = a + bX + cX^2$$

Dengan a, b dan c adalah titik taksiran secara berturut-turut untuk θ_1, θ_2 dan θ_3 .

2. Metoda Tangan Bebas

Metode Tangan Bebas adalah metode penentuan persamaan regresi kira-kira menggunakan diagram pencar. Jika fenomena meliputi sebuah variabel bebas X dan variabel tak bebas Y, maka data yang didapat digambarkan pada diagram dengan sumbu datar menyatakan X dan sumbu tegak menyatakan Y.

Jika letak titik-titik itu sekitar garis lurus, maka cukup beralasan untuk menduga regresi linier. Jika letak titik-titik itu sekitar garis lengkungan, maka diduga merupakan regresi nonlinier.

Jika letak titik-titik sekitar garis lurus maka untuk menentukan persamaan regresinya, dapat dicari dengan menggunakan dua titik yang dilalui garis tersebut, kemudian dicari persamaan garisnya.

Penentuan regresi dengan cara ini bersifat **tidak tunggal**, artinya tiap orang akan memberikan perkiraan yang berbeda bergantung pada pertimbangan pribadi masing-masing.

3. Metoda Kuadrat Terkecil untuk Regresi Linear

Seperti dikatakan sebelumnya regresi dengan variabel bebas X dan variabel takbebas Y dimana model regresi linier untuk populasi yaitu

$$\mu_{y.x} = \theta_1 + \theta_2 X$$

akan ditaksir harga-harga θ_1 dan θ_2 oleh a dan b sehingga didapat persamaan regresi menggunakan data sampel:

$$\hat{Y} = a + bX$$

Dengan

$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$
$$b = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

Dimana n adalah jumlah sampel, X_i adalah data variabel bebas ke – i dan Y_i adalah data variabel takbebas ke – i .

Jika terlebih dahulu dihitung koefisien b , maka koefisien a dapat pula ditentukan oleh rumus:

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

dimana $\bar{X}$ dan $\bar{Y}$ adalah rata-rata untuk masing-masing variabel X dan Y .

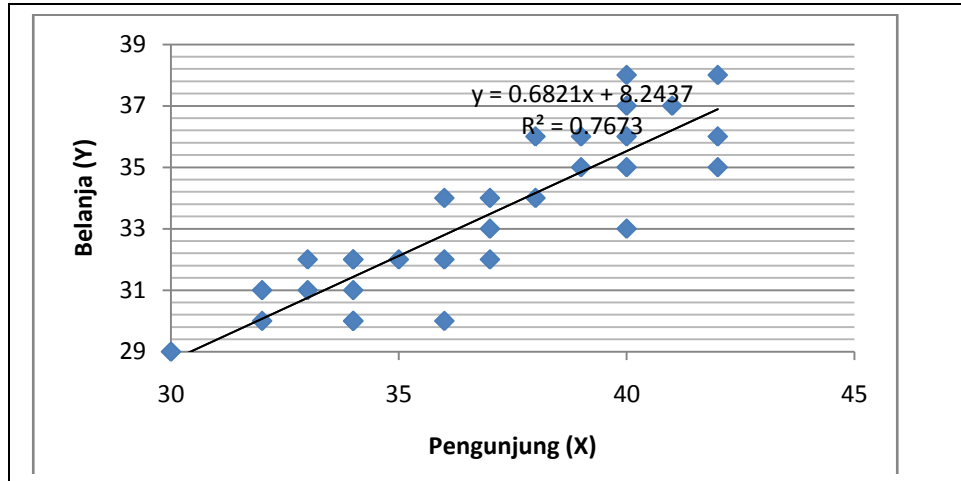
Koefisien b dinamakan **koefisien arah regresi linier** dan menyatakan perubahan rata-rata variabel Y untuk setiap perubahan variabel X sebesar satu unit. Perubahan ini merupakan pertambahan jika b bertanda positif dan penurunan atau pengurangan jika bertanda negatif.

Contoh:

Data berikut melukiskan hasil pengamatan mengenai banyak orang yang datang (X) dan banyak orang yang berbelanja (Y) di sebuah toko selama 30 hari.

Daftar 2.1
Banyak Pengunjung dan yang Berbelanja
Di Sebuah Toko Selama 30 Hari

Pengunjung (X_i)	Berbelanja (Y_i)	Pengunjung (X_i)	Berbelanja (Y_i)
34	32	42	38
38	36	41	37
34	31	32	30
40	38	34	30
30	29	36	30
40	35	37	33
40	33	36	32
34	30	37	34
35	32	39	35
39	36	40	36
33	31	33	32
32	31	34	32
42	36	36	34
40	37	37	32
42	35	38	34



Gambar 2.1

Jawab:

X_i	Y_i	$X_i Y_i$	X_i^2	X_i	Y_i	$X_i Y_i$	X_i^2
34	32	1088	1156	42	38	1596	1764
38	36	1368	1444	41	37	1517	1681
34	31	1054	1156	32	30	960	1024
40	38	1520	1600	34	30	1020	1156
30	29	870	900	36	30	1080	1296
40	35	1400	1600	37	33	1221	1369
40	33	1320	1600	36	32	1152	1296
34	30	1020	1156	37	34	1258	1369
35	32	1120	1225	39	35	1365	1521
39	36	1404	1521	40	36	1440	1600
33	31	1023	1089	33	32	1056	1089
32	31	992	1024	34	32	1088	1156
42	36	1512	1764	36	34	1224	1296
40	37	1480	1600	37	32	1184	1369
42	35	1470	1764	38	34	1292	1444

