ANALISIS REGRESI DAN KORELASI BERGANDA

Universitas Esa Unggul

ANALISIS REGRESI DAN KORELASI BERGANDA

Oleh Team Dosen
Universitas Esa Unggul

Contents

Pendahuluan

Analisis regresi merupakan salah satu teknik analisis data dalam statistika yang seringkali digunakan untuk mengkaji hubungan antara beberapa variabel dan meramal suatu variabel (Kutner, Nachtsheim dan Neter, 2004). Istilah "regresi" pertama kali dikemukakan oleh Sir Francis Galton (1822-1911), seorang antropolog dan ahli meteorologi terkenal dari Inggris. Dalam makalahnya yang berjudul "Regression towards mediocrity in hereditary stature", yang dimuat dalam Journal of the Anthropological Institute, volume 15, hal. 246-263, tahun 1885.

Galton menjelaskan bahwa biji keturunan tidak cenderung menyerupai biji induknya dalam hal besarnya, namun lebih medioker (lebih mendekati rata-rata) lebih kecil daripada induknya kalau induknya besar dan lebih besar daripada induknya kalau induknya sangat kecil (Draper dan Smith, 1992). Dalam mengkaji hubungan antara beberapa variabel menggunakan analisis regresi, terlebih dahulu peneliti menentukan satu variabel yang disebut dengan variabel tidak bebas dan satu atau lebih variabel bebas.

Jika ingin dikaji hubungan atau pengaruh satu variabel bebas terhadap variabel tidak bebas, maka model regresi yang digunakan adalah model regresi linier sederhana. Kemudian Jika ingin dikaji hubungan atau pengaruh dua atau lebih variabel bebas terhadap variabel tidak bebas, maka model regresi yang digunakan adalah model regresi linier berganda (multiple linear regression model). Kemudian untuk mendapatkan model regresi linier sederhana maupun model regresi linier berganda dapat diperoleh dengan melakukan estimasi terhadap parameter-parameternya menggunakan metode tertentu.

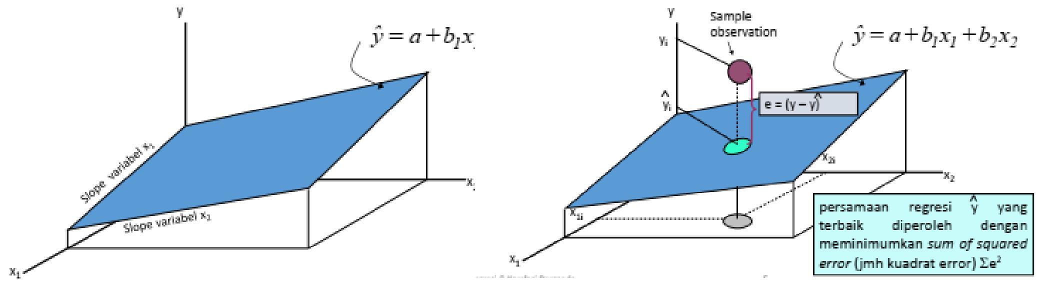
Adapun metode yang dapat digunakan untuk mengestimasi parameter model regresi linier sederhana maupun model regresi linier berganda adalah dengan metode kuadrat terkecil (ordinary least square/OLS) dan metode kemungkinan maksimum (maximum likelihood estimation/MLE) (Kutner et.al, 2004).

Analisis regresi berganda merupakan perluasan dari analisis regresi linier sederhana. Dalam regresi linier sederhana, dibuat analisis hubungan dua variabel (satu variabel independent dengan satu variabel dependent) yang dinyatakan dengan persamaan linier Y' = a + bX, dengan tujuan membuat prediksi tentang besarnya nilai Y (variabel dependent) berdasarkan nilai X (variabel independent) tertentu.

Prediksi perubahan variabel dependent (Y) akan menjadi lebih baik apabila dimasukkan lebih dari satu variabel independent dalam persamaan liniernya (X_1, X_2,……..X_n). Hubungan antara lebih dari satu variabel independent dengan satu variabel dependent inilah yang dibicarakan dalam analisis regresi linier berganda. Hubungan antara banyak variabel inilah yang sesungguhnya terjadi dalam dunia nyata, karena sebenarnya kebanyakan hubungan antar variabel dalam ilmu soisal merupakan hubungan statistikal, artinya bahwa perubahan nilai Y tidak mutlak hanya dipengaruhi oleh satu nilai X tertentu tetapi dipengaruhi oleh banyak nilai X.

