BAB II

METODE KUADRAT TERKECIL BIASA

Metode kuadrat terkecil biasa (ordinary least square) atau OLS adalah salah satu metode yang sering digunakan dalam teknik analisa regresi yang bertujuan untuk meminimumkam kuadrat kesalahan e_i sehingga nilai regresinya akan mendekati nilai yang sesungguhnya.

Di dalam menaksir suatu persamaan dalam sistem analisa regresi metode OLS sering dipakai karena selain mudah dan bentuknya sederhana juga karena metode memenuhi sifat BLUE. Dalam kaitannya dengan pembahasan metode 2-SLS dalam model persamaan simultan metode OLS ini kita pakai dengan maksud sebagai perbandingan dan juga karena pada dasarnya metode 2-SLS adalah perluasan dari metode OLS di mana dalam langkah pembahasanya memakai cara dua langkah.

2.1 Metode OLS pada regresi linier sederhana.

Model regresi linier sederhana biasanya dipakai untuk menggambarkan hubungan antara dua variabel yang secara matematik dapat ditulis sebagai berikut:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad (2.1)$$

di mana:

Y_i = variabel dependen ke-i

X_i = variabel independen ke-i

ε_i = variabel gangguan ke-i

α = konstanta

β = koefisien regresi

Metode kuadrat terkecil biasa(OLS) mempunyai beberapa asumsi yang harus dipenuhi yaitu:

asumsi 1

Nilai rata-rata kesalahan pengganggu nol, yaitu:

$$E(\varepsilon_i) = 0$$

asumsi 2

Varian(ε_i) = E(ε_j) = σ^2, sama untuk semua kesalahan pengganggu.

asumsi 3

Tidak ada otokorelasi antara kesalaham pengganggu, berarti kovarian (ε_i,ε_j) = 0, i=j

asumsi 4

Variabel bebas X_2, X_3, ... , X_k, konstan dalam sampling yang terulang dan bebas terhadap kesalahan pengganggu ε_i.

asumsi 5

Tidak ada koliniearitas ganda diantara variabel bebas X (Multikolinieritas).

asumsi 6

Kesalahan pengganggu mengikuti distribusi normal dengan rata-rata nol dan varian σ^2.

2.1.1 Penaksiran koefisien regresi pada regresi linier sederhana.

Pandang bentuk regresi linier berikut:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Model regresi linier tersebut di atas adalah

dalam bentuk populasi sedangkan dalam bentuk sampel akan berbentuk sbb:

$$Y_i = a + bX_i + \varepsilon_i \qquad (2.2)$$

akan ditaksir menggunakan persamaan penaksir:

$$Y_i = a + bX_i + e_i \qquad (2.3)$$

dimana persamaan garis regresi linier adalah:

$$\hat{Y}_i = a + bX_i \qquad (2.4)$$

sehingga persamaan (2.3) di atas dapat ditulis kembali sebagai berikut:

$$Y_i = \hat{Y}_i + e_i \qquad (2.5)$$

simpangannya adalah:

$$e_i = Y_i - \hat{Y}_i \qquad (2.6)$$

$$= Y_i - a - bX_i$$

yang sering pula disebut sebagai kesalahan pengganggu. Maka menurut metode OLS nilai taksiran dari masing-masing parameter a dan b adalah sebagai berikut:

$$a = \frac{\Sigma Y_i - b\,\Sigma X_i}{n}$$

$$a = \bar{Y} - b\bar{X} \qquad (2.7)$$

$$b = \frac{\Sigma X_i Y_i - \dfrac{\Sigma X_i \Sigma Y_i}{n}}{\Sigma X_i^2 - \dfrac{(\Sigma X_i)^2}{n}}$$

atau

$$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \qquad (2.8)$$

di mana $(X_i - \bar{X}) = x_i$ dan $(Y_i - \bar{Y}) = y_i$

2.1.2 Sifat-sifat penaksir kuadrat terkecil biasa (OLS).

Agar merupakan penaksir yang baik maka penaksir kuadrat terkecil (OLS) harus memiliki sifat BLUE (best linier unbiased estimator) sbagai berrikut:

a. Linier.

$$b = \frac{\Sigma((X_i - \bar{X})\ (Y_i - \bar{Y}))}{\Sigma(X_i - \bar{X})^2} \qquad \text{(dari 2.8)}$$

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

atau

$$b = \frac{\Sigma Y_i (X_i - \bar{X})}{\Sigma(X_i - \bar{X})^2} \qquad [\text{karena } \bar{Y}\ \Sigma(X_i - \bar{X}) = 0]$$

Sehingga

$$b = \frac{\Sigma Y_i x_i}{\Sigma x_i^2}$$

misalkan bahwa:

$$k_i = \frac{x_i}{\Sigma x_i^2} \qquad (i = 1, 2, \ldots, n)$$

maka

$$b = \Sigma k_i Y_i \qquad (2.9)$$

Begitu juga dengan (2.8) yang memberikan :

$$a = \bar{Y} - b\bar{X}$$

$$= (1/n)\Sigma Y_i - \bar{X}\ \Sigma k_i Y_i$$

Sehingga:

$$a = \Sigma[(1/n) - \bar{X}k_i]Y_i \qquad (2.10)$$

Jadi terlihat bahwa baik a maupun b adalah fungsi linier dari Y.

