BAB II

METODE SIMPLEKS UNTUK REGRESI MINMAD

Dalam beberapa masalah statistik, terdapat dua variabel atau lebih yang hubungannya tidak dapat dipisahkan. Analisa regresi adalah sebuah metode statistik yang digunakan untuk menyelidiki hubungan antara dua variabel atau lebih. Analisis regresi yang menyelidiki hubungan linier antara variabel respon Y dengan satu variabel independen X disebut analisis regesi linier sederhana. Sedangkan analisis regresi yang menyelidiki hubungan linier antara variabel respon Y dengan variabel independen X_1, X_2, ..., X_{p-1} disebut analisa regresi linier berganda.

Dalam regresi linier berganda, variabel respon Y merupakan variabel random pada saat variabel independen X_1, ..., X_{p-1} diukur dengan error yang dapat diabaikan dan biasanya dikontrol oleh pelaku percobaan. Hubungan linier antara variabel respon Y dengan variabel independen X_1, ..., X_{p-1} dapat dinyatakan sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} + d \qquad (2.1)$$

dimana d merupakan deviasi atau sesatan dan β_0, β_1, ..., β_{p-1} merupakan parameter atau koefisien regresi.

Misalkan terdapat n pengamatan dengan p-1 variabel independen ; $n \geq p$, dan x_{ij} menyatakan nilai pengamatan ke-i, dimana i = 1, 2, ..., n dari variabel independen ke-j, dimana j = 1, 2, ..., p-1, maka untuk pengamatan ke-i model (2.1) dapat juga ditulis sebagai :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{p-1} + d_i \qquad (2.2)$$

Model regresi linier berganda juga dapat dituliskan dalam bentuk matrik sebagai berikut :

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{d} \qquad (2.3)$$

dimana :

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np-1} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

Untuk mendapatkan harga parameter atau koefisien regresi dari model regresi linier berganda maka sesatannya perlu diminimalkan, sehingga akan diperoleh persamaan estimasinya sebagai berikut :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_{p-1} X_{p-1} \qquad (2.4)$$

Dalam analisa statistik klasik metode yang sering digunakan dalam mencari koefisien regresi adalah metode kuadrat terkecil, dimana yang diminimalkan adalah jumlah kuadrat sesatannya (JKS). Metode ini memerlukan asumsi kenormalan.

$$d = y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1})$$

$$JKS = \sum_{i=1}^{n} d_i^2$$

$$\text{JKS} = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{ip-1})^2 \qquad (2.5)$$

Untuk mendapatkan koefisien regresinya harus dipenuhi :

$$\frac{\partial \text{JKS}}{\partial \beta_0}\bigg|_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_{p-1}} = -2\sum_{i=1}^{n} (y_i - \hat\beta_0 - \sum_{j=1}^{p-1} \hat\beta_j x_{ij}) = 0$$

$$\frac{\partial \text{JKS}}{\partial \beta_j}\bigg|_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_{p-1}} = -2\sum_{i=1}^{n} (y_i - \hat\beta_0 - \sum_{j=1}^{p-1} \hat\beta_j x_{ij})\ \ x_{ij} = 0 \qquad (2.6)$$

Penyederhanaan persamaan (2.6) kita peroleh persamaan-persamaan normal kuadrat terkecil sebagai berikut :

$$\begin{array}{lllll}
n\hat\beta_0 & + \hat\beta_1 \sum_{i=1}^{n} x_{ij} & + \hat\beta_2 \sum_{i=1}^{n} x_{i2} & + \ldots + \hat\beta_{p-1} \sum_{i=1}^{n} x_{ip-1} & = \sum_{i=1}^{n} y_i \\
\hat\beta_0 \sum_{i=1}^{n} x_{i1} & + \hat\beta_1 \sum_{i=1}^{n} x_{i1}^2 & + \hat\beta_2 \sum_{i=1}^{n} x_{i1} x_{i2} & + \ldots + \hat\beta_{p-1} \sum_{i=1}^{n} x_{i1} x_{ip-1} & = \sum_{i=1}^{n} x_{i1} y_i \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\hat\beta_0 \sum_{i=1}^{n} x_{ip-1} & + \hat\beta_1 \sum_{i=1}^{n} x_{ip-1} x_{i1} & + \hat\beta_2 \sum_{i=1}^{n} x_{ip-1} x_{i2} & + \ldots + \hat\beta_{p-1} \sum_{i=1}^{n} x_{ip-1}^2 & = \sum_{i=1}^{n} x_{ip-1} y_i
\end{array}$$