Regresi Linier Berganda

Bentuk umum model regresi linier berganda dengan p variabel bebas adalah seperti pada persamaan berikut (Kutner, Nachtsheim dan Neter, 2004).



$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-} X_{ip-1} + \varepsilon_i$$

dengan:

Y_i =adalah variabel tidak bebas untuk pengamatan ke−i, untuk i = 1, 2, …, n

$\beta_0, \beta_1, \beta_2, .., \beta_{p-1}$ = adalah parameter.

$X_{i1}, X_{i2}, \dots, X_{ip-1}$ = adalah variabel bebas

ε_1 = adalah sisa (error) untuk pengamatan ke-i yang diasumsikan berdistribusi normal yang saling bebas dan identik dengan rata-rata 0 (nol) dan variansi σ^2.

Model regresi berganda dengan 1 variabel dependent (Y) dengan n variabel independent (X) adalah:

$Y' = a + b_1.X_1 + b_2.X_2 + \dots\dots + b_n.X_n + e$

Misalnya untuk n = 2, model regresinya adalah:

$Y' = a + b_1.X_1 + b_2.X_2 + e$

Dimana:

Y’	=	nilai Y prediksi
X_1	=	Variabel bebas 1
X_2	=	Variabel bebas 2
b_1	=	Koefisien regresi variabel bebas 1, adalah perubahan pada Y untuk setiap perubahan X_1 sebesar 1 unit dengan asumsi X_2 konstan
b_2	=	Koefisien regresi variabel bebas 2, adalah perubahan pada Y untuk setiap perubahan X_2 sebesar 1 unit dengan asumsi X_1 konstan
e	=	Kesalahan Prediksi (error)

Analisis regresi linier berganda, berdasarkan penelitian sampel dinyatakan dengan persamaan linier:

$Y' = a + b_1.X_1 + b_2.X_2 + \dots\dots + b_n.X_n$

Untuk kasus dua variabel independent, persamaan liniernya dinyatakan sebagai:

$Y' = a + b_1.X_1 + b_2.X_2$

Untuk mendapatkan nilai a, b_1 dan b_2 digunakan rumus-rumus sebagai berikut:

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

$$b_1 = \frac{\left(\sum X_2^2\right)\left(\sum X_1Y\right) - \left(\sum X_1X_2\right)\left(\sum X_2Y\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1X_2\right)^2}$$

$$b_2 = \frac{\left(\sum X_1^2\right)\left(\sum X_2Y\right) - \left(\sum X_1X_2\right)\left(\sum X_1Y\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1X_2\right)^2}$$

dimana:

- $\sum X_1^2 = \sum X_1^2 - \frac{\left(\sum X_1Y\right)^2}{n}$
- $\sum X_2^2 = \sum X_2^2 - \frac{\left(\sum X_2Y\right)^2}{n}$
- $\sum X_1Y = \sum X_1Y - \frac{\left(\sum X_1\right)\left(\sum Y\right)}{n}$
- $\sum X_2Y = \sum X_2Y - \frac{\left(\sum X_2\right)\left(\sum Y\right)}{n}$
- $\sum X_1X_2 = \sum X_1X_2 - \frac{\left(\sum X_1\right)\left(\sum Y\right)}{n}$
- $\sum Y^2 = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$
- $\bar{Y} = \sum \frac{Y}{n}$
- $\bar{X}_1 = \frac{\sum X_1}{n}$
- $\bar{X}_2 = \frac{\sum X_2}{n}$

Asumsi-Asumsi Model Regresi Linier Berganda

Menurut Gujarati (2003) asumsi-asumsi pada model regresi linier berganda adalah sebagai berikut:

1. Model regresinya adalah linier dalam parameter.
2. Nilai rata-rata dari error adalah nol.
3. Variansi dari error adalah konstan (homoskedastik).
4. Tidak terjadi autokorelasi pada error.
5. Tidak terjadi multikolinieritas pada variabel bebas.
6. Error berdistribusi normal.

