BAB II

MATERI PENUNJANG

2.1 Matriks dan Differesiansi Matriks

Oleh karena banyaknya konsep-konsep dasar dalam analisa statistik menggunakan matriks, sebab itu akan lebih baik bila ditinjau terlebih dahulu mengenai matriks aljabar beserta differensiasinya.

Theorema 1: (theorema decomposisi spektral)

Ambil matriks simetris A ukuran $(n \times n)$ maka ada matriks orthogonal P sedemikian sehingga:

$P'AP = \Lambda$ atau $A = P\Lambda P'$ dengan $P'P = PP' = I$,

dimana P merupakan matriks yang mempunyai kolom $P_1 \ldots P_n$, dan Λ matriks diagonal dengan diagonal $(\lambda_1, \lambda_2, \ldots \lambda_n)$ dan merupakan matriks eigenvalue dari A.

Maka

$$\mathrm{tr}\ A = \Sigma\ \lambda_i$$

Bukti:

$$\mathrm{tr}\ A = \mathrm{tr}\ P'AP = \mathrm{tr}\ \Lambda PP' = \mathrm{tr}\ \Lambda = \Sigma\ \lambda_i . \ \blacksquare$$

Theorema 2:

A adalah matriks simetris $(n \times n)$ dan D adalah definit non negatif, yaitu:

$$D \geq 0 \Leftrightarrow \text{tr } CD \geq 0 \quad \text{untuk semua } C \geq 0$$

Bukti:

Mengikuti theorema 1, maka dapat direpresentasikan $D = P\Lambda P' = \sum_{i=1}^{n} \lambda_i P_i P_i'$ dimana $\lambda_1, \lambda_2, \ldots \lambda_n$ adalah eigenvalue dari D ($\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$) dan $P = (P_1 \ldots P_n)$ adalah eigenvektor orthogonal yang sesuai.

Sehingga $\text{tr } (CD) = \text{tr}\left[\sum_{i=1}^{n} \lambda_i CP_i P_i'\right]$

$$= \sum_{i=1}^{n} \lambda_i P_i CP_i'$$

Syarat $P_i' CP_i$ adalah non negatif untuk semua $C \geq 0$. Jika $D \geq 0$, semua λ_i adalah non negatif sehingga didapat

$$\text{tr } CD \geq 0.$$

Sebaliknya, jika $\text{tr } CD \geq 0$ untuk semua $C \geq 0$ hal ini adalah benar terutama untuk $C = P_i P_i' \geq 0$. Kemudian didapat: $\text{tr } CD = \text{tr } \left[P_i P_i' \left(\sum_{j=1}^{n} \lambda_j P_j CP_j'\right)\right]$

$$= \lambda_i \geq 0 \qquad (i = 1, 2, \ldots n) \qquad \blacksquare$$

Selanjutnya jika $f(x)$ adalah sebuah fungsi skalar riil dari sebuah matriks $X = [x_{ij}]$ ukuran $(m \times n)$, maka differensi partial f terhadap X didefinisikan sebagai matriks $(m \times n)$ dari differensial parsial $\frac{\partial f}{\partial x_{ij}}$, yaitu:

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \partial f/\partial x_{11} & \partial f/\partial x_{12} & \cdots\cdots & \partial f/\partial x_{1n} \\ \partial f/\partial x_{21} & \partial f/\partial x_{22} & \cdots\cdots & \partial f/\partial x_{2n} \\ \vdots & & & \\ \partial f/\partial x_{m1} & \partial f/\partial x_{m2} & \cdots\cdots & \partial f/\partial x_{mn} \end{bmatrix}$$

Contoh 1:

$$y = \text{tr } \mathbf{AX} + \text{tr } \mathbf{XBX'}$$

yang mempunyai nilai maximum atau minimum terhadap $\mathbf{X}$. Jika $\mathbf{X}$ mempunyai m baris dan n kolom maka turunan $\partial y/\partial \mathbf{X}$ dengan elemen ke-(i,j)-nya adalah $\partial y/\partial x_{ij}$ (i=1,2,3.....m ;j = 1,2,3.....n).

