

ANALISIS REGRESI

1. PENDAHULUAN

Jika kita memiliki data yang terdiri atas dua atau lebih variabel, adalah sewajarnya untuk suatu cara bagaimana variabel-variabel tersebut berhubungan. Hubungan yang diperoleh pada umumnya dinyatakan dalam bentuk persamaan matematik yang menyatakan hubungan fungsional antara variabel-variabel. Misalnya berat orang dewasa laki-laki sampai usia tertentu bergantung pada tinggi badannya, tekanan semacam gas bergantung pada temperature, hasil produksi padi bergantung pada jumlah pupuk yang digunakan, banyak hujan, cuaca dan sebagainya. Studi yang menyangkut hubungan antar variabel dikenal dengan analisis regresi.

2. HUBUNGAN ANTAR VARIABEL

Pada analisis regresi terdapat dua jenis variabel yaitu variabel bebas (variabel predictor) dan variabel terikat (variabel respon). Penentuan jenis variabel dalam beberapa hal tidak mudah dapat dilaksanakan. Studi yang cermat, diskusi yang seksama, berbagai pertimbangan, kewajaran masalah yang dihadapi dan pengalaman akan membantu memudahkan penentuan. Variabel yang mudah didapat atau tersedia sering digolongkan ke dalam variabel bebas. Variabel bebas sering dinotasikan dengan $X_1, X_2, \dots, X_k$ ($k \geq 1$), sedangkan variabel terikat dinotasikan dengan Y . Misalnya untuk kasus hasil panen dan jumlah pupuk yang digunakan sebaiknya dipilih : Y = hasil panen, sedangkan jumlah pupuk dinotasikan dengan X .

Seperti dikatakan di atas, hubungan fungsional akan dinyatakan dalam bentuk persamaan matematik yang selanjutnya disebut persamaan regresi yang akan tergantung pada parameter – parameternya. Model untuk populasi, secara umum dapat ditukiskan dalam bentuk ;

$$Y = f(X_1, X_2, \dots, X_k | \theta_1, \theta_2, \dots, \theta_m) \quad (1)$$

Dengan $\theta_1, \theta_2, \dots, \theta_m$ merupakan parameter-parameter yang terdapat di dalam regresi. Sebuah contoh regresi regresi yang paling sederhana dengan sebuah variabel bebas adalah regresi linier sederhana, dengan model :

$$Y = \theta_1 + \theta_2 X \quad (2)$$

Dalam hal ini parameternya adalah θ_1, θ_2 .

Berdasarkan sample acak, persamaan regresi populasi (1) akan ditaksir, ini dilakukan dengan jalan menaksir parameter-parameter $\theta_1, \theta_2, \dots, \theta_m$. Sebagai contoh untuk regresi linier sederhana θ_1, θ_2 akan ditaksir oleh a dan b , sehingga persamaan regresi berdasarkan data sample adalah:

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

Dimana symbol $\hat{Y}$ dibaca Y topi. Koefisien b dinamakan koefisien arah regresi linier dan menyatakan perubahan rata-rata variabel Y untuk setiap perubahan variabel X sebesar satu unit. Perubahan tersebut merupakan pertambahan apabila b bertanda positif dan pengurangan jika b bertanda negative.

Regresi dengan X menyatakan variabel bebas dan Y variabel terikat sering dinamakan sebagai regresi Y atas X . Sebaliknya adalah regresi X atas Y , apabila hal ini mungkin dilakukan. Misalnya antara tinggi ayah dan anak, antara nilai matematik dan fisika dan lain-lain.

Model regresi pangkat dua atau parabola untuk sebuah variabel bebas dengan parameter $\theta_1, \theta_2, \theta_3$ adalah:

$$Y = \theta_1 + \theta_2 X + \theta_3 X^2$$

Dan berdasarkan data sample model tersebut ditaksir dengan:

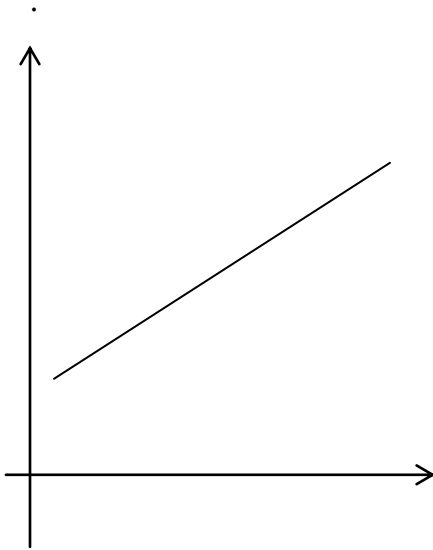
$$\hat{Y} = a + bX + cX^2$$

Dengan a, b dan c masing-masing diperoleh dari perhitungan berdasarkan data yang berturut-turut merupakan taksiran dari $\theta_1, \theta_2, \theta_3$.

