

---

## EFEKTIVITAS METODE *NEW STEPWISE* DALAM PEMILIHAN VARIABEL PADA MODEL REGRESI GANDA

**Thamrin Tayeb**

Fakultas Tarbiyah dan Keguruan UIN Alauddin Makassar  
Kampus II: Jalan Sultan Alauddin Nomor 36 Samata-Gowa  
Email: thamrin.tayeb@gmail.com

### **Abstrak:**

Metode *new stepwise* adalah suatu metode pemilihan variabel prediktor pada model regresi linear. Metode ini adalah suatu perluasan dari regresi komponen utama, dan terdiri atas pemilihan variabel prediktor asli secara iterasi pada waktu yang sama, himpunan bagian komponen utama dipilih secara berulang-ulang. Metode ini juga memiliki sifat-sifat dasar dari metode *stepwise*. Dengan demikian, kita akan mendapatkan gabungan yang terbaik dari metode pemilihan *stepwise* dan metode pemilihan komponen utama. Model yang diperoleh dengan menggunakan metode ini bercirikan PRESS yang bernilai rendah. Terapan dari metode ini tidak hanya pada model linear saja, tetapi dapat dikembangkan ke *generalized linear models*. Perbandingan dari kedua metode tersebut berdasarkan kriteria  $R^2$  dalam pemilihan variabel, diperoleh hasil nilai  $R^2$  yang hampir sama dari kedua model tersebut pada kasus data limbah padat, sehingga dengan memerhatikan banyaknya variabel prediktor yang masuk ke dalam model, maka dapat dikatakan bahwa metode *New stepwise* cenderung lebih baik dari pada Regresi komponen utama.

### **Abstract:**

New stepwise method is a method of selecting predictor variables in a linear regression model. This method is an extension of the principal component regression, and consists of the selection of the original predictor variables iteratively at the same time, a group of main subset component is selected repeatedly. This method has also the basic properties of the stepwise method. Thus we will get the best combination of stepwise selection and principal component selection methods. Model that is obtained by using this method characterizes a low-valued PRESS. The application of this method is not only for linear model, but also can be expanded to generalized linear models. The comparison of both methods are based on the  $R^2$  criteria in the variable selection, obtained  $R^2$  value results which are almost the same as those models in the case of solid waste of data, so having paid fully attention to the number of predictor variables entered into the models, it can be said that the new stepwise method tends to be better than the principal component regression.

### **Kata Kunci:**

Multikolinearitas; pemilihan variabel; komponen utama

**PEMILIHAN** variabel di dalam model regresi biasanya dilakukan jika terjadi amatan yang dikumpulkan dalam jumlah yang besar, sementara kita mengharapkan model yang terbaik dengan variabel prediktor yang sedikit. Pengurangan jumlah variabel prediktor adalah hal yang lazim dilakukan ketika di antara variabel prediktor saling berkorelasi.

---

Ada beberapa metode pemilihan variabel yang dikenal dan banyak digunakan orang. Metode tersebut dikelompokkan dalam dua kategori yaitu: (1) *Exhaustive search methods*, hal ini didasarkan pada metode *all possible subset of prediktors* dengan beberapa kriteria (yaitu  $R^2$ ,  $s^2$ ,  $C_p$  Mallow). (2) *Systematic selection algorithms*, seperti pada metode *forward*, *backward* dan *stepwise*.<sup>1</sup>

Di samping metode tersebut di atas cukup baik dalam memilih himpunan bagian variabel prediktor, juga mempunyai kekurangan. Pada *Exhaustive search methods*, prosedurnya membutuhkan komputasi yang banyak dan biayanya besar, serta tidak *feasible* dalam skala yang besar. Pada *Systematic selection algorithms*, sungguh pun komputasinya lebih efisien, kadang-kadang gagal menentukan himpunan variabel prediktor yang terbaik.