b. Tidak Bias.

$$b = \Sigma k_i Y_i \qquad \text{(dari 2.9)}$$

$$b = \Sigma k_i(\alpha + \beta X_i + \varepsilon_i)$$

$$b = \alpha\Sigma k_i + \beta\Sigma k_i X_i + \Sigma k_i \varepsilon_i \qquad (2.11)$$

karena $k_i = \dfrac{x_i}{\Sigma x_i^2}$ dan $\Sigma k_i = \dfrac{\Sigma x_i}{\Sigma x_i^2}$

maka

$$\Sigma k_i x_i = \Sigma k_i(X_i + \bar{X}) = \frac{\Sigma x_i^2}{\Sigma x_i^2} = 1 \qquad (2.12)$$

subtitusikan nilai-nilai ini ke dalam persamaan (2.11), akan diperoleh:

$$b = \beta + \Sigma k_i \varepsilon_i \qquad (2.13)$$

$$E(b) = \beta + \Sigma k_i E[\varepsilon_i]$$

$$= \beta \qquad \text{karena } E[\varepsilon_i] = 0$$

persamaan (2.12) memberikan :

$$a = \Sigma[(1/n) - \bar{X}k_i]Y_i$$

$$= \Sigma[(1/n) - \bar{X}k_i]\ (\alpha + \beta X_i + \varepsilon_i)$$

$$= \alpha + \beta(1/n)\Sigma X_i + (1/n)\Sigma\varepsilon_i - \alpha\bar{X}\Sigma k_i - \beta\bar{X}\Sigma k_i X_i - \bar{X}\Sigma k_i \varepsilon_i$$

$$= \alpha + \beta X_i + (1/n)\Sigma\varepsilon_i - \beta\bar{X} - \bar{X}\Sigma k_i \varepsilon_i$$

$$= \alpha + (1/n)\Sigma\varepsilon_i - \bar{X}\Sigma k_i \varepsilon_i \qquad (2.14)$$

sehingga

$$E[a] = \alpha + (1/n)\Sigma\varepsilon_i - \bar{X}\Sigma k_i \varepsilon_i$$

$E[a] = \alpha$

Jadi terbukti bahwa a dan b adalah penaksir yang tidak bias terhadap α dan β.

C. Penaksir a dan b mempunyai varian yang minimum.

Sekarang harus dibuktikan bahwa a dan b memiliki sampel terkecil dibanding dengan penaksir-penaksir linier yang tidak bias lainnya. Untuk itu, pertama-tama akan dicari varian b dan kemudian dibuktikan bahwa variannya minimum:

$Var(b)=E[(b - \beta)^2]$ (dari pers. 2.13)

$=E[k_1^2U_1^2+k_2^2U_2^2+..+k_n^2U_n^2+2k_1k_2U_1U_2+..+2k_{n-1}k_nU_{n-1}U_n]$

$=E[k_1^2U_1^2+k_2^2U_2^2+..+k_n^2U_n^2]+E[2k_1k_2U_1U_2+..+2k_{n-1}k_nU_{n-1}U_n]$

$=E[\Sigma(k_1^2U_1^2) + 2E[\Sigma k_ik_jU_iU_j]$

$=\Sigma k_1^2E[U_1^2] + 2\Sigma k_ik_jE[U_iU_j] = \sigma^2\Sigma k_i^2$ (karena $E[U_iU_j]=0$

padahal,

$$\Sigma k_i = \frac{\Sigma x_i}{\Sigma x_i^2} \quad \text{dan} \quad \Sigma k_i^2 = \frac{\Sigma x_i^2}{(\Sigma x_i^2)^2} = \frac{1}{\Sigma x_i^2}$$

maka

$$var\ (b) = \sigma^2\Sigma k_i^2 = \frac{\sigma^2}{\Sigma x_i^2} \qquad (2.15)$$

sedangkan:

$Var(a) = E[a - \alpha]^2$

$= E[\Sigma(1/n) - \bar{X}\Sigma k_i)U_i]^2$ (dari 2.14)

$= \sigma^2\Sigma((1/n) - X\bar{\Sigma}k_i)^2$

$= \sigma^2\Sigma(1/n^2- (2/n)\bar{X}\Sigma k_i) + \bar{X}^2\Sigma k_i^2)$

$= \sigma^2(n/n^2- (2/n)\bar{X}\Sigma k_i) + \bar{X}^2\Sigma k_i^2)$

$$= \sigma^2(1/n + (\overline{X}^2/\Sigma x_i^2) \quad (\text{karena } \Sigma k_i = 0 \text{ dan } \Sigma k_i^2 = \frac{1}{\Sigma x_i^2}$$

maka $$\frac{1}{n} + \frac{\overline{X}^2}{\Sigma x_i^2} = \frac{\Sigma x_i^2 + n\,\overline{X}^2}{n\,\Sigma x_i^2} = \frac{\Sigma X_i^2}{n\,\Sigma x_i^2}$$

dan

$$\text{var}\ (a) = \sigma^2(1/n + (\overline{X}^2/X_i^2) = \sigma^2 \frac{\Sigma x_i^2}{n\,\Sigma x_i^2} \qquad (2.16)$$

Apabila S_e^2 adalah penduga σ^2(varians Y atau varians ε, agar S_e^2 dapat merupakan penduga tak bias dari σ^2, yaitu $E(S_e^2)=\sigma^2$) harus dipergunakan rumus sbb:

$$S_e^2 = \Sigma \frac{e_i^2}{n-2} \qquad (2.17)$$

di mana $S_e = \sqrt{S_e^2}$ = kesalahan baku dari e.