Estimasi Parameter Model Regresi Linier Berganda

Estimasi parameter ini bertujuan untuk mendapatkan model regresi linier berganda yang akan digunakan dalam analisis. Pada materi pelatihan ini, metode yang digunakan untuk mengestimasi parameter model regresi linier berganda adalah metode kuadrat terkecil atau sering juga disebut dengan metode ordinary least square (OLS). Metode OLS ini bertujuan meminimumkan jumlah kuadrat error.

Pengujian Parameter Model Regresi Linier Berganda

Pengujian parameter ini bertujuan untuk mengetahui ada atau tidaknya pengaruh variabel bebas terhadap variabel tidak bebas, baik secara serentak maupun secara parsial.

Pengujian Parameter Secara Serentak (Simultan)

Prosedur pengujian parameter secara simultan adalah sebagai berikut:

1. Membuat hipotesis.
   $H_0 = \beta_1 = \beta_2 = \cdots \beta_{p-1} = 0$
   $H_0 \neq 0$
2. Menentukan tingkat signifikansi (α).
   Tingkat signifikansi (α) yang seringkali digunakan dalam penelitian adalah 5%.
3. Menentukan statistik uji.
   Statistik uji yang digunakan adalah:
   $$F = \frac{RKR}{RKE}$$
   dengan:
   RKR adalah rata-rata kuadrat regresi (dapat diperoleh dari Tabel Analisis Variansi).
   RKE adalah rata-rata kuadrat error (dapat diperoleh dari Tabel Analisis Variansi).
4. Menentukan daerah kritik (penolakan H0).
   Daerah kritik yang digunakan adalah H0 ditolak bila $F > F_{(a;p-1,\ n-p)}$.
   Dengan $F_{(a;p-1,\ n-p)}$ disebut dengan F tabel.
   Selain dari daerah kritik di atas, dapat juga digunakan daerah kritik yang lain yaitu jika nilai peluang (Sig.) < tingkat signifikansi (α) maka H0 ditolak.
5. Menarik kesimpulan.

Pengujian Parameter Secara Individu (Parsial)

Prosedur pengujian parameter secara parsial adalah sebagai berikut:

1. Membuat hipotesis.
   $H_0 = \beta_1 = \beta_2 = \cdots \beta_{p-1} = 0$
   $H_0 \neq 0$
2. Menentukan tingkat signifikansi (α).

Tingkat signifikansi (α) yang seringkali digunakan dalam penelitian adalah 5%.

3. Menentukan statistik uji.
   Statistik uji yang digunakan adalah:
   $t = \frac{b_k}{s(b_k)}$
   b_k = adalah nilai taksiran parameter ($b_k$yang diperoleh dari metode OLS).
   $s(b_k)$ = adalah standar deviasi nilai taksiran parameter b_k.
4. Menentukan daerah kritik (penolakan H0).
   Daerah kritik yang digunakan adalah H0 ditolak bila $t > t\ tabel$
5. Menarik kesimpulan.

Pelanggaran-Pelanggaran Terhadap Asumsi Regresi Linier Berganda

Dalam analisis regresi linier berganda terdapat beberapa pelanggaran-pelanggaran yang seringkali dilakukan terhadap asumsi-asumsinya, diantaranya diuraikan berikut ini.

Multikolinieritas

Multikolinieritas adalah terjadinya hubungan linier antara variabel bebas dalam suatu model regresi linier berganda (Gujarati, 2003). Hubungan linier antara variabel bebas dapat terjadi dalam bentuk hubungan linier yang sempurna (perfect) dan hubungan linier yang kurang sempurna (imperfect).

Adapun dampak adanya multikolinieritas dalam model regresi linier berganda adalah (Gujarati, 2003 dan Widarjono, 2007):

1. Penaksir OLS masih bersifat BLUE, tetapi mempunyai variansi dan kovariansi yang besar sehingga sulit mendapatkan taksiran (estimasi) yang tepat.
2. Akibat penaksir OLS mempunyai variansi dan kovariansi yang yang besar, menyebabkan interval estimasi akan cenderung lebih lebar dan nilai hitung statistik uji t akan kecil, sehingga membuat variabel bebas secara statistik tidak signifikan mempengaruhi variabel tidak bebas.
3. Walaupun secara individu variabel bebas tidak berpengaruh terhadap variabel tidak bebas melalui uji t, tetapi nilai koefisien determinasi (R2) masih bisa relatif tinggi.

