Jurnal Teknik Industri, Vol. 12, No. 2, Desember 2010, 95-100
ISSN 1411-2485

Pendeteksian *Outlier* dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode *Likelihood Displacement Statistic-Lagrange*

Makkulau[1], Susanti Linuwih[2], Purhadi[3], Muhammad Mashuri[4]

Abstract: There are several problems in industrial process for example problems associated with product quality. In statistics, observation which is significantly different to the average is called outlier. The outlier can give significant influence to the result of modeling, which can affect the decision making. This research develops the outlier detection method using the Likelihood Displacement Statistic method, called Likelihood Displacement Statistic-Lagrange (LDL) method. The LDL method is applied to sugar and molasses production data of Djombang Baru Sugar Factory, Jombang, East Java. The result of this research shows that factors influenced the sugar and molasses production are sugar cane with the dirt less than 5%, sugar cane with the dirt between 5% to 7%, sugar cane with the dirt higher than 7%, and imbibition water.

Keywords: Likelihood Displacement Statistic-Lagrange, multivariate linear models, outlier, Lagrange multipliers.

Pendahuluan

Outlier adalah pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya. Secara umum *outlier* dapat dibedakan menjadi dua, yaitu *outlier* pada pengamatan dan *outlier* pada model linear. Berdasarkan banyaknya variabel yang dipertimbangkan *outlier* dapat dibedakan menjadi outlier pada pengamatan univariat atau multivariat dan *outlier* pada model linear univariat atau multivariat. *Outlier* pada model linear multivariat dapat dibagi atas tiga kategori, yaitu *outlier* terhadap leverage dan residual ataupun keduanya.

Pendeteksian *outlier* pada pengamatan univariat telah dilakukan oleh Hawkins [8] yang melihat suatu *outlier* tunggal pada sampel dengan melihat pengamatan yang menyimpang dari pengamatan lain. Barnet dan Lewis [2] melihat pengaruh pengamatan yang tidak konsisten dengan pengamatan lain pada proses multivariat, Peña dan Prieto [10] menyajikan prosedur pendeteksian *outlier* didasarkan pada analisis proyeksi dari titik-titik sampel, dan Filzmoser [7] mengidentifikasi *outlier* didasarkan pada jarak Mahalanobis.

[1] Fakultas Matematika dan Ilmu Pengetahuan Alam, Jurusan Matematika, Universitas Haluoleo. Kampus Bumi Tridharma Anduonohu, Kendari 93232. Email: makkulau@statistika.its.ac.id,
[1,2,3,4] Fakultas Matematika dan Ilmu Pengetahuan Alam, Jurusan Statistika, Institut Teknologi Sepuluh Nopember Surabaya, Kampus Keputih Sukolilo, Surabaya 60111.
Email: susanti_l@statistika.its.ac.id, purhadi@statistika.its.ac.id, m_mashuri@statistika.its.ac.id

Diterima 13 Juli 2010; revisi1 7 Oktober 2010; revisi2 30 Oktober 2010; Diterima untuk dipublikasikan 8 November 2010.

Pendeteksian *outlier* pada model linear telah dilakukan oleh Cook [5] dengan memperkenalkan jarak Cook's sebagai ukuran untuk mendeteksi pengamatan berpengaruh dalam model linear univariat. Ukuran jarak Cook's (D_i) dirumuskan sebagai kombinasi dari *studentized residual*, variansi *residual*, dan variansi nilai prediksi. Adnan *et al.* [1] menguraikan suatu prosedur untuk mengidentifikasi berbagai *outlier* dalam regresi linear univariat. Prosedur yang digunakan adalah Metode *Least of Trimmed of Squares* (Metode LTS) dan Metode Pengelompokkan Pautan Tunggal (*The Single Linkage Clustering*) untuk memperoleh pengamatan yang berpotensi *outlier*. Peña dan Guttman [9] mendeteksi kasus *outlier* pada model linear univariat menggunakan pendekatan Bayesian dengan dua metode. Metode pertama adalah membatasinya ke model *null* (sederhana) untuk pembangkitan data, selanjutnya dilakukan identifikasi *outlier* pada modelnya. Metode kedua mempertimbangkan suatu model alternatif menggunakan model pergeseran rata-rata dan model pergeseran variansi. Diaz-Garcia *et al.* [6] mengusulkan pendeteksian pengamatan berpengaruh pada model linear multivariat dengan modifikasi jarak Cook's. Uji formal untuk mendeteksi sebuah *outlier* pada model linear multivariat telah dikembangkan oleh Srivastava dan von Rosen [14]. Xu *et al.* [16] mengembangkan jarak Cook's univariat untuk mendeteksi *outlier* pada model linear multivariat.