Σe_i^2 dapat di hitung langsung dari

$\Sigma(Y_i - \hat{Y})^2 = \Sigma\{Y_i - (a - bX_i)\}^2$, i = 1, 2 ,..., n.

$$\begin{aligned}
\Sigma e_i^2 &= \Sigma(y_i - bx_i)^2 \\
&= \Sigma(y_i^2 - 2bx_iy_i + b^2x_i^2) \\
&= \Sigma y_i^2 - 2b\Sigma x_iy_i + b^2\Sigma x_i^2 \\
&= \Sigma y_i^2 - 2b^2\Sigma x_i^2 + b^2\Sigma x_i^2 \\
&= \Sigma y_i^2 - b^2\Sigma x_i^2 \quad , \text{ ingat } b = \Sigma(x_iy_i) \ / \ \Sigma x_i^2
\end{aligned}$$

Jadi Persamaan (2.21) dapat ditulis sbb:

$$\Sigma Se^2 = \frac{\Sigma y_i^2 - b^2\Sigma x_i^2}{n-2} \quad \text{atau}$$

$$\Sigma Se^2 = \frac{\Sigma y_i^2 - b\ \Sigma x_iy_i}{n-2}$$

Untuk membuktikan bahwa b memiliki varian yang minimum, perlu dibandingkan varian b dengan varian beberapa penaksir β (kita sebut β^*) yang tidak bias.

Misalkan $\beta^* = \Sigma w_i Y_i$, di mana konstanta $w_i = k_i$, tetapi $w_i = k_i + c_i$ sehingga :

$$\beta^* = \Sigma w_i(\alpha + \beta X_i + \varepsilon_i)$$

$$= \alpha\Sigma w_i + \beta\Sigma w_i X_i + \Sigma w_i \varepsilon_i$$

dan $E[b^*] = \alpha\Sigma w_i + \beta\Sigma w_i X_i$ (karena $E[U_i] = 0$)

Karena β^*, diasumsikan penaksir yang tidak bias, berarti pada persamaan di atas $\Sigma w_i = 0$ dan $\Sigma w_i X_i = 1$.

Tetapi:

$$\Sigma w_i = \Sigma(k_i + c_i) = \Sigma k_i + \Sigma c_i$$

karena $\Sigma c_i = 0$, $\varkappa_i = \Sigma w_i = 0$

Maka:

$$\Sigma w_i X_i = \Sigma(k_i + c_i)X_i = \Sigma k_i X_i + \Sigma c_i X_i$$

karena $\Sigma w_i X_i = 1$, $\Sigma c_i X_i = 0$, dan $\Sigma k_i X_i = \Sigma k_i x_i = 1$

juga

$$\Sigma c_i X_i = \Sigma c_i X_i + X\Sigma c_i = 0$$

Jadi telah ditunjukkan, jika β^* merupakan penaksir yang tidak bias hasil-hasil berikut berlaku:

$$\Sigma w_i = 0,\ \Sigma w_i X_i = 1,\ \Sigma c_i = 0,\ \Sigma c_i X_i = \Sigma c_i x_i = 0 \qquad (2.18)$$

Varian dari β^* yang diasumsikan menjadi:

$$Var(\beta^*) = E[(\beta^* - \beta)^2]$$

$$= E[(\Sigma w_i U_i)^2] \qquad \text{(dari 2.14)}$$

$= \alpha^2 \Sigma w_i^2$ [dengan jalan pikiran yang sama persis ketika mendapatkan var(b)]

Jadi $Var(\beta^*) = \alpha^2 \Sigma w_i^2$

akan tetapi:

$$\Sigma w_i^2 = \Sigma(k_i + c_i)^2$$
$$= \Sigma k_i^2 + 2\Sigma k_i c_i + \Sigma c_i^2$$

sedangkan

$$= \Sigma c_i \frac{x_i}{\Sigma w_i^2} = \frac{\Sigma c_i x_i}{\Sigma x_i^2} = 0 \qquad \text{(dengan 2.18)}$$

Sehingga:

$$Var(\beta^*) = \sigma^2(\Sigma k_i^2 + \Sigma c_i^2) = \sigma^2 \Sigma k_i^2 + \sigma^2 \Sigma c_i^2$$
$$Var(\beta^*) = Var(\beta) + \sigma^2 \Sigma c_i^2$$

Oleh karena Σc_i^2 selalu positif, maka $Var(\beta^*) > Var(b)$. Hanya apabila $\Sigma c_i^2 = 0$ (tapi tidak mungkin), maka $Var(\beta^*)$ akan sama dengan Var(b). Hal ini menunjukkan b memiliki sifat varian minimum.

