

BAB II

PERSAMAAN REGRESI LINEAR

2.1 Model Regresi Linear Sederhana

Model regresi linear biasanya ditulis dalam bentuk

$$Y_i = \beta_0 + \beta_1 X + \varepsilon_i \quad (2.1)$$

Dimana β_0 merupakan jarak titik asal O dengan titik perpotongan antara sumbu Y dengan garis fungsi linearnya. β_1 merupakan koefisien arah, koefisien regresi atau besarnya pengaruh X terhadap Y. Sedangkan ε merupakan komponen error random.

Misalkan diberikan pasangan data $\{(x_1, y_1), \dots, (x_n, y_n)\}$ yang didapatkan dari suatu eksperimen. Dari data yang tersedia, dapat ditemukan harga-harga $\hat{\beta}_0$ dan $\hat{\beta}_1$ yang merupakan harga estimasi β_0 dan β_1 dengan metode kuadrat terkecil. Sehingga persamaan (2.1) berubah menjadi

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X \quad (2.2)$$

Tetapi sebelum menggunakan metode kuadrat terkecil, dibutuhkan suatu asumsi-asumsi agar didapatkan hasil yang akurat. Adapun asumsi-asumsi yang dibutuhkan adalah sebagai berikut :

1. Hubungan antara X dan Y adalah linear.
2. Error memiliki mean nol.
3. Error memiliki varian konstan.
4. Error tidak saling berkorelasi.

Setelah asumsi-asumsi diatas dipenuhi, dapatlah digunakan metode kuadrat terkecil.

Dari persamaan (2.1) untuk $i = 1, 2, 3, \dots, n$ jumlah kuadrat semua galat adalah

$$\begin{aligned} S(\beta_0, \beta_1) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \end{aligned} \quad (2.3)$$

$\hat{\beta}_0$ dan $\hat{\beta}_1$ sebagai nilai dugaan yang memiliki nilai, jika nilai tersebut disubtitusikan kedalam β_0 dan β_1 pada persamaan (2.3), maka akan menghasilkan nilai S yang paling kecil. Untuk menentukan $\hat{\beta}_0$ dan $\hat{\beta}_1$, persamaan (2.3) dideferensialkan terhadap β_0 dan kemudian β_1 , selanjutnya menyamakan hasil deferensial itu dengan nol.

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \quad (2.4a)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \quad (2.4b)$$

Sehingga nilai dugaan $\hat{\beta}_0$ dan $\hat{\beta}_1$ dapat diperoleh dari

$$\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \quad (2.5a)$$

$$\sum_{i=1}^n X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \quad (2.5b)$$

atau

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i Y_i \end{aligned} \quad (2.6)$$

Kedua persamaan ini disebut persamaan normal.

Penyelesaian kedua persamaan ini adalah:

$$\begin{aligned} n\hat{\beta}_0 &= \sum_{i=1}^n Y_i - \hat{\beta}_1 \sum_{i=1}^n X_i \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}\bar{X} \end{aligned} \quad (2.7)$$

dan

$$\begin{aligned} \hat{\beta}_1 &= \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n Y_i \sum_{i=1}^n X_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{aligned}$$

atau

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \quad (2.8)$$

Sehingga model Pendekatan regresi linear sederhana menjadi

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad (2.9)$$

Contoh 2.1

Suatu amatan terhadap banyaknya uap air (dalam pound) yang digunakan setiap bulan (X) dan suhu atmosfer rata-rata dalam derajat farenheit (Y)

| No | Peubah(Y) | Peubah(X) | No | Peubah(Y) | Peubah(X) |
|----|-----------|-----------|----|-----------|-----------|
| 1 | 10,89 | 35,3 | 16 | 9,58 | 48,5 |
| 2 | 11,13 | 29,7 | 17 | 10,09 | 59,3 |
| 3 | 12,51 | 30,8 | 18 | 8,11 | 70,0 |
| 4 | 8,40 | 58,8 | 19 | 6,83 | 70,0 |
| 5 | 9,27 | 61,4 | 20 | 8,88 | 74,5 |
| 6 | 8,72 | 71,3 | 21 | 7,68 | 72,1 |
| 7 | 6,36 | 74,4 | 22 | 8,47 | 58,1 |
| 8 | 8,50 | 76,7 | 23 | 8,86 | 44,6 |
| 9 | 7,82 | 70,7 | 24 | 10,36 | 33,4 |
| 10 | 9,14 | 57,5 | 25 | 11,08 | 28,6 |
| 11 | 8,24 | 46,4 | | | |
| 12 | 12,19 | 28,9 | | | |
| 13 | 11,88 | 28,1 | | | |
| 14 | 9,57 | 39,1 | | | |
| 15 | 10,94 | 46,8 | | | |