Xu *et al.* [16] telah mengembangkan pendeteksian *outlier* dalam model linear multivariat dengan Metode *Likelihood Displacement Statistic* (LD), Metode *Likelihood Ratio Statistic for a Mean Shift*

(LR), dan Metode *Multivariate Leverage* yang menggunakan elemen dari *the average diagonal* Q_{A_m}. Dalam mengestimasi parameter Metode LD dan LR, Xu *et al.* [16] menggunakan Metode *Maximum Likelihood Estimate* (MLE) yang bersifat umum, sehingga nilai optimum yang diperoleh bisa merupakan bukan nilai yang paling optimum.

Penelitian ini bertujuan mengembangkan metode pendeteksian *outlier* pada pengamatan dalam model linear multivariat yang disebut *Likelihood Displacement Statistic-Lagrange* (LDL). Metode ini merupakan pengembangan dari Metode LD dengan menggunakan pengganda *Lagrange*. Pengganda *Lagrange* yang digunakan adalah daerah kepercayaan dari vektor parameter yang bermanfaat untuk mengoptimalkan daerah kepercayaan yang diperoleh. Daerah kepercayaan tersebut didapat secara bertahap dengan menghilangkan data yang dianggap *outlier* secara numerik dengan menggunakan program nonlinear.

Sebagai studi kasus digunakan pengamatan proses produksi gula di Pabrik Gula Djombang Baru Jombang, Provinsi Jawa Timur (PGDB Jombang). Berdasarkan penelitian Saputri [13] dapat diketahui bahwa pengamatan proses produksi gula tersebut terdapat *outlier*. Adapun variabel terikat yang dipertimbangkan adalah banyaknya gula yang dihasilkan (Y_1) dan berat tetes yang dihasilkan (Y_2). Produksi gula dan tetes tebu yang dihasilkan diperoleh dari berat tebu dengan mutu A (X_1), berat tebu dengan mutu B_1 (X_2), berat tebu dengan mutu B_2 (X_3), berat tebu dengan mutu B_3 (X_4), dan banyaknya air imbibisi (X_5).

Metode Penelitian

Model Linear Multivariat

Apabila $X_1, ..., X_p$ adalah variabel bebas dan $Y_1, ..., Y_q$ adalah variabel terikat, jika diambil n sampel, maka model linear multivariat yaitu model linear dengan variabel terikat lebih dari satu (Christensen [4]), ditulis

$Y_h = \mathbf{X}_h\beta_h + \varepsilon_h, h = 1, \dots, q,$ dimana:
$Y_h = (y_{1h}, \dots, y_{nh})^T$; $\beta_h = (\beta_{0h}, \beta_{1h}, \dots, \beta_{ph})^T$; dan $\varepsilon_h = (\varepsilon_{1h}, \dots, \varepsilon_{nh})^T$

Model linear multivariat dari q model linear secara simultan dapat ditulis:

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon} \quad (1)$$

dengan $\mathbf{Y}_{n\times q} = (Y_1, \dots, Y_q)$, $\mathbf{X}_{n\times(p\times 1)} = (\mathbf{1}, X_1, \dots, X_q)$, $\mathbf{B}_{(p\times 1)\times q} = (\beta_0, \beta_1, \dots, \beta_q)$, dan $\boldsymbol{\varepsilon}_{n\times q} = (\varepsilon_1, \dots, \varepsilon_q)$, sebagai matriks acak. Diasumsikan bahwa $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ dan $\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$, $\otimes$ adalah perkalian Kronecker dan $\boldsymbol{\Sigma} = \sigma_{ab}$; $a, b = 1, ..., q$, dengan $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$.

Menurut Christensen [4], estimasi parameter $\mathbf{B}$ dan $\sum$ pada (1) menggunakan Metode MLE, diperoleh:

$$\widehat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (2)$$

$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})^T(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})$ adalah estimator bias untuk $\boldsymbol{\Sigma}$. $\mathbf{S} = \frac{1}{n-rank(\mathbf{X})}(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})^T(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})$ adalah estimator tak bias untuk $\boldsymbol{\Sigma}$.

