Regresi Linier Berganda

Program Studi Teknik Industri
Universitas Brawijaya

Ihwan Hamdala, ST., MT

Regresi Berganda

Menguji hubungan linier antara 1 variabel dependen (y) dan 2 atau lebih variabel independen (x_n)

Contoh

- Hubungan antara suhu warehouse dan viskositas cat dengan jumlah cacat foam mark pada produk
  Var. independen : suhu warehouse & viskositas cat
  Var. dependen : jumlah cacat foam mark
- Hubungan antara kecepatan pelayanan dan kualitas produk dengan kepuasan pelanggan
  Var. independen : kecepatan pelayanan & kualitas produk
  Var. dependen : kepuasan pelanggan

Model Regresi Berganda

Menguji hubungan linier antara 1 variabel dependen (y) dan 2 atau lebih variabel independen (x_n)

Model pd populasi:

Y-intercept | Population slopes | Random Error

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon$$

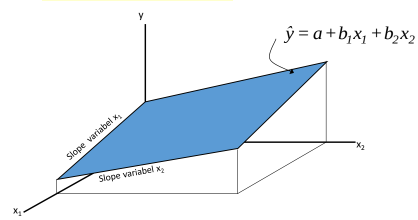
Estimasi model regresi berganda:

Estimasi (atau prediksi) Nilai y | Estimasi intercept | Estimasi koofisien slope

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

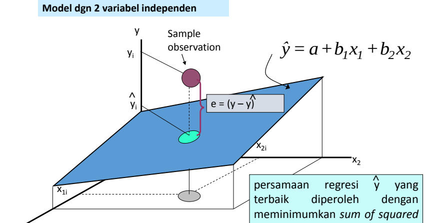
Model Regresi Berganda

Model dgn 2 variabel independen



Model Regresi Berganda

Model dgn 2 variabel independen



persamaan regresi $\hat{y}$ yang terbaik diperoleh dengan meminimumkan *sum of squared error* (jmh kuadrat error) Σe^2

Asumsi Regresi Berganda

Error (residual) dari model regresi:

$e = (y - \hat{y})$

- Error berdistribusi normal
- Mean dari error adalah nol
- Error memiliki variansi yang konstan
- Error bersifat independen

Regresi Berganda

- Tentukan tujuan apa yang diinginkan dan pilih variabel dependennya
- Tentukan sejumlah variabel independen
- Pengumpulan data sampel (observasi) untuk semua variabel

Mencari Persamaan Regresi Berganda

Dapat ditentukan dengan beberapa cara sbb:

1. Metode Kuadrat Terkecil
2. Persamaan Normal
3. Sistem Matriks

1. Metode Kuadrat Terkecil (dgn 2 var independen)

$$\hat{y} = a + b_1 x_1 + b_2 x_2$$

$$a = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2$$

$$\overline{Y} = \frac{\Sigma Y}{n}$$

$$\overline{X}_1 = \frac{\Sigma X_1}{n} \qquad \overline{X}_2 = \frac{\Sigma X_2}{n}$$

1. Metode Kuadrat Terkecil - lanjutan

b_1 dan b_2 → Koefisien regresi dicari dgn persamaan

$$b_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$b_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

1. Metode Kuadrat Terkecil - lanjutan

$$\Sigma y^2 = \Sigma Y^2 - n\overline{Y}^2$$

$$\Sigma x_1^2 = \Sigma X_1^2 - n\overline{X}_1^2$$

$$\Sigma x_2^2 = \Sigma X_2^2 - n\overline{X}_2^2$$

$$\Sigma x_1 y = \Sigma X_1 Y - n\overline{X}_1\overline{Y}$$

$$\Sigma x_2 y = \Sigma X_2 Y - n\overline{X}_2\overline{Y}$$

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - n\overline{X}_1\overline{X}_2$$

Contoh Soal

Internal Revenue Service mencoba mengestimasi pajak aktual yang tak terbayar tiap bulan di divisi Auditing. Dua faktor yang mempengaruhinya adalah jumlah jam kerja pegawai dan jumlah jam kerja mesin (komputer). Untuk menganalisis seberapa besar kedua faktor itu mempengaruhi besarnya pajak aktual tak terbayar tiap bulan, dilakukan pencatatan selama 10 bulan dengan data ditunjukkan pada tabel berikut.

