ISBN: 978-979-16353-3-2

S-33

KOEFISIEN DETERMINASI REGRESI FUZZY SIMETRIS UNTUK PEMILIHAN MODEL TERBAIK

Iqbal Kharisudin

Jurusan Matematika FMIPA Universitas Negeri Semarang

Email: iqbal_kh@staff.unnes.ac.id

Abstrak: Dalam analisis regresi biasa, indeks yang digunakan untuk membandingkan dekomposisi dari total jumlah kuadrat variabel dependen tegas adalah koefisien determinasi atau nilai *adjusted*-nya. Dalam konteks regresi fuzzy dengan variabel dependen fuzzy, diperlukan suatu kriteria pemilihan variabel independen yang menghasilkan model terbaik. Dibangun indeks berdasarkan dekomposisi dari total jumlah kuadrat variabel dependen fuzzy. Pada makalah ini dikaji kriteria pemilihan sub model terbaik dengan menggunakan koefisien determinasi dan nilai *adjusted*-nya. Selanjutnya diberikan simulasi data yang menggambarkan keefektivan kriteria tersebut.

Kata kunci: variabel fuzzy simetris, dekomposisi jumlah kuadrat, koefisien determinasi.

Pendahuluan

Salah satu pertimbangan penting dalam model regresi parametrik adalah berkaitan dengan pemilihan matriks desain $\mathbb{X}$. Misalkan dipunyai observasi variabel independen kuantitatif sebanyak *k* dengan *n* unit statistik. Model regresi linear dinyatakan dengan matriks desain $\mathbb{X} = \{\mathrm{f}_{ij}\}$, dengan baris generik dinyatakan dengan

$$\mathbf{f}_i = [1, \mathrm{x}_{i1}, \mathrm{x}_{i2}, \ldots, \ldots, \mathrm{x}_{ik}], \quad i = 1, \ldots, n.$$

Vektor desain di atas dapat dimodifikasi dengan beberapa cara, di antaranya: dengan menambahkan suku tak linear, dengan mengurangi banyaknya suku (mengeliminasi efek dari beberapa variabel), dengan memperkenalkan beberapa kelas

 ISBN: 978-979-16353-3-2

fungsi f_1 yang lebih umum, dan sebagainya. Tentu saja untuk setiap modifikasi tersebut menghasilkan vektor koefisien regresi yang berbeda-beda. Selanjutnya didefinisikan model parametrik yang sesuai, misalkan M, kemudian akan dicari satu model "terbaik" berdasarkan suatu kriteria tertentu. Dalam analisis regresi klasik, indeks yang digunakan untuk membandingkan dekomposisi dari total jumlah kuadrat variabel dependen tegas adalah koefisien determinasi R^2 atau nilai *adjusted*-nya.

Dalam domain data fuzzy, terdapat suatu model regresi dengan variabel dependen fuzzy dan variabel independen tegas. Model ini dikembangkan oleh D'Urso dan Gastaldi [6], [7], Coppi dan D'Urso [2], D'Urso [5], D'Urso dan Giordani [8,9], Coppi dkk. [3], D'Urso dan Santoro [11], [10]. Metode yang digunakan untuk menemukan model linear adalah meminimalkan fungsi jarak fuzzy antara variabel terobservasi dan variabel output yang didefinisikan dalam suatu ruang metrik tertentu. Beberapa sifat dari model ini telah dibahas dalam Kharisudin dan Subanar [14], Kharisudin [12]. Solusi dari model ini merupakan generalisasi dari model regresi linear biasa (Kharisudin [13]). Dalam konteks regresi fuzzy dengan variabel dependen fuzzy, pada makalah ini dikaji indeks R^2 dan nilai *adjusted*-nya $\overline{R}^2$ berdasarkan dekomposisi dari total jumlah kuadrat variabel dependen fuzzy.

1. Motivasi Regresi dalam Konsep Fuzzy

Penalaran statistik dipengaruhi oleh beberapa jenis sumber ketidakpastian, seperti: keacakan, ketidaktepatan, ketidakjelasan, ketidaktahuan sebagian, dan sebagainya. Dalam konteks analisis regresi, terdapat beberapa aspek ketidakpastian yang sering diperhatikan, yaitu ketidakpastian berkaitan dengan: (1) hubungan antara variabel dependen dengan variabel independen, (2) hubungan antara data terobservasi dengan "semesta" data yang mungkin, dan (3) ketidakpastian nilai-nilai variabel terobservasi (Coppi [1]).

Konsep ketidakpastian dalam konteks analisis regresi telah ditangani dengan sangat memuaskan melalui metode-metode model linear biasa. Namun demikian

 ISBN: 978-979-16353-3-2

ketidakpastian berkaitan dengan observasi data belum dipertimbangkan. Data yang digunakan dalam pendekatan regresi biasa merupakan data tegas (*crisp*), sehingga apabila data yang dianalisis adalah data atau variabel fuzzy maka metode tersebut belum dapat menyelesaikan permasalahan regresi.