Prosedur uji hipotesis parameter pada model linear sebagai berikut: $H_0: \Lambda^T\mathbf{B} = \mathbf{0}$, terhadap $H_1: \Lambda^T\mathbf{B} \neq \mathbf{0}$, dimana $\Lambda^T = \mathbf{P}^T\mathbf{X}$, $\Lambda = \frac{|\boldsymbol{\varepsilon}|}{|\boldsymbol{\varepsilon}+\mathbf{H}|}$ adalah nilai statistik uji Wilk's Lambda dan $\mathbf{P}$ adalah matriks ortogonal (Christensen [4]). Statistik penguji adalah: $\mathbf{H} \equiv \mathbf{Y}^T\mathbf{M}_{MP}\mathbf{Y} = (\Lambda^T\widehat{\mathbf{B}})^T(\Lambda^T(\mathbf{X}^T\mathbf{X})^{-1}\widehat{\mathbf{B}})^{-1}(\Lambda^T\widehat{\mathbf{B}})$, dimana $\mathbf{M}_{MP} = \mathbf{MP}(\mathbf{P}^T\mathbf{MP})^{-1}\mathbf{P}^T\mathbf{M}$; $\mathbf{M} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ adalah operator proyeksi dari $\mathbf{X}$; $\mathbf{M}_{MP}$ adalah operator proyeksi dari $\mathbf{X}$. H_0 ditolak jika nilai maksimum dari *likelihood* di bawah H_0 lebih besar dari nilai maksimum keseluruhan.

Pendeteksian *Outlier* dengan Metode LD

Pendeteksian *outlier* pada model linear multivariat pertama kali dikembangkan oleh Xu *et al.* [16]. Xu *et al.* [16] mengembangkan tiga metode, yaitu LD, LR, dan *Multivariate Leverage*. Metode LD adalah suatu metode yang dikembangkan dengan cara menghilangkan pengamatan yang diduga *outlier* secara bertahap, Metode LR adalah suatu metode yang didasarkan pada pergeseran rata-rata dalam model, sedangkan Metode *Multivariate Leverage* dikembangkan dengan menggunakan elemen dari *the average diagonal* Q_{A_m} (ADQ) untuk mengukur keekstriman dari m pengukuran pada variabel bebas, dimana Q_{A_m} adalah matriks proyeksi.

Pendeteksian *outlier* 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*, 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*.

Fungsi *likelihood* menurut Christensen [4]; Rencher dan Schaalje [11]):

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

$(\mathbf{Y} - \mathbf{XB})))$ (3)

Definisi 1 (Christensen [4]). *LD dari pengamatan yang ada outlier untuk* $\mathbf{B}$ *dengan diberikan* $\boldsymbol{\Sigma}$ *adalah*:

$$\mathrm{LD}_{A_m}(\mathbf{B}|\boldsymbol{\Sigma}) = 2\left(\ln L\left(\widehat{\mathbf{B}},\ \widehat{\boldsymbol{\Sigma}}\right) - \ln L\left(\widehat{\mathbf{B}}_{A_m}^{C}, \widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}_{A_m}^{C}\right)\right)\right) \quad (4)$$

dimana $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}_{A_m}^{C}\right)$ *adalah MLE dari* $\boldsymbol{\Sigma}$ *ketika* $\mathbf{B}$ *diestimasi oleh* $\widehat{\mathbf{B}}_{A_m}^{C}$ ■

Optimasi Nonlinear

Permasalahan optimasi disebut nonlinear jika fungsi tujuan dan fungsi kendalanya mempunyai bentuk nonlinear pada salah satu atau keduanya (Bazaara *et al.* [3]), dalam bentuk:
Maksimumkan (minimumkan): $\mathrm{g(X)} = \mathrm{g}(X_1, \ldots, X_m)$
Kendala: $h_p(\mathrm{X}) = b_p,\ \mathrm{X} = 0, \mathrm{p} = 1, \ldots, \mathrm{m}$
dapat diselesaikan dengan pengganda *Lagrange*:

$$L = \mathrm{g(X)} - \sum_{p=1}^{m} \boldsymbol{\lambda}_p \left(h_p(\mathrm{X}) - b_p\right).$$

dimana b_p adalah kendala dan $\boldsymbol{\lambda}_p$ adalah nilai eigen.

Proses Produksi Gula

Tanaman Tebu (*Saccharum Officanarum L*) merupakan tanaman perkebunan semusim yang di dalam batangnya terdapat zat gula. Tebu termasuk keluarga rumput-rumputan (*graminae*) seperti halnya padi, glagah, dan lain-lain (Rizaldi [12]). Sebagai bahan baku gula, tebu dapat dikategorikan menjadi 4 macam penilaian mutu, yaitu: (1) Tebu Mutu A (tebu sangat bersih), (2) Tebu Mutu B_1 (tebu bersih, kotoran < 5%), (3) Tebu Mutu B_2 (tebu dengan kotoran 5% sampai dengan 7%), (4) Tebu Mutu B_3 (tebu sangat kotor, kotoran > 7%)