Setelah dijumlahkan didapat:

$$\sum X_i = 1105, \sum Y_i = 1001, \sum X_i Y_i = 37094 \text{ dan } \sum X_i^2 = 41029$$

Maka diperoleh:

$$a = \frac{(1001)(41029) - (1105)(37094)}{30(41029) - (1105)^2} = 8,2437$$

$$b = \frac{30(37094) - (1105)(1001)}{30(41029) - (1105)^2} = 0,6821$$

Sehingga persamaan linier Y atas X adalah

$$\hat{Y} = 8,2437 + 0,6821X$$

Artinya untuk setiap X bertambah dengan seorang maka rata-rata pembeli Y bertambah dengan 0,68 orang.

4. Berbagai Varians Sehubungan Dengan Regresi Linear Sederhana

Untuk analisis selanjutnya tentang regresi linier sederhana, beberapa asumsi harus diambil.

Pertama, mengingat hasil pengamatan variabel takbebas Y belum tentu sama besarnya dengan harga diharapkan, yakni $\hat{Y}$ yang didapat dari regresi hasil pengamatan, maka terjadi perbedaan $e = |Y - \hat{Y}|$, biasa disebut kekeliruan prediksi atau galat prediksi. Dalam populasi, galat prediksi ini dimisalkan berbentuk variabel acak yang mengikuti distribusi normal dengan rata-rata nol dan varians σ_e^2 .

Kedua, untuk setiap harga X yang diberikan, variabel tak bebas Y independen dan berdistribusi normal dengan rata-rata $(\theta_1 + \theta_2 X)$ dan varians $\sigma_{Y.X}^2$. Varians $\sigma_{Y.X}^2$ dimisalkan sama untuk setiap X dan karenanya dapat dinyatakan oleh σ_e^2 yang biasa pula dinamakan *variens kekeliruan taksiran* sedangkan $\sigma_{y.x}$ dikenal dengan *kekeliruan baku taksiran*.

Berpegang kedua asumsi di atas, maka varians σ_e^2 ditaksir oleh rata-rata kuadrat penyimpangan sekitar regresi atau disebut juga rata-rata kuadrat residu, dinyatakan oleh varians s_e^2 yaitu

$$s_{Y.X}^2 = s_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{(n - 2)}$$

Dengan Y = variabel tak bebas hasil pengamatan dan $\hat{Y}$ = didapat dari regresi berdasarkan sampel, dan n = ukuran sampel.

Dapat ditulis juga

$$s_{Y.X}^2 = \left(\frac{n-1}{n-2} \right) (s_Y^2 - b^2 s_X^2)$$

Dengan s_Y^2 dan s_X^2 masing-masing menyatakan varians untuk variabel-variabel Y dan X.

Varians koefisien b:

$$s_b^2 = \frac{s_{Y.X}^2}{\sum (X_i - \bar{X})^2}$$

Varians koefisien a:

$$s_a^2 = s_{Y.X}^2 \left\{ \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right\}$$

Varians ramalan rata-rata Y untuk X_0 yang diketahui:

$$s_{\bar{Y}}^2 = s_{Y.X}^2 \left\{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right\}$$

Varians ramalan individu Y untuk X_0 yang diketahui

$$s_Y^2 = s_{Y.X}^2 \left\{ 1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right\}$$

Untuk rumus-rumus di atas $\sum (X_i - \bar{X})^2$ dapat diganti oleh $\left\{ \sum X_i^2 - \frac{(\sum x_i)^2}{n} \right\}$

Contoh:

Untuk dengan data dalam daftar 2.1, kita dapat menghitung varians-variens di atas. Kita perlu $\bar{X} = 36,8$; $s_X^2 = 11,32$; $s_Y^2 = 6,86$ dan $\sum (X_i - \bar{X})^2 = 328,2$, diperoleh $b = 0,68$ dan $n = 30$ didapat

$$s_{Y.X}^2 = \frac{30-1}{30-2} \{6,86 - (0,68)^2(11,32)\} = 1,684$$

$$s_b^2 = \frac{1,684}{328,2} = 5,13 \times 10^{-3}$$

$$s_a^2 = 1,684 \left\{ \frac{1}{30} + \frac{36,8^2}{328,2} \right\} = 7,005$$

Varians ramalan rata-rata Y untuk $X_0 = 40$ adalah

$$s_{\hat{Y}}^2 = 1,684 \left\{ \frac{1}{30} + \frac{(40 - 36,8)^2}{328,2} \right\} = 0,1087$$

Varians ramalan individu Y untuk $X_0 = 40$ adalah

$$s_{\hat{Y}}^2 = 1,684 \left\{ 1 + \frac{1}{30} + \frac{(40 - 36,8)^2}{328,2} \right\} = 1,7927$$