Regresi komponen utama dikenal sebagai suatu teknik yang baik dalam mengurangi variansi estimasi dalam analisis regresi, khususnya ketika ada multikolinearitas. Hanya saja kelemahannya adalah tidak ada kesesuaian pengurangan dalam jumlah variabel asli yang harus diukur.

*New stepwise* adalah menjadi solusi beberapa kelemahan dari beberapa metode pemilihan variabel pada model regresi linear ganda. Metode baru ini menggunakan regresi komponen utama untuk memilih variabel prediktor. Metode ini juga memiliki sifat-sifat dasar *Stepwise* dan hal ini didasarkan pada pengulangan kebalikan dari komponen utama ke variabel asli, sehingga diperoleh hasil yang merupakan gabungan dari kebaikan metode pemilihan *Stepwise* dan metode pemilihan variabel komponen utama.

Berdasarkan uraian tersebut di atas, maka penelitian ini difokuskan pada masalah sebagai berikut:

1. Bagaimana memilih variabel yang masuk dalam model regresi dengan menggunakan *New stepwise*.
2. Bagaimana memilih variabel yang masuk dalam model regresi dengan menggunakan regresi komponen utama.
3. Bagaimana perbandingan antara regresi komponen utama dengan metode *New stepwise* dalam pemilihan variabel pada model regresi ganda berdasarkan kriteria  $R^2$ .

## **METODE PEMILIHAN VARIABEL PADA MODEL REGRESI**

Pemilihan variabel terbaik di dalam membangun model, terkait dengan pemodelan regresi linear. Untuk mendapatkan model regresi yang terbaik, perlu memperhatikan kriteria pemilihan variabel. Selain itu, keberadaan kasus multikolinearitas sangat berpengaruh terhadap variabel yang terpilih ke dalam model.

### **Model Regresi Linear**

Asumsi yang mendasar dalam prosedur kuadrat terkecil dikemukakan oleh Myers<sup>2</sup> sebagai berikut:

- a.  $X$  adalah variabel prediktor yang merupakan variabel tetap;  
 b.  $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$ .

Dalam model regresi linear ganda, variabel prediktornya lebih dari satu. Jika variabel prediktornya berturut-turut adalah  $X_1, X_2, \dots, X_p$ , maka secara umum modelnya dapat digambarkan sebagai berikut:

$$y = S_0 + S_1 X_1 + S_2 X_2 + \dots + S_p X_p + v \quad (1)$$

Dengan  $p$  variabel prediktor dan  $S_j$  parameter,  $j = 0, 1, \dots, p$ , yang disebut sebagai koefisien regresi.

Bila dilakukan pengamatan  $(Y_i, x_{i1}, x_{i2}, \dots, x_{ip})$ , dimana  $i = 1, 2, \dots, n$ , maka model yang diperoleh adalah  $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$ .

Selanjutnya dapat ditulis dalam bentuk notasi matrik dan di

$$\tilde{Y} = \mathbf{X}\tilde{S} + \tilde{v} \quad (2)$$

$\tilde{Y}$  adalah vektor  $n \times 1$ ,  $\mathbf{X}$  adalah matrik  $n \times (p+1)$ ,  $\tilde{S}$  adalah vektor  $(p+1) \times 1$  dan  $\tilde{v}$  adalah vektor  $n \times 1$

Persamaan normalnya sebagai berikut:

$$\mathbf{X}^T \mathbf{X} \hat{S} = \mathbf{X}^T \tilde{Y} \quad (3)$$

Penaksir parameter  $\hat{S}$  sebagai berikut:

$$\hat{S} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{Y} \quad (4)$$

### Kriteria dalam Pemilihan Variabel

Ada beberapa kriteria dalam pemilihan variabel diantaranya adalah koefisien determinasi, *mean square error*,  $C_p$  Mallow. Namun dalam penelitian ini yang akan digunakan hanyalah koefisien determinasi.

