# BAB II

# REGRESI LINIER BERDASARKAN METODE MINMAD

Regresi merupakan suatu metode statistik yang berkenaan dengan studi ketergantungan variabel terikat pada variabel bebas. Analisis regresi berguna dalam menelaah hubungan dua variabel atau lebih dan hubungan yang berbentuk linier menghasilkan model regresi linier.

Masalah regresi linier biasanya diselesaikan dengan meminimalkan jumlah kuadrat deviasi antara nilai pengamatan dan nilai prediksi dari pengamatan. Metode ini dikenal sebagai *metode kuadrat terkecil* atau norm-$L_2$ minimal. Metode lain yang digunakan untuk mencari koefisien regresi linier adalah metode MINMAD atau norm-$L_1$ minimal yaitu metode dengan meminimalkan rata-rata deviasi mutlak antara nilai pengamatan dan nilai prediksi dari pengamatan.

Pada bagian ini akan dibahas mengenai model regresi linier berganda, estimator norm-$L_q$ minimal dan metode MINMAD. Tujuannya adalah menjelaskan model regresi MINMAD dan sifat estimasi parameter regresi MINMAD.

## 2.1. Model Regresi Linier Berganda

Beberapa permasalahan regresi dapat mencakup lebih dari satu variabel bebas. Model regresi linier yang menggunakan lebih dari satu variabel bebas disebut **model regresi linier berganda**.

Model regresi linier berganda dengan p-1 variabel bebas dinyatakan dengan :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} + \varepsilon$$

Parameter $\beta_j$ , j = 0, 1, ... , p-1 disebut koefisien regresi yang menyatakan rata-rata perubahan Y per unit akibat perubahan $X_j$ bilamana sisa variabel -variabel bebas $X_i$ ($i \neq j$) konstan.

Model dapat dinyatakan dalam bentuk matriks sebagai berikut :

$$\mathbf{Y = X\beta + \varepsilon}$$

dimana :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \text{dan} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Y adalah sebuah vektor (nx1) dari observasi-observasi, X adalah sebuah matriks (nxp) dari variabel-variabel bebas, β adalah sebuah vektor (px1) dari koefisien-koefisien regresi dan ε adalah sebuah vektor (nx1) dari error random.

## 2.2. Estimator Norm-$L_q$ Minimal

Berdasarkan masalah estimasi parameter dalam model regresi linier

$$Y = X\beta + \varepsilon$$

dimana :

Y adalah vektor pengamatan berukuran $n \times 1$

X adalah matriks pengamatan berukuran $n \times p$

β adalah vektor parameter berukuran $p \times 1$

ε adalah error berukuran $n \times 1$, dan $n > p$.

Estimasi parameter harus teliti untuk semua nilai pengamatan. Ketelitian persamaan regresi diukur dengan menggunakan sebuah norm dari vektor error $(Y - X\hat{\beta})$, dimana $\hat{\beta}$ adalah nilai estimasi dari β.

Norm-$L_q$ didefinisikan untuk vektor d dengan elemen-elemen $d_i$ sebagai berikut:

$$\|d\|_q = \left(\sum_i |d_i|^q\right)^{1/q}$$

Dibawah kriteria $L_q$ minimal, estimator $\hat{\beta}$ didapat dengan menyelesaikan masalah meminimalkan

$$\min \left\|Y - X\hat{\beta}\right\| = \left(\sum_i \left|Y_i - X_i\hat{\beta}\right|^q\right)^{1/q}$$

dimana $X_i$ adalah baris ke-i dari X.

Untuk q≥1, berguna sebagai ukuran ketepatan dari persamaan regresi untuk masalah estimasi. Untuk q=2 merupakan kriteria kuadrat terkecil dan untuk q=1, kriteria adalah nilai mutlak terkecil yang akan dikenal dengan MINMAD yaitu

$$\min \left\|Y - X\hat{\beta}\right\| = \sum_i \left|Y_i - X_i\hat{\beta}\right|.$$

## 2.3. Metode MINMAD

Metode MINMAD adalah metode yang digunakan untuk menyelesaikan masalah regresi dengan meminimalkan deviasi mutlak antara nilai pengamatan dan nilai prediksi dari pengamatan. Masalah MINMAD dikenal sebagai estimator norm-$L_1$ minimal, hal ini disebabkan oleh norm-$L_q$ minimal untuk q=1. Kelebihan dari estimator ini adalah dapat mengurangi terjadinya pencilan data dan terjadinya distribusi error ekor panjang.