Dengan cara yang sama pula dapat dibuktikan bahwa konstanta intercept a dari metode kuadrat terkecil (OLS) memiliki varian minimum juga.

Misalkan, α^* adalah sebuah penaksir baru yang diasumsikan fungsi linier dari Y_i, dengan bobot $\omega_i = k_i + c_i$, seperti sebelumnya.

Kuadrat terkecil dari a adalah:

$$a = \Sigma[(1/n) - \bar{x}k_i]Y_i \qquad \text{(dari 2.10)}$$

Dengan analogi,

$$a = \Sigma[(1/n) - \bar{x}w_i]Y_i = f(Y)$$

Karena diinginkan agar α^* menjadi penaksir tidak bias bagi α, maka $E[\alpha^*] = \alpha$. Subtitusikan α^* ke dalam persamaan $Y_i = \alpha + \beta X_i = \varepsilon_i$, tetapi dalam notasi α^*, dan

temukan nilai harapan dari α^*.

$$\alpha^* = \alpha(1 - x\Sigma w_i) + \beta(x - x\Sigma w_i X_i) + \Sigma(\frac{1}{n} - Xw_i)\varepsilon_i$$

$$E[\alpha^*] = \alpha(1-x\Sigma E[w_i]) + \beta(x - x\Sigma E[w_i X_i]) + \Sigma E(\frac{1}{n} - Xw_i)\varepsilon_i$$

sekarang, $E[\alpha^*] = \alpha$, jika dan hanya jika

$$E[w_i] = 0,\ E[w_i X_i] = (1/n),\ \text{dan}\ E[w_i \varepsilon_i] = 0$$

yaitu jika $\Sigma w_i = 0$, $\Sigma w_i X_i = 1$, dan $\Sigma w_i \varepsilon_i = 0$. Syarat ini mengandung arti bahwa $\Sigma c_i = 0$ dan $\Sigma c_i X_i = 0$. Maka varian dari penaksir α^* ini adalah:

$$\begin{aligned}
Var(\alpha^*) &= E[\alpha^* - \alpha]^2 \\
&= \alpha^2 \Sigma[\frac{1}{n} - \bar{X}w_i]^2 \\
&= \alpha^2 \Sigma[\frac{1}{n} - \bar{X}^2 w_i^2 - 2\frac{1}{n} Xw_i] \\
&= \alpha^2 [\frac{n}{n^2} + \bar{X}^2 \Sigma w_i^2 - 2X\frac{1}{n} w_i] \\
&= \alpha^2 \Sigma[\frac{1}{n} + \bar{X}^2 \Sigma w_i^2 - \frac{2}{n} \bar{x}\Sigma w_i]
\end{aligned}$$

Oleh karena $\Sigma w_i = 0$ dan $\Sigma w_i^2 = \Sigma k_i^2 + \Sigma c_i^2$, maka:

$$Var(\alpha^*) = \alpha^2[\frac{1}{n} + x^2(\Sigma k_i^2 + \Sigma c_i^2)]$$

$$Var(\alpha^*) = \alpha^2[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma x_i^2}] + [\alpha^2 x^2 \Sigma c_i^2]$$

Komponen pertama disebelah kanan tanda sama dengan adalah variabel dari a, sehingga:

$$Var(\alpha^*) = Var(a) + \alpha^2\ \bar{x}^2 \Sigma c_i^2$$

Tetapi, $\Sigma c_i^2 > 0$. Karena tidak semua c_i adalah nol.

Oleh karena itu $Var(\alpha^*) > var(a)$.

Jadi karena penaksir kuadrat terkecil memenuhi sifat-sifat BLUE yaitu:linier, tidak bias dan memiliki varian minimum maka penaksir kuadrat terkecil merupakan penaksir yang baik.

2.1.3 Koefisien korelasi.

Adapun untuk mengetahui hubungan keeratan antara dua variabel digunakan analisis korelasi. Analisis korelasi yang mencakup hubungan dua variabel X da Y disebut analisis *korelasi linier sederhana*, sedangkan yang mencakup lebih dari dua variabel disebut analisis *korelasi linier berganda*.

Interpretasi *koefisien korelasi* (nilai untuk mengukur kuatnya hubungan antar variabel) tergantung pada asumsi yang kita buat terhadap variabel X dan Y. Apabila variabel X dan Y bervariasi atau disebut variabel acak maka koefisien korelasi akan mengukur variasi bersamaan antara X dan Y.

Koefisien korelasi sebenarnya dari populasi adalah sebagai berikut:

$$\rho = \frac{Cov(X,Y)}{\sigma_{xy}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad , \rho\text{=rho}$$

Besarnya nilai ρ adalah $-1 \leq \rho \leq 1$ yang berarti:

jika $\rho = -1$ maka hubungan X dan Y adalah sempurna dan negatif,

jika $\rho = 0$ maka hubungan X dan Y lemah sekali (dianggap tak ada hubungan),

jika $\rho = 1$ maka hubungan antara X dan Y adalah sempurna dan positif.