Selanjutnya untuk mendeteksi adanya multikolinieritas dalam model regresi linier berganda dapat digunakan nilai variance inflation factor (VIF) dan tolerance (TOL) dengan ketentuan jika nilai VIF melebihi angka 10, maka terjadi multikolinieritas dalam model regresi. Kemudian jika nilai TOL sama dengan 1, maka tidak terjadi multikolinieritas dalam model regresi.

Heteroskedastisitas

Heteroskedastisitas adalah variansi dari error model regresi tidak konstan atau variansi antar error yang satu dengan error yang lain berbeda (Widarjono, 2007). Dampak adanya heteroskedastisitas dalam model regresi adalah walaupun estimator OLS masih linier dan tidak bias, tetapi tidak lagi mempunyai variansi yang minimum dan menyebabkan perhitungan standard error metode OLS tidak bisa dipercaya kebenarannya.

Selain itu interval estimasi maupun pengujian hipotesis yang didasarkan pada distribusi t maupun F tidak bisa lagi dipercaya untuk evaluasi hasil regresi. Akibat dari dampak heteroskedastisitas tersebut menyebabkan estimator OLS tidak menghasilkan estimator yang BLUE dan hanya menghasilkan estimator OLS yang linear unbiased estimator (LUE).

Selanjutnya dilakukan deteksi masalah heteroskedastisitas dalam model regresi. Salah satu cara yang dapat digunakan untuk mendeteksi adanya heteroskedastisitas dalam model regresi adalah dengan Metode Glejser. Glejser merupakan seorang ahli ekonometrika dan mengatakan bahwa nilai variansi variabel error model regresi tergantung dari variabel bebas.

Selanjutnya untuk mengetahui apakah pola variabel error mengandung heteroskedastisitas Glejser menyarankan untuk melakukan regresi nilai mutlak residual dengan variabel bebas. Jika hasil uji F dari model regresi yang diperoleh tidak signifikan, maka tidak ada heteroskedastisitas dalam model regresi (Widarjono, 2007).

Autokorelasi

Autokorelasi adalah terjadinya korelasi antara satu variabel error dengan variable error yang lain. Autokorelasi seringkali terjadi pada data *time series* dan dapat juga terjadi pada data *cross-section* tetapi jarang (Widarjono, 2007). Adapun dampak dari adanya autokorelasi dalam model regresi adalah sama dengan dampak dari heteroskedastisitas yang telah diuraikan di atas, yaitu walaupun estimator OLS masih linier dan tidak bias, tetapi tidak lagi mempunyai variansi yang minimum dan menyebabkan perhitungan standard error metode OLS tidak bisa dipercaya kebenarannya.

Selain itu interval estimasi maupun pengujian hipotesis yang didasarkan pada distribusi t maupun F tidak bisa lagi dipercaya untuk evaluasi hasil regresi. Akibat dari dampak adanya autokorelasi dalam model regresi menyebabkan estimator OLS tidak menghasilkan estimator yang BLUE dan hanya menghasilkan estimator OLS yang LUE (Widarjono, 2007).

Selanjutnya untuk mendeteksi adanya autokorelasi dalam model regresi linier berganda dapat digunakan metode Durbin-Watson. Durbin-Watson telah berhasil mengembangkan suatu metode yang digunakan untuk mendeteksi adanya masalah autokorelasi dalam model regresi linier berganda menggunakan pengujian hipotesis dengan statistik uji yang cukup popular.