### 2.3.1. Model Regresi MINMAD

Regresi MINMAD didefinisikan sebagai regresi linier yang didapat dengan meminimalkan rata-rata deviasi mutlak antara pengamatan $Y_i$ dan nilai prediksi $\hat{Y}_i$ untuk pengamatan ke-i. Dalam mengestimasi parameter $\beta$, meminimalkan rata-rata deviasi mutlak sama seperti meminimalkan $\sum \left| d_i \right|$ dimana $d_i$ adalah deviasi dari nilai pengamatan $Y_i$ dan nilai prediksi $\hat{Y}_i$ yaitu $\sum \left| Y_i - \hat{Y}_i \right|$.

Berdasarkan model regresi linier berganda :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} + \varepsilon$$

dari data X dan Y yang diberikan pada pasangan $(X_i , Y_i)$, $i = 1, 2, \ldots ,n$

Parameter $\beta_0 , \beta_1 , \beta_2 , \ldots , \beta_{p-1}$ diperoleh dengan meminimalkan :

$$\frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_{p-1} X_{p-1} \right| \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{(2.4)}$$

dikenal dengan rata-rata deviasi mutlak dari nilai pengamatan $Y_i$ dan nilai prediksi $\hat{Y}_i$. Meminimalkan (2.4) sama seperti meminimalkan

$$\sum_{i=1}^{n} \left| Y_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_{p-1} X_{p-1} \right|$$

Berdasarkan model regresi linier $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{d}$ , dimana **Y** adalah vektor dari variabel terikat dengan ukuran $n \times 1$ dan **X** adalah matriks berukuran $n \times p$, masalah regresi MINMAD dapat disusun seperti masalah program linier dengan n kendala dan p+2n variabel sebagai berikut :

Minimalkan : $\sum | d_i |$ .................. (2.5)

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

**d, β** tak dibatasi tanda

Untuk titik data ke-i, $d_{1i}$ dan $d_{2i}$ merupakan bagian positif dan negatif dari $d_i$, sesuai dengan deviasi atas dan deviasi bawah dari garis regresi. Karena $|d_i| = d_{1i} + d_{2i}$ dimana nilai $d_{1i}$ dan $d_{2i}$ tak negatif, $d_i = d_{1i} - d_{2i}$ kemudian $\mathbf{d_1}$ dan $\mathbf{d_2}$ masing-masing merupakan vektor untuk semua $d_{1i}$ dan $d_{2i}$ maka masalah regresi MINMAD dapat dinyatakan kembali sebagai berikut:

Minimalkan : $\sum d_{1i} + \sum d_{2i}$

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

**β** tak dibatasi tanda

$\mathbf{d_1, d_2 \geq 0}$ .................... (2.6)

Dalam bentuk matriks, (2.6) dapat disusun sebagai berikut:

Minimalkan : $\mathbf{e^1 d_1 + e^1 d_2}$

Dengan kendala : $\mathbf{X\beta + Id_1 - Id_2 = Y}$

$\mathbf{d_1, d_2 \geq 0}$ .................... (2.7) dimana e adalah vektor berukuran n×1 yang mempunyai elemen 1, $\beta^1 = (\beta_0, \ldots, \beta_{p-1})$, $d_1^1 = (d_{11}, \ldots, d_{1n})$, $d_2^1 = (d_{21}, \ldots, d_{2n})$, $Y^1 = (Y_1, \ldots, Yn)$ dan X adalah matriks n×p dengan n > p.

Diasumsikan bahwa rank kolom matriks X (rank X) = p. Dari asumsi rank dapat ditunjukkan bahwa sebuah basis optimal harus berisi n kolom dimana p kolom dari matriks X dan n-p kolom dari matriks I dan -I yang sesuai dengan 2n variabel $d_1$ dan $d_2$.

### 2.3.2. Sifat Estimasi Parameter Regresi MINMAD

Untuk menyatakan estimasi unbias dengan menggunakan regresi MINMAD, diasumsikan ε adalah simetri dan E(ε) = 0.

