BAB II

ESTIMASI PARAMETER REGRESI LINIER DENGAN METODE MAKSIMUM LIKELIHOOD

Penentuan atau pendugaan nilai parameter model adalah salah satu tahapan dalam proses pemodelan sebelum dilakukan pengujian terhadap model. Bab ini, berisi materi penunjang berupa estimasi parameter regresi linier berganda, dan pengujian untuk nyata model regresi linier berganda yang merupakan konsep dasar untuk pembahasan dalam Bab III.

2.1 Estimasi Parameter Regresi Linier Berganda

Regresi linier berganda merupakan salah satu cara yang digunakan dalam analisis data statistik. Untuk menggambarkan hubungan antara beberapa variabel bebas $X_1, X_2, \ldots, X_p$ dengan variabel terikatnya (Y) . Model tersebut dapat ditulis dalam bentuk :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon \quad \ldots..(2.1.1)$$

Model regresi di atas disebut model regresi linier berganda dengan p variabel bebas dan koefisien regresi β_j, j = 0,1,2,…,p. Model ini menggunakan asumsi bahwa variabel random ε_i saling bebas, berdistribusi normal dengan rata – rata 0 dan varian σ^2.

Untuk mengestimasi parameter dalam regresi linier berganda dapat digunakan metode kuadrat terkecil. Metode ini meminimalisasi jumlah dari kuadrat linier. Penyelesaian model ini lebih sederhana jika menggunakan bentuk matriks. Persamaan (2.1.1) dapat ditulis dalam bentuk matriks sebagai :

$$Y = X\beta + \varepsilon \qquad \ldots\ldots(2.1.2)$$

dimana :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} \ldots X_{1p} \\ 1 & X_{21} & X_2 \ldots X_{2p} \\ . & & \\ . & & \\ 1 & X_{n1} & X_{n2} \ldots X_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

y adalah vektor pengamatan pada variabel tak bebas berukuran Nx1, X adalah matriks variabel berukuran Nx(p+1), β adalah vektor koefisien regresi berukuran (p+1)x1, ε adalah vektor randam error berukuran Nx1.

Untuk mendapatkan vektor penaksir kuadrat terkecil $\hat{\beta}$ dilakukan dengan cara meminimumkan jumlah kuadrat galatnya, yaitu :

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon$$

$$= (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - \beta' X'Y - Y'X\beta + \beta' X' X\beta$$

$$= Y'Y - 2\beta' X'Y + \beta' X' X\beta$$

Karena $\beta' X'Y$ adalah sebuah vektor matriks skalar (1x1), dan $(\beta' X'Y)' = Y'X\beta$ adalah skalar yang sama, maka penaksir-penaksir kuadarat terkecil itu memenuhi :

$$\left.\frac{\partial L}{\partial \beta}\right|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

sehingga diperoleh

$$\hat{\beta} = (X'X)^{-1}X'Y \quad ....(2.1.3)$$

Selain metode kuadrat terkecil, metode lain untuk memperoleh estimator adalah metode maksimum likelihood. Misalkan X variabel random dengan distribusi probabilita $f(X,\theta)$, dimana parameter θ tidak diketahui. Misalkan $X_1, X_2, ..., X_n$ menjadi nilai yang diobservasi di dalam suatu sampel random yang besarnya n. Maka fungsi likelihood sampel tersebut adalah

$$L(\theta) = f(X_1,\theta).f(X_2,\theta)...f(X_n,\theta)$$

Estimator maksimum likelihood θ adalah nilai θ yang memaksimumkan fungsi likelihood $L(\theta)$ atau nilai θ yang memaksimumkan probabilita kejadian hasil sampel. Jika ε adalah variabel random yang diasumsikan berdistribusi normal dengan rata-rata nol dan varian σ^2 maka fungsi densitas gabungan untuk ε adalah

$$f(\varepsilon) = (2\pi)^{-n}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right)$$

dan fungsi regresi linier berganda adalah

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

Dengan menggunakan bentuk matriks fungsi likelihood dari regresi linier berganda dapat ditulis

$$L(X\beta,\sigma^2) = (2\pi)^{-n/2}\sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right]$$

log dari fungsi likelihoodnya adalah

$$\log L = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

log L adalah fungsi monoton dari L (A. Koutsosylannis ,1983). Maka nilai parameter yang memaksimumkan log L akan memaksimumkan L. Persamaan log L diturunkan terhadap β dan σ secara parsial kemudian hasilnya disamadengankan dengan nol. Sehingga didapatkan taksiran parameter $\hat{\beta}$ dan $\hat{\sigma}^2$

$$\frac{\partial \log L}{\partial \hat{\sigma}} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(Y - X\beta)'(Y - X\beta) = 0$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3}(Y - X\beta)'(Y - X\beta)$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$$

$$\frac{\partial \log L}{\partial \hat{\beta}} = -\frac{1}{2\sigma^2}(Y'Y - 2\beta' X'Y + \beta' X'X\beta)$$

$$= -\frac{1}{2\sigma^2}(-2X'Y + 2\beta X'X)$$

$$= \frac{X'Y}{\sigma^2} - \frac{\beta X'X}{\sigma^2} = 0$$

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

Prinsip dasar metode maksimum likelihood yang digunakan untuk mengestimasi parameter dalam regresi linier berganda sama dengan metode kuadrat terkecil.

2.2 Pengujian untuk Nyata Regresi dalam Regresi Linier Berganda

Untuk menentukan uji kecocokan pada regresi linier dengan variabel tak bebas Y dan variabel bebas $X_1, X_2, ..., X_p$ digunakan uji nyata regresi sebagai berikut.

Hipotesisnya :

$H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$

$H_1 : \beta_j \neq 0$ untuk paling sedikit satu j

Penolakan H_0 dengan suatu α yang diambil, menyatakan bahwa paling sedikit satu variabel bebas $X_1, X_2, ..., X_p$ memberikan sumbangan yang nyata pada model tersebut.

Statistik ujinya adalah

$$F_0 = \frac{SS_R / p}{SS_E / (n - p - 1)} = \frac{MS_R}{MS_E}$$

dengan $SS_E = y'y - \hat{\beta}'x'y$

$$SS_R = \hat{\beta}'x'y - n\bar{y}^2$$

n adalah besarnya sampel dan k adalah banyaknya parameter

dengan aturan keputusan tolak H_0 jika $F_0 > F_{\alpha,p,n-p-1}$

Kemudian untuk mengetahui apakah secara individu variabel bebas $X_1, X_2, \ldots, X_p$ berpengaruh terhadap variabel tak bebas Y digunakan uji koefisien regresi secara individual.

Hipotesisnya

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

Jika H_0 tidak ditolak dengan suatu α yang diambil, maka ini menunjukkan bahwa X_j dapat dihilangkan dari model tersebut. Sehingga statistik penguji adalah

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}}$$

dimana c_{jj} adalah elemen diagonal $(X'X)^{-1}$ yang berhubungan dengan $\hat{\beta}_j$. Hipotesis nol ditolak jika $|t_0| > t_{\alpha/2,n-p-1}$.