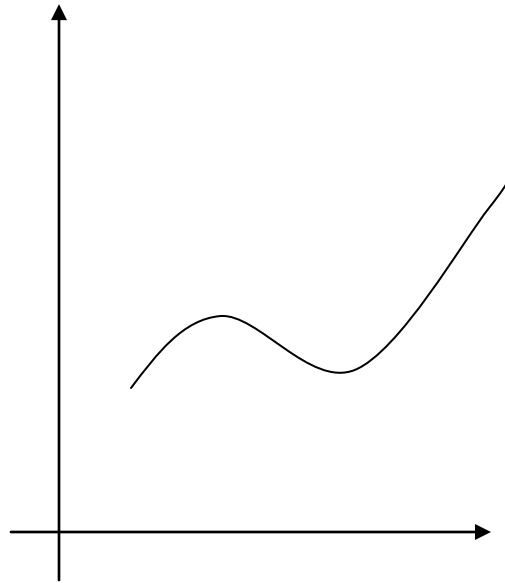
Bagaimana persamaan regresi dapat ditentukan, jika data telah ada? Ada dua metode yang dapat digunakan yaitu :

3. MOTODE TANGAN BEBAS

Jika fenomena meliputi sebuah variabel bebas X dan variabel tak bebas Y, maka data yang didapat dapat digambarkan dalam diagram cartesius dimana sumbu mendatar menyatakan X dan sumbu tegak menyatakan Y. Setelah titik-titik yang dinyatakan oleh absis X dan ordinat Y digambarkan, maka diperoleh suatu diagram pencar. Dengan memperhatikan letak titik-titik dalam diagram, bentuk regresi dapat diperkirakan. Jika letak titik-titik tersebut disekitar garis lurus, maka cukup beralasan untuk menduga regresi linier, jika letak titik-titik disekitar garis lengkung wajarlah untuk menduga bahwa regresi non linier. Sebagai ilustrasi perhatikan diagram pencar berikut:



Regresi linier ditarik secocok mungkin dengan letak titik – titik, kemudian persamaannya ditentukan dengan menggunakan dua titik yang dilakukan.



Gambar di atas memperlihatkan diagram pencar yang menunjukkan model lengkung, regresinya dapat digambarkan secocok mungkin dengan letak titik-titik sedangkan persamaannya masih harus difikirkan baik-baik apakah akan parabola, pangkat tiga atau lainnya.

Penentuan regresi dengan cara ini bersifat tidak tunggal, artinya setiap orang akan memberikan perkiraan yang berbeda bergantung kepada pertimbangan masing-masing. Hanya mereka yang betul-betul ahli mungkin dapat menentukan regresi yang baik dengan cara ini.

4. METODE KUADRAT TERKECIL UNTUK REGRESI LINIER

Metode tangan bebas dapat dipakai untuk menolong menentukan dugaan bentuk regresi apakah linier atau tidak. Persamaannya jika tidak betul-betul yakin lebih baik Untuk model regresi populasi

$$Y = \theta_1 + \theta_2 X$$

Akan ditaksir harga-harga θ_1 dan θ_2 oleh a dan b sehingga didapat persamaan regresi

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

Untuk keperluan tersebut, data hasil pengamatan harus disusun seperti pada table berikut:

Variabel Takbebas (Y)	Variabel bebas (X)
Y_1	X_1
Y_2	X_2
$\cdot$	$\cdot$
$\cdot$	$\cdot$
Y_n	X_n

Disini terdapat pasangan terurut X dan Y sebanyak n buah, dengan n menyatakan banyak data. Dengan menggunakan metode kuadrat terkecil (penurunannya akan saudara dapat pada mata kuliah analais regresi), ternyata koefisien regresi a dan b untuk regresi linier dapat ditentukan dengan rumus :

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

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

Jika koefisien ditentukan terlebih dahulu koefisien a dapat ditentukan dengan rumus :

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

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

Rumus –rumus di atas dipakai untuk menentukan koefisien – koefisien regresi Y atas X, untuk koefisien regresi X atas Y rumus yang sama dapat digunakan dengan terlebih dahulu mempertukarkan tempat untuk simbol-simbol X dan Y.

Contoh 1

Tabel berikut menunjukkan daya regang (Y) dan kekerasan aluminium(X) yang dinyatakan dalam satuan tertentu.

X	71	53	82	67	56	70	64	78	55	70	53	84
Y	354	313	322	334	247	377	308	340	301	349	293	368

Setelah data tersebut dibuat diagram perncarnya ternyata mendekati garis lurus, tentukan regresi linier Y atas X.

