BAB III

REGRESI BERGANDA

Analisis Regresi Berganda adalah perluasan dari Analisis Regresi Sederhana, yang mana variabel respons Y berhubungan dengan sejumlah variabel peramal $X_1, X_2, X_3, \ldots X_k$. Dan dalam bab ini analisis berganda didasarkan pada asumsi bahwa Y dinyatakan dengan model statistik linear. Pengartian model linear disini yaitu model yang parameter-parameternya linear.

III.1. Regresi Linear Berganda dengan k Buah Variabel Bebas

Model regresinya,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i \quad \ldots\ldots\ldots \; (3:1)$$

dimana,

$i = 1,2, \ldots n$ dan $k < n$

Asumsi-asumsi yang harus dipenuhi :

1. ε_i didistribusikan secara normal ,dengan rerata (mean) = 0 dan varians $= \sigma^2$, yaitu $E(\varepsilon_i) = 0$; $Var(\varepsilon_i^2) = \sigma^2$

   Secara ringkas dapat dinyatakan, $\varepsilon_i \sim N(0, \sigma^2)$
2. Tidak ada korelasi antara ε_i dan ε_j untuk $i \neq j$,yaitu $Cov(\varepsilon_i, \varepsilon_j) = 0$

III.1.1. Metode Kuadrat Terkecil (The Method of Least Squares)

Suatu persamaan prediksi yang didasarkan pada sejumlah variabel $X_1, X_2, X_3, \ldots, X_k$ dapat dilakukan dengan metode kuadrat terkecil. Jumlah kuadrat penyimpangan nilai Y yang diamati dari model yang disesuaikan (dicocokkan atau fitted) adalah,

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})^2$$

dimana,

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

Agar SSE minimum (terkecil) maka dengan membuat turunan parsial terhadap $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ dan menyamakannya dengan nol,

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = 0$$

$$2\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-1) = 0$$

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = 0$$

$$2\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-X_{1i}) = 0$$

$$\frac{\partial SSE}{\partial \hat{\beta}_2} = 0$$

$$2\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-X_{2i}) = 0$$

$$\vdots$$

$$\frac{\partial SSE}{\partial \hat{\beta}_k} = 0$$

$$2\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})(-X_{ki}) = 0$$

Setelah disederhanakan, diperoleh persamaan normal,

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} + \dots + \hat{\beta}_k \sum_{i=1}^{n} X_{ki} = \sum_{i=1}^{n} Y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n} X_{1i}+\hat{\beta}_1\sum_{i=1}^{n} X_{1i}^2+ \ldots + \hat{\beta}_k\sum_{i=1}^{n} X_{1i}X_{ki}=\sum_{i=1}^{n} X_{1i}Y_i$$

$$\vdots$$

$$\hat{\beta}_k\sum_{i=1}^{n} X_{ki}+\hat{\beta}_1\sum_{i=1}^{n} X_{1i}X_{ki}+ \ldots + \hat{\beta}_k\sum_{i=1}^{n} X_{ki}^2 = \sum_{i=1}^{n} X_{ki}Y_i$$

Dari persamaan normal, maka dapat dihitung harga koefisien regresi estimasi $\hat{\beta}_i$, i= 1,2,3,...,k. Apabila harga -harga $\hat{\beta}_i$ telah diperoleh, maka persamaan regresi estimasinya adalah,

$$\hat{Y} = \hat{\beta}_0+ \hat{\beta}_1X_1+\hat{\beta}_2X_2+ \ldots + \hat{\beta}_kX_k \quad \ldots\ldots\ldots (\ 3:2\ )$$

Selanjutnya parameter varians σ^2 diestimasi dengan s^2 yaitu,

$$s^2 = \frac{SSE}{n-(k+1)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\ 3:3\ )$$

$$= \frac{\sum_{i=1}^{n} (Y_i-\hat{Y}_i)^2}{n-(k+1)}$$

$$= \frac{\sum_{i=1}^{n} Y_i^2- \hat{\beta}_0\sum_{i=1}^{n} Y_i- \hat{\beta}_1\sum_{i=1}^{n} X_1Y_i- \ldots - \hat{\beta}_k\sum_{i=1}^{n} X_kY_i}{n-(k+1)}$$

Contoh III.1.

Tabel III.1. merupakan data yang dihasilkan dari 13 ulangan suatu eksperimen dengan menggunakan dua variabel independen (X_1 dan X_2), dan satu variabel respons Y. Dan ingin dicari persamaan regresi linear ganda estimasinya.

Tabel III.1.