Akan tetapi didalam prakteknya kita tidak mengetahui nilai ρ akan tetapi dapat diestimasi berdasarkan data sampel. Kalau r adalah penduga ρ, maka r

dihitung berdasarkan rumus berikut:

$$r = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2}\ \sqrt{\Sigma y_i^2}} \qquad \text{atau}$$

$$r = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{\sqrt{n\Sigma X_i^2 - (\Sigma X_i)^2}\ \sqrt{n\Sigma Y_i^2 - (\Sigma Y_i)^2}}$$

Disamping ada pendugaan r sebagai koefisien korelasi linier sederhana maka ada r^2 yang disebut koefisien determinasi sederhana yaitu: nilai untuk mengukur besarnya kontribusi X terhadap variasi (naik turunnya) Y.

Beberapa sifat dari r adalah sebagai berikut:

1. r dapat positif atau negatif, tandanya tergantung pada tanda faktor pembilang yang mengukur kovariasi dua sampel.
2. Nilainya terletak antara $-1 \leq r \leq 1$.
3. Sifat dasarnya simetris, yaitu koefisien korelasi antara X dan Y sama dengan koefisien korelasi antara Y dan X.
4. r hanyalah suatu ukuran hubungan linier atau ketergantungan linier saja; r tidak mempunyai arti untuk menggambarkan hubungan nonlinier.
5. Meski r adalah ukuran hubungan linier antara dua variabel, tetapi tidak perlu berarti hubungan sebab akibat.

2.2 Penaksiran kuadrat terkecil pada regresi linier berganda.

Dalam regresi linier berganda terdapat hubungan antara satu variabel tidak bebas (dependent variabel) Y dengan beberapa variabel lain yang bebas (dependent variabel) X_1, X_2, ..., X_k.

Untuk meramalkan Y, apabila semua nilai variabel bebas diketahui, dipergunakan persamaan regresi linier berganda. Hubungan Y dan X_1, X_2, ..., X_k yang sebenarnya adalah sebagai berikut:

$$Y_i = B_o + B_1X_{1i} + \ldots + B_kX_{ki} + \varepsilon_i \quad \text{(populasi)} \qquad (2.2.1)$$

$$Y_i = b_o + b_{1i}X \quad + \ldots + b_kX_{ki} + e_i \quad \text{(sampel)} \qquad (2.2.2)$$

di mana i = 1, 2, ..., n

b_o, b_1, ..., b_k , dan e_i adalah pendugaan atas B_o, B_1, ..., B_k, ε_k.

Salah satu cara penyelesaian persamaan regresi linier berganda adalah dengan cara matrik. Apabila disajikan dalam bentuk matrik maka bentuk persamaan di atas akan berbentuk :

$$\underline{Y} = \underline{X}\ \underline{B} + \underline{\varepsilon} \quad , \underline{Y}, \underline{B}, \underline{\varepsilon} = \text{vektor}$$

$$\underline{X} = \text{matrik}$$

Di mana:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_i \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \qquad B = \begin{bmatrix} B_o \\ B_1 \\ \cdot \\ \cdot \\ B_i \\ \cdot \\ \cdot \\ B_k \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_i \\ \cdot \\ \cdot \\ \varepsilon_k \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots\cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots\cdots & X_{k2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1i} & X_{2i} & & X_{ki} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1n} & X_{2n} & \cdots\cdots & X_{kn} \end{bmatrix}$$

Koefisien B harus diestimasi berdasarkan data hasil penelitian sampel acak. Prosedur estimasi tergantung asumsi mengenai variabel X dan kesalahan pengganggu ε. Beberapa asumsi penting tersebut adalah sebagai berikut:

1) Nilai harapan setiap kesalahan pengganggu sama dengan nol==>$E(\varepsilon_i)=0$, untuk semua i.

$$\begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ E(\varepsilon_i) \\ \cdot \\ \cdot \\ \cdot \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} = \underline{0} = \text{vektor nol}$$

Perhatikan bahwa nilai harapan suatu vektor/matrik adalah nilai harapan dari masing-masing komponen vektor/matriks tersebut.

2) Kesalahan pengganggu yang satu (ε_i) tidak berkorelasi (bebas) terhadap kesalahan pengganggu lainnya(ε_j), akan tetapi mempunyai varian yang sama.

$E(\varepsilon_i\varepsilon_j) = 0$, $i = j$, $E(\varepsilon_i^2) = \sigma^2$ untuk semua i.

Apabila dinyatakan dalam bentuk matriks adalah sbb:

$$E(\varepsilon\varepsilon^T) = \begin{bmatrix} E(\varepsilon_i^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \dots & E(\varepsilon_2\varepsilon_n) \\ E(\varepsilon_i\varepsilon_1) & E(\varepsilon_i\varepsilon_2) & \dots & E(\varepsilon_i\varepsilon_n) \\ \dots & & & \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 0 \dots & 0 \dots & 0 \\ 0 \ \sigma^2 \dots & 0 \dots & 0 \\ \dots & & \\ 0 \ 0 \dots & \sigma^2 \dots & 0 \\ \dots & & \\ 0 \ 0 \dots & 0 \dots & \sigma^2 \end{bmatrix} = \sigma^2 \underline{I} .$$

$\underline{\varepsilon}^T$ = tranpose dari vektor kolom $\underline{\varepsilon}$ atau ε^T merupakan vektor baris.