1.1. Bilangan Fuzzy dan Data Fuzzy. Bilangan fuzzy dapat didefinisikan berdasarkan konsep himpunan fuzzy, secara umum dengan menggunakan konsep himpunan fuzzy normal dan konveks maupun secara khusus dengan menggunakan fungsi keanggotaan. Bentuk khusus dari representasi bilangan fuzzy yang dapat meningkatkan efisiensi komputasional adalah bilangan fuzzy tipe LR. Bilangan fuzzy tipe LR paling banyak dan mudah digunakan untuk mendeskripsikan data.

Definisi 1.1.1. (Zimmermann [16]). Misalkan L (dan R) adalah fungsi berbentuk turun dari $\Re^+$ ke $[0,1]$ dengan $L(0) = 1$; $L(x) < 1$ untuk setiap $x > 0$; $L(x) > 0$ untuk setiap $x < 1$; $L(1) = 0$ atau ($L(x) > 0$ untuk setiap x dan $L(+\infty) = 0$). $\tilde{X}$ disebut bilangan fuzzy LR_1 jika untuk $m, \alpha > 0, \beta > 0$ dalam $\Re$, fungsi keanggotaan $\tilde{X}$ didefinisikan

$$\mu_{\tilde{X}}(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right), & \text{untuk } x \leq m, \\ R\left(\dfrac{x-m}{\beta}\right), & \text{untuk } x \geq m, \end{cases}$$

dengan m disebut nilai mean dari $\tilde{X}$ dan α dan β masing-masing disebut tepi (*spread*) kiri dan tepi kanan. Bilangan fuzzy $\tilde{X}$ dinyatakan dengan $\tilde{X} = (m, \alpha, \beta)_{LR}$.

Untuk merepresentasikan ketidakpastian dalam permasalahan kehidupan diperlukan data fuzzy. Pada dasarnya kita semua sering menggunakan data fuzzy, aturan samar, dan ketidaktepatan informasi untuk mengambil keputusan dalam situasi yang tidak menentu. Oleh karena itu model-model komputasional dari sistem real perlu juga bisa mengenali, merepresentasikan, memanipulasi, menginterpretasikan, dan menggunakan ketidakpastian (Bezdek (1993) dalam Coppi dkk. [4]). Kelas umum dari data fuzzy dinyatakan dengan (selanjutnya disebut dengan) data fuzzy LR. Data fuzzy LR dapat dinyatakan dengan matriks data fuzzy.

 ISBN: 978-979-16353-3-2

Definisi 1.1.2. (Coppi dkk. [4]). Matriks data fuzzy LR_2 (I unit observasi $\times$ J variabel (*fuzzy*)) didefinisikan sebagai $\mathbf{X} \equiv \left\{x_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR} : i = 1,2, \dots, I; j = 1,2, \dots, J\right\}$ dengan $x_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR}$ menyatakan variabel fuzzy terobservasi LR_2 ke-*j* pada unit observasi ke-*i*, m_{1ij} dan $m_{2ij}(m_{1ij} \leq m_{2ij})$ masing-masing menyatakan "pusat" kiri dan kanan, serta l_{ij} dan r_{ij} masingmasing menyatakan tepi kiri dan kanan, dengan fungsi keanggotaan dinyatakan sebagai:

$$\mu_{x_{ij}}(u_{ij}) = \begin{cases} L\left(\dfrac{m_{1ij} - u_{ij}}{l_{ij}}\right), & u_{ij} < m_{1ij}(l_{ij} > 0), \\ 1, & m_{1ij} \leq u_{ij} < m_{2ij}, \\ R\left(\dfrac{u_{ij} - m_{2ij}}{r_{ij}}\right), & u_{ij} \geq m_{2ij}(r_{ij} > 0), \end{cases}$$

dengan L (dan R) adalah fungsi berbentuk turun dari $\Re^+$ ke $[0,1]$ dengan $L(0) = 1$; $L(z_{ij}) < 1$ untuk setiap $z_{ij} > 0, \forall i, j$; $L(z_{ij}) > 0$ untuk setiap $z_{ij} < 1, \forall i, j$; $L(1) = 0$ (atau $L(z_{ij}) > 0$ untuk setiap z_{ij} dan $L(+\infty) = 0$).

Bilangan fuzzy $x_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR} : i = 1,2, \dots, I; j = 1,2, \dots, J$ berisi interval yang bergerak dari $m_{1ij} - l_{ij}$ ke $m_{2ij} + l_{ij}$ dan fungsi keanggotaan memberikan bobot-bobot yang berbeda terhadap masing-masing tepi kiri dan tepi kanan di sebelah kiri dan kanan dari pusat. Jika $m_{1ij} = m_{2ij}$, maka diperoleh bilangan fuzzy LR_1, dinotasikan dengan $x_{ij} = (m_{ij}, l_{ij}, r_{ij})_{LR}$, dengan m_{ij} menyatakan pusat, dan diperoleh matriks data fuzzy LR_1 $\mathbf{X} \equiv \left\{x_{ij} = (m_{ij}, l_{ij}, r_{ij})_{LR} : i = 1,2, \dots, I; j = 1,2, \dots, J\right\}$. Selanjutnya jika $l_{ij} = r_{ij}$, maka diperoleh bilangan fuzzy simetris LL_1, dinotasikan dengan $x_{ij} = (m_{ij}, l_{ij})_{LL}$, dan diperoleh matriks data fuzzy LL_1 simetris $\mathbf{X} \equiv \left\{x_{ij} = (m_{ij}, l_{ij})_{LL} : i = 1,2, \dots, I; j = 1,2, \dots, J\right\}$.