Proses produksi gula di PGDB Jombang dengan bahan baku tebu melalui proses dalam beberapa stasiun penggilingan, pemurnian, penguapan, masakan, putaran, dan penyelesaian. Tahapan proses produksi dimulai dengan memasukkan bahan baku gula dimasukkan ke stasiun penggilingan dan ditambahkan air, menghasilkan nirah merah dan ampas. Langkah berikutnya nirah merah dimasukkan ke stasiun pemurnian dengan ditambahkan kapur dan belerang, menghasilkan nirah encer dan blotong. Selanjutnya nirah encer dimasukkan ke stasiun penguapan, menghasilkan nira kental. Tahapan terakhir adalah memasukkan nirah kental ke stasiun masakan, lalu diteruskan ke stasiun pemutaran (stasiun penyelesaian), menghasilkan gula (sukrosa) dan tetes (molase). Hasil utama produksi PGDB Jombang adalah gula putih, sedangkan hasil lainnya di antaranya adalah ampas, blotong, dan tetes (Trubus [15]).

Pendeteksian Outlier pada Model Linear Multivariat dengan Metode LDL

Pendeteksian *outlier* pada model linear multivariat dengan Metode LDL dimulai dengan mengumpulkan m pengamatan yang diduga *outlier*. Kemudian menentukan $\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}$, $\widehat{\mathbf{B}}_{A_m}^{C}$ dan $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}_{A_m}^{C}\right)$ untuk mendapatkan $L\left(\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}\right)$ dan $L\left(\widehat{\mathbf{B}}_{A_m}^{C}, \widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}_{A_m}^{C}\right)\right)$. Selanjutnya menentukan LDL_{A_m} dan membandingkan dengan F_{tabel} untuk menentukan *outlier*.

Hasil dan Pembahasan

Bentuk umum model linear multivariat dengan variabel bebas sebanyak p dan variabel terikat sebanyak q dapat ditulis sebagai:

$$\begin{aligned}\mathbf{Y}_{n\times q} &= \left(J_{n\times 1} : \mathbf{X}_{1_{n\times(p\times 1)}}\right)\mathbf{B}_{(p\times 1)\times q} + \boldsymbol{\varepsilon}_{n\times q} \\ &= \mathbf{X}_{n\times(p\times 1)}\mathbf{B}_{(p\times 1)\times q} + \boldsymbol{\varepsilon}_{n\times q}\end{aligned} \quad (5)$$

dimana $\mathbf{Y}_{n\times q} = \left(Y_1, \ldots, Y_q\right)$, $J_{n\times 1} = (1, \ldots, 1)^T$, $\mathbf{X}_{n\times(p\times 1)} = \left(\mathbf{1}, X_1, \ldots, X_q\right)$, $\mathbf{B}_{(p\times 1)\times q} = \left(\beta_0, \beta_1, \ldots, \beta_q\right)$, dan $\boldsymbol{\varepsilon}_{n\times q} = \left(\varepsilon_1, \ldots, \varepsilon_q\right)$.

Pendeteksian *outlier* pada model linear multivariat dengan Metode LDL dimulai dengan memisalkan ada m pengamatan dari $Y_1, \ldots, Y_q$ yang diduga *outlier* disimbolkan $\mathbf{Y}_{A_m}$, sedangkan pengamatan yang tidak mengandung *outlier* disimbolkan $\mathbf{Y}_{A_m}^{C}$.

Untuk memudahkan analisis, Christensen [4] menggunakan vektorisasi matriks variabel terikat pada (5) dapat menjadi:

$\mathrm{Vec}(\mathbf{Y}) = \left(\mathbf{I}_q \otimes \mathbf{X}\right)\mathrm{Vec}(\mathbf{B}) + \mathrm{Vec}(\boldsymbol{\varepsilon})$, dimana:
$\mathrm{Vec}(\boldsymbol{\varepsilon}) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$ dan
$\mathrm{Vec}(\mathbf{Y}) \sim N_{nq}\left(\left(\mathbf{I}_q \otimes \mathbf{X}\right)\mathrm{Vec}(\mathbf{B}), \boldsymbol{\Sigma} \otimes \mathbf{I}_n\right)$ (6)

dengan menggunakan sifat hasil kali *kronecker* diperoleh:

$$\begin{aligned}\mathrm{Vec}(\widehat{\mathbf{B}}) &= \left(\mathbf{I}_q \otimes \mathbf{X}^T\right)\left(\mathbf{I}_q \otimes \mathbf{X}\right)^{-1}\left(\mathbf{I}_q \otimes \mathbf{X}^T\right)\mathrm{Vec}(\mathbf{Y}) \\ &= \left(\mathbf{I}_q \otimes (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)\mathrm{Vec}(\mathbf{Y})\end{aligned}$$

dimana $\mathrm{Vec}(\widehat{\mathbf{B}}) \sim N_{q(p+1)}\left(\mathrm{Vec}(\mathbf{B}), \boldsymbol{\Sigma} \otimes (\mathbf{X}^T\mathbf{X})^{-1}\right)$.

