## BAB II

## ESTIMASI PARAMETER LINEAR

Suatu model regresi adalah suatu cara yang dapat digunakan untuk :

a. Menyatakan kecenderungan berubah-ubahya variabel tak bebas Y terhadap variabel bebas X dalam cara yang sistematis.

b. Menyatakan berpencarnya pengamatan sekitar kurva yang menyatakan suatu hubungan statistik.

### 2.1. Model Regresi Linear Sederhana

Model regresi dengan satu variabel bebas dan linear, dapat dinyatakan sebagai :

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i \qquad (2.1)$$

dengan : $Y_i$ = harga variabel tak bebas pada trial ke-i

$X_i$ = harga variabel bebas pada trial ke-i

$\beta_0, \beta_1$ = parameter

$\varepsilon_i$ = suku galat dengan $E(\varepsilon_i)=0$; $i=1,2,\ldots,n$

Model diatas disebut model regresi linear sederhana. Arti linear disini ialah linear dalam parameter dan linear dalam variabel tak bebas. Nilai estimasi dari koefisien regresi $\beta_0$ dan $\beta_1$ dalam model (2.1) diperoleh dari data sampel. Data sampel didapat dengan cara melakukan eksperimen atau dengan cara survai. Dengan salah satu cara tersebut, diperoleh pasangan data :

$$\{( X_i, Y_i ) ; i = 1, 2, \ldots, n \}$$

untuk setiap pasangan amatan $(X_i,Y_i)$, pandang galat $\varepsilon_i$;

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_1) \qquad (2.2)$$

dan jumlah kuadrat galatnya adalah :

$$S(\beta_0,\beta_1) = \sum_{u=1}^{n} \varepsilon_i^{2}$$

$$= \sum_{u=1}^{n} (Y_i - \beta_0 - \beta_1 X_1)^2 \qquad (2.3)$$

Untuk mendapatkan estimasi fungsi regresi yang baik, parameter-parameter $\beta_0$ dan $\beta_1$ kita estimasi dari data sampel menggunakan metode kuadrat terkecil. Estimasi kuadrat terkecil untuk $\beta_0$ dan $\beta_1$ dicari sebagai hasil penyelesaian persamaan normal berikut :

$$\frac{\partial S(\beta_0,\beta_1)}{\partial \beta_0} = 0$$

$$- 2 \sum_{u=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial S(\beta_0,\beta_1)}{\partial \beta_1} = 0$$

$$- 2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

Nilai estimasi kuadrat terkecil untuk $\beta_0$ dan $\beta_1$ dilambangkan dengan $\hat{\beta}_0 = b_0$ dan $\hat{\beta}_1 = b_1$, maka persamaan-persamaan normalnya adalah :

$$n\ b_0 + b_1 \sum_{u=1}^{n} X_i = \sum_{u=1}^{n} Y_i \qquad (2.4)$$

$$b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^{\ 2} = \sum_{i=1}^{n} X_i \ Y_i \qquad (2.5)$$

dan jawabnya adalah :

$$\hat{\beta}_1 = b_1 = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2} \qquad (2.6)$$

$$= \frac{n \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X} \ ; \text{dimana} : \bar{X} = \frac{1}{n} \sum X_i$$

$$\bar{Y} = \frac{1}{n} \sum Y_i \qquad (2.7)$$

Jadi estimasi kuadrat terkecil garis regresinya adalah

$$E(Y) = E(\beta_0 + \beta_1 + \varepsilon)$$

atau :

$$\hat{Y} = b_0 + b_1 X$$

Sifat-sifat estimasi kuadrat terkecil : $b_0$ dan $b_1$

(i) $b_0$ dan $b_1$ fungsi linear dari $Y_i$

(ii) $b_0$ dan $b_1$ estimasi tak bias untuk $\beta_0$ dan $\beta_1$

Bukti :

(i) $$b_1 = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{\Sigma (X_i - \bar{X}) Y_i - \Sigma (X_i - \bar{X}) \bar{Y}}{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{\Sigma (X_i - \bar{X}) Y_i}{\Sigma (X_i - \bar{X})^2}$$

$$= \Sigma k_i Y_i \quad \text{dimana} \quad k_i = \frac{\Sigma (X_i - \bar{X})}{\Sigma (X_i - \bar{X})^2} \qquad (2.8)$$

$$b_o = \frac{1}{n} \Sigma Y_i - \bar{X} \Sigma k_i Y_i$$

$$= \Sigma b_i Y_i \; ; \; \text{dimana } b_i = \frac{1}{n} - \bar{X} k_i \qquad (2.9)$$

dari (2.8) dan (2.9) jelas bahwa $b_o$ dan $b_1$ fungsi linear dari $Y_i$.

