ESTIMASI PARAMETER DAN UJI HIPOTESIS PADA MODEL LINEAR MULTIVARIAT DENGAN METODE LDL

Makkulau (makkulau@statistika.its.ac.id)
Jurusan Matematika, FMIPA Universitas Haluoleo, Kendari

Susanti Linuwih, Purhadi, Muhammad Mashuri
Jurusan Statistika, Institut Teknologi Sepuluh Nopember

Rahmawati Pane
Jurusan Matematika, FMIPA Universitas Sumatera Utara, Medan

ABSTRACT

Outliers are observations (data) that lies in an abnormal distance from other observations. Outliers can be distinguished into outliers of univariate or multivariate observation and outliers of univariate or multivariate linear models. Multivariate linear model is a linear model with more than one dependent (response) variables. This research studied parameter estimation and hypothesis test for multivariate linear model using Likelihood Displacement Statistic-Lagrange Method called as LDL method for detecting outlier observations in multivariate linear models with the LDL_{Am} *statistical test.*

Keywords : estimation of parameter, likelihood displacement statistic-lagrange, multivariate linear models, outlier detection.

Pencilan (*outlier*) dibedakan atas *outlier* pada pengamatan (data) univariat atau multivariat dan *outlier* pada model linear univariat atau multivariat. Pendeteksian *outlier* pada pengamatan telah dilakukan antara lain oleh Hawkins (1980), Barnett dan Lewis (1994), Peña dan Prieto (2001), Filzmoser (2005), dan lain-lain. Pendeteksian *outlier* pada model linear telah dilakukan antara lain oleh Cook (1977), Rousseeuw (1984), Peña dan Guttman (1993), Srivastava dan von Rosen (1998), Diaz-Garcia, Gonzalez-Farias, dan Alvarado-Castro (2007), dan lain-lain. *Outlier* yang merupakan pengamatan yang menyimpang sedemikian jauh dari pengamatan lain, dapat mempunyai efek bagi pengambilan suatu kesimpulan atau keputusan pada penelitian.

Xu, Abraham, dan Steiner (2005) mengembangkan jarak univariat Cook untuk mendeteksi *outlier* pada model linear multivariat dengan Metode *Likelihood Displacement Statistic* disingkat Metode LD, seperti pada Makkulau, Linuwih, Purhadi, dan Mashuri (2007a), prosedurnya dapat dilihat pada Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2008) dan aplikasinya dapat dilihat pada Makkulau, Linuwih, Purhadi, dan Mashuri (2009); Metode lain yang digunakannya adalah Metode *Likelihood Ratio Statistic for a Mean Shift* disingkat Metode LR, seperti pada Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2007b); dan Metode *Multivariate Leverage* yang menggunakan elemen dari *the average diagonal* Q_{A_m} disingkat ADQ untuk mengukur keekstriman dari *m* pengukuran pada variabel independen. Xu, Abraham, dan Steiner (2005) dalam mengestimasi parameter dengan Metode LD dan Metode LR dari model linear multivariat tidak menggunakan fungsi pengganda *Lagrange*. Penelitian ini mengkaji estimasi parameter dan uji hipotesis pada model linear multivariat dengan Metode LDL.

Model Linear Multivariat

Misalkan $X_1, X_2, \cdots, X_p$ adalah variabel independen (variabel prediktor) dan $Y_1, Y_2, \cdots, Y_q$ adalah variabel dependen (variabel respon), jika dilakukan *n* pengamatan yang diambil pada setiap variabel dependen yang ditulis $y_{i1}, y_{i2}, \cdots, y_{iq}$ dimana $i = 1, 2, \cdots, n,$ atau $y_{iq} = \beta_{0h} + \beta_{1h}X_{i1} + \beta_{2h}X_{i2} + \cdots + \beta_{ph}X_{ip} + \underset{\sim}{\varepsilon}_h$ dimana $h = 1, 2, \cdots, q,$ dan misalkan $\underset{\sim}{Y}_h = \left[y_{1h}, y_{2h}, \cdots, y_{nh}\right]^T$, maka dapat ditulis sebagai $\underset{\sim}{Y}_h = \mathbf{X}_h \underset{\sim}{\beta}_h + \underset{\sim}{\varepsilon}_h,$ dimana:

$$\mathbf{X}_h = \begin{bmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{X}_q \end{bmatrix}$$ adalah matriks berukuran *n*x(*p*+1)

$\underset{\sim}{\beta}_h = \left[\beta_{0h}, \beta_{1h}, \beta_{2h}, \cdots, \beta_{ph}\right]^T$ adalah vektor parameter berukuran (*p*+1)x1 dan

$\underset{\sim}{\varepsilon}_h = \left[\varepsilon_{1h}, \varepsilon_{2h}, \cdots, \varepsilon_{nh}\right]^T$

Model linear multivariat yang terdiri dari *q* model linear secara simultan dapat ditulis sebagai:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{1}$$