Pendeteksian *outlier* pada model (6) dimulai dengan membuat fungsi *likelihood* dari sini kemudian $\widehat{\mathbf{B}}$ dan $\widehat{\boldsymbol{\Sigma}}$ dapat diestimasi. Untuk mengestimasi parameter $\mathbf{B}$ pada (5) dengan fungsi *likelihood* seperti pada (3), maka estimasi (5) dengan Metode MLE dimulai dengan melogaritmanaturalkan (3), sehingga:

$$\ln L(\mathbf{B},\boldsymbol{\Sigma}) = -\frac{nq}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{Y}-\mathbf{XB})^{\mathrm{T}}(\mathbf{Y}-\mathbf{XB}) \quad (7)$$

Kondisi optimal dicapai bila memenuhi kondisi berikut ini: Logaritma natural dari fungsi *likelihood* (7) diturunkan terhadap **B** dan disamakan dengan nol, maka:

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

diperoleh $\widehat{\mathbf{B}} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{Y}$

Persamaan (7) kemudian diturunkan terhadap **Σ** dan disama-kan dengan nol, sehingga didapat:

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

Metode LDL untuk memaksimumkan (3) dengan kendala:

$$\left(\mathrm{Vec}(\widehat{\mathbf{B}}) - \mathrm{Vec}(\mathbf{B})\right)^{T}\left(\mathrm{Var}\left(\mathrm{Vec}(\widehat{\mathbf{B}})\right)\right)^{-1}\left(\mathrm{Vec}(\widehat{\mathbf{B}}) - \mathrm{Vec}(\mathbf{B})\right) \le F_{v_1,v_2,\alpha}$$

dimana $F_{v_1,v_2,\alpha}$ adalah nilai tabel dari Distribusi *F*; $v_1 = p,\ v_2 = n-p-1$. Metode ini dimulai dengan membuat $\ln L(\mathbf{B},\boldsymbol{\Sigma})$ seperti pada (7), kemudian menentukan $\widehat{\mathbf{B}}^{C}_{A_m}$ dan $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)$ untuk mendapatkan $\ln L\left(\widehat{\mathbf{B}}^{C}_{A_m}|\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)\right)$. Langkah selanjutnya adalah membuat LDL_{A_m}.

Estimasi dari **B** setelah *outlier* dikeluarkan $\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)$ adalah: $\widehat{\mathbf{B}}^{C}_{A_m} = \widehat{\mathbf{B}} - (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\boldsymbol{\varepsilon}}_{A_m}$ dimana: $\mathbf{Q}_{A_m} = \mathbf{X}_{A_m}(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}_{A_m}$; $\hat{\boldsymbol{\varepsilon}}_{A_m} = \mathbf{Y}_{A_m} - \mathbf{X}^{T}_{A_m}\widehat{\mathbf{B}}$; $\mathbf{I} + \left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{Q}_{A_m} = \left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}$ dan $\widehat{\mathbf{B}}^{C}_{A_m} \sim N_P\left(\mathbf{B}, (\mathbf{X}^{T}\mathbf{X})^{-1} \otimes +(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathrm{Var}\left(\hat{\boldsymbol{\varepsilon}}_{A_m}\right).\mathbb{A}\right)$, dimana $\mathbb{A} = \left(\mathbf{X}^{T}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\right)^{T}$.

Serupa dengan estimasi **Σ** setelah *outlier* dikeluarkan $\left(\widehat{\boldsymbol{\Sigma}}^{C}_{A_m}\right)$ adalah:

$$\widehat{\boldsymbol{\Sigma}}^{C}_{A_m} = \frac{n}{n-m}\boldsymbol{\Sigma} - \frac{\mathbf{1}}{n-m}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\boldsymbol{\varepsilon}}_{A_m}.$$

Permasalahan di atas bersifat umum karena estimasi parameter diperoleh dengan Metode MLE yang masih 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.

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

$L\left(\widehat{\mathbf{B}}^{C}_{A_m}|\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)\right) =$
$(2\pi)^{-\frac{mn}{2}}|\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\mathbb{B}\right)$, dimana:
$\mathbb{B} = \left(\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)\right)^{-1}\left(\mathbf{Y}-\mathbf{X}^{C}_{A_m}\widehat{\mathbf{B}}^{C}_{A_m}\right)^{T}\left(\mathbf{Y}-\mathbf{X}^{C}_{A_m}\widehat{\mathbf{B}}^{C}_{A_m}\right)$
$\widehat{\mathbf{B}}^{C}_{A_m} = \widehat{\mathbf{B}} - (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\hat{\boldsymbol{\varepsilon}}_{A_m}$ dan
$\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right) = \widehat{\boldsymbol{\Sigma}} + \frac{1}{n}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}\mathbf{C}_{A_m}\hat{\boldsymbol{\varepsilon}}_{A_m}$.
$\mathbf{C}_{A_m} = \left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}\mathbf{Q}_{A_m}\left(\mathbf{I}-\mathbf{Q}_{A_m}\right)^{-1}$

