BAB II
METODE KUADRAT TERKECIL

2.1 TAKSIRAN KUADRAT TERKECIL

Model persamaan regresi linier secara umum dapat diberikan sebagai berikut :

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\ \dots\ +\beta_pX_{pi}+\varepsilon_i \qquad i = 1,2,\dots,n \quad (2.1.1)$$

dimana :

Y_i adalah variabel tak bebas (respons).

$X_2,X_3,\dots,X_p$ adalah variabel bebas (yang menjelaskan / regresor)

$\beta_1,\beta_2,\dots,\beta_p$ adalah parameter-parameter yang akan ditaksir

ε_i adalah kesalahan (error).

Persamaan (2.1.1) dapat pula dinyatakan dalam bentuk matriks berikut :

$$Y = X\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad (2.1.2)$$

dengan

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{21} & X_{31} \dots & X_{p1} \\ 1 & X_{22} & X_{32} \dots & X_{p2} \\ 1 & X_{23} & X_{33} \dots & X_{p3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} \dots & X_{pn} \end{bmatrix} \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{bmatrix} \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Matriks Var - Cov $(\underset{\sim}{\varepsilon})$ dinyatakan dalam bentuk sebagai berikut :

$$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \cdots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \cdots & E(\varepsilon_n^2) \end{bmatrix}$$

$$= \sigma^2 I \qquad (2.1.3)$$

Parameter - paramater yang belum diketahui $\beta = \{\beta_1, \beta_2, \beta_3, \ldots \beta_p\}$ dapat ditaksir dari persamaan (2.1.1) atau persamaan (2.1.2).

Namun untuk memudahkan pembahasan selanjutnya digunakan persamaan (2.1.2).

$$Y = X\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

Misalkan $\hat{\beta} = \{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots \hat{\beta}_p\}$ menunjukkan vektor kolom dari taksiran $\underset{\sim}{\beta}$, maka (2.1.2) dapat ditulis

$$Y = X\,\underset{\sim}{\hat{\beta}} + e \qquad (2.1.4)$$

dimana e adalah vektor kolom dari n residual $(Y - X\underset{\sim}{\hat{\beta}})$.

Dari (2.1.4) jumlah kuadrat residual adalah :

$$\sum_{i=1}^{n} e_i^2 = e'e$$

$$= (Y - X\underset{\sim}{\hat{\beta}})'\ (Y - X\underset{\sim}{\hat{\beta}})$$

$$= Y'Y - \underset{\sim}{\hat{\beta}}'X'Y - Y'X\underset{\sim}{\hat{\beta}} + \underset{\sim}{\hat{\beta}}'X'X\underset{\sim}{\hat{\beta}}$$

$$= Y'Y - 2\hat{\underset{\sim}{\beta}}'X'Y + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \qquad (2.1.5)$$

Taksiran kuadrat terkecil $\hat{\underset{\sim}{\beta}}$ diperoleh dengan meminimumkan jumlah kuadrat residual, dimana harus memenuhi :

$$\frac{\partial}{\partial \hat{\beta}}(e'e) = -2X'Y + 2X'X\hat{\underset{\sim}{\beta}} = 0$$

Sehingga didapat

$$X'X\,\hat{\underset{\sim}{\beta}} = X'Y \qquad (2.1.6)$$

Persamaan (2.1.6) dikenal sebagai **persamaan normal**. Persamaan normal tersebut diselesaikan dengan mengalikan kedua sisi persamaan (2.1.6) dengan invers dari **X'X**, sehingga diperoleh

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}\,X'Y \qquad (2.1.7)$$

Hal ini juga menunjukkan bahwa $\hat{\underset{\sim}{\beta}}$ adalah fungsi-fungsi linier dari Y.

2.2 SIFAT TAKSIRAN KUADRAT TERKECIL

Sehubungan dengan sifat taksiran kuadrat terkecil, Gauss - Markov menyatakannya dalam suatu teorema .