Dengan $\mathbf{Y}_{nxq} = \left[\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \cdots, \underset{\sim}{Y}_q\right], \quad \mathbf{X}_{nx(p+1)} = \left[\underset{\sim}{1}, \underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_p\right],$

$\mathbf{B}_{(p+1)xq} = \left[\underset{\sim}{\beta}_0, \underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \cdots, \underset{\sim}{\beta}_q\right]$, dan $\mathbf{E}_{nxq} = \left[\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2, \cdots, \underset{\sim}{\varepsilon}_q\right].$

Estimasi Parameter dan Uji Hipotesis Model Linear Multivariat

Pada model linear multivariat matriks *error* $\mathbf{E}_{nxq} = [\varepsilon_{ih}]$ merupakan matriks acak, dimana $i = 1, 2, \cdots, n$ dan $h = 1, 2, \cdots, q.$ Dengan mengestimasi parameter $\mathbf{B}$ pada (1), maka diperoleh

$$\hat{\mathbf{B}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} \tag{2}$$

Estimasi dari parameter $\mathbf{\Sigma}$ adalah:

$\hat{\mathbf{\Sigma}} = \frac{1}{n}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)^T\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)$; $\hat{\mathbf{\Sigma}}$ merupakan estimator bias untuk $\mathbf{\Sigma}$.

$\mathbf{S} = \frac{1}{n - \mathrm{rank}(\mathbf{X})}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)^T\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)$; $\mathbf{S}$ merupakan estimator tak bias untuk $\mathbf{\Sigma}$.

Vektorisasi matriks variabel dependen pada (1) ditulis Vec($\mathbf{Y}$) (Christensen, 1991) adalah

$\mathrm{Vec}(\mathbf{Y}) = \left(\mathbf{I}_q \otimes \mathbf{X}\right)\mathrm{Vec}(\mathbf{B}) + \mathrm{Vec}(\mathbf{E})$:

dimana: $\otimes$ adalah perkalian Kronecker

$$\mathrm{Vec}(\mathbf{E}) \sim N_p\left(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n\right) \text{ dan } \mathrm{Vec}(\mathbf{Y}) \sim N_{nq}\left(\left(\mathbf{I}_q \otimes \mathbf{X}\right)\mathrm{Vec}(\mathbf{B}), \mathbf{\Sigma} \otimes \mathbf{I}_n\right). \tag{3}$$

Dengan menggunakan sifat hasil kali *kronecker* diperoleh:

$$\text{Vec}\left(\hat{\mathbf{B}}\right)=\left(\left(\mathbf{I}_q\otimes\mathbf{X}^T\right)\left(\mathbf{I}_q\otimes\mathbf{X}\right)\right)^{-1}\left(\mathbf{I}_q\otimes\mathbf{X}^T\right)\text{Vec}(\mathbf{Y})=\left(\mathbf{I}_q\otimes\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\text{Vec}(\mathbf{Y})$$

dan distribusi $\text{Vec}\left(\hat{\mathbf{B}}\right)$ adalah $\text{Vec}\left(\hat{\mathbf{B}}\right)\sim N_{q(p+1)}\left(\text{Vec}(\mathbf{B}),\mathbf{\Sigma}\otimes\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)$.

Prosedur uji hipotesis parameter pada model linear multivariat adalah:

$H_0:\Lambda^T\mathbf{B}=0$ terhadap $H_1:\Lambda^T\mathbf{B}\neq 0$

dimana $\Lambda^T=\mathbf{P}^T\mathbf{X}$ dan $\mathbf{P}$ adalah matriks ortogonal (Christensen, 1991). Hipotesis nol ditolak jika nilai maksimum dari *likelihood* di bawah H_0 lebih besar dari nilai maksimum keseluruhan. Statistik uji rasio *likelihood* sering diarahkan pada *Wilk's* Λ.

Metode Pendeteksian *Outlier* pada Model Linear Multivariat

Outlier adalah pengamatan yang menyimpang sedemikian jauh dari pengamatan lain (Hawkins, 1980). *Outlier* dapat dibedakan atas *outlier* pada pengamatan univariat atau multivariat dan *outlier* pada model linear univariat atau multivariat. *Outlier* pada model linear multivariat dapat dibagi atas 3 kategori, yaitu *outlier* terhadap nilai *X* (*leverage outlier*); *outlier* terhadap nilai *Y* (*residual outlier*); dan *outlier* terhadap nilai *X* dan *Y* (*outlier* berpengaruh).

Rousseeuw dan Hubert (1997) mengklasifikasikan *outlier* ke dalam empat kelompok berdasarkan penyebabnya, yaitu observasi umum (*regular observations*); titik *leverage* baik (*good leverage points*); *outlier* vertikal (*vertical outliers*); dan titik *leverage* jelek (*bad leverage points* atau *X-Y-outliers*).