$\underline{\varepsilon}^T = (\varepsilon_1 \dots \varepsilon_2 \dots \varepsilon_i \dots \varepsilon_n)$ = matriks identitas, karena setiap kesalahan pengganggu mempunyai varians yang sama (perhatikan tanda σ^2 pada diagonal matriks).

3) X_{1i}, X_{2i}, ..., X_{ki} merupakan bilangan riil.Dengan kata lain matriks merupakan himpunan angka-angka konstan.

4) Matriks X mempunyai rank $k < n$ (ada k kolom dari matriks X yang bebas linier). Banyaknya observasi n harus lebih banyak dari banyaknya variabel, atau lebih banyak dari koefisien regresi parsial yang akan diestimasi.

2.2.1 Penaksiran koefisien regresi pada regresi linier berganda.

Persamaan regresi linier berganda perkiraan adalah:

$$\hat{Y} = b_o + b_1X_1 + \ldots + b_kX_k \qquad (2.2.3)$$

maka $Y = \hat{Y} + e$

Misalkan b sebagai penduga $\underline{B}$ merupakan vektor kolom dengan k baris sebagai berikut:

$$b = \begin{bmatrix} b_o \\ b_1 \\ \vdots \\ b_k \end{bmatrix} \Longrightarrow \underline{Y} = \underline{Xb} + e \quad \Longrightarrow e = \underline{Y} - \underline{Xb}$$

$$\begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_i \\ \cdot \\ \cdot \\ e_k \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_i \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} - \begin{bmatrix} 1 & X_{11} & X_{21} & \ldots\ldots & X_{k1} \\ 1 & X_{12} & X_{22} & \ldots\ldots & X_{k2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1i} & X_{2i} & & X_{ki} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{1n} & X_{2n} & \ldots\ldots & X_{kn} \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ \cdot \\ \cdot \\ b_i \\ \cdot \\ \cdot \\ b_k \end{bmatrix}$$

$$e_i = Y_i - b_o - b_1X_{1i} - \ldots - b_kX_{ki}$$

$$\Sigma e_i^2 = \Sigma(Y_i - b_o - b_1X_{1i} - \ldots - b_kX_{ki})^2$$

$$\underline{e}^T\underline{e} = (\underline{Y} - \underline{Xb})^T(\underline{Y} - \underline{Xb}) = \underline{Y}T\ \underline{Y} - 2\underline{b}^T\underline{X}^T\underline{Y} + \underline{b}^T\underline{X}^T\ \underline{Xb}$$

$\underline{b}^T\underline{X}^T\underline{Y}$ = suatu skalar, maka dari itu sama dengan transposnya $\underline{Y}^T\ \underline{Xb}$.

Estimasi vektor $\underline{B}$ dengan menggunakan metode kuadrat terkecil, ialah vektor $\underline{b}$ sedemikian rupa sehingga jumlah kuadrat kesalahan pengganggu, $e^Te = \Sigma e_i^2$ = minimum. Caranya ialah dengan melakukan penurunan secara parsial Σe_i^2 terhadap setiap komponen dari vektor b dan menyamakannya

dengan nol.

$$\frac{\delta \Sigma e_i^2}{\delta b_o} = 2\Sigma(Y_i - b_o - b_1X_{1i} - \ldots - b_kX_{ki})\ (-1) = 0$$

$$\frac{\delta \Sigma e_i^2}{\delta b_1} = 2\Sigma(Y_i - b_o - b_1X_{1i} - \ldots - b_kX_{ki})\ (-X_{1i}) = 0$$

....

$$\frac{\delta \Sigma e_i^2}{\delta b_k} = 2\Sigma(Y_i - b_o - b_1X_{1i} - \ldots - b_kX_{ki})\ (-X_{ki}) = 0$$

Persamaan di atas, setelah disederhanakan akan menjadi:

$$b_on + b_1\Sigma X_{1i} + \ldots + b_k\Sigma X_{ki} = \Sigma Y_i$$

$$b_o\Sigma X_{1i} + b_1\Sigma X_{1i}^2 + \ldots + b_k\Sigma X_{1i}X_{ki} = \Sigma X_{1i}Y_i$$

. . . . . .

. . . . . .

$$b_o\Sigma X_{ki} + b_1\Sigma X_{1i}X_{ki} + \ldots + b_k\Sigma X_{ki}^2 = \Sigma X_{ki}Y_i$$

Persamaan di atas disebut persamaan normal.

Jika dinyatakan dalam bentuk matriks, maka persamaan normal di atas akan menjadi $\underline{X}^T\underline{X}\underline{b} = \underline{X}^T\underline{Y}$. Dengan demikian, $\underline{b}$ sebagai penduga $\underline{B}$ dapat diperoleh melalui persamaan berikut:

$$b = (\underline{X}^T\underline{X})^{-1}\ \underline{X}^T\underline{Y} \qquad X \text{ dengan rank } k < n$$

$$(\underline{X}^T\underline{X})^{-1} = \text{inverse } \underline{X}^T\underline{X}$$