Didefinisikan bahwa estimator dari β yang dinyatakan dengan $\beta_0$ adalah antisimetris jika $\beta - \beta_0(\varepsilon) = -(\beta - \beta_0(-\varepsilon))$ untuk sebarang nilai ε.

Estimator kuadrat terkecil $\beta_0$ untuk parameter β adalah $\left(X^1X\right)^{-1} X^1Y$

Sehingga,

$$\beta_0(\varepsilon) = (X^1X)^{-1} X^1 (X\beta + \varepsilon)$$

$$= \beta + (X^1X)^{-1} X^1\varepsilon \ , \qquad \text{dan}$$

$$\beta_0(-\varepsilon) = (X^1X)^{-1} X^1 (X\beta - \varepsilon)$$

$$= \beta - (X^1X)^{-1} X^1\varepsilon$$

Karena $\beta - \beta_0(\varepsilon) = -(\beta - \beta_0(-\varepsilon))$ maka $\beta_0$ adalah antisimetris.

Jika $\varepsilon$ berdistribusi simetri disekitar 0 maka estimator antisimetris $\beta_0$ adalah unbias.

Masalah MINMAD yang disusun seperti persamaan (2.7) adalah sebagai berikut :

Masalah $P_1$ : Minimalkan : $e^l d_1 + e^l d_2$

Dengan kendala : $X\beta_1 + Id_1 - Id_2 = Y - X\beta_0$

$d_1, d_2 \geq 0$

Masalah $P_2$ : Minimalkan : $e^l d_1 + e^l d_2$

Dengan kendala : $-X\beta_1 - Id_1 + Id_2 = -Y + X\beta_0$

$d_1, d_2 \geq 0$

Prosedur selanjutnya adalah sebagai berikut :

1. Tentukan $\beta_0$
2. Seleksi salah satu masalah $P_1$ atau $P_2$ dengan probabilitas masing-masing ½
3. Nyatakan sebuah solusi optimal $\beta_1^*$ untuk masalah yang diseleksi
4. Estimasi $\beta$ dengan $\hat{\beta} = \beta_1^* + \beta_0$

Telah diketahui bahwa

$$Y(-\varepsilon) - X\beta_0(-\varepsilon) = X\beta - \varepsilon - X\beta_0(-\varepsilon)$$

$$= -\varepsilon + X[\beta - \beta_0(-\varepsilon)]$$

$$= -\varepsilon - X[\beta + \beta_0(-\varepsilon)] \quad \text{karena } \beta_0 \text{ adalah antisimetris}$$

$$= -Y(\varepsilon) + X\beta_0(\varepsilon)$$

masalah $P_1(\varepsilon)$ dan $P_2(-\varepsilon)$ adalah identik kecuali untuk tanda $\beta_1$ dan juga identik untuk $P_1(-\varepsilon)$ dan $P_2(\varepsilon)$.

Selanjutnya ,

$$\beta_1^*[P_1(\varepsilon)] = -\beta_1^*[P_2(-\varepsilon)] \text{ dan } \beta_1^*[P_1(-\varepsilon)] = -\beta_1^*[P_2(\varepsilon)]$$

Sehingga estimasi $\beta$ dengan $\hat{\beta} = \beta_1^* + \beta_0$ adalah

$$\begin{aligned} E(\hat{\beta}) &= E(\beta_0) + E(\beta_1^*) \\ &= \beta + E[(\beta_1^* \mid \varepsilon) + (\beta_1^* \mid -\varepsilon)] \\ &= \beta + E[(\beta_1^* \mid \varepsilon, P_1)\tfrac{1}{2} + (\beta_1^* \mid \varepsilon, P_2)\tfrac{1}{2} + (\beta_1^* \mid -\varepsilon, P_1)\tfrac{1}{2} \\ &\quad + (\beta_1^* \mid -\varepsilon, P_2)\tfrac{1}{2}] \\ &= \beta \end{aligned}$$

maka dari sini diperoleh bahwa $\hat{\beta}$ adalah unbias.

Untuk menyelesaikan masalah regresi MINMAD akan dijelaskan pada bab III, dimana akan digunakan pendekatan program linier yaitu metode simpleks. Penggunaan metode tersebut bertujuan untuk mencari koefisien regresi $\beta$ dengan cara meminimalkan deviasi mutlak sehingga akan didapat persamaan regresi MINMAD $\hat{Y} = X\hat{\beta}$ dengan deviasi mutlak minimal.