Cari persamaan regresi linier bergandanya!

Contoh Soal-lanjutan

Bulan	X_1	X_2	Y (Rp 1000)
	Jam kerja pegawai	Jam kerja mesin/komputer	Pajak aktual yang tidak dibayar
Januari	45	16	29
Pebruari	42	14	24
Maret	44	15	27
April	45	13	25
Mei	43	13	26
Juni	46	14	28
Juli	44	16	30
Agustus	45	16	28
September	44	15	28
Oktober	43	15	27

Jawab

n ke	X_1	X_2	Y	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
1	45	16	29	1.305	464	720	2.025	256	841
2	42	14	24	1.008	336	588	1.764	196	576
3	44	15	27	1.188	405	660	1.936	225	729
4	45	13	25	1.125	325	585	2.025	169	625
5	43	13	26	1.118	338	559	1.849	169	676
6	46	14	28	1.288	392	644	2.116	196	784
7	44	16	30	1.320	480	704	1.936	256	900
8	45	16	28	1.260	448	720	2.025	256	784
9	44	15	28	1.232	420	660	1.936	225	784
10	43	15	27	1.161	405	645	1.849	225	729
Rata2	**44,1**	**14,7**	**27,2**						
Total	**441**	**147**	**272**	**12.005**	**4.013**	**6.485**	**19.461**	**2.173**	**7.428**

Jawab - lanjutan

$$\Sigma y^2 = \Sigma Y^2 - n\overline{Y}^2 = 7.428 - (10)(27{,}2)^2 = 29{,}6$$

$$\Sigma x_1^2 = \Sigma X_1^2 - n\overline{X_1}^2 = 19.461 - (10)(44{,}1)^2 = 12{,}9$$

$$\Sigma x_2^2 = \Sigma X_2^2 - n\overline{X_2}^2 = 2.173 - (10)(14{,}7)^2 = 12{,}1$$

$$\Sigma x_1 y = \Sigma X_1 Y - n\overline{X_1}\,\overline{Y} = 12.005 - (10)(44{,}1)(27{,}2) = 9{,}8$$

$$\Sigma x_2 y = \Sigma X_2 Y - n\overline{X_2}\,\overline{Y} = 4.013 - (10)(14{,}7)(27{,}2) = 14{,}6$$

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - n\overline{X_1}\,\overline{X_2} = 6.485 - (10)(44{,}1)(14{,}7) = 2{,}3$$

Jawab - lanjutan

$$b_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = \frac{(12{,}1)(9{,}8) - (2{,}3)(14{,}6)}{(12{,}9)(12{,}1) - (2{,}3)^2} = 0{,}564$$

$$b_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2} = \frac{(12{,}9)(14{,}6) - (2{,}3)(9{,}8)}{(12{,}9)(12{,}1) - (2{,}3)^2} = 1{,}099$$

$$a = \overline{Y} - b_1\overline{X_1} - b_2\overline{X_2} = 27{,}2 - (0{,}564)(44{,}1) - (1{,}099)(14{,}7) = -13{,}828$$

Sehingga diperoleh persamaan regresi linier berganda yaitu:
$Y = -13{,}828 + 0{,}564X_1 + 1{,}099X_2$

Interpretasi persamaan regresi berganda

Persamaan regresi linier berganda
$Y = -13{,}828 + 0{,}564X_1 + 1{,}099X_2$

Nilai a = -13,828
Jika jam kerja pegawai (X1) dan jam kerja mesin (X2) keduanya bernilai nol, maka estimasi besarnya pajak tertunda (Y) sebesar -13,828

Nilai b1 = + 0,564
- Hubungan antara jam kerja pegawai (X1) dengan pajak tertunda (Y)
- Jika jam kerja mesin (X2) adalah konstan, maka setiap kenaikan nilai jam kerja pegawai (X1) sebesar satu satuan akan meningkatkan pajak tertunda (Y) sebesar 0,564 satuan,

Nilai b2 = + 1,099
- Hubungan antara jam kerja mesin (X2) dengan pajak tertunda (Y)
- Jika jam kerja pegawai (X1) adalah konstan, maka setiap kenaikan nilai jam kerja mesin (X2) sebesar satu satuan akan meningkatkan pajak tertunda (Y) sebesar 1,099 satuan