Dimisalkan $y = \text{tr } Z$; dimana $Z = \mathbf{AX} + \mathbf{XBX'}$

Sekarang $\dfrac{\partial X}{\partial x_{ij}} = E_{ij} = e_i e_j'$

dan $\dfrac{\partial X'}{\partial x_{ij}} = E_{ji} = e_j e_i'$

dimana $\mathbf{E}_{ij}$ adalah matriks dengan elemen pada posisi ke (i,j) adalah satu dan elemen ditempat lain nol. Vektor-vektor $\mathbf{e}_i$ dan $\mathbf{e}_j$ didefinisikan secara sama.

Jadi:

$$Z = \mathbf{AX} + \mathbf{XBX'}$$

$$\frac{\partial Z}{\partial x_{ij}} = \mathbf{AE}_{ij} + \mathbf{E}_{ij}\mathbf{BX'} + \mathbf{XBE}_{ji}$$

Oleh karena itu,

$$\frac{\partial y}{\partial x_{ij}} = \text{tr } \frac{\partial Z}{\partial x_{ij}}$$

$$= \text{tr } (\mathbf{AE}_{ij} + \mathbf{E}_{ij}\mathbf{BX'} + \mathbf{XBE}_{ji})$$

$$= \text{tr } (\mathbf{Ae}_i\mathbf{e}_j' + \mathbf{e}_i\mathbf{e}_j'\mathbf{BX'} + \mathbf{XBe}_j\mathbf{e}_i')$$

$$\frac{\partial y}{\partial x_{ij}} = \mathrm{tr}\ (e_j'Ae_i + e_j'BX'e_i + e_i'XBe_j)$$

$$= (A)_{ji} + (BX')_{ji} + (XB)_{ij}$$

Hasil ini sama dengan elemen ke-(i,j) dari (A'+XB'+XB).

Contoh 2:

$$y = \mathrm{tr}\ Z \quad ; \text{ dimana } Z = AXBXCX'$$

Maka $\dfrac{\partial Z}{\partial x_{ij}} = AE_{ij}BXCX' + AXBE_{ij}CX' + AXBXCE_{ji}$

Oleh karena itu:

$$\frac{\partial y}{\partial x_{ij}} = \mathrm{tr}\ \frac{\partial Z}{\partial x_{ij}}$$

$$= \mathrm{tr}\ (AE_{ij}BXCX' + AXBE_{ij}CX' + AXBXCE_{ji})$$

$$= \mathrm{tr}\ (Ae_ie_j'BXCX' + AXBe_ie_j'CX' + AXBXCe_je_i')$$

$$= \mathrm{tr}\ (e_j'ABXCX'e_i + e_j'AXBCX'e_i + e_i'AXBXCe_j)$$

$$= (ABXCX')_{ji} + (AXBCX')_{ji} + (AXBXC)_{ij}$$

Jadi $\dfrac{\partial y}{\partial X} = X\ C'X'B'A' + XC'B'X'A' + AXBXC$

Catatan:

$$\frac{\partial}{\partial X} AB = \frac{\partial A}{\partial X} B + A \frac{\partial B}{\partial X}$$

Theorema 3:

Jika $y = \mathrm{tr}\ AC'WC$ maka:

a) $\dfrac{\partial y}{\partial C} = WCA + W'CA'$

b) Jika W dan A adalah simetris, maka: $\dfrac{\partial y}{\partial C} = 2WCA$

Bukti:

Misal $AC'WC = Z$

Jadi, $y = \mathrm{tr}\ AC'WC = \mathrm{tr}\ Z$

Sehingga:

$$\frac{\partial y}{\partial c_{ij}} = \mathrm{tr}\ \frac{\partial Z}{\partial c_{ij}}$$

$$= \mathrm{tr}\ (\ AE_{ji}WC + AC'WE_{ij}\ )$$

$$= \mathrm{tr}\ (\ Ae_j e_i' WC + AC'We_i e_j'\ )$$

$$= \mathrm{tr}\ (\ e_i' AWCe_j + e_j' AC'We_i\ )$$

$$= (AWC)_{ij} + (AC'W)_{ji}$$

Jadi,

$$\frac{\partial y}{\partial C} = (AWC) + (AC'W)'$$

$$= WCA + W'CA'$$

untuk W dan A simetris, maka

$$\frac{\partial y}{\partial C} = 2WCA$$

2.2 Model Linier dan Estimator-estimatornya

Model standar yang digunakan dalam tulisan ini adalah:

$$y = X\beta + u \quad ;\text{dengan } u \sim (0,\Sigma) \qquad \text{<2.2.1>}$$

dimana:

y : vektor $(n{\times}1)$ dari variabel dependen, yang bersifat random dan terobservasi

X : matriks $(n{\times}p)$ dari variabel penjelas, yang bersifat tetap dan terobservasi

β : vektor (p×1) dari koefisien regresi, yang bersifat tetap dan tak terobservasi

u : vektor (n×1) pengganggu, yang bersifat random dan tak terobservasi

Σ : matriks penyebaran (n×n), yang bersifat tetap dan tak terobservasi.