(ii) $$E(b_1) = E \; \frac{\Sigma (X_i - \bar{X}) Y_i}{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{1}{\Sigma (X_i - \bar{X})^2} \times \Sigma (X_i - \bar{X}) E(Y)$$

$$= \frac{1}{\Sigma (X_i - \bar{X})^2} \times \Sigma (X_i - \bar{X})(\beta_o + \beta_1 X_i)$$

$$= \frac{1}{\Sigma (X_i - \bar{X}) X_i} \times \; \beta_o \Sigma (X_i - \bar{X}) + \beta_1 \Sigma X_i (X_i - \bar{X})$$

$$= \frac{1}{\Sigma (X_i - \bar{X}) X_i} \quad 0 + \beta_1 \Sigma X_i (X_i - \bar{X})$$

$$= \beta_1$$

$$
\begin{aligned}
E(b_0) &= E(\bar{Y} - b_1\bar{X}) \\
&= E(\bar{Y}) - E(b_1\bar{X}) \\
&= \frac{1}{n} \Sigma Y_i - \bar{X}\, E(b_1) \\
&= \frac{1}{n} \Sigma (\beta_0 + \beta_1 X_i) - (\frac{1}{n} \Sigma X_i)\, \beta_1 \\
&= \frac{1}{n}\, n\, \beta_0 = \beta_0
\end{aligned}
$$

## 2.2. Analisis Regresi Linear Ganda

Dalam penelitian yang menggunakan analisis regresi, seringkali digunakan lebih dari satu variabel bebas dalam model yang lebih kompleks. Jika modelnya linear, maka kita mempunyai model regresi linear ganda yang dapat dianggap sebagai perluasan dari model regresi linear sederhana. Jika kita mempunyai p variabel bebas, maka $X = (X_1, X_2, \ldots, X_p)$ maka himpunan data amatannya adalah :

$$\{(X_{i1}, X_{i2}, \ldots, X_{ip}, Y_i) \;;\; i = 1,2,\ldots,n\}$$

Sebagai perluasan model regresi linear sederhana, kita dapat menyatakan hubungan fungsional antara X dan Y sebagai $Y = f(X) = f(X_1, X_2, \ldots, X_p)$. Contoh hubungan antara variabel tak bebas Y dengan variabel bebas $X_1, X_2, \ldots, X_p$ :

1. Misalkan p = 2.

   Hubungan antara berat badan anak (Y) dengan umur anak ($X_1$) dan tinggi badan anak ($X_2$).

   Misalkan p = 3.

2. Hubungan antara kadar lemak dalam darah (Y)

dengan umur ($X_1$) berat badan ($X_2$) dan tekanan darah ($X_3$) seorang pasien.

Model regresi linear ganda seringkali dapa digunakan sebagai model pendekatan untuk struktur yang kompleks dan merupakan representasi yang cukup baik dengan rentang (range) tertentu variabel-variabel bebas yang digunakan.

Model regresi linear ganda :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i \qquad (2.10)$$
$$i = 1, 2, \ldots, n$$

dimana : $Y_i$ = harga variabel tak bebas pada trial ke-i

$X_{ip}$ = harga variabel bebas ke-p pada trial ke-i

$\beta_i$ = parameter

$\varepsilon_i$ = suku galat dengan pemisalan $\varepsilon_i$ saling bebas, berdistribusi $N(0,\sigma^2)$.

Pandang fungsi regresi linear ganda :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon_i \qquad (2.11)$$

Seperti pada fungsi regresi linear sederhana, untuk mendapatkan estimasi fungsi linear yang baik, kita akan menaksir parameter $\beta_0, \beta_1, \ldots, \beta_p$ berdasarkan data sampel dengan menggunakan metode kuadrat terkecil.