2. Persamaan Normal

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

Contoh (dari soal sebelumnya)

n ke	X_1	X_2	Y	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
1	45	16	29	1.305	464	720	2.025	256	841
2	42	14	24	1.008	336	588	1.764	196	576
3	44	15	27	1.188	405	660	1.936	225	729
4	45	13	25	1.125	325	585	2.025	169	625
5	43	13	26	1.118	338	559	1.849	169	676
6	46	14	28	1.288	392	644	2.116	196	784
7	44	16	30	1.320	480	704	1.936	256	900
8	45	16	28	1.260	448	720	2.025	256	784
9	44	15	28	1.232	420	660	1.936	225	784
10	43	15	27	1.161	405	645	1.849	225	729
Rata2	**44,1**	**14,7**	**27,2**						
Total	**441**	**147**	**272**	**12.005**	**4.013**	**6.485**	**19.461**	**2.173**	**7.428**

Jawab

1) Dalam 3 persamaan normal :

$$272 = 10\,a + 441\,b_1 + 147\,b_2 \quad \text{..................(1)}$$
$$12.005 = 441\,a + 19.461\,b_1 + 6.485\,b_2 \quad \text{..................(2)}$$
$$4.013 = 147\,a + 6.485\,b_1 + 2.173\,b_2 \quad \text{..................(3)}$$

2) Menghilangkan nilai a dengan menjumlahkan persamaan (1) dengan (2). Persamaan (1) x -441 dan persamaan (2) x 10 :

$$(1)\ x\ (-441)\ :\ -119.952 = -4410\,a - 194.481\,b_1 - 64.827\,b_2$$
$$(2)\ x\ (10)\ :\ 120.050 = 4410\,a + 194.610\,b_1 + 64.850\,b_2$$
$$(4)\ :\ 98 = 129\,b_1 + 23\,b_2$$

3) Persamaan (1) kalikan dengan –147 dan persamaan (3) dengan 10. Jumlahkan persamaan (1) dan (3) :

$$(1)\ x\ (-147)\ :\ -39.984 = -1470\,a - 64.827\,b_1 - 21.609\,b_2$$
$$(3)\ x\ (10)\ :\ 40.130 = 1470\,a + 64.850\,b_1 + 21.730\,b_2$$
$$(5)\ :\ 146 = 23\,b_1 + 121\,b_2$$

Jawab – lanjutan

4) Kalikan persamaan (4) dengan –23 dan (5) dengan 129. Jumlahkan (4) dan (5) untuk menduga nilai b :

$$(4)\ x\ (-23)\ :\ -2.254 = -2.967\,b_1 - 529\,b_2$$
$$(5)\ x\ (129)\ :\ 18.834 = 2.967\,b_1 + 15.609\,b_2$$
$$(6)\ :\ 16.580 = 15.080\,b_2$$

Maka, b_2 = 1,099

5) Cari nilai penduga b_1 dari persamaan (4) :

$$(4)\ :\ 98 = 129\,b_1 + 23\,b_2$$
$$98 = 129\,b_1 + (23)\,(1.099)$$
$$98 = 129\,b_1 + 25.277$$

Maka, b_2 = 0,564

Diperoleh persamaan:
$Y = -13,828 + 0,564X_1 + 1,099X_2$

7) Cari nilai a dari persamaan :

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$
$$= 27.2 - (0.564)\,(44.1) - (1.099)\,(14.7)$$
$$= 27.2 - (24.8724) - (16.1553)$$
$$= \mathbf{-\,13.8277} = \mathbf{-13.828}$$

3. Sistem Matriks

$$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1X_2 \\ \Sigma X_2 & \Sigma X_1X_2 & \Sigma X_2^2 \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1Y \\ \Sigma X_2Y \end{bmatrix}$$

Dari persamaan normal disusun dalam bentuk matriks

$$A = \begin{pmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{pmatrix} \qquad A_2 = \begin{pmatrix} n & \sum Y & \sum X_2 \\ \sum X_1 & \sum X_1Y & \sum X_1X_2 \\ \sum X_2 & \sum X_2Y & \sum X_2^2 \end{pmatrix}$$

