BAB II

MODEL REGRESI LINIER GANDA

2.1. Model Regresi Linier Ganda.

Seringkali dijumpai dalam penelitian yang menggunakan analisa regresi, digunakan lebih dari satu variabel pediktor. Model regresi linier ganda pada pengamatan ke i bagian sesatan normal adalah :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_{p-1} X_{p-1} + \epsilon_i \qquad (2.1)$$

dengan

Y_i = harga variabel respon pada trial ke i

X_{i1} = harga variabel prediktor ke 1 pada trial ke i.

X_{i2} = harga variabel prediktor ke 2 pada trial ke i.

⋮

X_{ip} = harga variabel prediktor ke p pada trial ke i.

$\beta_0\ \beta_1, \ldots, \beta_{p-1}$ = parameter-parameter yang tak diketahui.

ϵ_i = suku sesatan random yang saling bebas dan berdistribusi normal dengan mean 0 dan variansi σ^2.

Model (2.1) bisa juga dinyatakan dengan bentuk matriks, dan ditulis, sebagai berikut :

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-2} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ 1 & X_{31} & X_{32} & & \\ \vdots & \vdots & \vdots & & \\ 1 & X_{n1} & X_{n2} & & X_{n,p-1} \end{bmatrix} \quad \underset{\sim}{\beta} = \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{vmatrix} \quad \underset{\sim}{\epsilon} = \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \\ \epsilon_n \end{vmatrix}.$$

atau

$$\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times p}{X} \; \underset{p\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\epsilon}} \qquad (2.2)$$

dengan

$E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$, $E[\underset{\sim}{\epsilon}\,\underset{\sim}{\epsilon}^1] = \sigma^2\, I_n$, I_n matriks indentitas n x p

$E(\underset{\sim}{Y}) = X\,\underset{\sim}{\beta}$ var $(\underset{\sim}{Y}) = \sigma^2\, I$.

2.2. Estimator-estimator Kuadrat Terkecil

Kita bisa mendapatkan vektor estimator-estimator kuadrat terkecil $\underset{\sim}{\beta}$, dengan meminimumkan :

$$S(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \epsilon_i^2 = \underset{\sim}{\epsilon}'\underset{\sim}{\epsilon} = (\underset{\sim}{Y} - X\underset{\sim}{\beta})'(\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

Perhatikan bahwa $S(\underset{\sim}{\beta})$ bisa dinyatakan sebagai :

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

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\,\underset{\sim}{\beta}'X'\underset{\sim}{Y} + \underset{\sim}{\beta}'X'X\underset{\sim}{\beta} \qquad (2.3)$$

Karena $\underset{\sim}{\beta}'X'\underset{\sim}{Y}$ adalah sebuah matriks (1 x 1), dan transpose $(\underset{\sim}{\beta}'X'\underset{\sim}{Y})' = \underset{\sim}{Y}'X\underset{\sim}{\beta}$, maka merupakan skalar yang sama. Estimator-estimator kuadrat terkecil harus memenuhi :

$$\frac{\delta S}{\delta \beta}\bigg|_{\hat{\beta}} = -2\,X\,\underset{\sim}{Y} + 2\,X\,X\,\underset{\sim}{\hat{\beta}} = 0$$

dengan bentuk sederhana :

$$X\ X\ \underset{\sim}{\hat{\beta}} = X'\underset{\sim}{Y} \qquad (2.4)$$

Persamaan (2.4) adalah persamaan-persamaan normal kuadrat terkecil sehingga estimator kuadrat terkecil dari $\underset{\sim}{\beta}$ adalah :

$$\underset{\sim}{\hat{\beta}} = (X'X)^{-1}\ X'\underset{\sim}{Y} \qquad (2.5)$$

Bentuk matriks dari persamaan normal (2.4) adalah bisa ditulis dengan bentuk skalar :

$$\begin{vmatrix} n & \sum_{i=1}^{n} X_{i1} & & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^{2} & & \sum_{i=1}^{n} X_{i1}X_{ik} \\ & & & \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{ii}Y_{ik} & & \sum_{i=1}^{n} X_{ik}^{2} \end{vmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i1}Y_i \\ \\ \sum_{i=1}^{n} X_{ik}Y_i \end{bmatrix} \qquad (2.6)$$

Tampak bahwa $(X'X)$ adalah sebuah matrik simetris (pxp) dan $X'Y$ adalah sebuah vektor kolom (px1). Elemen-elemen diagonal $(X'X)$ adalah jumlahan kuadrat elemen-elemen pada kolom X dan elemen-elemen lainnya adalah jumlahan perkalian elemen-elemen pada kolom X. Selanjutnya elemen-elemen $X'Y$ adalah jumlahan perkalian kolom X dan pengamatan-pengamatan ke Y_i.