Teorema Gauss - Markov

Dengan melihat asumsi model regresi linier sederhana, penaksir kuadrat terkecil, dalam kelas penaksir linier tak bias, mempunyai varians minimum yaitu penaksir tadi **BLUE (Best Linear Unbiased Estimator).**

Akan ditunjukkan bahwa $\hat{\underset{\sim}{\beta}}$ adalah penaksir linier tak bias dari $\underset{\sim}{\beta}$. Sebelumnya akan diberikan definisi sebagai berikut

Definisi :

Suatu penaksir $\hat{\theta}$ dikatakan penaksir tak bias dari θ jika nilai yang diharapkan dari $\hat{\theta}$ sama dengan θ sebenarnya.

Atau $$E(\hat{\theta}) = \theta$$

Dari (2.1.7) diketahui

$$\begin{aligned}\hat{\underset{\sim}{\beta}} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) \\ &= (X'X)^{-1}X'X\underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon} \\ &= \underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon} \end{aligned} \qquad (2.2.1)$$

$$\begin{aligned} E(\hat{\underset{\sim}{\beta}}) &= E\,[\,\underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}\,] \\ &= E(\underset{\sim}{\beta}) + E\,[(X'X)^{-1}X'\underset{\sim}{\varepsilon}\,] \\ &= E(\underset{\sim}{\beta}) + (X'X)^{-1}X'\,E(\underset{\sim}{\varepsilon}) \end{aligned}$$

$$E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta} \qquad (2.2.2)$$

Terbukti $\hat{\underset{\sim}{\beta}}$ adalah penaksir linier tak bias dari $\underset{\sim}{\beta}$.

Sekarang akan dibuktikan bahwa penaksir ini mempunyai varians minimum. Sebelumnya diberikan definisi sebagai berikut,

Definisi :

$\hat{\theta}_1$ dikatakan sebagai penaksir dengan varians minimum dari θ, jika

$$\mathrm{Var}\ (\hat{\theta}_1) \le \mathrm{Var}\ (\hat{\theta}_2)$$

dimana $\hat{\theta}_2$ merupakan penaksir lain yang manapun dari θ

Berikut ini akan ditunjukkan varians dari taksiran kuadrat terkecil $\hat{\underset{\sim}{\beta}}$.

$$E\ [(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})']$$

$$= \begin{bmatrix} E(\hat{\beta}_1 - \beta_1)^2 & E(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2) & \cdots & E(\hat{\beta}_1-\beta_1)(\hat{\beta}_p-\beta_p) \\ E(\hat{\beta}_2-\beta_2)(\hat{\beta}_1-\beta_1) & E(\hat{\beta}_2 - \beta_2)^2 & \cdots & E(\hat{\beta}_2-\beta_2)(\hat{\beta}_p-\beta_p) \\ \vdots & \vdots & & \vdots \\ E(\hat{\beta}_p-\beta_p)(\hat{\beta}_1-\beta_1) & E(\hat{\beta}_p-\beta_p)(\hat{\beta}_2-\beta_2) & \cdots & E(\hat{\beta}_p - \beta_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{Var}\ (\hat{\beta}_1) & \mathrm{Cov}\ (\hat{\beta}_1,\hat{\beta}_2) & \cdots & \mathrm{Cov}\ (\hat{\beta}_1,\hat{\beta}_p) \\ \mathrm{Cov}\ (\hat{\beta}_2,\hat{\beta}_1) & \mathrm{Var}\ (\hat{\beta}_2) & \cdots & \mathrm{Cov}\ (\hat{\beta}_2,\hat{\beta}_p) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}\ (\hat{\beta}_p,\hat{\beta}_1) & \mathrm{Cov}\ (\hat{\beta}_p,\hat{\beta}_2) & \cdots & \mathrm{Var}\ (\hat{\beta}_p) \end{bmatrix} \tag{2.2.3}$$

Elemen-elemen diagonal pada matriks diatas adalah *Varians* dari $\hat{\underset{\sim}{\beta}}$ dan elemen-elemen di luar diagonal adalah *Kovarians* dari $\hat{\beta}$. Selanjutnya matriks (2.2.3) dinyatakan dengan $\mathrm{Var}(\hat{\underset{\sim}{\beta}})$. Dari (2.2.1) diperoleh

$$(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) = (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