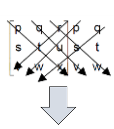
$$A_1 = \begin{pmatrix} \sum Y & \sum X_1 & \sum X_2 \\ \sum X_1Y & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2Y & \sum X_1X_2 & \sum X_2^2 \end{pmatrix} \qquad A_3 = \begin{pmatrix} n & \sum X_1 & \sum Y \\ \sum X_1 & \sum X_1^2 & \sum X_1Y \\ \sum X_2 & \sum X_1X_2 & \sum X_2Y \end{pmatrix}$$

$$a = \frac{\det A_1}{\det A} \qquad b_1 = \frac{\det A_2}{\det A} \qquad b_2 = \frac{\det A_3}{\det A}$$

Mencari Determinan Matriks

Untuk mencari determinan matriks berordo 3 x 3 dapat dengan beberapa metode, salah satunya dengan metode Sarrus. Misal ada sebuah matriks B.

$$B = \begin{bmatrix} p & q & r \\ s & t & u \\ v & w & x \end{bmatrix}$$



Maka

$$\det(B) = |B| = \begin{vmatrix} p & q & r \\ s & t & u \\ v & w & x \end{vmatrix} = ptx + quv + rsw - rtv - qsx - puw$$

Persamaan regresi berganda dengan 3 variabel bebas

$\hat{Y} = a_0 + b_1X_1 + b_2X_2 + b_3X_3$

Keterangan :

$\hat{Y}$	= variabel terikat (nilai duga Y)
X_1, X_2, X_3	= variabel bebas
a, b_1, b_2, b_3	= koefisien regresi linier berganda
a	= nilai Y, apabila $X_1 = X_2 = X_3 = 0$

b_1	= besarnya kenaikan/penurunan Y dalam satuan, jika X_1 naik/turun satu satuan dan X_2 dan X_3 konstan
b_2	= besarnya kenaikan/penurunan Y dalam satuan, jika X_2 naik/turun satu satuan dan X_1 dan X_3 konstan
b_3	= besarnya kenaikan/penurunan Y dalam satuan, jika X_3 naik/turun satu satuan dan X_1 dan X_2 konstan

Persamaan regresi berganda dengan 3variabel bebas

$$\sum Y = a.n + b_1 \sum X_1 + b_2 \sum X_2 + b_3 \sum X_3$$
$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + b_3 \sum X_1 X_3$$
$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 + b_3 \sum X_2 X_3$$
$$\sum X_3 Y = a \sum X_3 + b_1 \sum X_1 X_3 + b_2 \sum X_2 X_3 + b_3 \sum X_3^2$$

atau dalam bentuk deviasi dari mean :

$$\sum x_1 y = b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + b_3 \sum x_1 x_3$$
$$\sum x_2 y = b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + b_3 \sum x_2 x_3$$
$$\sum x_3 y = b_1 \sum x_1 x_3 + b_2 \sum x_2 x_3 + b_3 \sum x_3^2$$
$$a = \frac{\sum Y - b_1 \sum X_1 - b_2 \sum X_2 - b_3 \sum X_3}{n}$$

Persamaan regresi berganda dengan 3 variabel bebas

$$\sum x_1 y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{n}$$
$$\sum x_2 y = \sum X_2 Y - \frac{(\sum X_2)(\sum Y)}{n}$$
$$\sum x_3 y = \sum X_3 Y - \frac{(\sum X_3)(\sum Y)}{n}$$
$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$
$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$
$$\sum x_3^2 = \sum X_3^2 - \frac{(\sum X_3)^2}{n}$$
$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n}$$
$$\sum x_1 x_3 = \sum X_1 X_3 - \frac{(\sum X_1)(\sum X_3)}{n}$$
$$\sum x_2 x_3 = \sum X_2 X_3 - \frac{(\sum X_2)(\sum X_3)}{n}$$