Kemudian Durbin-Watson berhasil menurunkan nilai kritis batas bawah (dL) dan batas atas (dU) sehingga jika nilai d hitung dari persamaan (6.1) terletak di luar nilai kritis ini, maka ada atau tidaknya autokorelasi baik positif atau negatif dapat diketahui. Deteksi autokorelasi pada model regresi linier berganda dengan metode Durbin-Watson adalah seperti pada Tabel berikut:

Nilai Statistik Durbin-Watson	Hasil
$0 < d < d_L$	Menolak hipotesis nol; ada autokorelasi positif
$d_L \leq d \leq d_U$	Daerah keragu-raguan; tidak ada keputusan
$d_U \leq d \leq 4 - d_U$	Menerima hipotesis nol; tidak ada autokorelasi positif/negatif
$4 - d_U \leq d \leq 4 - d_L$	Daerah keragu-raguan; tidak ada keputusan
$4 - d_L \leq d \leq 4$	Menolak hipotesis nol; ada autokorelasi positif

Sumber : Widarjono (2007)

Salah satu keuntungan dari uji Durbin-Watson yang didasarkan pada error adalah bahwa setiap program komputer untuk regresi selalu memberi informasi statistik d. Adapun prosedur dari uji Durbin-Watson adalah (Widarjono, 2007):

1. Melakukan regresi metode OLS dan kemudian mendapatkan nilai errornya.
2. Menghitung nilai d dari persamaan (kebanyakan program komputer secara otomatis menghitung nilai d).
3. Dengan jumlah observasi (n) dan jumlah variabel bebas tertentu tidak termasuk konstanta (p-1), kita cari nilai kritis. d_L dan d_U di statistik Durbin-Watson.
4. Keputusan ada atau tidaknya autokorelasi dalam model regresi didasarkan pada Tabel diatas.

Selain Kriteria uji seperti pada Tabel, dapat juga digunakan kriteria lain untuk mendeteksi adanya autokorelasi dalam model regresi linier berganda adalah sebagai berikut (Santoso, 2000):

1. Jika nilai $d < -2$, maka ada autokorelasi positif.
2. Jika nilai $-2 \leq d \leq 2$ maka tidak ada autokorelasi.
3. Jika nilai. $d > 2$ maka ada autokorelasi negative.

Analisis Korelasi Berganda

Analisis korelasi berganda merupakan perluasan dari analisis korelasi sederhana. Dalam analisis korelasi berganda bertujuan untuk mengetahui bagaimana derajat hubungan antara beberapa variabel independent (Variabel X_1, X_2, ……., X_k) dengan variabel dependent (Variabel Y) secara bersama-sama.

Asumsi-asumsi sehubungan dengan analisis regresi berganda tersebut adalah:

1. Variabel-Variabel independent dan variabel dependent mempunyai hubungan linier
2. Semua variabel, baik variabel-variabel independent maupun variabel dependent, merupakan variabel-variabel random kontinyu.
3. Distribusi kondisional nilai masing-masing variabel berdistribusi normal (multivariate normal distribution)
4. Untuk berbagai kombinasi nilai variabel yang satu dengan yang lain tertentu, varaince dari distribusi kondisional masing-masing variabel adalah homogen (asumsu homoscedasticity berlaku untuk semua variabel)

5. Untuk masing-masing variabel, nilai observasi yang satu dengan yang lain, tidak berkaitan.

Berdasarkan korelasi berganda, yang diberi notasi $R_{Y.12.....n}$ dihitung melalui jalur terjadinya hubungan antara beberapa variabel independent (X_1, X_2, ……., X_n) dengan satu variabel dependent (Y), yakni yang berupa regresi linier berganda Y' = a + $b_1.X_1$ + $b_2.X_2$ + …… + $b_n.X_n$.

Berdasarkan adanya regresi berganda tersebut, koefisien korelasi linier berganda tersebut dihitung dengan menggunakan rumus sebagai berikut:

$$R_{Y.12} = \sqrt{\frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y + ...+ b_n \sum X_n Y}{\sum Y^2}}$$

Berdasarkan contoh tentang hubungan antara penghasilan keluarga (X_1) dan besar keluarga (X_2) dengan pengeluaran untuk bahan makanan (Y), koefisien korelasi linier bergandanya dinyatakan dengan:

$$R_{Y.12} = \sqrt{\frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2}}$$

dimana,

b_1 = 0,0544
b_2 = 0,02304
$\Sigma X_1 Y$ = 16,858
$\Sigma X_2 Y$ = 9,96
ΣY^2 = 3,777