2.2.1 Ordinary Least Square (OLS)

Notasi $u \sim (0,\Sigma)$ berarti bahwa **u** vektor pengganggu mempunyai ekspektasi $\mathbb{E}[u]=0$ dan penyebaran (kovarian) $\mathbb{E}[uu']=\Sigma$. Dan model diatas disingkat oleh RAO (1971) dengan tiga simbol yaitu $(y,X\beta,\Sigma)$. Dalam kasus khusus dimana $\Sigma=\sigma^2 I$, sehingga elemen-elemen dari vektor pengganggu tidak berkorelasi, maka estimator standar dari β adalah estimator Ordinary Least Square (OLSE):

$$\hat{\beta} = (X'X)^{-1}X'y \qquad \text{<2.2.2>}$$

Yang dapat ditulis sebagai:

$$\hat{\beta} = T^{-1}X'y \quad \text{atau} \quad \hat{\beta} = GX'y$$

dengan $T = X'X$ dan $G = T^{-1}$, serta model <2.2.1> disebut *regresi klasik* atau *regresi dengan vektor pengganggu bervarian sama.*

2.2.2 Generalized Least Square

GLS dapat digunakan pada situasi dimana syarat pengganggu (elemen-elemen **u**) tidak semuanya berdistribusi sama, atau jika variabel-variabel tersebut berkorelasi.

Estimator GLS (GLSE) dari β didefinisikan sebagai:

$$\mathbf{b} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y} \qquad \langle 2.2.3\rangle$$

Catatan, ketika $\Sigma=\sigma^2 I$ maka estimator diatas akan menjadi estimator OLS seperti yang telah dibicarakan pada bagian sebelumnya.

Beberapa ide diatas mungkin akan menjadi jelas jika diambil contoh pada kasus univariat berikut:

$$y_i = \mu + u_i$$

yang menggambarkan kumpulan observasi $y_1, y_2, \ldots\ y_n$, yang masing-masing mempunyai rata-rata μ. Persamaan ke-n dapat dituliskan kembali oleh:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{u}$$

dimana $\mathbf{1}$ adalah vektor satuan yang berukuran (n×1) yang jelas berbentuk:

$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$, dengan $\mathbf{X}=\mathbf{1}$ dan vektor β yang berukuran (p×1) adalah sebuah skalar μ.

OLSE untuk model diatas adalah:

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

$$\hat{\beta} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{y}$$
$$= \frac{1}{n} n\bar{y}$$
$$= \bar{y}$$

Dengan kata lain, OLSE untuk kasus ini adalah benar-benar merupakan rata-rata sampel.

2.2.3 Estimator Linier

Jika $\boldsymbol{y}$ adalah sebuah vektor random, dan $\mathbf{C}$ serta $\mathbf{d}$ adalah matriks dan vektor yang berukuran sesuai, maka vektor:

$$\mathbf{bo} = \mathbf{C}\boldsymbol{y} + \mathbf{d} \qquad \text{<2.2.4>}$$

dapat disebut "estimator linier" dari vektor β. Dalam kasus khusus dimana $\mathbf{d}$=0, dikatakan bahwa $\mathbf{bo}$ adalah estimator "homogen". Ketika $\mathbf{d}\neq 0$, βo disebut "estimator heterogen". Dan OLSE yang diberikan oleh persamaan <2.2.2> adalah kasus khusus dari estimator linier homogen, didapat dengan mengambil $\mathbf{C}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ dan $\mathbf{d}$=0 dalam persamaan <2.2.4>.

2.2.4 Estimator Ridge

Sebagai perluasan dari estimator OLS, HOERL dan KENNARD (1970) mengusulkan suatu estimator yang disebut "estimator ridge" dan didefinisikan sebagai :

$$\beta_k = (X'X + kI)^{-1} X'y \qquad \text{<2.2.5>}$$

Jelaslah bahwa <2.2.3> adalah kasus khusus dari <2.2.5> yang didapat dengan menempatkan k = 0.