Jika k=3 maka $\hat{Y} = b_o + b_1X_{1i} + b_2X_{2i}$ akan disajikan dalam bentuk matriks sbb:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ . & . & . \\ 1 & X_{31} & X_{32} \end{bmatrix} \quad b = \begin{bmatrix} b_o \\ b_1 \\ b_2 \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix}$$

dari bentuk di atas dibentuk persamaan sbb:

$$X^T X b = X^T Y$$

$$\begin{bmatrix} 1 & 1 \ldots 1 \\ X_{11} & X_{21} \ldots X_{k1} \\ X_{12} & X_{22} \ldots X_{kn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & & \\ 1 & X_{k1} & X_{kn} \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 1 \ldots 1 \\ X_{11} & X_{21} \ldots X_{k1} \\ X_{12} & X_{22} \ldots X_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

akan diperoleh:

$$\underbrace{\begin{bmatrix} n & \Sigma X_{1i} & \Sigma X_{2i} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 & \Sigma X_{1i} X_{2i} \\ \Sigma X_{2i} & \Sigma X_{2i} X_{1i} & \Sigma X_{2i}^2 \end{bmatrix}}_{\underline{A}} \underbrace{\begin{bmatrix} b_o \\ b_1 \\ b_2 \end{bmatrix}}_{\underline{B}} = \underbrace{\begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1i} Y_i \\ \Sigma X_{2i} Y_i \end{bmatrix}}_{\underline{H}}$$

$$\underline{A}\ \underline{b} = \underline{H}$$

$$\underline{b} = \underline{A}^{-1} \underline{H}$$

$$\underline{A}^{-1} = \text{inverse } \underline{A}$$

$$A^{-1} = \frac{1}{\det(\underline{A})} \operatorname{adj}(\underline{A}) = \frac{1}{|A|} \operatorname{adj}(\underline{A}) = \frac{K^T}{|A|}$$

$\underline{K}^T$= transpose matriks kofaktor $\underline{K}$.

Matriks kofaktor K tersebut adalah:

$$\underline{K} = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix}$$

di mana:

$K_{11} = \Sigma X_{1i}^2 \Sigma X_{2i}^2 - (\Sigma X_{1i} \Sigma X_{2i})^2$

$K_{12} = \Sigma X_{1i} \Sigma X_{2i}^2 - \Sigma X_{2i} (\Sigma X_{1i} \Sigma X_{2i})$

$K_{13} = \Sigma X_{1i} \Sigma X_{2i} X_{1i} - \Sigma X_{2i} \Sigma X_{1i}^2$

$K_{21} = \Sigma X_{1i} \Sigma X_{2i}^2 - \Sigma X_{2i} (\Sigma X_{1i} \Sigma X_{2i})$

$K_{22} = n \Sigma X_{2i}^2 - (\Sigma X_{2i})^2$

$K_{23} = n \Sigma X_{2i} X_{1i} - \Sigma X_{1i} X_{2i}$

$K_{31} = \Sigma X_{1i} \Sigma X_{2i} X_{1i} - \Sigma X_{2i} \Sigma X_{1i}^2$

$K_{32} = n \Sigma X_{2i} X_{1i} - \Sigma X_{1i} X_{2i}$

$K_{33} = n \Sigma X_{1i}^2 - (\Sigma X_{1i})^2$

$\det(\underline{A}) = a_{11} K_{11} + a_{12} K_{12} + a_{13} K_{13}$

padahal diketahui $\underline{b} = \underline{A}^{-1} H$, maka

$$\begin{bmatrix} b_o \\ b_1 \\ b_2 \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

Dengan menyelesaikan persamaan matriks di atas akan dapat diketahui harga-harga b_o, b_1, dan b_2.

Dan dapat ditentukan persamaan regresi linier berganda yaitu: $\hat{Y} = b_o + b_1 X_{1i} + b_2 X_{2i}$.

Selanjutnya untuk menghitung simpangan baku atau kesalahan pengganggu S_e kita gunakan persamaan berikut ini:

$$S_e^2 = \frac{\underline{e}^T \underline{e}}{n - k} \quad , \quad \underline{e}^T \underline{e} = \Sigma \underline{e}_i^2 \ , \ n = \text{banyaknya observasi}$$

k = banyaknya variabel

$$S_e = \sqrt{\frac{\underline{e}^T \underline{e}}{n - k}} = \sqrt{\frac{\Sigma \underline{e}_i^2}{n - k}}$$

Se = mengukur variasi Y terhadap garis regresi $\hat{Y}$, sebab e $= Y - \hat{Y}$.

$$\underline{e}^T\underline{e} = (\underline{Y} - \underline{X}b)^T(\underline{Y} - \underline{X}b)$$
$$= \underline{Y}^T\underline{Y} - 2b^T\underline{X}^T\underline{Y} + b^T\underline{X}^T\underline{X}b$$
$$= \underline{Y}^T\underline{Y} - 2b^T\underline{X}^T\underline{Y} + b^T\underline{X}^T\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{Y}$$
$$= \underline{Y}^T\underline{Y} - 2b^T\underline{X}^T\underline{Y} + b^T\underline{X}^T\underline{Y}$$
$$= \underline{Y}^T\underline{Y} - b^T\underline{X}^T\underline{Y}$$