Untuk mendeteksi *outlier* pada model linear multivariat, Xu, Abraham, dan Steiner, (2005) mengembangkan jarak univariat Cook. Metode yang digunakan adalah Metode LD, yaitu suatu metode yang menghilangkan pengamatan yang diduga *outlier* pada model seperti pada Makkulau, Linuwih, Purhadi, dan Mashuri (2007a) prosedurnya dapat dilihat pada Makkulau, Linuwih, Purhadi, dan Mashuri (2008) dan aplikasinya dapat dilihat pada Makkulau, Linuwih, Purhadi, dan Mashuri (2009); Metode LR, yaitu suatu metode dengan cara pergeseran rata-rata pada model seperti pada Makkulau, Linuwih, Purhadi, dan Mashuri (2007b); dan Metode *Multivariate Leverage* yang menggunakan elemen dari *the average diagonal* Q_{A_m} disingkat ADQ untuk mengukur keekstriman dari *m* pengukuran pada variabel independen. Fungsi *likelihood* dalam model linear multivariat ditulis sebagai berikut (Christensen, 1991 serta Rencher dan Schaalje, 2008):

$$L(\mathbf{B},\mathbf{\Sigma})=(2\pi)^{-\frac{1}{2}nq}|\mathbf{\Sigma}|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XB})^T(\mathbf{Y}-\mathbf{XB})\right)\right). \qquad (4)$$

Pendeteksian *outlier* pada model linear multivariat dengan Metode LD dilakukan dengan cara menghilangkan pengamatan yang diduga *outlier* pada model. Misalkan ada *m* pengamatan dikumpulkan pada himpunan tertentu, dengan *m* pengamatan diduga *outlier*. Indeks A_m adalah kumpulan dari *m* pengamatan yang diduga *outlier*. Dengan kata lain, indeks A_m artinya ada *outlier*, sehingga:

$\mathbf{Y}_{A_m}$ adalah himpunan $\mathbf{Y}$ dengan pengamatan yang ada *outlier*.

$\mathbf{Y}_{A_m}^C$ adalah himpunan $\mathbf{Y}$ dengan pengamatan tanpa *outlier*.

Definisi 1 (Christensen, 1991)
Likelihood Displacement Statistic (LD) dengan pengamatan yang ada *outlier* adalah:

$$\mathrm{LD}_{A_m}\left(\mathbf{B},\mathbf{\Sigma}\right)=2\left(\ln L\left(\hat{\mathbf{B}},\hat{\mathbf{\Sigma}}\right)-\ln L\left(\hat{\mathbf{B}}_{A_m}^{C},\hat{\mathbf{\Sigma}}_{A_m}^{C}\right)\right) \quad (5)$$

Definisi 2 (Christensen, 1991)
LD dengan pengamatan yang ada *outlier* dan bersyarat adalah:

$$\mathrm{LD}_{A_m}\left(\mathbf{B}_1,\mathbf{\Sigma}_1|\mathbf{B}_2,\mathbf{\Sigma}_2\right)=2\left(\ln L\left(\hat{\mathbf{B}},\hat{\mathbf{\Sigma}}\right)-\ln L\left(\left(\hat{\mathbf{B}}_{1A_m}^{C},\hat{\mathbf{\Sigma}}_{1A_m}^{C}\right),\left\{\left(\hat{\mathbf{B}}_{2A_m}^{C},\hat{\mathbf{\Sigma}}_{2A_m}^{C}\right)\right\}\left(\hat{\mathbf{B}}_{A_m}^{C},\hat{\mathbf{\Sigma}}_{1A_m}^{C}\right)\right)\right) \quad (6)$$

dimana $\{\ \}$ menotasikan suatu fungsi.

Estimator dari $\mathbf{B}$ dengan pengamatan tanpa *outlier* yaitu $\hat{\mathbf{B}}_{A_m}^{C}$ adalah:

$\hat{\mathbf{B}}_{A_m}^{C}=\hat{\mathbf{B}}-\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^{T}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\mathbf{E}}_{A_m}$; dimana $\mathbf{Q}_{A_m}=\mathbf{X}_{A_m}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^{T}$; $\hat{\mathbf{E}}_{A_m}=\mathbf{Y}_{A_m}-\mathbf{X}_{A_m}^{T}\hat{\mathbf{B}}$.

dan estimator dari $\mathbf{\Sigma}$ dengan pengamatan tanpa *outlier* yaitu $\hat{\mathbf{\Sigma}}_{A_m}^{C}$ adalah:

$$\hat{\mathbf{\Sigma}}_{A_m}^{C}=\frac{n}{n-m}\hat{\mathbf{\Sigma}}-\frac{1}{n-m}\hat{\mathbf{E}}_{A_m}^{T}\left(\mathrm{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\mathbf{E}}_{A_m},$$

sehingga fungsi *likelihood* dalam model linear multivariat dengan pengamatan tanpa *outlier* adalah

$$\mathrm{LD}_{A_m}\left(\mathbf{B},\mathbf{\Sigma}\right)=2(\ln L\left(\hat{\mathbf{B}},\hat{\mathbf{\Sigma}}\right)-\ln L(\hat{\mathbf{B}}_{A_m}^{C},\ \Sigma\left(\hat{\mathbf{B}}_{A_m}^{C}\right)).$$