Terdapat p persamaan normal. Untuk menyelesaikan p persamaan di atas digunakan pendekatan matriks, sehingga diperoleh :

$$X'X\hat\beta = X'y$$

$$\hat\beta = (X'X)^{-1}X'y \qquad (2.7)$$

dimana :

$$
X'X = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip-1} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ip-1} & \sum_{i=1}^{n} x_{ip-1}x_{i1} & \sum_{i=1}^{n} x_{ip-1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip-1}^2 \end{bmatrix}, \quad X'y = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ip-1}y_i \end{bmatrix}
$$

Terdapat metode lain untuk mendapatkan persamaan estimasi yaitu dengan menggunakan pendekatan pemrograman matematik. Metode ini tidak memerlukan asumsi kenormalan. Untuk mendapatkan koefisien regresinya digunakan metode simpleks, dimana yang menjadi fungsi tujuannya adalah harga jumlah sesatan absolutnya. Persamaan regresi yang terbentuk disebut Regresi MINMAD.

2.1. Regresi MINMAD (*Minimizing Mean Absolute Deviation*)

Regresi MINMAD merupakan regresi yang koefisien regresinya diperoleh dengan cara meminimalkan rata-rata deviasi absolut (MAD).

$$
\text{MAD} = \frac{1}{n}\sum_{i=1}^{n} |d_i|
$$

Atau untuk lebih memudahkan perhitungan, meminimalkan MAD sama saja dengan meminimalkan jumlah sesatan absolut (SAD = Sum of Absolut Deviation).

$$SAD = \sum_{i=1}^{n} |d_i|$$

Koefisien regesinya diperoleh dengan menggunakan metode simpleks. Metode simpleks merupakan salah satu metode yang digunakan untuk memecahkan persoalan pemrograman linier (PL) yang mempunyai variabel dan pembatas yang besar.

2.2. Metode Simpleks Untuk Regresi MINMAD

Dalam menyelesaikan persoalan pemrograman linier diperlukan definisi-definisi sebagai berikut :

Definisi 2.2.1 :

Variabel keputusan adalah variabel yang menguraikan secara lengkap keputusan-keputusan yang akan dibuat.

Definisi 2.2.2 :

Fungsi tujuan adalah fungsi dari variabel keputusan yang akan dimaksimalkan atau diminimalkan.

Definisi 2.2.3 :

Pembatas adalah kendala yang dihadapi sehingga kita tidak dapat menentukan harga-harga variabel keputusan secara sebarang.

Definisi 2.2.4 :

Pembatas tanda adalah pembatas yang menjelaskan apakah variabel keputusannya diasumsikan hanya berharga nonnegatif atau boleh berharga positif dan boleh juga berharga negatif (tidak terbatas dalam tanda).

Definisi 2.2.5 :

Daerah fisibel adalah himpunan dari semua titik yang memenuhi seluruh pembatas, termasuk pembatas tanda.