1.2. Jarak dan Ruang Metrik Bilangan Fuzzy. Misalkan $S_{LL}(\Re)$ menyatakan himpunan semua bilangan fuzzy simetris.

 ISBN: 978-979-16353-3-2

Definisi 1.2.1. (Yang dan Ko [23]). Misalkan $\tilde{X} = (m_x, \alpha_x, \beta_x)_{LR}$ dan $\tilde{Y} = (m_y, \alpha_y, \beta_y)_{LR}$ adalah bilangan fuzzy LR di dalam $S_{LL}(\Re)$. Jarak antara dua bilangan fuzzy $\tilde{X}$ dan $\tilde{Y}$ didefinisikan dengan

$$d_{LR}^2(\tilde{X}, \tilde{Y}) = (m_x - m_y)^2 + [(m_x - \lambda\alpha_x) - (m_y - \lambda\alpha_y)]^2 + [(m_x + \rho\beta_x) - (m_y + \rho\beta_y)]^2$$

dengan $\lambda = \int_0^1 L^{-1}(\omega)\, d\omega$ dan $\rho = \int_0^1 R^{-1}(\omega)\, d\omega$.

Nilai λ dan ρ menyatakan pengaruh bentuk dari fungsi keanggotaan terhadap jarak antara dua bilangan fuzzy. Nilai λ dan ρ memiliki peran ganda, yaitu berhubungan dengan variabilitas fungsi keanggotaan dan menurunkan penekanan pada tepi, karena pada kenyataannya bobot pusat lebih besar daripada bobot tepi. Selanjutnya pada definisi 1.2.1, jika kedua bilangan adalah bilangan fuzzy simetris ($\alpha_x = \beta_x$, $\alpha_y = \beta_y$, dan $\lambda = \rho = \int_0^1 L^{-1}(\omega)\, d\omega$), maka diperoleh jarak antara dua bilangan fuzzy simetris $\tilde{X}$ dan $\tilde{Y}$, yaitu:

$$d_{LL}^2(\tilde{X}, \tilde{Y}) = 3(m_x - m_y)^2 + 2\lambda^2(\alpha_x - \alpha_y)^2.$$

2. Regresi Fuzzy dengan Variabel Dependen Fuzzy Simetris

Ide dasar analisis regresi fuzzy yang dikembangkan adalah memodelkan pusat (*center*) dari variabel dependen fuzzy simetris dengan mengadopsi model regresi klasik, selanjutnya secara simultan memodelkan tepi variabel dependen fuzzy melalui regresi linear sederhana. Hubungan antara $\tilde{Y}$ (variabel dependen fuzzy simetris) dengan $X_1, \ldots, X_k$ (variabel independen tegas) dinyatakan dengan model ([11],[6]):

$$\boldsymbol{m} = \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$

$$\boldsymbol{m} - \boldsymbol{l} = \boldsymbol{\mu} - \boldsymbol{\delta} + \boldsymbol{\varepsilon}_L;$$

$$\boldsymbol{m} + \boldsymbol{l} = \boldsymbol{\mu} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_R;$$

 ISBN: 978-979-16353-3-2

$$\text{dengan } \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{a} \text{ dan } \boldsymbol{\delta} = \boldsymbol{\mu} b + \mathbf{1}d, \tag{2.0.1}$$

dengan $\mathbf{1}$ adalah vektor 1-an berukuran $(n \times 1)$, $\mathbf{X}$ matriks berukuran $(n \times (k+1))$ berisi vektor $\mathbf{1}$ dan variabel input $[\mathbf{x}_1, \dots, \mathbf{x}_k]$; $\mathbf{m}$, $\boldsymbol{\mu}$ masing-masing adalah vektor pusat terobservasi dan vektor pusat interpolasi berukuran $(n \times 1)$; $\mathbf{l}$, $\boldsymbol{\delta}$ masing-masing adalah vektor tepi terobservasi dan vektor tepi interpolasi berukuran $(n \times 1)$; $\mathbf{a}$ vektor koefisien/parameter regresi untuk $\mathbf{m}$ berukuran $((k+1) \times 1)$; *b* dan *d* koefisien/parameter regresi untuk model tepi; serta $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}_L$, dan $\boldsymbol{\varepsilon}_R$ adalah vektor residual.

Model regresi tersebut di bangun atas tiga model linear. Pertama interpolasi pusat dari observasi fuzzy, kedua dan ketiga adalah model untuk batas bawah (pusat – tepi) dan model untuk batas atas (pusat + tepi) yang dibangun berdasarkan model pertama. Dalam kasus variabel output adalah simetris, maka tepi kiri sama dengan tepi kanan, sehingga model kedua dan model ketiga mempunyai estimasi tepi yang sama.