Cara lain pendeteksian *outlier* pada model linear multivariat dengan Metode LR dilakukan dengan cara pergeseran rata-rata pada model, yang ditulis dalam bentuk model:

$\mathbf{Y}=\mathbf{XB}+\mathbf{Z}_{A_m}^{C}\mathbf{\Psi}+\mathbf{E}$; dimana: $\mathbf{Z}_{A_m}^{C}=\left[\underset{\sim}{z}_{i1}\ \underset{\sim}{z}_{i2}\ \cdots\ \underset{\sim}{z}_{im}\right]$, dan $\underset{\sim}{z}_j$, $\underset{\sim}{j}=i_1, i_2, \cdots, i_m$

$\mathbf{\Psi}_{mxq}$: yaitu matriks pergeseran yang berhubungan dengan pengamatan di himpunan A_m.

$\mathrm{E}\left(\mathbf{Y}\right)=\mathbf{XB}+\mathbf{Z}_{A_m}^{C}\mathbf{\Psi}$ adalah pergeseran rata-rata. Menurut Xu *et al*. (2005), untuk sampel besar:

Jika $\mathbf{\Lambda}^{\frac{2}{n}}=\dfrac{\left|n\hat{\mathbf{\Sigma}}_H\right|}{\left|n\hat{\mathbf{\Sigma}}_H+n\left(\hat{\mathbf{\Sigma}}-\hat{\mathbf{\Sigma}}_H\right)\right|}$, maka $-2\ln\mathbf{\Lambda}=-n\ln\dfrac{\left|n\hat{\mathbf{\Sigma}}_H\right|}{\left|n\hat{\mathbf{\Sigma}}_H+n\left(\hat{\mathbf{\Sigma}}-\hat{\mathbf{\Sigma}}_H\right)\right|}\sim\chi_q^2$

METODE ANALISIS

Estimasi parameter dan uji hipotesis pada model linear multivariat dengan Metode LDL, berdasarkan langkah-langkah sebagai berikut:

i. Mengumpulkan *m* pengamatan yang diduga *outlier*.
ii. Mendeteksi *outlier* pada pengamatan dalam model linear multivariat dengan asumsi $\mathrm{Vec}\left(\mathbf{E}\right)\sim N_p\left(\mathbf{0},\mathbf{\Sigma}\otimes\mathbf{I}_n\right)$ dimulai dengan membuat fungsi *likelihood* $L\left(\mathbf{B},\mathbf{\Sigma}\right)$, lalu menentukan $\hat{\mathbf{B}}$ dan $\hat{\mathbf{\Sigma}}$ dengan menggunakan Metode *Maximum Likelihood Estimate* disingkat Metode MLE untuk mendapatkan $L\left(\hat{\mathbf{B}},\hat{\mathbf{\Sigma}}\right)$.

iii. Memaksimumkan fungsi *likelihood* $L(\mathbf{B},\boldsymbol{\Sigma})$ dengan kendala jika ada *m* buah pengamatan adalah *outlier* menggunakan pengganda *Lagrange*, lalu membuat $\ln L(\mathbf{B},\boldsymbol{\Sigma})$ dan menentukan $\hat{\mathbf{B}}_{A_m}^C, \hat{\boldsymbol{\Sigma}}_{A_m}^C$ untuk mendapatkan $L\left(\hat{\mathbf{B}}_{A_m}^C, \hat{\boldsymbol{\Sigma}}_{A_m}^C\right)$.

iv. Menentukan LDL_{A_m}.

HASIL DAN PEMBAHASAN

Penelitian ini dibatasi hanya pada pendeteksian *outlier* pada model linear multivariat dengan Metode LDL. Pendeteksian *outlier* pada model linear multivariat dimulai dengan memisalkan ada *m* pengamatan yang diduga *outlier* (A_m) dari $\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \cdots, \underset{\sim}{Y}_h, \cdots, \underset{\sim}{Y}_q$, sehingga $\mathbf{Y}_{A_m}$ adalah himpunan $\mathbf{Y}$ dengan pengamatan yang ada *outlier* dan $\mathbf{Y}_{A_m}^C$ adalah himpunan $\mathbf{Y}$ dengan pengamatan tanpa *outlier*. Sebelumnya, jika dipunyai variabel independen sebanyak *p* dan variabel dependen sebanyak *q*, maka model linear multivariat (1) secara simultan dapat ditulis sebagai:

$$\mathbf{Y}_{nxq} = \left(\underset{\sim}{J}_{nx1} \ \vdots \ \mathbf{X}_{1_{nx(p+1)}}\right)\mathbf{B}_{(p+1)xq} + \mathbf{E}_{nxq} = \mathbf{X}_{nx(p+1)}\mathbf{B}_{(p+1)xq} + \mathbf{E}_{nxq} \tag{7}$$