Untuk meminimalkan jumlah sesatan absolut digunakan pendekatan pemrograman linier, oleh sebab itu masalahnya dapat diformulasikan sebagai berikut :

Minimalkan : $Z = \sum |d_i|$

Dengan kendala : $X\beta + \mathbf{d} = Y$

$\mathbf{d}, \beta$ tidak terbatas dalam tanda (2.8) dimana $d_i = d_{1i} - d_{2i}$ dengan d_{1i} merupakan sesatan atas dan d_{2i} yang merupakan sesatan bawah, dan $d_{1i} \geq 0$, $d_{2i} \geq 0$. Untuk setiap i yang diberikan, d_{1i} dan d_{2i} tidak boleh positif kedua-duanya, maksudnya jika $d_{1i} > 0$ maka $d_{2i} = 0$ atau jika $d_{2i} > 0$ maka $d_{1i} = 0$, sehingga berakibat, jika $d_{2i} > 0$ maka $d_i = - d_{2i} < 0$ dan jika $d_{1i} > 0$ maka $d_i = d_{1i} > 0$. Oleh sebab itu $|d_i| = d_{1i} + d_{2i}$ sehingga (2.8) dapat juga ditulis sebagai berikut :

Minimalkan : $Z = \Sigma d_{1i} + \Sigma d_{2i} = \Sigma\ (d_{1i} + d_{2i})$

Dengan kendala : $X\beta + \mathbf{d_1} - \mathbf{d_2} = \mathbf{Y}$

$\mathbf{d_1}, \mathbf{d_2} \geq 0$

β tidak terbatas dalam tanda (2.9)

Untuk menyelesaikan persoalan (2.9) diperlukan definisi-definisi sebagai berikut :

Misalkan :

A = (X, I, -I) merupakan matrik berukuran nxp+2n.

$$\mathbf{A} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p-1} & 1 & \dots & 0 & -1 & \dots & 0 \\ 1 & x_{21} & x_{22} & \dots & x_{2p-1} & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np-1} & 0 & \dots & 1 & 0 & \dots & -1 \end{bmatrix}$$

W = (β, **d**$_1$, **d**$_2$) merupakan vektor berukuran p+2nx1

$$\mathbf{W} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \\ d_{11} \\ d_{12} \\ \vdots \\ d_{1n} \\ d_{21} \\ d_{22} \\ \vdots \\ d_{2n} \end{bmatrix}$$

Definisi 2.2.6 :

Sebarang **W** yang memenuhi **AW** = **Y** disebut sebagai solusi, dimana **I** merupakan matrik identitas berukuran nxn.

Definisi 2.2.7 :

Misalkan $\mathbf{C}'$ merupakan vektor $(\mathbf{0},\mathbf{e}',\mathbf{e}')$ dimana **0** merupakan vektor 1xp dan $\mathbf{e}' = (1,\ldots,1)$ merupakan vektor 1xn, maka $\mathbf{C}'\mathbf{W}$ disebut sebagai fungsi tujuan.

Definisi 2.2.8 :

Sebarang solusi W jika memenuhi :

$$W_j \geq 0 \qquad , j = p+1, \ldots, p+2n$$

maka solusi tersebut disebut sebagai solusi fisibel.

Definisi 2.2.9 :

Himpunan n kolom bebas linier dari matrik **A** disebut basis dari **A**.

Misal **B** = $\lfloor a_{i_1}, \ldots, a_{i_n} \rfloor$ merupakan basis maka terdapat n+p kolom yang tidak terdapat dalam **B** yang disebut kolom nonbasis.

Definisi 2.2.10 :

Solusi optimal adalah suatu titik pada daerah fisibel dengan nilai fungsi tujuan yang paling menguntungkan. Untuk masalah maksimasi, solusi optimalnya adalah suatu titik pada daerah fisibel dengan nilai fungsi tujuan terbesar dan pada masalah minimasi, solusi optimalnya adalah suatu titik pada daerah fisibel dengan nilai fungsi tujuan terkecil.

Metode simpleks merupakan prosedur aljabar yang bersifat iteratif yang bergerak selangkah demi selangkah, dimulai dari suatu titik ekstrim pada daerah fisibel menuju ke titik ekstrim yang optimum. Metode simpleks dalam regresi MINMAD digunakan untuk mencari koefisien regresi atau parameter regresi dan yang menjadi fungsi tujuannya adalah jumlahan deviasi absolut pada harga pengamatan ke-i, i =1,2, ..., n.