Sehingga

$$\mathrm{Var}\ (\hat{\underset{\sim}{\beta}}) = E\ [(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})']$$

$$= E\ [(X'X)^{-1}X'\underset{\sim}{\varepsilon}\ \underset{\sim}{\varepsilon}'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X' \; E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') \; X(X'X)^{-1}$$

$$= (X'X)^{-1}X' \; \sigma^2 I \; X(X'X)^{-1}$$

$$\mathrm{Var}\,(\hat{\underset{\sim}{\beta}}) = \sigma^2 \; (X'X)^{-1} \qquad (2.2.4)$$

Di antara penaksir-penaksir linier tak bias dari parameter $\underset{\sim}{\beta}$ akan ditunjukkan bahwa taksiran kuadrat terkecil $\hat{\underset{\sim}{\beta}}$ mempunyai varians minimum.

Pandang suatu fungsi linier parametris $C'\underset{\sim}{\beta}$ dimana C adalah vektor kolom (p x 1) dari konsanta yang diketahui. Penaksir yang mungkin dari $C'\underset{\sim}{\beta}$ adalah $C'\hat{\underset{\sim}{\beta}}$, dimana

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

Penaksir tersebut tak bias karena

$$E(C'\hat{\underset{\sim}{\beta}}) = C'\underset{\sim}{\beta}$$

dan varians dari $C'\hat{\underset{\sim}{\beta}}$ adalah

$$\mathrm{Var}\,(C'\hat{\underset{\sim}{\beta}}) = E\,[(C'\hat{\underset{\sim}{\beta}} - C'\underset{\sim}{\beta})(C'\hat{\underset{\sim}{\beta}} - C'\underset{\sim}{\beta})']$$

$$= E\,[(C'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})'C\,]$$

$$\mathrm{Var}\,(C'\hat{\underset{\sim}{\beta}}) = \sigma^2 \; C'(X'X)^{-1}C \qquad (2.2.5)$$

Misalkan b = a'Y adalah penaksir liner tak bias lain dari $C'\underset{\sim}{\beta}$.

Maka $\quad E(b) = E(a'X\underset{\sim}{\beta} + a'\underset{\sim}{\varepsilon})$

$$= a'X\underset{\sim}{\beta}$$

$$E(b) = C'\underset{\sim}{\beta}$$

dimana $a'X = C'$ (2.2.6)

maka $b = C'\underset{\sim}{\beta} + a'\underset{\sim}{\varepsilon}$

jadi

$$\begin{aligned} \text{Var } (b) &= E\,[(b - C'\underset{\sim}{\beta})(b - C'\underset{\sim}{\beta})'] \\ &= E\,[a'\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'a] \\ &= \sigma^2 a'a \end{aligned}$$

Dengan menggunakan (2.2.6), var $(C'\hat{\underset{\sim}{\beta}})$ pada (2.2.5) menjadi

$$\text{Var } (C'\hat{\underset{\sim}{\beta}}) = \sigma^2 a'X(X'X)^{-1}X'a$$

sehingga

$$\begin{aligned} \text{Var } (b) - \text{Var } (C'\hat{\underset{\sim}{\beta}}) &= \sigma^2 a' \,[\, I - X(X'X)^{-1}X' \,]a \\ &= \sigma^2 \, a'M\,a \end{aligned} \qquad (2.2.7)$$

dimana

$$M = I - X(X'X)^{-1}X' \qquad (2.2.8)$$

adalah matriks simetris dan idempoten. Dan karena **M** definit positif maka

$$\text{Var } (b) - \text{Var } (C'\hat{\underset{\sim}{\beta}}) \geq 0 \qquad (2.2.9)$$

Artinya penaksir kuadrat terkecil dari $C'\underset{\sim}{\beta}$ mempunyai varians yang kecil atau sama dengan varians penaksir linier tak bias lainnya.