Estimator-estimator kuadrat terkecil untuk persamaan (2.5) adalah tak bias dengan variansi minimum, yaitu $E(\underset{\sim}{\hat{\beta}}) = \underset{\sim}{\beta}$.

Bukti :

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

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

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

$= E\,[I\,\underset{\sim}{\beta}\,]$

$= \underset{\sim}{\beta}$

Karena $E\,(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$ dan $(X\,X)^{-1}X'X = I$, jadi $\hat{\underset{\sim}{\beta}}$ adalah estimator tak bias untuk $\underset{\sim}{\beta}$, dan dengan variansi minimum dibuktikan :

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

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

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

$= E\,[(X'X)^{-1}X\,[\underset{\sim}{Y}\underset{\sim}{Y}' - 2\underset{\sim}{Y}\underset{\sim}{\beta}X' + X\underset{\sim}{\beta}\underset{\sim}{\beta}'X']^{-1}\,X(X'X)^{-1}]$

$= (X\,X)^{-1}\,X\,E\,(\underset{\sim}{\epsilon}\,\underset{\sim}{\epsilon}')\,X(X'X)^{-1}$

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

Dimana $(XX)^{-1}$ adalah sebuah matriks (pxp) yang elemen diagonal ke j adalah variansi dari $\hat{\beta}_j$ dan elemen diagonal ke ij adalah kovariansi antara $\hat{\beta}_i$ dan $\hat{\beta}_j$. Selanjutnya jika $C = (X^1X)^{-1}$, maka variansi $\hat{\beta}_j$ adalah $\sigma^2 C_{jj}$ dab kovarian antara $\hat{\beta}_i$ dan $\hat{\beta}_j$ adalah $\sigma^2 C_{ij}$.

Seperti dalam regresi linier sederhana, kita bisa mengembangkan sebuah estimator σ^2 dari jumlah kuadrat residual (SSE) :

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

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

Subtitusi $\underset{\sim}{e} = \underset{\sim}{Y} - X\, \hat{\underset{\sim}{\beta}}$, kita dapatkan

$$SSE = \underset{\sim}{Y}'\underset{\sim}{Y} - 2\, \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{Y} + \hat{\underset{\sim}{\beta}}'X'X\, \hat{\underset{\sim}{\beta}}$$

Karena $X'X\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{Y}$, persamaan ini menjadi

$$SSE = \underset{\sim}{Y}'\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{Y} \qquad (2.7)$$

Jumlah kuadrat residual mempunyai nxp derajat bebas yang berhubungan denganhya, karena parameter p merupakan estimasi didalam model regresi.

Rata-rata kuadrat residual adalah

$$MSE = \frac{SSE}{n-p} \qquad (2.8)$$

merupakan estimator tak bias untuk σ^2, sedemikian hingga

$$E\ (MSE) = \sigma^2 \qquad (2.9)$$

Bukti :

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

$$= (\hat{\underset{\sim}{\beta}}'X' + \underset{\sim}{\epsilon}')\ (X\, \hat{\underset{\sim}{\beta}} + \underset{\sim}{\epsilon}) -$$

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

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

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

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

$$\underset{\sim}{\epsilon}'X\ (X'X)^{-1}\ X'X\, \hat{\underset{\sim}{\beta}} -$$

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

Sehingga

$$E(SSE) = E(\underset{\sim}{\epsilon}' \underset{\sim}{\epsilon}) - E(\underset{\sim}{\epsilon} X (X'X)^{-1} X \underset{\sim}{\epsilon}) \qquad (2.10)$$

Perhatikan sekarang $\underset{\sim}{\epsilon}' X (X'X)^{-1} X' \underset{\sim}{\epsilon}$. Jika $X (X'X)^{-1} X' = A$ maka bentuk ini bisa ditulis sebagai bentuk kuadrat $\underset{\sim}{\epsilon} A \underset{\sim}{\epsilon}$, Maka

$$E(\underset{\sim}{\epsilon} A \underset{\sim}{\epsilon}) = E(\sum_i \sum_j \epsilon_i \epsilon_j a_{ij})$$

$$= E(\sum_i \epsilon_j^2 a_{ii}) - E(\sum_{i=j} \sum a_{ij} \epsilon_i \epsilon_j) \qquad (2.11)$$

Kalau a_{ij} merupakan unsur A pada baris ke-i dan kolom ke-j. Akan tetapi diketahui bahwa :

$$E(\epsilon_i \epsilon_j) = \begin{cases} 0 & \text{untuk } i \neq j \\ \sigma^2 & \text{untuk } i = j \end{cases}$$

Sehingga :

$$E(\underset{\sim}{\epsilon} A \underset{\sim}{\epsilon}) = \sigma^2 \Sigma_i a_{ii} + 0 = \sigma^2 \operatorname{Tr} A = \sigma^2 \operatorname{Tr} X(X'X)^{-1} X'$$

Untuk mengetahui berapa nilai Tr A, baik dikemukakan dahulu dalil berikut :

Kalau $A = (a_{ij})_{mn}$ dan $B = (b_{ij})_{nm}$, maka ukuran Tr AB = Tr BA.