Pendeteksian *outlier* pada model linear multivariat dengan asumsi $\mathrm{Vec}(\mathbf{E}) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}\otimes\mathbf{I}_n)$ seperti pada (3) dimulai dengan membuat fungsi *likelihood* untuk populasi $L(\mathbf{B},\boldsymbol{\Sigma})$, lalu menentukan $\hat{\mathbf{B}}$ dan $\hat{\boldsymbol{\Sigma}}$. Untuk mengestimasi parameter $\mathbf{B}$ pada (7) dengan fungsi *likelihood* seperti pada (4), maka estimasi (7) dengan Metode MLE dimulai dengan me-ln-kan (4), sehingga:

$$\ln L(\mathbf{B},\boldsymbol{\Sigma}) = -\frac{nq}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XB})^T(\mathbf{Y}-\mathbf{XB})\right) \tag{8}$$

Kemudian (8) diturunkan terhadap $\mathbf{B}$:

$$\frac{\partial \ln L(\mathbf{B},\boldsymbol{\Sigma})}{\partial \mathbf{B}} = -\frac{1}{2}\mathrm{tr}\left(-2\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)\mathbf{X}^T\right) = -\frac{1}{2}\mathrm{tr}\left(-2\boldsymbol{\Sigma}^{-1}\mathbf{X}^T\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)\right) = \mathbf{0},$$ diperoleh:

$$\hat{\mathbf{B}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Kemudian (8) diturunkan terhadap $\boldsymbol{\Sigma}$, maka:

$$\frac{\partial \ln L(\mathbf{B},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = -\frac{1}{2}\mathrm{tr}\left(n\hat{\boldsymbol{\Sigma}}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)^T\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)\right) = \mathbf{0}$$, diperoleh:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)^T\left(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\right)$$

Dari (7) diperoleh pula $\mathbf{Y} \sim N_p(\mathbf{XB}, \boldsymbol{\Sigma}\otimes\mathbf{I})$ dan dalam bentuk vektor ditulis:

$\mathrm{Vec}(\mathbf{Y}) = \left(\mathbf{I}_q \otimes \mathbf{X}\right)\mathrm{Vec}(\mathbf{B}) + \mathrm{Vec}(\mathbf{E})$, dengan $\mathrm{Vec}(\mathbf{E}) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}\otimes\mathbf{I}_n)$.

Untuk memaksimumkan fungsi *likelihood* $L(\mathbf{B},\mathbf{\Sigma})$ dengan kendala $\left(\mathrm{Vec}(\hat{\mathbf{B}})-\mathrm{Vec}(\mathbf{B})\right)^T$ (Var $\left(\mathrm{Vec}(\mathbf{B})\right)^{-1}$) $\left(\mathrm{Vec}(\hat{\mathbf{B}})-\mathrm{Vec}(\mathbf{B})\right)$ jika ada *m* buah variabel dependen adalah *outlier* dengan menggunakan pengganda *Lagrange*, dimulai dengan membuat $\ln L(\mathbf{B},\mathbf{\Sigma})$ dan menentukan $\hat{\mathbf{B}}_{A_m}^C$, $\hat{\mathbf{\Sigma}}_{A_m}^C$, sehingga didapat $L\left(\hat{\mathbf{B}}_{A_m}^C, \hat{\mathbf{\Sigma}}_{A_m}^C\right)$. Kemudian membuat

$$\mathrm{LDL}_{A_m} = 2\left(\ln L\left(\hat{\mathbf{B}},\hat{\mathbf{\Sigma}}\right)-\ln L\left(\hat{\mathbf{B}}_{A_m}^C,\hat{\mathbf{\Sigma}}_{A_m}^C\right)\right).$$

Fungsi *likelihood* untuk $\mathbf{B}$ yaitu (4), sehingga fungsi *likelihood* untuk $\hat{\mathbf{B}}_{A_m}^C$ adalah:

$$L\left(\mathbf{B}_{A_m}^C,\mathbf{\Sigma}_{A_m}^C\right)=(2\pi)^{-\frac{1}{2}(n-m)p}\left|\mathbf{\Sigma}_{A_m}^C\right|^{-\frac{(n-m)}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\mathbf{\Sigma}_{A_m}^C\right)^{-1}\left(\mathbf{Y}_{A_m}^C-\mathbf{X}_{A_m}^C\hat{\mathbf{B}}_{A_m}^C\right)^T\left(\mathbf{Y}_{A_m}^C-\mathbf{X}_{A_m}^C\hat{\mathbf{B}}_{A_m}^C\right)\right)\right)$$

Setelah didapatkan $\hat{\mathbf{B}}$ dan $\hat{\mathbf{\Sigma}}$ seperti di atas, selanjutnya ditentukan $\hat{\mathbf{B}}_{A_m}^C$ dan $\hat{\mathbf{\Sigma}}_{A_m}^C$.