Sekarang pandang vektor

$$C_j' = [\,0 \;\ldots\; 0 \;1\; 0 \;\ldots\; 0\,]$$

dengan nilai 1 pada posisi ke-j dan nilai 0 untuk lainnya, maka (2.2.9) menjadi

$$\text{Var } (\hat{\underset{\sim}{\beta}}_j) \leq \text{Var } (b\;) \qquad (2.2.10)$$

Terbukti bahwa penaksir kuadrat terkecil $\hat{\beta}_j$ (j= 1,2,...,p) merupakan penaksir linier tak bias terbaik dari parameter β_j (j = 1,2,...,p), karena mempunyai varians minimum.

Dari persamaan (2.2.4) tampak bahwa varians penaksir bergantung pada varians kesalahan σ^2. Dan karena σ^2 tidak diketahui berarti σ^2 harus ditaksir. Taksiran σ^2 yaitu $\hat{\sigma}^2$ diperoleh dari jumlah kuadrat residual (residual sum of squares = SS_e).

$$SS_e = \sum_{i=1}^{n} e_i^2$$

$$= e'e$$

Dari (2.1.4) didapat

$$e = Y - X\hat{\underset{\sim}{\beta}}$$

dengan menggunakan (2.1.2) dan (2.1.7) diperoleh

$$e = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} - X\,[(X'X)^{-1}X'(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon})]$$

$$= \underset{\sim}{\varepsilon} - X(X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

$$= [I - X(X'X)^{-1}X']\,\underset{\sim}{\varepsilon}$$

dari (2.2.8) diperoleh

$$e = M\underset{\sim}{\varepsilon} \qquad (2.2.11)$$

yang menunjukkan bahwa residual merupakan fungsi linier dari kesalahan.

Nilai Ekspektasinya adalah

$$E(e'e) = E(\underset{\sim}{\varepsilon}'M'M\underset{\sim}{\varepsilon}) = E(\underset{\sim}{\varepsilon}'M\underset{\sim}{\varepsilon})$$

karena $\underset{\sim}{\varepsilon}'M\underset{\sim}{\varepsilon}$ skalar, maka :

$$
\begin{aligned}
E(e'e) &= E[tr(\underset{\sim}{\varepsilon}'M\underset{\sim}{\varepsilon})] = E[tr(M\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon})] \\
&= tr\ E(M\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}) = tr[M(E(\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon})] \\
&= tr\ [M(\sigma^2\ I)] = \sigma^2\ tr\ (M) \\
&= \sigma^2 tr\ [I - X(X'X)^{-1}X'] \\
&= \sigma^2\ [tr(I) - tr(X(X'X)^{-1}X')] \\
&= \sigma^2\ [tr(I) - tr(X'X\ (X'X)^{-1}] \\
&= \sigma^2\ [tr(I_n) - tr(I_p)] \\
&= \sigma^2(n-p) \qquad (2.2.12)
\end{aligned}
$$

Taksiran σ^2 yaitu $\hat{\sigma}^2$ adalah :

$$\hat{\sigma}^2 = \frac{SS_e}{n-p} \qquad (2.2.13)$$

Terlihat bahwa $\hat{\sigma}^2$ merupakan penaksir tak bias dari $\hat{\sigma}^2$, karena

$$
\begin{aligned}
E(\hat{\sigma}^2) &= E\left(\frac{e'e}{n-p}\right) \\
&= \frac{\sigma^2}{n-p}(n-p) \\
&= \sigma^2
\end{aligned}
$$

2.3 INTERVAL KEPERCAYAAN PARAMETER

Dari (2.2.1) diketahui :

$$\underset{\sim}{\hat{\beta}} = \underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

dimana diasumsikan bahwa $\varepsilon_i \infty\ N(0,\sigma^2)$. Maka $\underset{\sim}{\hat{\beta}}$ pun

berdistribusi normal, dengan mean $\underset{\sim}{\beta}$ dan varians $\sigma^2(X'X)^{-1}$. Dengan demikian distribusi dari $\hat{\beta}_j$ ($j = 1,2,\ldots,p$) adalah $N(\beta_j, \hat{\sigma}^2 C_{jj})$, dimana $\hat{\sigma}^2$ adalah taksiran dari σ^2 dan C_{jj} adalah elemen diagonal ke-j dari matriks $(X'X)^{-1}$.
Maka

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \, C_{jj}}} \qquad (\, j = 1,2,\ldots,p) \qquad (2.3.1)$$

berdistribusi t dengan derajat kebebasan (n-p). Sehingga diperoleh (1 - α) 100 % interval kepercayaan untuk parameter β_j, yaitu

$$\hat{\beta}_j - t_{\alpha/2,n-p} \sqrt{\hat{\sigma}^2 \, C_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2,n-p} \sqrt{\hat{\sigma}^2 \, C_{jj}} \qquad (2.3.2)$$

Nilai distribusi t didapat dari Tabel 4 pada lampiran, berdasarkan tingkat signifikan α yang digunakan dan derajat bebas n-p.