Maka :

$$S(\alpha,\beta) = \sum_{u=1}^{n} \varepsilon_i^2$$

$$= \sum_{u=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_p X_{ip})^2 \quad (2.12)$$

Estimasi kuadrat terkecil untuk $\beta_0$, $\beta_1$, ..., $\beta_p$ dilambangkan $b_0$, $b_1$, ..., $b_p$ dicari dari hasil penyelesaian persamaan normal :

$$\frac{\partial S}{\partial \beta_0} = 0$$

$$\frac{\partial S}{\partial \beta_1} = 0 \qquad \text{(p+1) buah persamaan normal}$$

$$\vdots$$

$$\frac{\partial S}{\partial \beta_p} = 0$$

didapat :

$$n\, b_0 + b_1 \sum X_{i1} + \dots + b_p \sum X_{ip} = \sum Y_i$$

$$b_0 \sum X_{i1} + b_1 \sum X_{i1}^2 + \dots + b_p \sum X_{i1} X_{ip} = \sum X_{i1} Y_i$$

$$\vdots$$

$$b_0 \sum X_{ip} + b_1 \sum X_{ip} X_{i1} + \dots + b_p \sum X_{ip}^2 = \sum X_{ip} Y_i$$

....(2.13)

Jika parameter $\beta$ diestimasi, maka estimasi kuadrat terkecil persamaan regresi linear ganda (2.12) adalah :

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p \qquad (2.14)$$

## 2.3. Pendekatan Matriks terhadap Regresi Linear

Model (2.10) dapat ditulis dalam bentuk matriks

$$Y = X\beta + \varepsilon \qquad (2.15)$$

di mana :

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

$$\beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{vmatrix} \qquad \varepsilon = \begin{vmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{vmatrix}$$

$Y$ adalah vektor amatan yang berukuran (n x 1)

$X$ adalah matriks variabel bebas berukuran (n x p)

$\beta$ adalah vektor parameter yang berukuran (p x 1)

$\varepsilon$ adalah vektor galat yang berukuran (n x 1)

dan $E(\varepsilon) = 0$ , $V(\varepsilon) = I\sigma^2$, jadi unsur-unsur $\varepsilon$ tidak berkorelasi. Karena $E(\varepsilon) = 0$, model (2.15) dapat ditulis :

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

Dengan demikian jumlah kuadrat galatnya adalah :

$$\begin{aligned} \varepsilon^T \varepsilon &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - \beta^T X^T Y - Y^T X\beta - \beta^T X^T X\beta \\ &= Y^T Y - 2\beta^T X^T Y - \beta^T X^T X\beta \qquad (2.16) \end{aligned}$$

Nilai estimasi kuadrat terkecil $\beta$ kita lambangkan b yang bila disubstitusikan ke dalam persamaan (2.16) akan meminimumkan $\varepsilon^T\varepsilon$. Nilai estimasi itu dapat diperoleh dengan pendiferensialan persamaan (2.16) terhadap $\beta$. Ini akan menghasilkan persamaan normal :

$$X^T X b = X^T Y \qquad (2.17)$$

Dua kemungkinan muncul disini; Persamaan (2.17) terdiri atas p persamaan yang bebas dalam p unsur yang tidak diketahui, atau sebagian tidak bebas terhadap sebagian lainnya sehingga persamaan yang bebas kurang dari p. Bila persamaan normal tidak bebas terhadap sebagian lainnya, maka matriks $X^TX$ bersifat singular sehingga $(X^T X)^{-1}$ tidak ada. Bila p persamaan normal itu bebas $X^TX$ bersifat tidak singular, sehingga inversnya ada. Dalam hal demikian, estimasi kuadrat terkecil untuk parameter-parameter $\beta_0$, $\beta_1$, ..., $\beta_p$ adalah :

$$\hat{\beta} = b = \begin{vmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{vmatrix} = (X^T X)^{-1} X^T Y$$

Perhatikan bahwa nilai estimasi $\hat{Y}$ diperoleh dengan menghitung :

$$\hat{Y} = X b$$