2.1. Solusi Model. Berdasarkan kriteria kuadrat terkecil, parameter dari model (2.0.1) diestimasi dengan meminimalkan kuadrat jarak antara variabel dependen terobservasi $\tilde{Y}$ dengan nilai teoritis yang berkorespondensi $\tilde{Y}^*$ yang didefinisikan melalui model (2.0.1). Untuk tujuan ini, digunakan konsep jarak Euclid untuk bilangan fuzzy LR (seperti pada definisi 1.2.1), yaitu:

$$\Delta_{LL}^2 = 3(\mathbf{m} - \boldsymbol{\mu})'(\mathbf{m} - \boldsymbol{\mu}) + 2\lambda^2(\mathbf{l} - \boldsymbol{\delta})'(\mathbf{l} - \boldsymbol{\delta}). \tag{2.1.1}$$

Berdasarkan model (2.0.1), basis jarak (2.1.1) dapat ditulis menjadi

$$\begin{aligned}\Delta_{LL}^2 = 3(&\mathbf{m}'\mathbf{m} - 2\mathbf{m}'\mathbf{X}\mathbf{a} + \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a}) \\ &+ 2\lambda^2(\mathbf{l}'\mathbf{l} - 2\mathbf{l}'\mathbf{X}\mathbf{a}b - 2\mathbf{l}'\mathbf{1}d + \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a}b^2 + 2\mathbf{1}'\mathbf{X}\mathbf{a}bd + nd^2).\end{aligned}$$

Dengan demikian fungsi objektif kuadrat terkecil menjadi

$$\min_{\mathbf{a},b,d}[3(\mathbf{m}'\mathbf{m} - 2\mathbf{m}'\mathbf{X}\mathbf{a} + \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a}) + 2\lambda^2(\mathbf{l}'\mathbf{l} - 2\mathbf{l}'\mathbf{X}\mathbf{a}b - 2\mathbf{l}'\mathbf{1}d + \mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a}b^2 + 2\mathbf{1}'\mathbf{X}\mathbf{a}bd + nd^2)].$$

(2.1.2)

 ISBN: 978-979-16353-3-2

Untuk menentukan solusi masalah (2.1.2), dicari turunan parsial Δ^2_{LL} terhadap parameter **a**, *b*, dan *d* untuk nilai sama dengan nol, sehingga diperoleh sistem persamaan sebagai berikut.

$$\mathbf{a} = (3 + 2\lambda^2 b^2)^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[3\mathbf{m} + 2\lambda^2(\mathbf{l}b - \mathbf{1}bd)]. \quad (2.1.3)$$

$$b = (\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a})^{-1}\mathbf{a}'\mathbf{X}'(\mathbf{l} - \mathbf{1}d). \quad (2.1.4)$$

$$d = n^{-1}\mathbf{1}'(\mathbf{l} - \mathbf{X}\mathbf{a}b). \quad (2.1.5)$$

Solusi iteratif dari sistem persamaan di atas diperoleh dengan mengasumsikan bahwa **X** mempunyai rank penuh. Prosedur optimisasi dengan menggunakan algoritma iteratif berdasarkan persamaan (2.1.3) - (2.1.5) tidak dijamin diperolehnya minimum global, hanya minimum lokal saja. Dengan demikian, sangat disarankan untuk menggunakan algoitma iterasi dengan beberapa nilai awal untuk mengetahui stabilitas solusi ([11], [3]).

Selanjutnya dapat dilihat bahwa pada kasus variabel dependen tegas (*crisp*) yaitu $\mathbf{l} = \mathbf{0}$ dan $b = d = 0$ maka estimasi $\mathbf{a}$ yang termuat dalam (2.1.3) akan menghasilkan solusi kuadrat terkecil biasa yaitu $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{m}$. Dengan demikian model dan solusi pada sistem persamaan di atas merupakan generalisasi dari model regresi linear klasik, jika variabel dependen memuat ketidakpastian.

2.2. Sifat Solusi Model. Solusi kuadrat terkecil iteratif (2.1.3) s.d. (2.1.5) dari model (2.0.1) mempunyai beberapa sifat penting (penjelasan dan bukti dapat dilihat pada [3], [11], [14], [13]). Berkaitan dengan model (2.0.1), selanjutnya estimasi kuadrat terkecil iteratif dari $\boldsymbol{\mu}$ dan $\boldsymbol{\delta}$ masing-masing dinyatakan dengan $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\mathbf{a}}$ dan $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\mu}}\hat{b} + \mathbf{1}\hat{d}$.

Proposisi 2.2.1. Hubungan berikut berlaku:

$$(\mathbf{m} - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\mu}} = 0 \quad (2.2.1)$$

yaitu residual $(\mathbf{m} - \hat{\boldsymbol{\mu}})$ tidak berkorelasi dengan estimasi pusat $\hat{\boldsymbol{\mu}}$.