Berdasarkan (5) dan (6) diperoleh:

$$\ln L\left(\left(\hat{\mathbf{B}}_{1A_m}^C,\hat{\mathbf{\Sigma}}_{1A_m}^C\right),\left\{\left(\hat{\mathbf{B}}_{2A_m}^C,\hat{\mathbf{\Sigma}}_{2A_m}^C\right)\right\}\left(\hat{\mathbf{B}}_{1A_m}^C,\hat{\mathbf{\Sigma}}_{1A_m}^C\right)\right)=\max_{(\boldsymbol{\beta}_2,\mathbf{\Sigma}_2)}\ln L\left(\hat{\mathbf{B}}_{1A_m}^C,\hat{\mathbf{\Sigma}}_{1A_m}^C\right),\left(\hat{\mathbf{B}}_2,\hat{\mathbf{\Sigma}}_2\right)$$

dengan $\left(\mathbf{B}_1,\mathbf{\Sigma}_1\right)=\left(\hat{\mathbf{B}}_{1A_m}^C,\hat{\mathbf{\Sigma}}_{1A_m}^C\right)$, maka $\hat{\mathbf{B}}_1=\hat{\mathbf{B}}_{A_m}^C$ dan $\hat{\mathbf{\Sigma}}_1=\hat{\mathbf{\Sigma}}_{A_m}^C$.

Berdasarkan (7), maka $\mathbf{Y}_{A_m}^C=\mathbf{X}_{A_m}^C\mathbf{B}_{A_m}^C+\mathbf{E}_{A_m}^C$ dan $\mathrm{Vec}\left(\mathbf{Y}_{A_m}^C\right)\sim N_p\left(\mathbf{0},\mathbf{I}_{n-m}\otimes\mathbf{\Sigma}\right)$, dan dari (2) diperoleh $\hat{\mathbf{B}}_{A_m}^C=\left(\left(\mathbf{X}_{A_m}^C\right)^T\mathbf{X}_{A_m}^C\right)^{-1}\left(\mathbf{X}_{A_m}^C\right)^T\mathbf{Y}_{A_m}^C$ dimana $\left(\mathbf{X}_{A_m}^C\right)^T\mathbf{X}_{A_m}^C=\mathbf{X}^T\mathbf{X}-\mathbf{X}_{A_m}^T\mathbf{X}_{A_m}$.

Dengan demikian dapat ditulis: $\left(\mathbf{X}_{A_m}^C\right)^T\mathbf{Y}_{A_m}^C=\mathbf{X}^T\mathbf{Y}-\mathbf{X}_{A_m}^T\mathbf{Y}_{A_m}$.

Estimasi dari $\mathbf{B}$ setelah *outlier* dikeluarkan $\left(\hat{\mathbf{B}}_{A_m}^C\right)$ adalah:

$$\hat{\mathbf{B}}_{A_m}^C=\left(\mathbf{X}^T\mathbf{X}-\mathbf{X}_{A_m}^T\mathbf{X}_{A_m}\right)^{-1}\left(\mathbf{X}^T\mathbf{Y}-\mathbf{X}_{A_m}^T\mathbf{Y}_{A_m}\right)=\hat{\mathbf{B}}-\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^T\left(\left(\mathbf{I}+\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{Q}_{A_m}\right)\mathbf{Y}_{A_m}-\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{X}_{A_m}^T\hat{\mathbf{B}}\right)$$

Sehingga:, $\hat{\mathbf{B}}_{A_m}^C=\hat{\mathbf{B}}-\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^T\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\mathbf{E}}_{A_m A_m}$

dimana $\mathbf{Q}_{A_m}=\mathbf{X}_{A_m}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^T$; $\hat{\mathbf{E}}_{A_m}=\mathbf{Y}_{A_m}-\mathbf{X}_{A_m}^T\hat{\mathbf{B}}$

diperoleh pula $\hat{\mathbf{B}}_{A_m}^C\sim N_p\left(\mathbf{B},\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\otimes\mathbf{\Sigma}+\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^T\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\right)\mathrm{Var}\left(\hat{\mathbf{E}}_{A_m}\right)\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^T\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\right)^{\mathrm{T}}\right)$

Permasalahan di atas bersifat umum, sehingga nilai optimal yang diperoleh bisa saja bukan nilai yang paling optimal. Oleh karena itu digunakan pengganda *Lagrange,* sehingga nilai optimal yang diperoleh diharapkan merupakan nilai yang paling optimal pada daerah kepercayaan yang telah ditentukan.