#### Koefisien Determinasi

Koefisien determinasi dirumuskan oleh Montgomery and Peck<sup>3</sup> sebagai berikut:

$$R^2 = \frac{SS_R}{S_{YY}} = 1 - \frac{SS_E}{S_{YY}} \quad (5)$$

di mana  $R^2$  adalah koefisien determinasi,  $SS_R$  adalah simbol jumlah kuadrat regresi,  $SS_E$  adalah jumlah kuadrat sisaan dan  $S_{YY}$  adalah simbol dari jumlah kuadrat total.

Beberapa analis lebih menyukai penggunaan  $R^2$  *adjusted*.  $R^2$  *adjusted* tidak terpengaruh oleh derajat bebas dari  $SS_E$  maupun derajat bebas  $S_{YY}$ .<sup>4</sup> Penambahan variabel prediktor ke dalam model tidak selalu menyebabkan bertambahnya nilai  $R^2$  *adjusted*, sehingga model yang terbaik akan diperoleh dengan melihat kriteria  $R^2$  *adjusted* yang tertinggi. Selanjutnya dirumuskan sebagai berikut:

$$R^2_{adj} = 1 - \frac{SS_E / (n - p)}{S_{YY} / (n - 1)} = 1 - \left( \frac{n - 1}{n - p} \right) (1 - R^2) \quad (6)$$

$$R^2_{adj} = \text{koefisien determinasi } adjusted^5$$

#### Metode Regresi Komponen Utama

Model regresi komponen utama dirumuskan oleh Montgomery and Peck<sup>6</sup> sebagai berikut:

$$\tilde{y} = \mathbf{Z}\tilde{\mathbf{b}}_z + \tilde{v} \quad (7)$$

di mana  $\mathbf{Z}$  adalah variabel prediktor hasil transformasi,  $\tilde{\mathbf{b}}_z$  adalah koefisien regresi yang diperoleh dari penggunaan komponen utama,  $\tilde{\mathbf{b}}_z = \mathbf{U}^T \tilde{\mathbf{S}}$  dengan mengingat bahwa,  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  adalah matrik diagonal nilai eigen yang berasal dari matrik korelasi  $\mathbf{X}^T \mathbf{X}$  yang berukuran  $p \times p$  dan  $\mathbf{U} = [\tilde{U}_1, \tilde{U}_2, \dots, \tilde{U}_p]$  adalah matrik orthogonal vektor eigen yang bersesuaian dengan nilai eigen  $\lambda_1, \lambda_2, \dots, \lambda_p$ .

Sebelum membentuk model regresi komponen utama, pertama-tama yang dilakukan adalah mentransformasi variabel prediktor asli ke komponen utama yakni  $\mathbf{Z} = [\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_p]$ . Selanjutnya di rumuskan oleh Jackson<sup>7</sup> sebagai berikut:

$$\mathbf{Z}_{p \times p} = \mathbf{X}_{p \times p} \mathbf{U}_{p \times p} \quad (8)$$

di mana  $\mathbf{U}^T \mathbf{U} = \mathbf{I}_p$ ,  $\mathbf{Z}^T \mathbf{Z} = \Lambda$  dan  $\mathbf{X} = \mathbf{Z} \mathbf{U}^T$  adalah sifat-sifat dari komponen utama.<sup>8</sup>

Penaksir parameter dari regresi komponen utama digunakan rumusan oleh Jackson<sup>9</sup> sebagai berikut:

$$\hat{\tilde{\mathbf{b}}}_z = [\mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U}]^{-1} \mathbf{U}^T \mathbf{X}^T \tilde{\mathbf{y}} \quad (9)$$

### Metode New Stepwise

Metode ini adalah pengembangan dari regresi komponen utama sehingga digunakan orthogonal komponen utama sebagai lawan dari non-orthogonal variabel asli. Prosedur awal dari metode ini, pada dasarnya sama dengan prosedur yang dilakukan pada Regresi komponen utama, yaitu melakukan transformasi variabel prediktor asli ke komponen utama. Selanjutnya akan memeriksa komponen utama yang mempunyai nilai eigen yang terkecil, dan tidak diikutkan dalam prosedur pemilihan komponen utama. Dalam pemilihan komponen utama akan digunakan metode pemilihan variabel *Stepwise*.