Proposisi 2.2.2. Jumlahan (dan juga mean) dari *n* residual pusat $(\mathbf{m} - \hat{\boldsymbol{\mu}})$ dan jumlahan (dan juga mean) dari *n* residual tepi $(\mathbf{l} - \hat{\boldsymbol{\delta}})$ adalah nol, yaitu

 ISBN: 978-979-16353-3-2

$$\mathbf{1}'(\mathbf{m}-\hat{\boldsymbol{\mu}})=0, \text{ dan } \mathbf{1}'(\mathbf{l}-\hat{\boldsymbol{\delta}})=0. \tag{2.2.2}$$

Proposisi 2.2.3. Hubungan berikut berlaku:

$$(\mathbf{l}-\hat{\boldsymbol{\delta}})'\hat{\boldsymbol{\delta}}=0. \tag{2.2.3}$$

yaitu residual $(\mathbf{l}-\hat{\boldsymbol{\delta}})$ tidak berkorelasi dengan estimasi tepi $\hat{\boldsymbol{\delta}}$.

3. Koefisien Determinasi Model Regresi

Dalam analisis regresi klasik, indeks yang digunakan untuk membandingkan dekomposisi dari total jumlah kuadrat variabel dependen tegas adalah koefisien determinasi R^2 atau nilai *adjusted*-nya. Dalam konteks regresi fuzzy dengan variabel dependen fuzzy, akan dibangun indeks R^2 berdasarkan dekomposisi dari total jumlah kuadrat variabel dependen fuzzy.

3.1. Dekomposisi Jumlah Kuadrat Variabel Dependen. Untuk mengukur kebaikan model regresi berganda dengan variabel dependen fuzzy, didefinisikan koefisien determinasi (R^2) dan nilai *adjusted*-nya $(\bar{\mathrm{R}}^2)$.

Definisi 3.1.1. Jumlah Kuadrat Total (JKT) dari variabel dependen fuzzy didefinisikan

$$\mathrm{JKT}=3\|\mathbf{m}-\mathbf{1}\bar{m}\|^2+2\lambda^2\left\|\mathbf{l}-\mathbf{1}\bar{l}\right\|^2$$

dengan $\bar{m}$ adalah nilai rata-rata dari observasi pusat m_i dan $\bar{l}$ adalah rata-rata dari observasi tepi $l_i\ (i=1,\dots,n)$.

Sebagai catatan bahwa definisi 3.1 di atas menyatakan penyimpangan total (*total deviance*) yaitu sama dengan definisi jarak Euclid antara pasangan vektor variabel fuzzy $\tilde{\mathbf{y}}=(\mathbf{m},\mathbf{l})$ dengan $\bar{\tilde{\mathbf{y}}}=(\mathbf{1}\bar{m},\mathbf{1}\bar{l})$.

Definisi 3.1.2. Jumlah Kuadrat Regresi (JKR) yaitu variasi yang dihitung oleh model, didefinisikan dengan

$$\text{JKR} = 3\|\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m}\|^2 + 2\lambda^2\left\|\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right\|^2$$

dengan $\bar{m}$ adalah nilai rata-rata dari observasi pusat m_i dan $\bar{l}$ adalah rata-rata dari observasi tepi l_i $(i = 1, \ldots, n)$.

Definisi 3.1.3. Jumlah Kuadrat Error (JKE) yaitu variasi yang tidak dihitung oleh model, didefinisikan dengan

$$\text{JKE} = 3\|\mathbf{m} - \hat{\boldsymbol{\mu}}\|^2 + 2\lambda^2\left\|\mathbf{l} - \hat{\boldsymbol{\delta}}\right\|^2.$$

Proposisi 3.1.4. Jumlah Kuadrat Total (JKT) sama dengan Jumlah Kuadrat Regresi (JKR) ditambah Jumlah Kuadrat Error (JKE), yaitu

$$\text{JKT} = \text{JKR} + \text{JKE}. \tag{3.34}$$

Bukti. Persamaan Jumlah Kuadrat Total (JKT) dapat ditulis menjadi

$$\begin{aligned}
\text{JKT} &= 3\|\mathbf{m} - \mathbf{1}\bar{m}\|^2 + 2\lambda^2\left\|\mathbf{l} - \mathbf{1}\bar{l}\right\|^2 \\
&= 3\|(\mathbf{m} - \hat{\boldsymbol{\mu}}) + (\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m})\|^2 + 2\lambda^2\left\|\left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right) + \left(\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right)\right\|^2 \\
&= 3\|\mathbf{m} - \hat{\boldsymbol{\mu}}\|^2 + 2\lambda^2\left\|\mathbf{l} - \hat{\boldsymbol{\delta}}\right\|^2 + 3\|\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m}\|^2 + 2\lambda^2\left\|\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right\|^2 \\
&\qquad + 3[2(\mathbf{m} - \hat{\boldsymbol{\mu}})'(\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m})] + 2\lambda^2\left[2\left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right)'\left(\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right)\right] \\
&= \text{JKE} + \text{JKR} + 3[2(\mathbf{m} - \hat{\boldsymbol{\mu}})'(\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m})] + 2\lambda^2\left[2\left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right)'\left(\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right)\right]
\end{aligned}$$