Misalkan **B** merupakan basis yang sesuai dengan solusi basis fisibel **W**, $\mathbf{W_B}$ menunjukkan vektor variabel W_j yang sesuai dengan kolom basis **B** dan koefisien yang sesuai dengan C_j dinyatakan sebagai $\mathbf{C_B}$. Bentuk tabel dari metode simpleks yang digunakan untuk menyelesaikan masalah regresi MINMAD dapat dilihat pada tabel 2.2.1.

Tabel 2.2.1. Tabel awal dari metode simplek untuk regresi MINMAD

C_B	Vektor basis	W_B	α_1	α_2	...	α_j	...	α_{p+2n}
1	a_{p+1} atau a_{p+n+1}	$\lvert Y_1 \rvert$	α_{11}	α_{12}	...	α_{1j}	...	$\alpha_{1.p+2n}$
⋮	⋮	⋮	⋮	⋮		⋮		⋮
1	a_{p+n} atau a_{p+2n}	$\lvert Y_n \rvert$	α_{n1}	α_{n2}	...	α_{nj}	...	$\alpha_{n.p+2n}$
	$C_k - Z_k$	$Z = \Sigma \lvert Y_i \rvert$				$C_j - Z_j$		

2.2.1. Algoritma Metode Simpleks Untuk Regresi MINMAD

Langkah-langkah yang dilakukan dalam metode simpleks adalah sebagai berikut :

1. Menentukan basis awal dan solusi basis fisibel :
   - Untuk α_j, j = 1, ..., p, bukan merupakan basis untuk tabel awal
   - Untuk α_j, j = p+1, ..., p+2n , merupakan calon basis untuk tabel awal

   Untuk menentukan kolom yang menjadi basis dan solusi basis fisibel dalam tabel awal dilakukan cara sebagai berikut :

   - Jika $1 \leq j \leq p$ maka $W_j = 0$
   - Jika $p+1 \leq j \leq p+n$ maka :

$$W_j = \begin{cases} Y_r & , Y_r > 0 \\ 0 & , \text{yang lain} \end{cases}$$

Dalam hal ini yang menjadi kolom basis adalah α_j, $p+1 \le j \le p+n$

- Jika $p+n+1 \le j \le p+2n$ maka :

$$W_j = \begin{cases} -Y_r & , Y_r < 0 \\ 0 & , \text{yang lain} \end{cases}$$

Dalam hal ini yang menjadi kolom basis adalah α_j, $p+n+1 \le j \le p+2n$

2. Menentukan C_B awal :

   Karena $C' = (0, e', e')$ maka :

   - untuk $1 \le j \le p$, $C_j = 0$
   - untuk $p+1 \le j \le p+2n$, $C_j = 1$

   Jadi untuk C_B awal, koefisiennya berharga 1.

3. Menentukan baris $C_k - Z_k$

   Karena C_B awal, koefisiennya berharga 1 maka $Z = \Sigma\ |Y_i|$ dan untuk mencari $C_j - Z_j$ digunakan rumus :

$$C_j - Z_j = C_j - \sum_{i=1}^{n} C_{B_i} \alpha_{ij}$$

   Untuk kolom yang menjadi basis, $C_j - Z_j = 0$

4. Menentukan kolom kunci atau vektor yang akan masuk basis.

   a. Misal $C_{j1} - Z_{j1} = \max\limits_{\substack{k \\ C_k - Z_k > 0}} C_k - Z_k$

b. Misal $|C_{j2} - Z_{j2}| = \max\limits_{\substack{k \\ C_k - Z_k < 0}} C_k - Z_k$, pilih j sehingga :

$$|C_j - Z_j| = \max\,[C_{j1} - Z_{j1}\ ,\ |C_{j2} - Z_{j2}|]$$

Jika j_1 dan j_2 tidak ditemukan, maka lanjutkan ke langkah 8. Jika ditemukan, lanjutkan ke langkah 5.