Y	X_1	X_2
25,5	1,74	5,30

Y	X_1	X_2
.	.	.
31,2	6,32	5,42
25,9	6,22	8,41
38,4	10,52	4,63
18,4	1,19	11,60
26,7	1,22	5,85
26,4	4,10	6,62
25,9	6,32	8,72
32,0	4,08	4,42
25,2	4,15	7,60
39,7	10,15	4,83
25,7	1,72	3,12
25,5	1,70	5,30

Penyelesaian :

Hasil perhitungan adalah sebagai berikut,

$n = 13$ $\quad \Sigma Y = 367,5$ $\quad \Sigma Y^2 = 11400,15$

$\Sigma X_1 = 59,43$ $\quad \Sigma X_1^2 = 394,7255$

$\Sigma X_2 = 81,82$ $\quad \Sigma X_2^2 = 576,7264$

$\Sigma X_1 Y = 1877,567$

$\Sigma X_2 Y = 2246,661$

$\Sigma X_1 X_2 = 360,6621$

Sehingga persamaan normal dapat ditulis :

$$13\,\hat{\beta}_0 + 59,43\,\hat{\beta}_1 + 81,82\,\hat{\beta}_2 = 367,5$$
$$59,43\,\hat{\beta}_0 + 394,7255\,\hat{\beta}_1 + 360,6621\,\hat{\beta}_2 = 1877,567$$
$$81,82\,\hat{\beta}_0 + 360,6621\,\hat{\beta}_1 + 576,7264\,\hat{\beta}_2 = 2246,661$$

Sistem persamaan dapat diselesaikan untuk mendapatkan harga -harga $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, dan diperoleh,

$$\hat{\beta}_0 = 36{,}694 \qquad \hat{\beta}_1 = 1{,}031 \qquad \hat{\beta}_2 = -1{,}870$$

Sehingga persamaan regresi estimasinya adalah,

$$\hat{Y} = 36{,}694 + 1{,}031\ X_1 - 1{,}870\ X_2$$

Adapun varians estimasinya dapat dihitung,

$$s^2 = \frac{SSE}{n-(k+1)}$$

$$\begin{aligned} SSE &= 11400{,}15 + (36{,}694)(367{,}5) + (1{,}031)(1877{,}567) \\ &\quad + (-1{,}870)(2246{,}661) \\ &= 40{,}01 \end{aligned}$$

Maka,

$$s^2 = \frac{40{,}01}{10} = 4{,}001$$

III.2. Regresi Polinomial

Model regresinya,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_p X_i^p + \varepsilon_i \qquad (3:4)$$

dimana p adalah derajad polinomial dan ε_i adalah random error yang berkaitan dengan respons Y_i.

Model regresi polinomial dipandang sebagai kasus dari model umum regresi linear ganda, yaitu dengan memandang $X_j = X^j$ dimana, $j = 1, 2, \dots, p$

III.2.1. Metode Kuadrat Terkecil

(The Method of Least Squares)

Koefisien regresi estimasi, β_0, β_1, β_2, yaitu dengan metode kuadrat terkecil,

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2 - \ldots - \hat{\beta}_p X_i^p)^2 \quad (3:5)$$

Persamaan normal diperoleh, yaitu dengan meminimumkan SSE dengan membuat turunan parsial terhadap $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, ... , $\hat{\beta}_p$ dan menyama dengankan nol,

$$\frac{\partial\, SSE}{\partial \hat{\beta}_0} = 0 \quad , \frac{\partial\, SSE}{\partial \hat{\beta}_1} = 0 \quad , \frac{\partial\, SSE}{\partial \hat{\beta}_2} = 0 \quad , \frac{\partial\, SSE}{\partial \hat{\beta}_p} = 0$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2 - \ldots - \hat{\beta}_p X_i^p)(-1) = 0$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2 - \ldots - \hat{\beta}_p X_i^p)(-X_i) = 0$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2 - \ldots - \hat{\beta}_p X_i^p)(-X_i^2) = 0$$

$$\vdots$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2 - \ldots - \hat{\beta}_p X_i^p)(-X_i^p) = 0$$

Dengan menyederhanakan, diperoleh persamaan normal,

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 + \ldots + \hat{\beta}_p \sum_{i=1}^{n} X_i^p = \sum_{i=1}^{n} Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3 + \ldots + \hat{\beta}_p \sum_{i=1}^{n} X_i^{p+1} = \sum_{i=1}^{n} X_i Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4 + \ldots + \hat{\beta}_p \sum_{i=1}^{n} X_i^{p+2} = \sum_{i=1}^{n} X_i^2 Y_i$$

$$\vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i^p + \hat{\beta}_1 \sum_{i=1}^{n} X_i^{p+1} + \hat{\beta}_2 \sum_{i=1}^{n} X_i^{p+2} + \ldots + \hat{\beta}_p \sum_{i=1}^{n} X_i^{2p} = \sum_{i=1}^{n} X_i^p Y_i$$