Dengan mensubstitusi (2.2.1), (2.2.2), dan (2.2.3) diperoleh

$$\begin{aligned}
&3[2(\mathbf{m} - \hat{\boldsymbol{\mu}})'(\hat{\boldsymbol{\mu}} - \mathbf{1}\bar{m})] + 2\lambda^2\left[2\left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right)'\left(\hat{\boldsymbol{\delta}} - \mathbf{1}\bar{l}\right)\right] \\
&= 6[(\mathbf{m} - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\mu}} - (\mathbf{m} - \hat{\boldsymbol{\mu}})'\mathbf{1}\bar{m}] + 4\lambda^2\left[\left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right)'\hat{\boldsymbol{\delta}} - \left(\mathbf{l} - \hat{\boldsymbol{\delta}}\right)'\mathbf{1}\bar{l}\right] \\
&= 6[0 - 0\bar{m}] + 4\lambda^2[0 - 0\bar{l}] = 0.
\end{aligned}$$

Jadi terbukti $\text{JKT} = \text{JKR} + \text{JKE}.$ ■

3.2. Koefisien Determinasi. Berdasarkan dekomposisi di atas, dapat dibangun suatu ukuran atau indeks *goodness of fit* dari model regresi fuzzy. Indeks *goodness of fit* menjelaskan variasi regresi (JKR) dibandingkan dengan variasi total. Selanjutnya indeks *goodness of fit* disebut koefisien determinasi (R^2) dan koefisien determinasi *adjusted* $(\bar{\mathrm{R}}^2)$ yang dinyatakan dalam definisi di bawah ini.

Definisi 3.2.1 (Koefisien Determinasi R^2). Koefisien determinasi model (2.0.1) didefinisikan dengan

$$\mathrm{R}^2 = 1 - \frac{\mathrm{JKE}}{\mathrm{JKT}} = 1 - \frac{3\|\mathbf{m}-\hat{\boldsymbol{\mu}}\|^2 + 2\lambda^2\|\mathbf{l}-\hat{\boldsymbol{\delta}}\|^2}{3\|\mathbf{m}-\mathbf{1}\bar{m}\|^2 + 2\lambda^2\|\mathbf{l}-\mathbf{1}\bar{l}\|^2}.$$

Definisi di atas menyatakan rasio antara variasi dari variabel dependen fuzzy simetris yang dihitung oleh model regresi dengan total variasi dari variabel dependen fuzzy simetris. Berdasarkan proposisi 3.1.4, dapat dilihat bahwa nilai R^2 berkisar pada interval [0,1]. $\mathrm{R}^2 = 0$ apabila model tidak menjelaskan apapun dari variabilitas variabel dependen fuzzy. $\mathrm{R}^2 = 1$ menyatakan kasus sempurna, dalam arti bahwa model menginterpolasi seluruh observasi secara sempurna, sehingga R^2 mewakili variabilitas dari variabel dependen fuzzy. Pada kenyataannya dua kejadian eksrim tersebut sangat jarang ditemuai pada penerapan nyata. Dengan demikian, sebagai konsekuensinya, model dikatakan memuaskan apabila nilai koefisien determinasi mendekati satu $(\mathrm{R}^2 \approx 1)$.

Pada definisi 3.2.1, tidak dimasukkan banyaknya variabel independen (*k*) dan banyaknya parameter dalam model (2.0.1). Selain alasan tersebut, karena R^2 adalah fungsi tak turun dari *k*, maka dengan menggunakan kriteria koefisien determinasi saja tidak mungkin mendapatkan model “terbaik” dalam kelas M. Oleh karena itu untuk menjawab masalah tersebut, didefinisikan koefisien determinasi *adjusted*.

Definisi 3.2.2 (Koefisien Determinasi Adjusted). Koefisien determinasi adjusted dari model (2.0.1) didefinisikan dengan

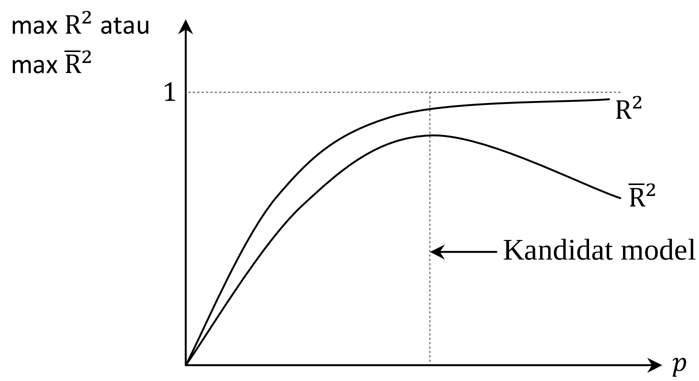
$$\bar{\mathrm{R}}^2 = 1 - \frac{\frac{\mathrm{JKE}}{n-(k+3)}}{\frac{\mathrm{JKT}}{n-1}} = 1 - (1-\mathrm{R}^2)\frac{n-1}{n-k-3}.$$