Penyajian dalam bentuk matriks adalah sbb:

$$Y^TY = [Y_1 \; Y_2 \; \ldots \; Y_i] \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_i \end{bmatrix} = Y_1^2 + Y_2^2 + \ldots + Y_i^2 = \Sigma Y_i^2$$

$$\underline{X}^TY = \begin{bmatrix} 1 & 1 & \ldots 1 \\ X_{11} & X_{21} & \ldots X_{k1} \\ & & \\ X_{1i} & X_{2i} & \ldots X_{ki} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_i \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1i}Y_i \\ \\ \Sigma X_{ki}Y_i \end{bmatrix}$$

$$b^T\underline{X}^T\underline{Y} = (b_o, b_1, \ldots, b_k) \begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1i}Y_i \\ \\ \Sigma X_{ki}Y_i \end{bmatrix}$$

$$= b_o\Sigma Y_i + b_1\Sigma X_{1i}Y_i + \ldots + b_k\Sigma X_{ki}Y_i$$

maka:

$$\underline{e}^T\underline{e} = \Sigma\underline{e}_i^2 = \Sigma Y_i^2 - b_o\Sigma Y_i + b_1\Sigma X_{1i}Y_i + \ldots + b_k\Sigma X_{ki}Y_i$$

kalau hubungan yang terjadi adalah antara tiga variabel maka:

$$\Sigma e_i^2 = \Sigma Y_i^2 - b_o\Sigma Y_i + b_1\Sigma X_{1i}Y_i + b_2\Sigma X_{ki}Y_i$$

$$S_e^2 = \frac{\underline{e}^T \underline{e}}{n - k}$$

Apabila $D = \underline{A}^{-1}$ dan $S_{bj}^2 = S_e^2 \; d_{jj}$, di mana d_{jj}= elemen matrik dari baris j dan kolom j yang terletak pada diagonal pokok.

Harga masing-masing simpangan baku dapat ditentukan dari persamaan berikut:

$$S_{bo}^2 = S_e^2 \; d_{11}/D \quad ==> \quad S_{bo} = \sqrt{S_e^2 \; d_{11}/D}$$

$$S_{b1}^2 = S_e^2 \; d_{22}/D \quad ==> \quad S_{b1} = \sqrt{S_e^2 \; d_{22}/D}$$

$$S_{b2}^2 = S_e^2 \; d_{33}/D \quad ==> \quad S_{b2} = \sqrt{S_e^2 \; d_{33}/D}$$

S_{bo}, S_{b1}, dan S_{b2} merupakan kesalahan baku dari penduga b_o, b_1, dan b_2 , dan biasanya ditulis dibagian bawah masing-masing penduga tersebut. Makin kecil kesalahan baku penduga tersebut, maka makin baiklah (teliti) penduga tersebut. Penerapan metode kuadrat terkecil akan memberikan kesalahan baku yang minimum bagi setiap penduga. Artinya metode lain tidak akan menghasilkan kesalahan baku yang lebih kecil atau paling-paling sama, apabila dibandingkan dengan metode kuadrat terkecil.

2.2.2 Koefisien determinasi berganda.

Didalam regresi linier berganda untuk menyatakan hubungan keeratan antara variabel Y dan veriabel X_1, X_2, ...X_n akan dinyatakan dengan koefisien determinasi berganda.

Untuk pembahasan ini akan dibatasi untuk 3 variabel saja, yaitu: Y (variabel tak bebas) dan dua variabel bebas

X_1 dan X_2. Bentuk persamaan regresi perkiraan adalah:

$\hat{Y} = b_o + b_1X_{1i} + b_2X_{2i} + e$.

Besarnya koefisien determinasi berganda (R^2) dari persamaan regresi linier berganda yaitu:

$$R^2 = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$$

$$= \frac{\Sigma y_i^2 - \Sigma e_i^2}{\Sigma y_i^2}$$

di mana:

$$\Sigma e_i^2 = \Sigma(y_i - b_1x_{1i} - b_2x_{2i})^2$$

$$= \Sigma e_i(y_i - b_1x_{1i} - b_2x_{2i}) \text{ karena } e_i = y_i - b_1x_{1i} - b_2x_{2i}$$

$$= \Sigma e_i y_i - b_1\Sigma e_i x_{1i} - b_2\Sigma e_i x_{2i}$$

$$= \Sigma e_i y_i \quad (\text{karena } b_1\Sigma e_i x_{1i} = b_2\Sigma e_i x_{2i} = 0)$$

$$= \Sigma y_i(y_i - b_1x_{1i} - b_2x_{2i})$$

$$= \Sigma y_i^2 - b_1\Sigma x_{1i}y_i - b_2\Sigma x_{2i}y_i$$

$$R^2 = \frac{\Sigma y_i^2 - (\Sigma y_i^2 - b_1\Sigma x_{1i}y_i - b_2\Sigma x_{2i}y_i)}{\Sigma y_i^2}$$

$$R^2 = \frac{b_1\Sigma x_{1i}y_i + b_2\Sigma x_{2i}y_i}{\Sigma y_i^2}$$

di mana nilai R^2 berada dalam interval $0 \leq R^2 \leq 1$.

Adapun makin besar nilai R^2 berarti makin tepat suatu garis regresi linier digunakan sebagai suatu pendekatan. Dan apabila nilai R^2 sama dengan 1(satu) berarti pendekatan tersebut benar-benar sempurna.