Dari perhitungan diperoleh sebagai berikut:

$$n = 25 \quad \sum_{i=1}^n X_i Y_i = 11.821,432$$

$$\sum_{i=1}^n Y_i = 235,6 \quad \bar{Y} = 235,6 / 25 \\ = 9,424$$

$$\sum_{i=1}^n X_i^2 = 76.323,42 \quad \sum_{i=1}^n x_i = 1.315$$

$$\bar{X} = 52,6$$

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n X_i Y_i - [\left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n Y_i\right)] / n}{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2 / n} \\ &= \frac{11.821,432 - (1.315)(235,6) / 25}{76.323,42 - (1.315)^2 / 25} \\ &= -0,079829\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ &= 9,424 - (-0,079829)(52,6) \\ &= 13,23005\end{aligned}$$

Jadi persamaan regresinya adalah

$$\begin{aligned}\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X \\ &= 13,623005 - 0,079829 X\end{aligned}$$

2.2 Model regresi linear berganda

2.2.1 Metode kuadrat terkecil dengan matrik

Bentuk umum persamaan regresi linear berganda dengan k variabel bebas adalah

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.10)$$

Untuk mengestimasi koefisien-koefisien regresinya digunakan metode kuadrat terkecil dengan asumsi bahwa $E(\varepsilon) = 0$ dan $\text{Var } (\varepsilon) = \sigma^2$. Dimisalkan tersedia n observasi dengan $n > p$ sebagai berikut:

| Observasi i | Y_i | X_{i1} | X_{i2} | X_{i3} | | X_{ik} |
|-------------|-------|----------|----------|----------|-------|----------|
| 1 | Y_1 | X_{11} | X_{12} | X_{13} | | X_{1k} |
| 2 | Y_2 | X_{21} | X_{22} | X_{23} | | X_{2k} |
| 3 | Y_3 | X_{31} | X_{32} | X_{33} | | X_{3k} |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| n | Y_n | X_{n1} | X_{n2} | X_{n3} | | X_{nk} |

Sesuai dengan model regresi linear pada persamaan (2.10) maka dari tabel dapat dibentuk persamaan :

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon \quad (2.11) \\ &= \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon, \quad i=1, 2, \dots, n \end{aligned}$$

Untuk mendapatkan estimasi-estimasi parameternya digunakan metode kuadrat terkecil:

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_p) &= \sum_{i=1}^n \varepsilon^2 \quad (2.12) \\ &= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij})^2 \end{aligned}$$

Kemudian fungsi S diturunkan ke $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ sebagai berikut:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij}) \quad (2.13)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij}) X_{i1} \quad (2.14)$$

Dan seterusnya, dengan $i = 1, 2, \dots, n$; $j = 1, 2, \dots, k$.

Selanjutnya dari persamaan (2.13) dan (2.14) diperoleh

$$\sum_{i=1}^n Y_i - n\hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij} = 0$$

$$\sum_{i=1}^n Y_i = n\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j X_{ij}$$

$$\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij}) X_{ij} = 0$$

$$\sum_{i=1}^n X_{i1} Y_i = \hat{\beta}_0 \sum_{i=1}^n X_{i1} + \hat{\beta}_1 \sum_{i=1}^n X_{i1}^2 + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i1} X_{ik}$$

.

.

.

$$\sum_{i=1}^n X_{ik} Y_i = \hat{\beta}_0 \sum_{i=1}^n X_{ik} + \hat{\beta}_1 \sum_{i=1}^n X_{ik} X_{i1} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik}^2$$

(2.15)

Solusi dari persamaan (2.15) merupakan estimator kuadrat terkecil parameter - parameter β -nya. Dan penyelesaian persamaan tersebut akan menjadi lebih jelas dan cepat jika disajikan dalam bentuk matrik.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ 1 & X_{31} & X_{32} & \dots & X_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_n \end{bmatrix} \quad \text{dan} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Keterangan

Y adalah vektor observasi ($n \times 1$).