2.3 Sesatan Kuadrat Rata-Rata Pada Model Linier Umum

Vektor Bias dan matriks Kovarian (penyebaran) dari $\hat{\beta}$ dalam model $(y, X\beta, \sigma^2 I)$ adalah:

$$\mathbb{B}_{\hat{\beta}} = \mathbb{E}[\hat{\beta}] - \beta = 0 \quad \text{dan} \quad \mathbb{V}_{\hat{\beta}} = \sigma^2 (X'X)^{-1} \qquad \text{<2.3.1>}$$

Karena itu matriks SKR dari model tersebut adalah:

$$\mathrm{SKR}_{\hat{\beta}} = \mathbb{E}\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right]$$

$$\mathrm{SKR}_{\hat{\beta}} = \sigma^2 (X'X)^{-1} \qquad \text{<2.3.2>}$$

Secara umum, jika **bo** adalah sebuah estimator linier heterogen yang diberikan oleh persamaan <2.2.4>, maka :

$$\begin{aligned} \mathbb{B}_{\mathbf{bo}} &= \mathbb{E}[\mathbf{bo}] - \beta \\ &= \mathbf{C}\mathbf{X}\beta + \mathbf{d} - \beta \\ &= (\mathbf{CX} - \mathbf{I})\beta + \mathbf{d} \end{aligned} \qquad \text{<2.3.3>}$$

dan $\mathbb{V}_{\mathbf{bo}} = \sigma^2 CC'$ <2.3.4>

Matriks SKR dari **bo** adalah:

$$SKR_{bo} = \sigma^2 CC' + \left[(CX - I)\beta + d\right]\left[(CX - I)\beta + d\right]' \qquad \text{<2.3.5>}$$

Ekspresi terakhir ini merupakan rumus "Sesatan Kuadrat Rata-rata = Varian + Bias kuadrat" yang diperluas untuk matriks. Dan ketika $C = (X'X)^{-1}X'$ dan $d=0$ (syarat akhir dalam persamaan <2.3.3> adalah nol), maka keseluruhan ekspresi ekuivalen dengan persamaan <2.3.2>. Catatan bahwa ketika b_o merupakan estimator homogen dan $d=0$, syarat akhir dalam persamaan <2.3.5> adalah :

$$(CX - I)\beta\beta'(CX - I)' \qquad \text{<2.3.6>}$$

Perlu dibedakan antara SKR dalam bentuk matriks dengan definisi lain, yaitu SKR dalam bentuk skalar (SKR skalar) yaitu :

$$\text{SKR skalar } (\hat{\beta}) = \mathbb{E}\left[[\hat{\beta}-\beta]'[\hat{\beta}-\beta]\right]$$

Aplikasi dari operator trace didapat persamaan

$$\text{SKR skalar } (\hat{\beta}) = \text{tr } \mathbb{E}\left[[\hat{\beta}-\beta]'[\hat{\beta}-\beta]\right]$$
$$= \text{tr SKR skalar } (\hat{\beta})$$

Dalam tulisan ini akan dicari estimator yang meminimalkan harga dari persamaan <2.3.5>. Secara umum, tujuannya adalah memperbaiki estimasi berdasarkan kerugian kuadrat rata-rata dengan ukuran kerugian estimator $\hat{\beta}$ oleh fungsi kerugian

$$\mathbb{L}(\hat{\beta},A) = [\hat{\beta}-\beta]'A[\hat{\beta}-\beta]$$

dimana A adalah matriks simetris definit positif yang sesuai dengan order p×p.

Catatan:

A > 0 jika A definit positif

A ≥ 0 jika A definit non negatif

Sedangkan resiko dari estimator $\hat{\beta}$ didefinisikan sebagai:

$$\mathbb{R}(\hat{\beta},A) = \mathbb{E}[\mathbb{L}(\hat{\beta},A)] = \mathbb{E}\Big[[\hat{\beta}-\beta]'A[\hat{\beta}-\beta]\Big] \quad \langle 2.3.7\rangle$$

Aplikasi dari operator trace didapatkan hubungan antara resiko dan (harga matriks) SKR dari estimator yaitu :

$$\mathbb{R}(\hat{\beta},A) = \text{tr } A\ \mathbb{E}\ [\hat{\beta}-\beta]'[\hat{\beta}-\beta] = \text{tr } A\ \text{SKR}(\hat{\beta}) \quad \langle 2.3.8\rangle$$

Rumus ini bersama-sama theorema 1 dan theorema 2 menunjukkan kepadanan untuk mengestimasi resiko $\mathbb{R}(\hat{\beta},A)$ atau SKR $(\hat{\beta})$untuk estimator-estimator $\hat{\beta}$ yang ada dalam kelas tetap dari estimator yang mungkin.