Berangkat dari model regresi linear  $\tilde{Y} = \mathbf{X}\tilde{\mathbf{S}} + \tilde{v}$ , di mana  $\tilde{Y}$  adalah vektor variabel respon yang berukuran  $n \times 1$ ,  $\mathbf{X} = [\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p]$  adalah matrik variabel prediktor yang berukuran  $n \times p$ ,  $\tilde{\mathbf{S}}$  adalah vektor dari parameter yang akan ditaksir berukuran  $p \times 1$ , dan  $\tilde{v}$  adalah vektor sisaan yang acak berukuran  $n \times 1$  yang tidak saling berkorelasi dengan  $E(\varepsilon_i) = 0$  dan  $\text{Var}(\varepsilon_i) = \sigma^2$ .

Misalkan  $\mathbf{V} = [\tilde{V}_1, \tilde{V}_2, \dots, \tilde{V}_p]$  adalah matrik  $p \times p$  yang berisi vektor eigen yang telah dinormalisir dari matrik korelasi  $\mathbf{X}^T \mathbf{X}$  yang bersesuaian dengan nilai eigen  $\lambda_1, \lambda_2, \dots, \lambda_p$ .  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  adalah matrik diagonal dari nilai eigen.  $\tilde{\mathbf{W}} = [\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_p] = \mathbf{X} \mathbf{V}$  adalah matrik komponen utama dari  $\mathbf{X}$ . Penaksir parameter  $\tilde{\mathbf{u}}$  adalah

$$\hat{\tilde{\mathbf{u}}} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \tilde{\mathbf{Y}} \quad (10)$$

## METODE PENELITIAN

### Bahan dan Data

- Bahan diperoleh dari Jurnal dan Referensi yang terkait dengan materi bahasan.
- Data yang digunakan dalam penelitian ini adalah data Limbah Padat Pabrik Gula Asembagus Situbondo.

---

Pada penelitian ini variabel yang digunakan adalah:

Variabel respon:  $Y$  = Berat limbah padat/blotong (kuintal)

Variabel prediktor:

$X_1$  = Berat Kapur tohor (kuintal)

$X_2$  = Berat Sulfur (kg)

$X_3$  = Berat Flokulan (kg)

$X_4$  = Berat Tebu (kuintal)

$X_5$  = Berat Fosfat (kg)

### Metode Pemilihan Variabel Regresi Komponen Utama

Tahapan metode pemilihan variabel regresi komponen utama adalah:

- Dapatkan komponen utama  $[\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_p]$  dari  $[\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p]$
- Menentukan matrik korelasi  $\tilde{Z}^T \tilde{Z}$ .
- Berdasarkan matrik korelasi, kita tentukan nilai eigen yakni  $\lambda_1, \lambda_2, \dots, \lambda_r$  ( $r = p-1$ )
- Dari setiap nilai eigen, kita menentukan vektor eigen yaitu  $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_r$ .
- Menentukan model  $\tilde{W}_j = \tilde{x}_1 \tilde{Z}_1 + \tilde{x}_2 \tilde{Z}_2 + \dots + \tilde{x}_r \tilde{Z}_r$

### Metode Pemilihan Variabel *New Stepwise*

Tahapan metode pemilihan variabel *new stepwise* adalah sebagai berikut:

#### *Pemilihan variabel pertama*

- Dapatkan komponen utama  $[\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_p]$  dari  $[\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p]$
- Regresikan variabel respon  $y$  pada  $[\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_p]$  untuk mendapatkan model  $\tilde{y} = \mathbf{W}\tilde{u} + \tilde{v}$ .
- Ambil  $W_{(s)}$  yang menjadi himpunan bagian dari  $[\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_p]$  yang berisi komponen utama yang mempunyai koefisien regresi  $\hat{u}_j$  yang signifikan pada level  $\alpha$ .
- Jika  $W_{(s)}$  adalah himpunan kosong, maka proses pemilihan dihentikan dengan kesimpulan bahwa tidak ada variabel prediktor yang harus dimasukkan ke dalam model. Sebaliknya ambil  $SS_{E_j}$ ,  $j=1, \dots, p$ , simbol dari jumlah kuadrat sisaan ketika  $\tilde{X}_j$  diregresikan pada  $W_{(s)}$ . Variabel pertama yang terpilih adalah variabel yang mempunyai  $SS_{E_j}$  minimal.

### *Perbandingan Metode New Stepwise dengan Metode Regresi Komponen Utama*

Perbandingan kedua metode tersebut diatas dikelompokkan sebagai berikut:

#### *Berdasarkan Teori*

Langkah-langkahnya adalah sebagai berikut:

- Menentukan penaksir parameter dari metode regresi komponen utama dan metode *New Stepwise*.
- Menentukan  $E(\hat{S})$  dan  $\text{Var}(\hat{S})$  dari masing-masing metode regresi komponen utama dan metode *New Stepwise*.
- Membanding  $\text{MSE}(\hat{S})$  dari metode regresi komponen utama dan metode *New Stepwise*.

---

### Berdasarkan Kasus

Langkah-langkahnya adalah sebagai berikut:

- Menentukan nilai  $R^2$  dari masing-masing metode regresi komponen utama dan metode *New Stepwise*, pada kasus data Limbah Padat Pabrik Gula Asebagus Situbondo.
- Membandingkan nilai  $R^2$  dari masing-masing metode regresi komponen utama dan metode *New Stepwise* pada kasus data Limbah Padat Pabrik Gula Asebagus Situbondo.

### EFEKTIVITAS METODE NEW STEPWISE

#### Perbandingan MSE Penaksir Parameter dari Regresi Komponen Utama dengan Metode New Stepwise

$$\begin{aligned}\frac{\widehat{Var}(\hat{b}_Z)}{\widehat{Var}(\hat{u})} &= \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} \left( \frac{SS_E}{n-p} \right)}{(\mathbf{W}^T \mathbf{W})^{-1} \left( \frac{SS_E^*}{n-p-1} \right)} \\ &= \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} (SS_E)(n-p-1)}{(\mathbf{W}^T \mathbf{W})^{-1} (SS_E^*)(n-p)} \\ &= \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} (n-p-1) \left( \frac{SS_E}{SS_E^*} \right)}{(\mathbf{W}^T \mathbf{W})^{-1} (n-p)}\end{aligned}$$

dengan berdasar pada proses pemilihan variabel pada Regresi komponen utama dan metode *New stepwise*, maka dari kedua metode tersebut memperlihatkan bahwa nilai eigen yang masuk ke dalam model adalah sama, maka:

$$\begin{aligned}Z &= XU \\ &= X [ \tilde{U}_1, \tilde{U}_2, \dots, \tilde{U}_p ] \\ &= X \left[ \frac{x_1}{\sqrt{x_1^T x_1}}, \frac{x_2}{\sqrt{x_2^T x_2}}, \dots, \frac{x_p}{\sqrt{x_p^T x_p}} \right] \\ &= X V \\ Z &= W,\end{aligned}$$

Sehingga  $SS_E = SS_E^*$ , jadi:

$$= \frac{(\mathbf{Z}^T \mathbf{Z})^{-1} (n-p-1)}{(\mathbf{W}^T \mathbf{W})^{-1} (n-p)} \leq 1$$

Untuk  $n$  yang besar, maka  $\frac{\widehat{Var}(\hat{b}_Z)}{\widehat{Var}(\hat{u})} \cong 1$ , yang berarti bahwa variansi penaksir parameter dari Regresi komponen utama hampir sama dengan variansi penaksir parameter dari metode *New stepwise*.