X adalah matriks ($n \times p$) merupakan matrik dari regresor-regresornya dengan $p = k + 1$.

β adalah vektor ($p \times 1$) merupakan koefisien-koefisien regresi yang harus dicari, dan

ε adalah vektor ($n \times 1$) dari random error.

Sehingga persamaan (2.15) bila disajikan dalam bentuk matrik menjadi

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

Untuk mencari estimator β akan dimisalkan fungsi residu

/

$$\begin{aligned}
 S(\beta) &= \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon \\
 &= (Y - X\beta)' (Y - X\beta) \\
 &= Y' Y - \beta' X' Y - Y' X\beta + \beta' X' X\beta \\
 &= Y' Y - \beta' X' Y - \beta' X' Y + \beta' X' X\beta \\
 &= Y' Y - 2\beta' X' Y + \beta' X' X\beta
 \end{aligned} \tag{2.16}$$

Dari turunan pertama dari persamaan (2.16) adalah

$$\begin{aligned} -2X'Y + 2X'X\beta &= 0 \\ X'X\beta &= X'Y \end{aligned} \quad (2.17)$$

Selanjutnya dari persamaan (2.17) dikalikan invers dari matriks $X'X$ sehingga diperoleh

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

Hasil estimasi β -nya adalah

$$\beta = (X'X)^{-1} X'Y \quad (2.18)$$

2.2.2 Sifat - Sifat Estimator Kuadrat Terkecil

Metode kuadrat terkecil memiliki sifat - sifat yang baik. Untuk menyelidikinya, pandang model umum persamaan regresi linear

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

Disini diasumsikan bahwa ε , bebas satu sama lain dan $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ untuk setiap $i = 1, 2, 3, \dots, n$. Dalam lambang matrik, ini berarti $\Sigma_{\varepsilon} = \sigma^2 I$ bila I menyatakan matrik satuan berukuran $n \times n$. Dengan demikian maka

$$E(Y) = X\beta$$

dan

$$\Sigma_Y = \sigma^2 I \quad (2.19)$$

Sifat estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

1. Tak bias

$$\begin{aligned} E(\hat{\beta}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= \beta \end{aligned}$$

2. Variansi minimum

$$\begin{aligned} \Sigma_{\hat{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma_{\mathbf{Y}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

bahwa $\Sigma_{\hat{\beta}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ merupakan variansi terkecil dari semua estimator linear tak bias dijamin oleh teorema gaus-markov berikut.

Teorema Gaus-markov

Estimator kuadrat terkecil $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ mempunyai variansi terkecil dalam himpunan semua estimator linear tak bias.

Bukti:

Misalkan $\hat{\beta}^*$ estimator linear lain dari β yang juga tak bias. Karena $\hat{\beta}^*$ estimator linear maka dapat dimisalkan bentuknya sebagai $\hat{\beta}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{U}] \mathbf{Y}$ untuk suatu matrik \mathbf{U} yang merupakan fungsi dari \mathbf{X} .

$$\begin{aligned} \text{Jadi } E(\hat{\beta}^*) &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{U}] E(\mathbf{Y}) \\ &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{U}] (\mathbf{X}\beta) \end{aligned}$$

$$\begin{aligned}
 &= \hat{\beta} + UX\beta \\
 &= (I + UX)\hat{\beta}
 \end{aligned}$$

Agar $\hat{\beta}^*$ estimasi tak bias dari β maka haruslah $UX = 0$,
jadi

$$\begin{aligned}
 \hat{\Sigma}_{\beta^*} &= [(X'X)^{-1} X' + U] \Sigma_Y [U' + X (X'X)^{-1}] \\
 &= \sigma^2 [(X'X) X' U' + UU' + (X'X) + UX (X'X)^{-1}]
 \end{aligned}$$

karena $UX = X'U' = 0$ maka

$$\hat{\Sigma}_{\beta^*} = \sigma^2 [(X'X)^{-1} + UU']$$

Matrik UU' adalah definit tak negatif, karena semua unsur diagonalnya berbentuk kuadrat. Jadi terbukti bahwa variansi, dari setiap unsur dari vektor $\hat{\beta}^*$ selalu lebih besar, atau paling kecil sama, dengan variansi unsur $\hat{\beta}$ yang sesuai.