Jawab:

Untuk keperluan tersebut terlebih dahulu akan dikitung besaran-besaran yang diperlukan, seperti ditunjukkan oleh table berikut:

X_i	Y_i	$X_i Y_i$	X_i^2
71	354	25134	5041
53	313	16589	2809
82	322	26404	6724
67	334	22378	4489
56	247	13832	3136
70	377	26390	4900
64	308	19712	4096
78	340	26520	6084
55	301	16555	3025
70	349	24430	4900
53	293	15529	2809
84	368	30912	7056

Dari tabel di atas diperoleh nilai:

$$\sum X_i = 803, \quad \sum Y_i = 3906, \quad \sum X_i Y_i = 264385, \quad \sum X_i^2 = 55069 \quad \text{dan} \quad \sum Y_i^2 = 1285802$$

Dengan metode kuadrat terkecil diperoleh nilai-nilai berikut:

$$a = \frac{(3906)(55069) - (803)(264385)}{12(55069) - (803)^2} = 174,69$$

$$b = \frac{12(264385) - (803)(3906)}{12(55069) - (803)^2} = 2,25$$

Dengan demikian persamaan regresi linier Y atas X untuk masalah di atas adalah :

$$\hat{Y} = 174,69 + 2,25 X$$

Tanda $\hat{Y}$ menyatakan bahwa kita berhadapan dengan Y yang diperoleh dari regresi untuk membedakannya dengan Y dari hasil pengamatan. Karena koefisien $b = 2,25$ (bertanda positif) sehingga dapat dikatakan bahwa jika X (= kekuatan alumunium) bertambah satu satuan, maka rata-rata daya regang (Y) bertambah 2,25 satuan.

Regresi yang diperoleh selanjutnya digunakan untuk keperluan peramalan, apabila nilai variabel bebas diketahui. Misalnya jika $X = 80$, maka dengan memasukan nilai tersebut kepada persamaan regresi di atas diperoleh nilai:

$$\hat{Y} = 174,69 + 2,25(80) = 354,69$$

Diperkirakan rata – rata daya regang alumunium akan samadengan 354,69 jika kekuatan alumunium 80. Nilai ramalan Y dengannilai X yang diambil dari ruang gerak X hasil pengamatan maka proses tersebut disebut interpolasi., sedangkan jika diambil di luar itu proses dinamakan ekstrapolasi. Jika ekstrapolasi harus dilakukan, maka hendaknya dilakukan dengan hati-hati dan penuh pertimbangan.

5. UJI INDEPENDENSI ANTAR VARIABEL PADA REGRESI LINIER

Untuk menyelidiki apakah variabel X independent atau tidak dengan variabel Y, perlu dilakukan pengujian dengan hipotesis:

$$H_0 : \theta_2 = 0$$

$$H_1 : \theta_2 \neq 0$$

Bentuk hipotesis di atas dapat juga disajikan dalam bentuk lain, yaitu:

$$H_0 : \text{Variabel X dan Y saling independen}$$

$$H_1 : \text{Variabel X dan Y tidak saling independen}$$

Dalam hal ini pengujian dilakukan dengan analisis variansi. Untuk ini jumlah kuadrat-kuadrat semua nilai individu Y yaitu $\sum Y^2$ dipecah menjadi tiga sumber variasi berbentuk:

$$\sum Y^2 = \frac{(\sum Y_i)^2}{n} + b(\sum X_i - \bar{X})(Y_i - \bar{Y}) + \sum (Y_i - \hat{Y}_i)^2$$

$$\sum Y^2 = \frac{(\sum Y_i)^2}{n} + JK(b|a) + JK_{(Res)}$$

Untuk memudahkan, satuan satuan tersebut disusun dalam sebuah daftar sehingga diperoleh daftar analisis varians, disingkat ANAVA, seperti ditunjukkan berikut ini:

DAFTAR I
DAFTAR ANALISIS VARIANSI UNTUK REGRESI LINIER

Sumber Variasi	Dk	JK	RJK	F
Regresi (a)	1	$(\sum Y_i)^2/n$	$(\sum Y_i)^2$	$\frac{S_{reg}^2}{S_{res}^2}$
Regresi (b a)	1	$JK(b a)$	$S_{reg}^2 = JK(b a)$	
Residu	n-2	$\sum (Y_i - \hat{Y})^2$	$S_{res}^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$	
Jumlah	N	$\sum Y_i^2$		

Dengan;

$$JK(b|a) = b \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$= b \left\{ \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} \right\}$$

Hasil bagi $F = \frac{S_{reg}^2}{S_{res}^2}$ ternyata berdistribusi F dengan dk(derajat kebebasan) pembilang satu dan dk penyebut (n-2). Berdasarkan hal tersebut hipotesis $H_0 : \theta_2 = 0$ ditolak jika $F \geq F_{(1-\alpha)(1,n-2)}$.