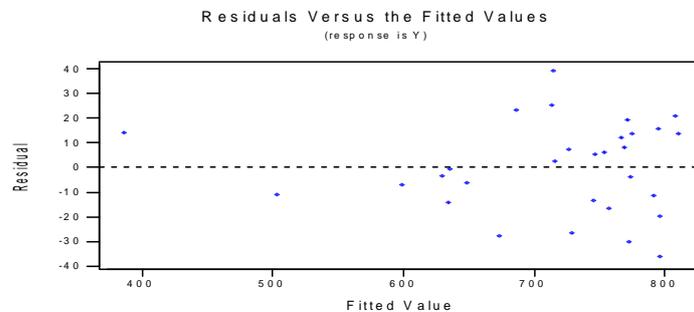
---

## Perbandingan Regresi Komponen Utama dengan Metode *New stepwise* dalam Pemilihan Variabel Berdasarkan Kriteria $R^2$ pada Kasus Data Limbah Padat

Sebelum melakukan analisis terhadap kedua metode yang digunakan dalam penelitian, maka terlebih dahulu melakukan pemeriksaan asumsi yang mendasar dari persamaan regresi. Adapun data yang digunakan dalam penelitian ini, telah digunakan oleh Mahmud, M.E. dan telah melakukan pengujian asumsi, karena model yang diperoleh penulis berbeda dengan model yang dihasilkan oleh Mahmud, M.E., maka penulis melakukan pengujian asumsi kembali sebagai mana berikut ini.

Asumsi dari model yang dihasilkan oleh Regresi komponen utama adalah sebagai berikut:

### *Asumsi Indentik*

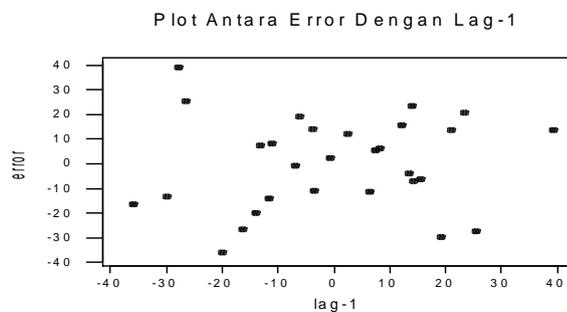


Grafik 1

Berdasarkan plot sisaan terhadap  $Y$  taksiran pada grafik 1, menunjukkan tidak ada pola tertentu atau pola yang menyerupai pita lurus yang mendatar.

Jika menggunakan Score test,  $S = SS_{reg}/2$ ,<sup>10</sup> dengan memerhatikan lampiran 18, hasil regresi  $u_i$  ( $u_i = e_i^2 / \sigma^2$ ,  $\sigma^2 = \sum e_i^2 / n$ ) pada seluruh variabel prediktor  $X$ , maka nilai score test  $S = 7,833$ . Karena nilai score test  $S$  lebih kecil dari  $F_{(0.05;5)}^2 = 11,070$ , maka dapat disimpulkan bahwa variansi sisaan ini adalah konstan.