Jadi koefisien korelasi berganda dari contoh tersebut adalah:

$$R_{Y.12} = \sqrt{\frac{0{,}0544(16{,}858) + 0{,}2304(9{,}96)}{3{,}777}} = \sqrt{\frac{3{,}2118592}{3{,}777}} = 0{,}92$$

Sedang koefisien determinasi berganda (R^2) dari contoh tersebut adalah:

$$R^2_{Y.12} = \frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2} \quad atau \quad \frac{SSR}{SST}$$

$$R^2_{Y.12} = \frac{0{,}0544(16{,}858) + 0{,}2304(9{,}96)}{3{,}777} = \frac{3{,}2118592}{3{,}777} = 0{,}85 \; atau \; (0{,}92)^2$$

Angka tersebut menunjukkan bahwa sekitar 85% dari variasi pengeluaran untuk bahan makanan (Y) dijelaskan oleh kombinasi dari penghasilan keluarga (X_1) dan besar keluarga (X_2). Sisanya yakni 15% dijelaskan oleh variabel independent lainnya yang tidak teramati.

Dari kasus korelasi linier berganda, peneliti dapat menghitung koefisien korelasi parsialnya. Korelasi parsial (partial correlation) adalah korelasi antara sebuah variabel dependent (Y) dengan sebuah variabel independent (X), sementara sejumlah variabel independent lainnya konstan.

Apabila variabel independentnya ada dua buah yaitu X_1 dan X_2, maka koefisien parsial yang ada ialah r_{Y12} dan r_{Y21}, yang masing-masing menunjukkan koefisien korelasi antara Y dengan X_1 apabila X_2 konstan dan koefisien korelasi antara Y dengan X_2 apabila X_1 konstan. Seperti dalam contoh tersebut dimuka, r_{Y12} menunjukkan korelasi antara penghasilan keluarga (X_1) dengan pengeluaran untuk bahan makanan (Y) apabila besar keluarga (X_2) konstan. Dan r_{Y21} menunjukkan korelasi antara besar keluarga (X_2) dengan pengeluaran untuk bahan makanan (Y) apabila penghasilan keluarga (X_1) konstan. Rumus-rumusnya adalah:

$$r_{Y.1(2)} = \frac{rX_1Y - (rX_2Y)(rX_1X_2)}{\sqrt{(1-r^2X_2Y)(1-r^2X_1X_2)}}$$

$$r_{Y.2(1)} = \frac{rX_2Y - (rX_1Y)(rX_1X_2)}{\sqrt{(1-r^2X_1Y)(1-r^2X_1X_2)}}$$

Untuk menghitung koefisien korelasi parsialnya terlebih dahulu harus dihitung koefisien korelasi sederhana antara X_1 dengan Y, X_2 dengan Y dan antara X_1 dengan X_2. Berdasarkan contoh diawal:

$$r_{X_1Y} = \frac{\sum X_1Y}{\sqrt{(\sum X_1^2)(\sum Y^2)}} = \frac{16{,}858}{40{,}693} = 0{,}41$$

$$r_{X_2Y} = \frac{\sum X_2Y}{\sqrt{(\sum X_2^2)(\sum Y^2)}} = \frac{9{,}96}{13{,}797} = 0{,}72$$

$$r_{X_1X_2} = \frac{\sum X_2X_2}{\sqrt{(\sum X_1^2)(\sum X_2^2)}} = \frac{-30{,}38}{\sqrt{(438{,}416)(50{,}4)}} = \frac{-30{,}38}{148{,}65} = -0{,}20$$

Koefisien korelasi parsial:

$$r_{Y.1(2)} = \frac{0{,}41-(0{,}72)(-0{,}20)}{\sqrt{(1-(0{,}72)^2)(1-(0{,}20)^2)}} = \frac{0{,}554}{(0{,}694)(0{,}98)} = 0{,}81$$

$$r_{Y.2(1)} = \frac{0{,}72-(0{,}41)(-0{,}20)}{\sqrt{(1-(0{,}41)^2)(1-(0{,}20)^2)}} = \frac{0{,}802}{(0{,}912)(0{,}98)} = 0{,}90$$

Koefisien determinasi dan pengertiannya:

1. $r^2 X_1Y = (0{,}41)^2 = 0{,}17$

   Sekitar 17% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam penghasilan keluarga (X_1) dimana faktor lain tidak dipertimbangkan.
2. $r^2 X_2Y = (0{,}72)^2 = 0{,}52$

Sekitar 52% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam besar keluarga (X_2) dimana faktor lain tidak dipertimbangkan.