Dari persamaan tersebut maka dapat dihitung koefisien regresi estimasi $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$. Sehingga persamaan regresi estimasinya adalah,

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + \ldots + \hat{\beta}_p X^p \quad \ldots\ldots\ldots\ldots (3:6)$$

Untuk parameter varians σ^2 diestimasi dengan s^2 yaitu,

$$s^2 = \frac{SSE}{n-(p+1)} \quad \ldots\ldots\ldots\ldots (3:7)$$

$$= \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-(p+1)}$$

$$= \frac{\sum_{i=1}^{n} Y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 Y_i - \ldots - \hat{\beta}_p \sum_{i=1}^{n} X_i^p Y_i}{n-(p+1)}$$

Contoh III.2.

Data sampel dari 10 rumah di kota B yaitu yang berhubungan dengan ukuran rumah (X, meter persegi) dan penggunaan aliran listrik bulanan (Y, killowatt-jam), hasilnya diberikan dalam Tabel III.2.

X	Y
1.290	1.182
1.350	1.172
1.470	1.264
1.600	1.493
1.710	1.571
1.840	1.711
1.980	1.804
2.230	1.840
2.400	1.956
2.930	1.954

Penyelesaian :

Dari hasil perhitungan didapat,

$n = 10$ $\quad \Sigma Y = 15947$ $\quad \Sigma Y^2 = 26277083$

$\Sigma X = 18800$ $\quad \Sigma X^2 = 37755400$ $\quad \Sigma X^3 = 8093,9 \times 10^7$

$\Sigma X^4 = 1843 \times 10^{14}$

$\Sigma XY = 31283250$ $\quad \Sigma X^2Y = 6,53069 \times 10^{10}$

Dengan mensubstitusikan nilai-nilai tersebut dalam persamaan normal, yaitu :

$$10\ \hat{\beta}_0 + 18800\ \hat{\beta}_1 + 37755400\ \hat{\beta}_2 = 15947$$

$$18800\ \hat{\beta}_0 + 37755400\ \hat{\beta}_1 + 8093,9 \times 10^7 \hat{\beta}_2 = 31283250$$

$$37755400 \hat{\beta}_0 + 8093,9 \times 10^7 \hat{\beta}_1 + 1843 \times 10^{14} \hat{\beta}_2 = 6,53069 \times 10^{10}$$

maka dapat dihitung harga-harga $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$,yaitu :

$\hat{\beta}_0 = -1216,14389$

$\hat{\beta}_1 = 2,39893$

$\hat{\beta}_2 = -0,00045$

Jadi persamaan regresi estimasi (prediksi) adalah:

$$\hat{Y} = -1216,14389 + 2,39893\ X - 0,00045\ X^2$$

Varians estimasi adalah :

$$s^2 = \frac{SSE}{n - (p+1)}$$

$$SSE = 26277083 - (-1216,1439)(15947) - (2,3989)(31283250) - (-0,00045)(6,53069 \times 10^{10})$$

$$= 13646,35$$

Maka,

$$s^2 = \frac{13646,35}{10-3}$$

$$= \frac{13646,35}{7}$$

$$= 1949,4786$$

III.3. Koefisien Determinasi Ganda

Untuk melihat pengaruh semua variabel independen secara bersama terhadap variabel dependen maka dipergunakannya koefisien determinasi ganda, R^2

Didefinisikan,

$$R^2 = \frac{\text{Total SS} - \text{SSE}}{\text{Total SS}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3:8)$$

$$= \frac{\text{SSR}}{\text{Total SS}}$$

dimana,

$$0 \leq R^2 \leq 1$$

Nilai R^2 yang kecil berarti bahwa $X_1, X_2, X_3, \ldots, X_k$ menyumbang sangat sedikit informasi untuk peramalan Y. Nilai R^2 mendekati 1 berarti bahwa $X_1, X_2, X_3, \ldots, X_k$ memberikan hampir semua informasi yang diperlukan untuk meramalkan Y. Jadi R^2 merupakan ukuran yang memberikan kecocokkan (kesesuaian) dari model regresi.

Contoh III.3.

Untuk data dari tabel III.1.

Hitung koefisien determina si ganda.