Untuk kasus spesial $\boldsymbol{\theta}_1$ dari $\boldsymbol{\theta}$, maka LD dapat dimodifikasi sebagai:

$$\mathrm{LDL}_{A_m}\left(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2\right)=2\left(\ln L\left(\hat{\boldsymbol{\theta}}\right)-\ln L\left(\hat{\boldsymbol{\theta}}_{1A_m}^C,\hat{\boldsymbol{\theta}}_2\left(\hat{\boldsymbol{\theta}}_{1A_m}^C\right)\right)\right)$$

dimana $\boldsymbol{\theta}=\left(\boldsymbol{\beta},\boldsymbol{\Sigma}\right)$, $\boldsymbol{\theta}_1=\boldsymbol{\beta},\ \ \boldsymbol{\theta}_2=\boldsymbol{\Sigma}$, sehingga $\hat{\boldsymbol{\theta}}=\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}}\right)$

$$\mathrm{LDL}_{A_m}\left(\mathbf{B}\,|\,\boldsymbol{\Sigma}\right)=2(\ln L\left(\hat{\mathbf{B}},\hat{\boldsymbol{\Sigma}}\right)-1\ln L\left(\hat{\mathrm{B}}_{A_m}^{C},\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)\right)$$

Fungsi *likelihood* dengan kendala sebanyak *m* pengamatan yang diduga *outlier* adalah:

$$L\left(\hat{\mathrm{B}}_{A_m}^{C},\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)\right)=\frac{1}{\left(2\pi\right)^{\frac{mn}{2}}\left|\boldsymbol{\Sigma}\left(\mathbf{B}_{A_m}^{C}\right)\right|^{\frac{n}{2}}}\,e^{-\frac{1}{2}tr\left\{\left(\boldsymbol{\Sigma}\left(\mathbf{B}_{A_m}^{C}\right)\right)^{-1}\left(\mathbf{Y}-\mathbf{X}_{A_m}^{C}\mathbf{B}_{A_m}^{C}\right)^{T}\left(\mathbf{Y}-\mathbf{X}_{A_m}^{C}\mathbf{B}_{A_m}^{C}\right)\right\}}$$

$$\hat{\mathbf{B}}_{A_m}^{C}=\hat{\mathbf{B}}-\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}_{A_m}^{T}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\mathbf{E}}_{A_m}$$

$$\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)=\frac{1}{n}\left(\mathbf{Y}-\mathbf{X}_{A_m}\hat{\mathbf{B}}_{A_m}\right)^{T}\left(\mathbf{Y}-\mathbf{X}_{A_m}\hat{\mathbf{B}}_{A_m}\right)=\frac{1}{n}\left\{\hat{\mathbf{E}}^T\hat{\mathbf{E}}+\left(\hat{\mathbf{B}}-\hat{\mathbf{B}}_{A_m}\right)^{T}\mathbf{X}^T\mathbf{X}\left(\hat{\mathbf{B}}-\hat{\mathbf{B}}_{A_m}\right)\right\}.$$

sehingga diperoleh

$$\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)=\hat{\boldsymbol{\Sigma}}+\frac{1}{n}\hat{\mathbf{E}}_{A_m}^{T}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{Q}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\mathbf{E}}_{A_m}\ \text{dan}$$

$$\mathrm{LDL}_{A_m}=\mathrm{LDL}_{A_m}\left(\mathbf{B}\,|\,\boldsymbol{\Sigma}\right)=2(\ln L\left(\hat{\mathbf{B}},\hat{\boldsymbol{\Sigma}}\right)-1\ln L\left(\hat{\mathrm{B}}_{A_m}^{C},\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)\right)\tag{9}$$

Dengan me-ln-kan dan membuang $\hat{\mathbf{B}}_{A_m}$ pada persamaan (9), maka diperoleh:

$$\mathrm{LDL}_{A_m}=2(-\frac{mn}{2}\ln\left(2\pi\right)-\frac{n}{2}\ln\left|\hat{\Sigma}\right|-\frac{n}{2}-\left(-\frac{mn}{2}\ln\left(2\pi\right)\right)-\frac{n}{2}\ln\left|\Sigma\left(\hat{\mathrm{B}}_{A_m}^{C}\right)-\frac{\mathrm{n}}{2}\right|)=n\ln\left\{\frac{\left|n\hat{\Sigma}+\frac{1}{n}\hat{\mathbf{E}}_{A_m}^{T}\mathbf{C}_{A_m}\hat{\mathbf{E}}_{A_m}\right|}{n\left|\hat{\Sigma}\right|}\right\}$$

dimana $\mathbf{C}_{A_m}=\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{Q}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}$.

Selanjutnya menentukan nilai eigen dari $\mathbf{C}_{A_m}$ didapat $\lambda_1,\lambda_2,\cdots,\lambda_m$.

Jika LDL_{A_m} didekati dengan $\mathrm{LDL}_A=\sum_{i=1}^{m}\lambda_i \underset{\sim}{Z}_i^2$, sehingga $\mathrm{LDL}_A\sim\chi_v^2$, dimana $v=m.q;\ \underset{\sim}{Z}_i^2\sim\chi_q^2$

maka $\mathrm{LDL}_A=\lambda.Z^2\sim\lambda.\chi_q^2$,

dimana $\lambda=\dfrac{h_{ii}}{\left(1-h_{ii}\right)^2}$; h_{ii} elemen diagonal ke-*i* dari $\mathbf{H}$ dimana $\mathbf{H}=\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}$.