 ISBN: 978-979-16353-3-2

Indeks pada definisi 3.2.2 di atas berisi faktor penyesuaian yang didasarkan pada banyaknya parameter dalam model regresi. Nilai *k* menyatakan banyaknya variabel independen, $k + 1$ menyatakan banyaknya parameter regresi dari model pusat, dan dua parameter dari model tepi. Berbeda dengan R^2, nilai $\bar{R}^2$ tidak selalu naik, jika k bertambah. Dengan kata lain $\bar{R}^2$ adalah fungsi yang tak monoton naik. Fungsi $\bar{R}^2$ naik jika peningkatan variabilitas regresi lebih besar dari pada banyaknya variabel. Nilai maksimum $\bar{R}^2$ adalah 1 yang menggambarkan kasus sempurna, akan tetapi dapat pula bernilai negatif apabila model sangat buruk. Penyebut pada faktor penyesuaian koefisien determinasi *adjusted* yaitu $n - k - 3$ menyebabkan nilai $\bar{R}^2$ lebih besar dari penyebut nilai koefisien determinasi *adjusted* pada model klasik (*crisp*). Oleh karena itu, jika banyaknya observasi sedikit, maka dapat digunakan alternatif versi koefisien determinasi *adjusted* yang lain yaitu dengan hanya memperhatikan nilai *k* yang menyatakan banyaknya koefisien regresi dari model pusat saja (D'Usro dan Santoro [11]).

3.3. Kriteria Pemilihan Model.

Kriteria seleksi model berdasarkan atau . Berdasarkan prosedur ini, perlu dilakukan penetapan semua model yang mungkin, kemudian hasil yang ada di rangking untuk mempermudah identifikasi model "terbaik". Pertama dievaluasi model yang mungkin dengan banyaknya variabel independen untuk *p* = 1, 2, 3, dan seterusnya. Selanjutnya nilai-nilai dan ditabulasi atau diplot. Nilai selalu naik seiring bertambahnya variabel independen sedangkan suatu saat turun. Banyaknya variabel independen yang optimal dipilih jika mulai bergerak mendatar atau mencapai maksimum (lihat gambar 3.1 (D'Usro dan Santoro [11]).



Gambar 1. Plot maksimum R^2 dan maksimum $\bar{R}^2$ suatu model dengan p input

4. Penerapan dalam Pemilihan Model Terbaik

Pada bagian ini ditunjukkan hasil analisis regresi dengan data simulasi. Dilakukan simulasi dengan 6 variabel independen masing-masing sebanyak 25 unit sampel dan untuk setiap unit dibangkitkan variabel dependen fuzzy, seperti dirangkum dalam tabel 1. Pada kasus ini, diasumsikan *slope* fungsi keanggotaan dari variabel dependen fuzzy adalah fungsi keanggotaan segitiga simetris, yaitu diambil nilai $\lambda = 0.5$. Berdasarkan tabel 1 diharapkan variabel dependen fuzzy hanya bergantung pada tiga variabel independen yang pertama, yaitu X_1, X_2, dan X_3, sedangkan variabel independen yang lain tidak relevan.

Untuk menentukan banyaknya variabel independen yang sesuai (signifikan), diestimasi model regresi fuzzy untuk setiap nilai $p \leq k$. Untuk setiap p, diperhatikan kombinasi yang mungkin dengan p variabel independen dari 6 variabel independen. Pada tabel 2 didaftar nilai-nilai minimum JKE dan nilai maksimum R^2 dan $\bar{R}^2$ yang diperoleh untuk setiap model dengan p variabel independen.

Hasil analisis seperti terlihat pada tabel 2. Berdasarkan kriteria $\bar{R}^2$ diperleh nilai maksimum untuk $p = 4$ yaitu dengan variabel independen X_1, X_2, X_3 dan X_5, Di lain pihak, terlihat juga nilai R^2 menuju stasioner pada $p = 3$, lihat gambar 2. Hasil estimasi berdasarkan model dengan empat variabel independen X_1, X_2, X_3, dan X_5 adalah

$\hat{\mathbf{a}} = (1{:}9785 \quad 2{:}0303 \quad 0{:}9647 \quad 2{:}3308 \quad 0{:}$, $\hat{b} = 0{:}\,($, dan

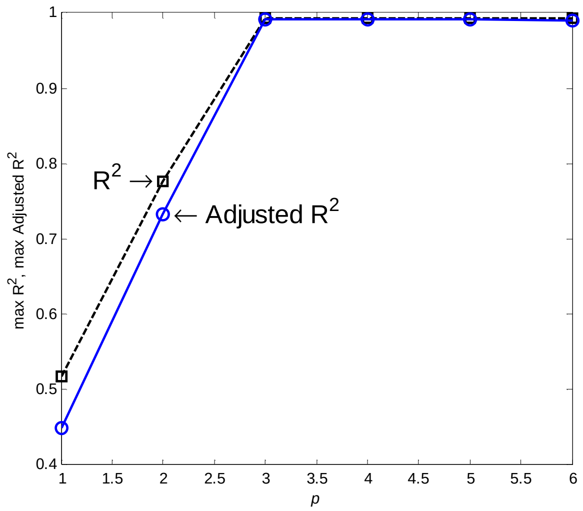
$\hat{d} = 1{:}\,9$.