Kesalahan Baku & Koefisien Regresi Berganda

Kesalahan baku : nilai yang menyatakan seberapa jauh menyimpangnya nilai regresi terhadap nilai yang sebenarnya

$$S_e = \sqrt{\frac{\sum y^2 - \left(b_1\left(\sum x_1 y\right) + b_2\left(\sum x_2 y\right)\right)}{n - m}}$$

m = k+1
k = jmh var bebas

$$Sb_1 = \frac{S_e}{\sqrt{\left(\Sigma X_1^2 - n\overline{X}_1^2\right)\left(1 - r_{Y.1}^2\right)}} \qquad Sb_2 = \frac{S_e}{\sqrt{\left(\sum X_2^2 - n\overline{X}_2^2\right)\left(1 - r_{Y.1}^2\right)}}$$

$$r_{Y.1} = \frac{n\sum X_1 X_2 - \sum X_1 \sum X_2}{\sqrt{\left(n\sum X_1^2 - \left(\sum X_1\right)^2\right)\left(n\sum X_2^2 - \left(\sum X_2\right)^2\right)}}$$

Koefisien Korelasi antara X1 dan X2

Kesalahan Baku & Koefisien Regresi Berganda

Pada contoh soal sebelumnya

$$\Sigma y^2 = \Sigma Y^2 - n\overline{Y}^2 = 7.428 - (10)(27,2)^2 = 29,6$$
$$\Sigma x_1^2 = \Sigma X_1^2 - n\overline{X}_1^2 = 19.461 - (10)(44,1)^2 = 12,9$$
$$\Sigma x_2^2 = \Sigma X_2^2 - n\overline{X}_2^2 = 2.173 - (10)(14,7)^2 = 12,1$$
$$\Sigma x_1 y = \Sigma X_1 Y - n\overline{X}_1\overline{Y} = 12.005 - (10)(44,1)(27,2) = 9,8$$
$$\Sigma x_2 y = \Sigma X_2 Y - n\overline{X}_2\overline{Y} = 4.013 - (10)(14,7)(27,2) = 14,6$$
$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - n\overline{X}_1\overline{X}_2 = 6.485 - (10)(44,1)(14,7) = 2,3$$

$$S_e = \sqrt{\frac{\sum y^2 - \left(b_1\left(\sum x_1 y\right) + b_2\left(\sum x_2 y\right)\right)}{n - m}}$$

$$S_e = \sqrt{\frac{29,6 - (0,56(9,8) + 1,10(14,6))}{10 - 3}} = 1,071$$

Dgn persamaan pd slide sebelumnya bisa diperoleh nilai Sb1 dan Sb2:

Interval Keyakinan Bagi penduga B_1 dan B_2

Pengujian menggunakan distribusi t dengan derajat bebas (db) = n – m, Dengan contoh soal sebelumnya, dgn ∝ =5%, db = n – m = n – k -1 = 10 – 2 - 1 = 7, maka:

Interval keyakinan bagi penduga B_1 adalah

$b_1 - t_{(\alpha/2,\ n-k-1)}.S_{b1} < B_1 < b_1 + t_{(\alpha/2,\ n-k-1)}.S_{b1}$

$0,564 - (2,365)(0,303) < B_1 < 0,564 + (2,365)(0,303)$

$-0,153 < B_1 < 1,281$

Interval keyakinan bagi penduga B2 adalah

$B_2 - t_{(\alpha/2,\ n-k-1)}.S_{b2} < B_2 < b_2 + t_{(\alpha/2,\ n-k-1)}.S_{b2}$

$1,099 - (2,365)(0,313) < B_2 < 1,099 - (2,365)(0,313)$

$0,359 < B_2 < 1,839$

Pengujian Parameter Koefisien Regresi Berganda

Bertujuan untuk menentukan apakah ada sebuah hubungan linear antar variabel tidak bebas Y dengan variabel bebas X_1, X_2,… ,X_k.

Ada 2 bentuk pengujian hipotesis bagi koefisien regresi berganda:
1. Pengujian hipotesis serentak
2. Pengujian hipotesis individual

Pengujian Hipotesis Serentak
Merupakan pengujian hipotesis koefisien regresi berganda dengan B_1 dan B_2 serentak atau secara bersama-sama mempengaruhi Y.

Pengujian Hipotesis individual
Merupakan pengujian hipotesis koefisien regresi berganda dengan hanya satu B (B_1 atau B_2) yang mempengaruhi Y.

Latihan Soal

12-6. The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature (x_1), the number of days in the month (x_2), the average product purity (x_3), and the tons of product produced (x_4). The past year's historical data are available and are presented in the following table:

y	x_1	x_2	x_3	x_4
240	25	24	91	100
236	31	21	90	95
270	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2.