Bukti :

$AB = (\sum_{t=1}^{n} a_{rt} b_{tp})_{mn}$, sehingga $\operatorname{Tr} AB = \sum_{r=1}^{m} \sum_{t=1}^{n} a_{rt} b_{tr}$.

$$BA = (\sum_{k=1}^{n} a_{ik} b_{kj})_{nn} \text{ , sehingga } Tr\ AB = \sum_{i=1}^{n} \sum_{k=1}^{m} b_{ik}\ a_{ki}$$

Akan tetapi

$$Tr\ BA = \sum_{i=1}^{n} \sum_{k=1}^{m} b_{ik}\ a_{ki} = \sum_{i=1}^{m} \sum_{k=1}^{n} a_{ik}\ b_{ki}$$

$$= \sum_{r=1}^{n} \sum_{t=1}^{m} a_{rt}\ b_{tr} = Tr\ AB$$

Dalil ini dapat digunakan terhadap matriks A

$$Tr \underset{nxp}{X} \underset{pxp}{(X'X)^{-1}} \underset{pxn}{X'} = Tr \underset{pxn}{X'} \underset{nxp}{X} \underset{pxp}{(X'X)^{-1}}$$

Oleh karena Tr A = Tr I = p, sehingga

$$E\ (SSE) = n\ \sigma^2 - p\ \sigma^2 = (n-p)\ \sigma^2 \quad (2.12)$$

Akibatnya estimator kuadrat terkecil bagi σ^2 adalah :

$$MSE = \frac{SSE}{n-p}\ \sigma^2 \qquad (2.13)$$

nilai n-p dinamakan derajad bebas dari SSE.

2.3. Koefisien Korelasi Parsial.

Jika koefisien determinan ganda mengukur poporsi variansi Y yang disebabkan oleh hubungan dengan semua variabel X_1 , X_2 , ... , X_k.

Secara serentak, maka koefisien korelasi parsial mengukur proporsi dari satu variabel prediktor pada Y, jika variabel-variabel prediktor lain sudah ada dalam model. Pengertian koefisien antara dua variabel prediktor juga dapat diartikan sebagai ukuran korelasi antara dua variabel prediktor sedangkan variabel-variabel lain dianggap konstan. Pandang model regresi ganda yaitu :

$$Y_i^0 = b_1 w_{i1} + b_2 w_{i2} + \dots + b_k w_{ik} + \varepsilon_i \text{ ,}$$

$$i = 1 , 2 , 3 , \dots , n \qquad (2.14)$$

Kita definisikan fungsi variabel sebagai :

$$W_{ij} = \frac{X_{ij} - X_j}{S_{jj}^{1/2}} \qquad \left[\begin{array}{l} i = 1 , 2 , \dots , n \\ \\ j = 1 , 2, \dots , k \end{array} \right. \qquad (\ 2.15\ )$$

$$Y_j^0 = \frac{Y_i - \bar{Y}}{S_{yy}^{1/2}} \qquad i = 1 , 2 , \dots , n \qquad (\ 2.16\ )$$

dimana

$$S_{jj} = \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2$$

$$S_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Vektor koefisien regresi kuadrat terkecil adalah :

$$\hat{\underline{b}} = (w^1 w)^{-1} \, w^1 y^0 \qquad (2.17)$$

Dan matriks $w^1 w$ merupakan bentuk dari sebuah matrik korelasi yaitu :

$$W^1 W = \begin{bmatrix} 1 & r_{12} & r_{13} \cdots & r_{1k} \\ r_{12} & 1 & r_{23} \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \\ \vdots & \vdots & \vdots & \\ r_{1k} & r_{2k} & r_{3k} & 1 \end{bmatrix}$$

dimana

$$r_{ij} = \sum_{y=1}^{n} \frac{(X_{ui} - \bar{X}_i)(X_{uj} - \bar{X}_j)}{(S_{ii}\ S_{jj})^{1/2}} = \frac{S_{ij}}{(S_{ii} \cdot S_{jj})^{1/2}}$$

adalah korelasi sederhana antara variabel prediktor X_i dan X_j , secara sama w^1y^0adalah :

$$W^1Y^0 = \begin{bmatrix} r_1y \\ r_2y \\ r_3y \\ \vdots \\ r_ny \end{bmatrix}$$

dimana

$$r_{jy} = \sum_{y=1}^{n} \frac{(X_{ui} - \bar{X}_i)(X_{uj} - \bar{X}_j)}{(S_{ii}\ S_{jj})^{1/2}} = \frac{S_{ij}}{(S_{ii}\ S_{jj})^{1/2}}$$

adalah korelasi sederhana antara variabel prediktor X_j dan variabel respons Y.