Tabel 1. Pembangkitan data simulasi

Variabel independen tegas	Nilai X_1 diekstrak dari v.r. uniform pada interval [0,10] Nilai X_2 diekstrak dari v.r. uniform pada interval [30,55] Nilai X_3 diekstrak dari v.r. uniform pada interval [10,25] Nilai X_4 diekstrak dari v.r. uniform pada interval [25,50] Nilai X_5 diekstrak dari v.r. uniform pada interval [50,60] Nilai X_6 diekstrak dari v.r. uniform pada interval [0,350] Catatan v.r. : variabel random
Variabel dependen fuzzy	Nilai pusat dan tepi dari variabel dependen fuzzy $\tilde{Y}$ dibangkitkan dari: $\mathbf{m} = \mathbf{Xa} + N_{25\times1}(0,1)$ dan $\mathbf{l} = \mathbf{Xa}b + \mathbf{1}d + N_{25\times1}(0,1)$ dimana $\mathbf{X}$ adalah matriks berukuran 25×7 yang berisi vektor kolom $\mathbf{1}_{25\times1}$ dan nilai-nilai variabel independen tegas hasil simulasi; $N_{25\times1}(0,1)$ adalah vektor 25×1 variabel random normal dengan mean 0 dan standar deviasi 1. Parameter yang diharapkan dari model adalah $\mathbf{a} = (10 \quad 2 \quad -1 \quad 2.3 \quad 0 \quad 0 \quad 0)', b = 0.05$, dan $d = 2$.

Tabel 2. Kandidat model

p	min JKE	max R^2	max $\bar{\mathrm{R}}^2$	Variabel dependen
	JKT = 12214:0979			
1	5883.6364	0.518291	0.449476	X_3
2	2725.9308	0.776821	0.732185	X_2, X_3
3	83.6425	0.993152	0.991350	X_1, X_2, X_3
4	78.5394	0.993570	0.991426	X_1, X_2, X_3, X_5
5	78.2509	0.993593	0.990955	X_1, X_2, X_3, X_5, X_6
6	78.0891	0.993607	0.990410	$X_1, X_2, X_3, X_4, X_5, X_6$



Gambar 2. Plot dan

Daftar Pustaka

1. R. Coppi, *Management of uncertainty in statistical reasoning: The case of regression analysis*, International Journal of Approximate Reasoning **47** (2008), 284-305.
2. R. Coppi and P. D'Urso, *Regression analysis with fuzzy informational paradigm: a least squares approach using membership function information*, Int. J. Pure Appl. Math. **8** (2003), no. 3, 279-306.
3. R. Coppi, P. D'Urso, P. Giordani, and A. Santoro, *Least squares estimation of a linear regression model with LR fuzzy response*, Computational Statistics & Data Analysis **51** (2006), 267-286.
4. R. Coppi, P. Giordani, and P. D'Urso, *Component models for fuzzy data*, Psychometrika **71** (2006), no. 4, 733-761.
5. P. D'Urso, *Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data*, Computational Statistics & Data Analysis **42** (2003), 47-72.
6. P. D'Urso and T. Gastaldi, *A least-squares approach to fuzzy linear regression analysis*, Computational Statistics & Data Analysis **34** (2000), 427-440.
7. -------, *An "orderwise" polynomial regression procedure for fuzzy data*, Fuzzy Sets and Systems **130** (2002), 1-19.
8. P. D'Urso and P. Giordani, *Fitting of fuzzy linear regression models with multivariate response*, Int. Math. J. **3** (2003), no. 6, 655-664.
9. -------, *A weighted fuzzy c-means clustering model for fuzzy data*, Computational Statistics & Data Analysis **50** (2006), no. 6, 1496-1523.

10. P. D'Urso and A. Santoro, *Fuzzy clusterwise linear regression analysis with symmetrical fuzzy output variable*, Computational Statistics & Data Analysis **51** (2006), 287-313.
11. -------, *Goodness of fit and variable selection in the fuzzy multiple linear regression*, Fuzzy Sets and Systems **157** (2006), 2627-2647.
12. I. Kharisudin, *Bentuk fungsi keanggotaan pada model regresi dengan variabel dependen fuzzy simetris*, Prosiding Seminar Nasional Statistika IX, Jurusan Statistika FMIPA ITS Surabaya, 2009.
13. -------, *Generalisasi solusi kuadrat terkecil pada model regresi fuzzy simetris*, Prosiding Seminar Nasional V, Jurusan Matematika FMIPA UNNES Semarang, 2009.
14. I. Kharisudin and Subanar, *Fuzzy regression analysis with symmetrical fuzzy dependent variable*, submitted to The Proceeding of IICMA 2009, Yogyakarta, October 12-13, 2009.
15. M.-S. Yang and C.-H. Ko, *On a class of fuzzy c-numbers clustering procedures for fuzzy data*, Fuzzy Sets and Systems **84** (1996), 49-60.
16. H. J. Zimmermann, *Fuzzy set theory and its applications*, Kluwer Academic Publisher, Boston, 1991.