Setelah mendapatkan $\widehat{\mathbf{B}}^{C}_{A_m}$ dan $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)$ selanjutnya menentukan fungsi *Likelihood* untuk pengamatan yang ada *outlier* yaitu:

$$\mathrm{LDL}_{A_m} = \mathrm{LDL}_{A_m}(\mathbf{B}|\boldsymbol{\Sigma}) = 2\left(\ln L\left(\widehat{\mathbf{B}},\widehat{\boldsymbol{\Sigma}}\right) - \ln L\left(\widehat{\mathbf{B}}^{C}_{A_m},\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{B}}^{C}_{A_m}\right)\right)\right) \quad (8)$$

dengan melogaritmanaturalkan dan membuang $\widehat{\mathbf{B}}^{C}_{A_m}$ pada (8), maka:

$$\mathrm{LDL}_{A_m} = n\left(\ln\frac{|\widehat{\boldsymbol{\Sigma}}+\frac{1}{n}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}\mathbf{C}_{A_m}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}|}{|\widehat{\boldsymbol{\Sigma}}|}\right) = n\ln\left(\ln\frac{|n\widehat{\boldsymbol{\Sigma}}+\frac{1}{n}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}\mathbf{C}_{A_m}\hat{\boldsymbol{\varepsilon}}^{T}_{A_m}|}{|n\widehat{\boldsymbol{\Sigma}}|}\right) \quad (9)$$

Selanjutnya menentukan nilai eigen dari $\mathbf{C}_{A_m}$ didapat $\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_m$. LDL_{A_m} didekati dengan $\mathrm{LDL}_A = \sum_{p=1}^{m}\boldsymbol{\lambda}_p\tilde{Z}^{T}_{p}\tilde{Z}_p$, dimana $\boldsymbol{\lambda}_p$ adalah nilai eigen dari $\mathbf{C}_{A_m}$, $\tilde{Z}_p = Z_1,\ldots,Z_m$, sehingga: $\mathrm{LDL}_A \sim F_{v_1,v_2,\alpha}$.

Statistik uji yang dipakai untuk mendeteksi adanya *outlier* pada pengamatan dalam model linear multivariat dengan Metode LDL adalah LDL_{A_m} seperti pada (9).

Penentuan *outlier* dengan membandingkan LDL_{A_m} dan $F_{v_1,v_2,\alpha}$ dengan:
H_0: A_m bukan *outlier* dan H_1: A_m adalah *outlier*.
Jika $\mathrm{LDL}_{A_m} > \lambda.F_{tabel}$, maka tolak H_0, artinya pengamatan ke-*i* adalah *outlier*.

Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu

Pendeteksian *outlier* dengan Metode LDL menggunakan pengganda *Lagrange* berupa daerah kepercayaan dari vektor parameter yang mana pengamatan yang diduga *outlier* telah dihilangkan. Untuk mengetahui faktor-faktor yang mempengaruhi produksi gula dan tetes tebu di PGDB

Jombang dengan adanya kasus *outlier*, dimulai dengan menetapkan variabel bebas, yaitu X_1 adalah berat tebu dengan mutu A dalam kuintal, X_2 adalah berat tebu dengan mutu B_1 dalam kuintal, X_3 adalah berat tebu dengan mutu B_2 dalam kuintal, X_4 adalah berat tebu dengan mutu B_3 dalam kuintal, dan X_5 adalah banyaknya air imbibisi yang digunakan dalam kuintal. Setelah variabel bebas ditetapkan maka variabel terikat ditetapkan variabel terikat, yaitu Y_1 adalah banyaknya gula yang dihasilkan dalam kuintal dan Y_2 adalah berat tetes yang dihasilkan dalam kuintal. Selanjutnya mendeteksi adanya kasus multikolinearitas, menguji korelasi antara kedua variabel terikat, dan melakukan pemilihan model terbaik, serta menginterpretasikan model tersebut.