Penyelesaian :

$$\text{Total SS} = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}$$

$$= 11400,15 - (13)(367,5)^2$$

$$= 438,13$$

Dan telah dihitung,

$$SSE = 40,01$$

maka,

$$SSR = 438,13 - 40,01$$

$SSR = 398,12$

Sehingga koefisien determinasi ganda,

$$R^2 = \frac{398,12}{438,13}$$

$$= 0,9087$$

Ini berarti bahwa pengaruh X_1 dan X_2 bersama-sama terhadap Y adalah kira-kira 90,87 %.

III.4. Pengujian Kegunaan Model Regresi

Pemisahan total SS ke dalam SSE dan SSR dinamakan analisis varians. Nama ini digunakan karena jika $X_1, X_2, \ldots, X_k$ tidak menyumbangkan informasi apapun untuk peramalan Y, masing-masing besaran SSR dan SSE memberikan suatu penduga yang bebas (independen) untuk σ^2, yaitu varians dari Y untuk nilai $X_1, X_2, \ldots, X_k$ tertentu. Penduga ini disebut kuadrat rerata (mean squares).

Jadi,

$$MSR = \text{kuadrat rerata untuk regresi} = \frac{SSR}{v_1}$$

$$MSE = s^2 = \text{Kuadrat rerata untuk kesalahan} = \frac{SSE}{v_2}$$

dimana,

v_1 = satu lebih kecil dari banyaknya parameter β dalam model = k

v_2 = n - (banyaknya parameter β dalam model)

= n - (k + 1)

Jadi untuk setiap model regresi berganda yang mengandung k variabel peramal,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

menggunakan MSR dan MSE untuk menguji hipotesis nol bahwa $X_1, X_2, \ldots, X_k$ tidak menyumbang informasi untuk peramalan Y. Ini serupa dengan menghipotesiskan bahwa,

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

Jika data itu memberikan informasi yang cukup untuk menolak hipotesis ini, yang berarti bahwa sekurang-kurangnya satu dari variabel peramal $X_1, X_2, \ldots, X_k$ menyumbang informasi untuk peramalan Y.

Statistik uji untuk pengujiannya adalah,

$$F = \frac{MSR}{MSE} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3:9)$$

$$= \frac{MSR}{s^2}$$

Statistik ini memiliki distribusi F dengan kebebasan (db) v_1 dan v_2.

Contoh III.4.

Ujilah data dari tabel III.1.

Penyelesaian :

Telah dihitung SSE = 40,01 dengan db = 10

SSR = 398,13 dengan db= 2

Sehingga $MSR = \frac{398,13}{2}$

$= 199,065$

$MSE = \frac{40,01}{10}$

$= 4,001$

Jadi,

$$F = \frac{MSR}{MSE}$$

$$= \frac{199,065}{4,001}$$

$F = 49,754$

Sedangkan nilai kritis $F(2;10;0,01) = 7,56$,sehingga H_o ditolak, yaitu karena nilai F yang dihitung jauh lebih besar dari nilai kritis F.

Maka disimpulkan persamaan regresi signifikan.

III.5. Menggunakan Persamaan Prediksi untuk Pendugaan (Estimation) dan **Peramalan** (Prediction)

Dalam kasus yang manapun, nampaknya untuk sebagian besar usaha penelitian, pengujian hipotesis atau pendugaan parameter model secara individual nilainya jauh lebih sedikit daripada penggunaan persamaan prediksi secara lengkap. Persamaan prediksi akan ada nilainya jika memenuhi tiga hal :

1. Persamaan prediksi tersebut dapat digunakan untuk menduga nilai rerata (mean) Y untuk nilai-nilai tertentu dari variabel-variabel peramal.
2. Persamaan itu dapat digunakan untuk meramal nilai Y untuk beberapa waktu yang akan datang untuk nilai-nilai X_1, X_2, X_3, . . . ,X_k tertentu.
3. Jika suatu persamaan prediksi memberikan kesesuaian yang baik (good fit) terhadap serangkaian data (R^2 besar) dan banyaknya variabel peramal tidak terlalu besar ,maka persamaan itu bisa membantu dalam memahami proses yang diselidiki.

Penduga untuk nilai rerata Y untuk nilai-nilai X_1, X_2, X_3, . . . ,X_k tertentu atau peramal nilai Y tertentu (akan teramati dimasa yang akan datang) untuk nilai-nilai X_1, X_2, X_3, . . . ,X_k tertentu dapat diperoleh dengan mensubstitusikan nilai numerik (berbentuk angka/bilangan) X_1, X_2, . . . ,X_k kedalam persamaan prediksi.