3. $r^2_{Y_{1(2)}} = (0{,}81)^2 = 0{,}66$

   Sekitar 66% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam penghasilan keluarga (X1), apabila pengaruh dari besar keluarga (X2) dianggap konstan.

4. $r^2_{Y_{2(1)}} = (0{,}90)^2 = 0{,}81$

   Sekitar 81% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam besar keluarga (X2), apabila pengaruh dari penghasilan keluarga (X1) dianggap konstan.

5. $R^2_{Y_{12}} = (0{,}92)^2 = 0{,}85$

   Sekitar 85% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh kombinasi dari penghasilan keluarga (X_1) dan besar keluarga (X_2).

Contoh Soal

Untuk memberikan gambaran bagaimana membuat analisis regresi linier berganda, diberikan contoh sebagai berikut; Misalnya kita hendak memprediksi besarnya pengeluaran untuk bahan makanan per bulan (variabel Y) berdasarkan penghasilan keluarga per bulan (variabel X_1) dan banyaknya/besar keluarga (variabel X_2). Berdasarkan sampel random 15 keluarga diperoleh informasi sebagai berikut:

Penghasilan Kelurga (Ratusan Ribu Rp/Bulan) X_1	Besar Keluarga X_2	Pengeluaran Bahan Makanan (Ratusan Ribu Rp/Bulan) Y
5,5	1	0,8
8,9	1	1,0
21,8	1	1,7
6,8	2	1,4
7,5	2	1,2
17,2	2	1,8
22,1	2	1,9
19,0	3	2,3
12,0	3	1,7
14,0	4	1,5
10,9	4	1,8
7,5	5	2,0
14,0	5	2,2
13,7	6	2,8
6,0	7	2,1

Untuk memperoleh persamaan garis regresi linier tentang hubungan antara variabel penghasilan keluarga (X_1) dan besar keluarga (X_2) dengan variabel pengeluaran untuk bahan makanan (Y) periksa tabel berikut:

Tabel Komputasi Persamaan Regresi Linier Berganda Berdasarkan Hubungan variable Penghasilan Keluarga (X1) dan Variabel Besar Keluarga (X2) dengan Variabel Pengeluaran Untuk Bahan Makanan (Y).

X_1	X_2	Y	X^2_1	X^2_2	Y^2	X_1Y	X_2Y	X_1X_2
5,5	1	0,8	30,25	1	0,64	4,4	0,8	5,5
8,9	1	1	79,21	1	1	8,9	1	8,9
21,8	1	1,7	475,24	1	2,89	37,06	1,7	21,8
6,8	2	1,4	46,24	4	1,96	9,52	2,8	13,6
7,5	2	1,2	56,25	4	1,44	9	2,4	15
17,2	2	1,8	295,84	4	3,24	30,96	3,6	34,4
22,1	2	1,9	488,41	4	3,61	41,99	3,8	44,2
19	3	2,3	361	9	5,29	43,7	6,9	57
12	3	1,7	144	9	2,89	20,4	5,1	36
14	4	1,5	196	16	2,25	21	6	56
10,9	4	1,8	118,81	16	3,24	19,62	7,2	43,6
7,5	5	2	56,25	25	4	15	10	37,5
14	5	2,2	196	25	4,84	30,8	11	70
13,7	6	2,8	187,69	36	7,84	38,36	16,8	82,2
6	7	2,1	36	49	4,41	12,6	14,7	42
186,9	48	26,2	2767,19	204	49,54	343,31	93,8	567,7
Mean								
12,46	3,2	1,74						