Statistik uji yang dipakai untuk mendeteksi adanya *outlier* dalam model linear multivariat dengan Metode LDL adalah: $\mathrm{LDL}_{A_m}=n\ln\left(\dfrac{\left|n\hat{\Sigma}+\frac{1}{n}\hat{\mathbf{E}}_{A_m}^{T}\mathbf{C}_{A_m}\hat{\mathbf{E}}_{A_m}\right|}{n\left|\hat{\Sigma}\right|}\right)$

Penentuan *outlier* dilakukan dengan membandingkan LDL_{A_m} dengan χ_v^2 dimana:

H_0: A_m bukan *outlier* dan H_1: A_m adalah *outlier.*

Jika $\mathrm{LDL}_{A\,hitung} > \lambda.\chi^2_{tabel}$, maka tolak H_0, artinya pengamatan tersebut adalah *outlier*.

KESIMPULAN

Estimasi parameter dan uji hipotesis pada model linear multivariat dengan Metode LDL dapat digunakan untuk mendeteksi *outlier* pada model linear multivariat dengan statistik uji

$$\mathrm{LDL}_{A_m} = n\ln\left(\frac{\left|n\hat{\Sigma} + \frac{1}{n}\hat{\mathbf{E}}^T_{A_m}\mathbf{C}_{A_m}\hat{\mathbf{E}}_{A_m}\right|}{n\left|\hat{\Sigma}\right|}\right)$$

. Jika $\mathrm{LDL}_{A\,hitung} > \lambda.\chi^2_{tabel}$, maka pengamatan tersebut adalah *outlier*.

REFERENSI

Barnett, V. & Lewis, T. (1994). *Outliers in statistical data*. (3^{rd} ed). Great Britain: John Wiley.

Christensen, R. (1991). *Linear model for multivariate, time series, and spatial data*. New York: Springer-Verlag.

Cook, R.D. (1977). Detection of influential observation in linear regression. *Technometrics*, Februari 2000, 42, no. (1), 65-68.

Diaz-Garcia, J.A., Gonzalez-Farias, G. & Alvarado-Castro, V. (2007). Exact distributions for sensitivity analysis in linear regression. *Applied Mathematical Sciences*, 22;1083-1100.

Filzmoser, P. (2005). Identification of multivariate outliers: A performance study. *Austrian Journal of Statistics*. 2;127-138.

Hawkins, D.M. (1980). *Identifications of outliers*. New York: Chapman and Hall.

Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2007a). Outlier detection for the value of *Y* variable (residual outlier) in multivariate regression models. *Proceeding International Conference and Workshop on Basic and Applied Science*, Universitas Airlangga, Agustus 2007, Surabaya.

Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2007b). Pendeteksian o*utlier* pada model linear multivariat dengan pergeseran rata-rata. *Prosiding Seminar Nasional Statis-tika VIII*, Jurusan Statistika FMIPA ITS, November 2007, Surabaya.

Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2008). Prosedur pendeteksian o*utlier* pada model linear multivariat dengan metode l*ikelihood displacement statistic*. *Prosiding Seminar Nasional Matematika IV*, Jurusan Matematika FMIPA ITS, Desember 2008, Surabaya.

Makkulau, Linuwih, S., Purhadi, & Mashuri, M. (2009). Pendeteksian o*utlier* model linear multivariat pada produksi gula dan tetes tebu. *Prosiding Seminar Nasional Matematika*, Jurusan Matematika FMIPA Universitas Jember, Februari 2009, Jember.

Peña, D. & Guttman, I. (1993). Comparing probabilistic methods for outlier detection in linear models. *Biometrika*, Technometrics, August 2001. 3;603-610.

Peña, D. & Prieto, F.J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *American Statistical Association and the American Society for Quality*, *Technometrics*. 43, no (3).

Rencher, A.C. & Schaalje, G.B. (2008). *Linear models in Statistics*, (2nd ed). John Wiley & Sons: New York.

Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association*. 79, 871-880.

Rousseeuw, P.J. & Hubert, M. (1997). Recent developments in PROGRESS, dalam L1-Statistical procedure and related topics, edited by Y. Dodge, *Institute of Mathematical Statistics Lecture Notes and Monograph Series*, Hayward, California, Vol. 31, 201-214.

Srivastava, M.S. & von Rosen, D. (1998). Outliers in multivariate regression models, *Journal of Multivariate Analysis*. 65, 195-208.

Xu, J., Abraham, B., & Steiner, S.H. (2005). *Outlier detection methods in multivariate regression models.* Diambil tanggal 4 April 2007, dari http://www.bisrg.uwaterloo.ca/archive/RR-06-07.pdf.