Untuk lebih jelasnya perhatikan daftar analisis variansi di atas untuk menguji hipotesis $H_0 : \theta_2 = 0$, dengan data seperti pada contoh 1 diperoleh:

$$\frac{(\sum Y_i)^2}{n} = \frac{(3906)^2}{12} = 1271403$$

$$b \left\{ \sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n} \right\} = 2,25 \left\{ 264385 - \frac{803 \cdot 3906}{12} \right\} = 6769,125$$

$$\sum (Y_i - \hat{Y}_i)^2 = 1285802 - 1271403 - 6769,125 = 7629,875$$

Sehingga diperoleh daftar analisis variansi berikut:

DAFTAR 2
DAFTAR ANALISIS VARIANSI UNTUK UJI INDEPENDEN
DALAM REGRESI LINIER

Sumber Variasi	Dk	JK	RJK	F
Regresi (a)	1	1271403	1271403	8,87
Regresi (b a)	1	6769,125	6769,125	
Residu	10	7629,875	762,9875	
Jumlah	12	1285802		

Dari daftar Anava tersebut diperoleh $F = 8,87$, sementara dari table F diperoleh $F_{0,95(1,10)} = 4,96$ sehingga kita menolak H_0 , dengan kata lain kita menerima H_1 artinya

variabel X dan Y tidak saling independent atau variabel Y dipengaruhi oleh variabel X, dapat pula dikatakan koefisien arah regresi cukup berarti.

6. UJI LINIERITAS REGRESI

Setelah bentuk regresi linier diperoleh, selanjutnya diperlukan suatu pengujian apakah model linier yang telah diambil betul-betul cocok dengan keadaan atau tidak? Jika hasil pengujian mengatakan model linier kurang cocok maka selayaknya diambil model lain yang non linier. Untuk keperluan tersebut JK_{res} dipecah menjadi dua bagian yaitu:

- JK kekeliruan eksperimen atau galat eksperimen ($JK(E)$)
- JK tuna cocok model linier ($JK(TC)$)

Bagian kedua digunakan untuk menguji linieritas regresi, yakni menguji apakah model linier yang telah diambil betul-betul cocok dengan keadaannya atau tidak. Jika ternyata model tidak cocok maka perlu dicari model lain yang mungkin tidak linier. Supaya JK_{res} dapat dipecah seperti maksud di atas maka perlu dihitung jumlah kuadrat kekeliruan, dengan rumus :

$$JK(E) = \sum_x \left\{ \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right\}$$

Dimana tanda jumlah yang pertama diambil untuk semua harga X. jumlah kuadrat-kuadrat untuk tuna cocok model linier disingkat dengan $JK(TC)$, didapat dengan mengurangkan JK_{res} oleh $JK(E)$.

Untuk selanjutnya pengujian linieritas regresi disajikan pada table berikut:

DAFTAR II
DAFTAR ANALISIS VARIANSI UNTUK LINIERITAS REGRESI

Sumber Variasi	Dk	JK	RJK	F
Regresi (a)	1	$(\sum Y_i)^2/n$	$(\sum Y_i)^2$	S_{reg}^2
Regresi (b a)	1	$JK(b a)$	$S_{reg}^2 = JK(b a)$	S_{res}^2
Residu	n-2	$\sum (Y_i - \hat{Y})^2$	$S_{res}^2 = \frac{\sum (Y_i - \hat{Y})^2}{n-2}$	
Tuna Cocok	k - 2	$JK(TC)$	$S_{TC}^2 = \frac{JK(TC)}{k-2}$	$\frac{S_{TC}^2}{S_e^2}$
Kekeliruan	n - k	$JK(E)$	$S_e^2 = \frac{JK(E)}{n-k}$	

Dari daftar diatas diperoleh dua hasil:

- $F = S_{reg}^2/S_{res}^2$ untuk uji independen seperti telah diuraikan pada bagian sebelumnya.
- $F = S_{TC}^2/S_e^2$ yang akan digunakan untuk menguji kelinieran regresi. Dalam hal ini kita menolak hipotesis model regresi linier jika $F \geq F_{(1-\alpha)(k-2, n-k)}$.

Contoh:

7. Regresi Non linier

Jika ternyata model regresi tidaklinier, maka perlu dicari model lain yang sesuai dengan data yang ada, model tersebut adalah model non linier. Beberapa model regresi non linier diantaranya:

- a. Parabola Kuadratik $Y = a + bX + cX^2$
- b. Parabola Kubik $Y = a + bX + cX^2 + dX^3$
- c. Eksponensial $Y = ab^X$
- d. Geometrik $Y = aX^b$