Pendeteksian *outlier* dengan Metode LDL dilakukan pada pengamatan selama periode giling kedua hingga kesembilan bulan Juni sampai dengan September 2007 di PGDB Jombang (Saputri [13]). Jika $n = 122$, $p = 5$, dan $q = 2$, maka dengan menggunakan Matlab 7.9, diperoleh:

$$\widehat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} -45{,}57 & -0{,}001 & -0{,}015 & -0{,}275 & 0{,}001 & 0{,}229 \\ 101{,}50 & -0{,}005 & -0{,}037 & -0{,}0096 & 0{,}058 & 0{,}172 \end{pmatrix}$$

dan

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n}\left(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\right)^T\left(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\right) = \begin{pmatrix} 0{,}6788 \times 10^{-4} & 0{,}3489 \times 10^{-4} \\ 0{,}3489 \times 10^{-4} & 2{,}9726 \times 10^{-4} \end{pmatrix}$$

Selanjutnya dicari nilai dari LDL_{A_m} dan nilai F_{tabel}, lalu membandingkannya. Dari pengamatan di PGDB Jombang dengan Metode LDL, dimana: $F_{v_1,v_2,\alpha} = 9{,}155$; $\alpha = 0{,}05$; dan $\lambda = 0{,}5$, diperoleh nilai-nilai LDL_{A_m}. Dari perhitungan diperoleh nilai $\text{LDL}_{A_3} = 11{,}483$; $\text{LDL}_{A_{71}} = 9{,}591$ dan $\text{LDL}_{A_{116}} = 11{,}349$; dan nilai ini lebih besar dari 9,155; berarti dari 122 pengamatan, ada 3 pengamatan yang dapat dianggap sebagai *outlier* yaitu pengamatan ke-3, ke-71, dan ke-116.

Selanjutnya melakukan pemilihan model terbaik berdasarkan Metode Eliminasi *Backward* Multivariat. Hasilnya terlihat pada Tabel 1.

Dalam Tabel 1 diperlihatkan bahwa variabel X_1 memberikan pengaruh yang tidak signifikan (P-*Value* = 0,893) terhadap variabel Y_1 dan Y_2, sedangkan variabel X_2, X_3, X_4, dan X_5 memberikan pengaruh yang signifikan terhadap variabel Y_1 dan Y_2 dengan P-*Value* berturut-turut adalah 0,004, 0,000, 0,000, dan 0,000.

Dari Tabel U diketahui nilai $U^{0,05}(2, 6, 116) = 0{,}835$. Didapat nilai $\Lambda < U^{0,05}(2, 6, 116)$, sehingga dapat disimpulkan bahwa model yang telah diduga sudah signifikan. Hasil pengujian untuk mengetahui pengaruh seluruh variabel bebas terhadap seluruh variabel terikat secara serentak terlihat pada Tabel 2.

Tabel 1. Pemilihan model terbaik berdasarkan metode eliminasi *backward* multivariat

Model	Variabel bebas	P-*Value*	Keterangan
1	X_1	0,893	X_1 dikeluarkan
	X_2	0,000	
	X_3	0,000	
	X_4	0,000	
	X_5	0,000	
2	X_2	0,004	Prosedur eliminasi *Backward* berhenti
	X_3	0,000	
	X_4	0,000	
	X_5	0,000	

Tabel 2. Uji untuk mengetahui pengaruh seluruh variabel bebas terhadap seluruh variabel terikat secara serentak

Variabel Bebas	Wilk's Lambda	P-*Value*
X_2	0,924	0,015
X_3	0,818	0,000
X_4	0,941	0,039
X_5	0,345	0,000

Dalam Tabel 2 diperlihatkan bahwa seluruh P-*Value* signifikan pada α=5%, sehingga disimpulkan bahwa berat tebu mutu B_1, B_2, B_3, dan air imbibisi mempengaruhi secara simultan terhadap produksi gula dan tetes tebu.

Selanjutnya berdasarkan ke-4 variabel yang signifikan tersebut diperoleh model linear dengan pengamatan tanpa *outlier* sebagai berikut:

$$\hat{Y}^C_{1A_m} = -36{,}0104 - 0{,}0143X_2 + 0{,}315X_3 - 0{,}0024X_4 + 0{,}2256X_5$$
$$\hat{Y}^C_{2A_m} = 133{,}990 - 0{,}026X_2 + 0{,}001X_3 - 0{,}031X_4 + 0{,}159X_5$$

dimana $\hat{Y}^C_{1A_m}$ adalah produksi gula dan $\hat{Y}^C_{2A_m}$ adalah tetes yang dihasilkan; X_1, X_2, X_3, X_4 dan X_5 berturut-turut adalah berat tebu mutu B_1, B_2, B_3, dan air imbibisi yang digunakan. Model pertama tersebut memiliki nilai koefisien determinasi sebesar 71,09%. Hal ini menunjukkan bahwa variabilitas produksi gula yang mampu dijelaskan oleh variabel berat tebu mutu B_1, B_2, B_3, dan air imbibisi sebesar 71,09%. Model kedua menunjukkan hubungan antara tetes tebu dengan variabel berat tebu mutu B_1, B_2, B_3, dan air imbibisi. Model kedua tersebut memiliki nilai koefisien determinasi sebesar 30,26%. Hal ini menunjukkan bahwa variabilitas tetes tebu yang mampu dijelaskan oleh variabel berat tebu mutu B_1, B_2, B_3, dan air imbibisi sebesar 30,26%. Perbedaan koefisien determinasi tersebut cukup