2.4 UJI HIPOTESA

Setelah diperoleh penaksir parameter-parameter model regresi linier (2.1.2) beserta variansnya, maka perlu dilakukan uji hipotesa terhadap parameter-parameter tersebut untuk mengetahui seberapa jauh keberhasilan model tersebut dalam menjelaskan variabel respons.
Untuk model ini, terdapat dua prosedur pengujian hipotesa.

2.4.1 Uji Signifikan dari Regresi

Uji ini menentukan ada atau tidaknya hubungan linier antara variabel respons dengan sekumpulan variabel penjelas.

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_p = 0$$

$$H_1 : \beta_j \neq 0 \text{ (sekurang-kurangnya satu j, dimana } j=2,\ldots,p). \qquad (2.4.1)$$

Statistik pengujinya adalah

$$F_{hitung} = \frac{SS_r / k}{SS_e / n-p} = \frac{MS_r}{MS_e} \qquad (2.4.2)$$

berdistribusi F dengan derajat bebas k dan n-p, (k=p-1). Disini jumlah kuadrat total S_{YY} dipartisi menjadi jumlah kuadrat regresi dan jumlah kuadrat residual, yaitu :

$$S_{YY} = SS_r + SS_e \qquad (2.4.3)$$

Dari (2.1.5) diketahui

$$e'e = SS_e = Y'Y - 2\hat{\underset{\sim}{\beta}}'X'Y + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}$$

$$= Y'Y - \hat{\underset{\sim}{\beta}}'X'Y$$

dan karena

$$S_{YY} = Y'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} \qquad (2.4.4)$$

Maka SS_e dapat ditulis sebagai

$$SS_e = Y'Y - \hat{\underset{\sim}{\beta}}'X'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} + \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

$$= Y'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} - \left[\hat{\underset{\sim}{\beta}}'X'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right] \qquad (2.4.5)$$

atau

$$SS_e = S_{YY} - SS_r$$

Berarti

$$SS_r = \hat{\underset{\sim}{\beta}}'X'Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n} \qquad (2.4.6)$$

Aturan keputusan :

Jika $F_{hitung} \geq F_{\alpha(k,n-p)}$, H_o ditolak pada tingkat signifikan α.

Jika $F_{hitung} < F_{\alpha(k,n-p)}$, H_o tidak ditolak pada tingkat signifikan α.

Nilai distribusi F diperoleh dari Tabel 5 pada lampiran, berdasarkan tingkat signifikan α yang digunakan dengan derajat kebebasan k dan n-p.

2.4.2 Uji Untuk Masing-masing Parameter.

Hipotesa untuk uji masing-masing parameter ini adalah :

$$H_o : \beta_j = \beta_o$$
$$H_1 : \beta_j \neq \beta_o \qquad (j=1,2,\ldots p) \qquad (2.4.7)$$

Statistik untuk hipotesa ini adalah

$$t_{hitung} = \frac{\hat{\beta}_j - \beta_o}{\sqrt{\hat{\sigma}^2 \, C_{jj}}} \qquad (2.1.8)$$

berdistribusi t dengan derajat bebas n-p.

Aturan keputusan :

Jika $t_{hitung} \geq t_{\alpha/2,n-p}$, H_o ditolak pada signifikan α.

Jika $t_{hitung} < t_{\alpha/2,n-p}$, H_o tidak ditolak pada tingkat signifikan α.

Nilai distribusi t didapat dari Tabel 4 pada lampiran, berdasarkan tingkat signifikan α yang digunakan dengan derajat bebas n-p.