Theorema 4:

Ambil dua estimator β_1 dan β_2. Maka dua pernyataan berikut adalah ekuivalen:

a) $\text{SKR}(\beta_1) - \text{SKR}(\beta_2) \geq 0$

b) $\mathbb{R}(\beta_1, A) - \mathbb{R}(\beta_2, A) \geq 0$

Bukti:

Dari persamaan <2.3.8> didapat:

$$\mathbb{R}(\beta_1, A) - \mathbb{R}(\beta_2, A) = \text{tr } A\{SKR(\beta_1) - SKR(\beta_2)\}$$
$$= \text{tr } A\Delta, \text{ katakan.}$$

Dengan theorema 2 didapatkan:

tr $A\Delta \geq 0$ untuk semua $A \geq 0$ jika hanya jika $\Delta \geq 0$. ■

Theorema 5:

Jika **A** adalah definit positif maka **P'AP** adalah definit non negatif untuk semua **P**. Jika **P** adalah matriks bujur sangkar dan mempunyai rank penuh, maka **P'AP** adalah definit positif.

Bukti:

Jika $x'P'APx = y'Ay$ dimana $y = Px$, maka $x'P'APx \geq 0$ untuk semua x. Jika **P** merupakan matriks bujur sangkar dan memiliki rank penuh, maka $y = 0$ hanya jika $x = 0$, sehingga $x'P'APx > 0 \ \forall x \neq 0$. ■

2.4 Estimator Admissible

Untuk dua estimator $\hat{\beta}(x_1, x_2, x_3, \ldots x_n)$ dan $\underline{\beta}(x_1, x_2, x_3, \ldots x_n)$, estimator $\underline{\beta}$ didefinisikan sebagai estimator yang lebih baik daripada $\hat{\beta}$ jika dan hanya jika:

jika:

$$\mathbb{R}(\underline{\beta}) \le \mathbb{R}(\hat{\beta}) \quad \text{untuk semua } \beta \text{ dalam } \mathbb{H}$$

dengan tanda lebih kecil berlaku paling sedikit untuk satu β dalam $\mathbb{H}$. Maka estimator $\underline{\beta}(x_1,x_2,x_3, \ldots x_n)$ dikatakan *admissible* jika dan hanya jika tidak ada estimator yang lebih baik.

2.5 Pengganda/Pengali Lagrange

Akan ditentukan nilai stasioner (maksimum atau minimum relatif) suatu fungsi $f(\theta_1,\theta_2,\ldots,\theta_m)$ yang tersusun atas m peubah $\theta_1,\theta_2,\ldots,\theta_m$ dengan kendala-kendala terhadap θ_i berupa $g_j(\theta_1,\theta_2,\ldots,\theta_m)=0$ $(j = 1,2,\ldots,q)$. Untuk mengerjakan masalah ini, dibuat suatu fungsi:

$$F = f - \sum_{j=1}^{q} \lambda_j g_j$$

dimana $\lambda_1,\lambda_2,\ldots,\lambda_j$ tidak diketahui. Dengan membuat turunan pertamanya terhadap setiap θ_i dan menyamakannya dengan nol, yaitu:

$$\frac{\partial F}{\partial \theta_i} \equiv \frac{\partial f}{\partial \theta_i} - \sum_{j=1}^{q} \lambda_j \frac{\partial g_j}{\partial \theta_i} = 0 \qquad (i = 1,2, \ldots m)$$

Dengan memandang

$$F = f + \Lambda G \qquad (\text{dengan } G = \sum_{j=1}^{q} g_j \ )$$

maka didapatkan

$$\frac{\partial f}{\partial \theta_i} + \Lambda \frac{G}{\partial \theta_i} = 0$$

Konstanta Λ inilah yang dinamakan suatu *pengali Lagrange* (*Lagrange Multiplier*). Seringkali λ_j berhasil dieliminasi dan tidak dicari nilainya, karena alasan inilah λ_j juga disebut "pengganda yang tidak ditentukan" (*undertermined multiplier*).