ΣX_1^2 = 2767,19 – $(186,9)^2$/15 = 438,416

ΣX_2^2 = 204 – $(48)^2$/15 = 50,4

ΣY^2 = 49,54 – $(26,2)^2$/15 = 3,777

ΣX_1Y = 343,31 – (186,9) (26,2)/15 = 16,858

ΣX_2Y = 93,8 – (48) (26,2)/15 = 9,96

ΣX_1X_2 = 567,7 – (186,9) (48)/15 = -30,38

Koefisien regresinya adalah:

$$b_1 = \frac{(50{,}4)(16{,}858)-(-30{,}38)(9{,}96)}{(438{,}416)(50{,}4)-(-30{,}38)^2} = \frac{849{,}6432+302{,}5848}{22.096{,}1664-922{,}9444} = \frac{1.152{,}228}{21.173{,}222} = 0{,}05442$$

$$b_1 = \frac{(438{,}416)(9{,}96)-(-30{,}38)(16{,}858)}{(438{,}416)(50{,}4)-(-30{,}38)^2} = \frac{4.366{,}62336+512{,}14604}{22.096{,}1664-922{,}9444} = \frac{4.828{,}7694}{21.173{,}222} = 0{,}23042$$

Intersepnya adalah:

a = 1,74 – 0,05442 (12,46) – 0,23042 (3,2)
a = 1,74 – 0,6780732 – 0,737344

a = 1,74 – 1,4154172 = 0,3245828

Persamaan regresi linier bergandanya adalah:

Y' = 0,3246 + 0,0544.X_1 + 0,2304.X_2

Pengertian persamaan tersebut adalah : Pertama, apabila X_2 konstan, pertambahan satu unit pada X_1 akan mempunyai pengaruh menaikkan 0,0544 unit pada Y. Kedua, apabila X_1 kostan, pertambahan satu unit pada X_2, akan mempunyai pengaruh menaikkan 0,2304 unit pada Y. Ketiga, apabila X_1 dan X_2 sama dengan nol, besarnya Y adalah 0,3246 satuan.

Berdasarkan persamaan tersebut dapat dibuat prediksi/ramalan nilai-nilai Y berdasarkan kombinasi nilai X_1 dan X_2 tertentu misalnya nilai X_1 = 5,5 dan X_2 = 1, maka nilai Y adalah Y = 0,3246 + 0,0544 (5,5) + 0,2304 (1) = 0,8542

Standard error of estimates dinyatakan dengan rumus:

$$S_{y12} = \sqrt{\frac{\sum Y^2 - b_1 \sum X_1 Y - b_2 \sum X_2 Y}{n-3}}$$

dimana,

n = jumlah observasi

3 = banyak koefisien

Berdasarkan contoh tersebut dimuka, besarnya standard error of estimate adalah:

Rumus (1):

$$S_{Y12} = \sqrt{\frac{0{,}565654}{15-3}} = 0{,}217$$

Rumus (2):

$$S_{y12} = \sqrt{\frac{3{,}777 - 0{,}0544(16{,}858) - 0{,}2304(9{,}96)}{15-3}}$$

$$S_{y12} = \sqrt{\frac{3{,}777 - 0{,}9170752 - 2{,}294784}{12-3}} = \sqrt{\frac{3{,}777 - 3{,}2118592}{12}} = \sqrt{\frac{0{,}56514}{12}}$$

$$= \sqrt{0{,}047095} = 0{,}217$$

Daftar Pustaka

Draper, N. dan Smith, H. 1992. Analisis Regresi Terapan. Edisi Kedua. Terjemahan Oleh Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.

Gujarati, N.D. 2003. Basic Econometrics. 4th ed. New York: McGraw-Hill Companies,Inc.

Kutner, M.H., C.J. Nachtsheim., dan J. Neter. 2004. Applied Linear Regression Models. 4^{th} ed. New York: McGraw-Hill Companies, Inc.

Santoso, S. 2000. Buku Latihan SPSS Statistik Parametrik. Jakarta: Elex Media Komputindo.

Sembiring, R.K. 2003. Analisis Regresi. Edisi Kedua. Bandung: Institut Teknologi Bandung.

Widarjono, A. 2007. Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis. Edisi Kedua. Yogyakarta: Ekonisia Fakultas Ekonomi Universitas Islam Indonesia.