5. Memilih elemen pivot dengan cara :

   - Jika $C_j - Z_j > 0$, pilih r sebagai berikut :

$$\frac{W_{B_r}}{\alpha_{rj}} = \max_{i \notin R_B}\left[\frac{W_{B_i}}{\alpha_{ij}}\right],\ \alpha_{ij} < 0$$

   dimana R_B menunjukkan himpunan indeks variabel terbatas pada basis.

   - Jika $C_j - Z_j < 0$, pilih r sebagai berikut :

$$\frac{W_{B_r}}{\alpha_{rj}} = \min_{i \notin R_B}\left[\frac{W_{B_i}}{\alpha_{ij}}\right]\alpha_{ij} > 0$$

6. Lakukan operasi baris elementer dan akan diperoleh tabel baru , dimana :

$$\hat{W}_{B_r} = \frac{W_{B_r}}{\alpha_{rj}} \qquad\qquad \hat{\alpha}_{rl} = \frac{\alpha_{rl}}{\alpha_{rj}}$$

dimana α_{rj} merupakan elemen pivot.

$$\hat{W}_{B_i} = W_{B_i} - \frac{W_{B_r}}{\alpha_{rj}}\alpha_{ij} \qquad , i \neq r$$

$$\hat{\alpha}_{il} = \alpha_{il} - \frac{\alpha_{ij}\alpha_{rl}}{\alpha_{rj}} \qquad , l = 1, \ldots, p+2$$

$$\hat{Z} = Z + \left(C_j - Z_j\right)\frac{W_{B\,r}}{\alpha_{rj}}$$

$$\hat{C}_l - \hat{Z}_l = (C_l - Z_l) - \left(C_j - Z_j\right)\frac{\alpha_{rl}}{\alpha_{rj}}$$

Kolom $\mathbf{C_B}$ berubah yaitu jika kolom yang masuk dalam vektor basis adalah kolom $\alpha_1, \ldots, \alpha_p$ maka koefisien $\mathbf{C_B}$ berharga 0 selain itu $\mathbf{C_B}$ berharga 1 .

7. Jika masih ada $C_k - Z_k$ yang negatif maka kembali ke langkah 4. Jika tidak maka lanjutkan ke langkah 8.

8. Proses berhenti, solusi telah optimal dimana $C_j - Z_j \geq 0$ untuk semua variabel non basis dan $C_j - Z_j = 0$ untuk semua variabel tak terbatas W_j .

Hasil akhir dari algoritma simpleks untuk regresi MINMAD dapat dilihat pada tabel 2.2.2 .Dari tabel akhir dapat dilihat bahwa untuk vektor basis a_1, a_2, ..., a_p dengan $\mathbf{C_B}$ berharga 0 diperoleh nilai parameter atau koefisien regresi $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{p-1}$ dan diperoleh harga fungsi tujuan Z =. $\sum_{r=1}^{n}(d_{1r} + d_{2r})$

Tabel. 2.2.2. Tabel akhir dari metode simpleks untuk regresi MINMAD

C_B	Vektor basis	W_B	α_1	…	α_p	α_{p+i}	…	α_{p+2n}
0	a_1	β_0	1	…	0	$\hat{\alpha}_{1.p+i}$	…	$\hat{\alpha}_{1.p+2n}$
⋮	⋮	⋮	⋮		⋮	⋮		⋮
0	a_p	β_{p-1}	0	…	1	$\hat{\alpha}_{p.p+i}$	…	$\hat{\alpha}_{p.p+2n}$
1	a_{p+r} atau a_{p+n+r}	d_{1r} atau d_{2r}	0	…	0	$\hat{\alpha}_{r.p+i}$	…	$\hat{\alpha}_{r.p+2n}$
		$Z = \Sigma (d_{1r} + d_{2r})$	0	…	0		$C_j - Z_j \geq 0$	