masuk akal mengingat produksi gula merupakan produksi utama, sedangkan tetes tebu merupakan produksi sampingan. Berdasarkan hasil analisis dengan uji serentak (multivariat) dapat disimpulkan bahwa faktor-faktor yang mempengaruhi produksi gula dan tetes tebu adalah tebu bersih dengan kotoran lebih kecil dari 5%, tebu dengan kotoran 5% sampai dengan 7%, tebu sangat kotor dengan kotoran lebih besar dari 7%, dan air imbibisi.

Simpulan

Metode LDL dapat mendeteksi adanya *outlier* pada pengamatan produksi gula dan tetes tebu pada PGDB Jombang. Berdasarkan 122 pengamatan yang berhasil dikumpulkan dapat diidentifikasi *outlier* pada pengamatan ke-3, ke-71, dan ke-116. Berdasarkan hasil analisis dapat disimpulkan bahwa faktor-faktor yang mempengaruhi produksi gula dan tetes tebu adalah tebu bersih dengan kotoran lebih kecil dari 5%, tebu dengan kotoran 5% sampai dengan 7%, tebu sangat kotor dengan kotoran lebih besar dari 7%, dan air imbibisi.

Ucapan Terima Kasih

Terima kasih kepada DP2M DIKTI yang telah mendanai Penelitian Disertasi Doktor dengan Nomor: 1109/D3/PL/2010, 4 Juni 2010.

Daftar Pustaka

1. Adnan R., Mohamad, M. N., and Setan, H., Multiple Outliers Detection Procedures in Linear Regression, *Matematika*, 1**,** 2003, pp. 29-45.
2. Barnett, V., and Lewis, T., *Outliers in Statistical Data*, 3rd ed., John Wiley, Great Britain, 1994.
3. Bazaara, M. S., Sherali, H. D., and Shetty, C. M., *Nonlinear Programming*: *Theory and Algorithms*, 2nd ed., John Wiley & Sons, New York, 1993.
4. Christensen, R., *Linear Model for Multivariate, Time Series, and Spatial Data*, Springer-Verlag, New York, 1991.
5. Cook, R. D., Detection of Influential Observation in Linear Regression, *Technometrics*, 42(1), 2000, pp. 65-68.
6. Diaz-Garcia, J. A., Gonzalez-Farias, G., and Alvarado-Castro, V., Exact Distributions for Sensitivity Analysis in Linear Regression, *Applied Mathematical Sciences*, 22, 2007, pp. 1083-1100.
7. Filzmoser, P., Identification of Multivariate Outliers: A Performance Study, *Austrian Journal of Statistics*, 2, 2005, pp. 127-138.
8. Hawkins, D. M., *Identifications of Outliers*, Chapman and Hall, New York, 1980.
9. Peña, D., and Guttman, I., Comparing Probabilistic Methods for Outlier Detection in Linear Models, *Biometrika*, Technometrics, August 3**,** 2001, pp. 603-610.
10. Peña, D., and Prieto, F. J., Multivariate Outlier Detection and Robust Covariance Matrix Estimation, *American Statistical Association and the American Society for Quality*, *Technometrics*, 43(3), 2001, pp.286-310.
11. Rencher, A. C., and Schaalje, G. B., *Linear Models in Statistics*, 2nd ed., John Wiley & Sons, New York, 2008.
12. Rizaldi, D., *Profil* Tebu, 2003, retrieved from http://www.kppbumn.depkeu.go.id on 11 September 2007.
13. Saputri, A. C., *Model Linear Multivariat pada Produksi Gula dan Tetes Tebu di P.G. Djombang Baru Jombang*, Tugas Akhir, Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, 2008.
14. Srivastava, M. S., and von Rosen, D., Outliers in Multivariate Regression Models, *Journal of Multivariate Analysis*, 65, 1998, pp. 195-208.
15. Trubus, *Metamorfosis Limbah Tetes* Tebu, 2007, retrieved from http://www.trubus-online.com on 11 September 2007.
16. Xu, J., Abraham, B., and Steiner, S. H., Outlier Detection Methods in Multivariate Regression Models, 2005, retrieved from http://www.bisrg.uwaterloo.ca/archive/RR-06-07.pdf on 04 April 2007.