### *Asumsi Independen*



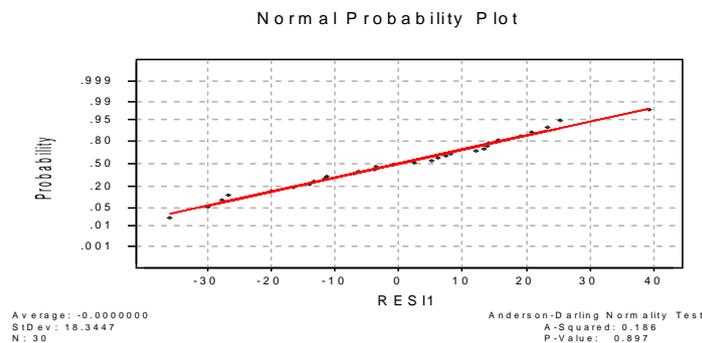
Grafik 2

---

Berdasarkan hasil plot antara sisaan dengan lag-1 pada Grafik 2, menunjukkan bahwa tidak ada pola atau kecenderungan sebaran antara sisaan dengan lag-1.

Berdasarkan Uji Durbin-Watson dengan hipotesis  $H_0: \rho_s=0$  dan  $H_1: \rho_s \neq 0$ , diperoleh Statistik Durbin-Watson ( $d$ ) sebesar  $1,82 > d_L=0,89$  dan  $(4-d) > 1,73$ , sehingga tidak cukup alasan untuk menolak  $H_0$ . Oleh karena itu, dapat disimpulkan bahwa tidak terjadi korelasi serial sisaan pada lag-1 atau dengan kata lain asumsi independen dapat terpenuhi.

### ***Asumsi Kenormalan***



Grafik 3

Berdasarkan plot normal sisaan Grafik 3, menunjukkan pola plot sisaan berupa garis lurus yang merupakan ciri bila sebaran data berdistribusi normal. Adapun hipotesis dari Anderson-Darling Normality Test adalah:

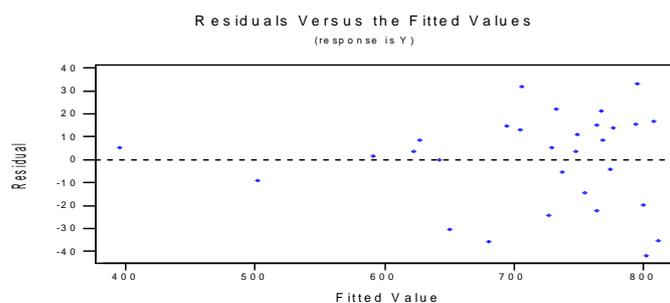
$H_0$ : Data sisaan berdistribusi normal

$H_1$ : Data sisaan tidak berdistribusi normal

Pada plot tersebut, juga ditunjukkan bahwa p-value sebesar 0,897 dari Anderson-Darling Normality Test yang cukup besar, menjadi alasan untuk tidak menolak  $H_0$ , sehingga dapat disimpulkan bahwa data sisaan berdistribusi normal.

Asumsi dari model yang dihasilkan oleh metode *New stepwise* adalah sebagai berikut:

### ***Asumsi Indentik***

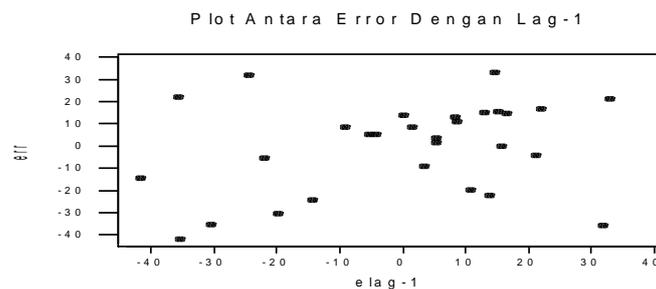


Grafik 4

Berdasarkan plot sisaan terhadap  $Y$  taksiran pada grafik 4, menunjukkan tidak ada pola tertentu atau pola yang menyerupai pita lurus yang mendatar.

Jika menggunakan Score test  $S$ , maka nilai score test  $S = 4,280$ . Karena nilai score test  $S$  lebih kecil dari  $F_{(0,05;3)}^2 = 7,81$ , maka dapat disimpulkan bahwa variansi sisaan ini konstan.

### *Asumsi Independen*

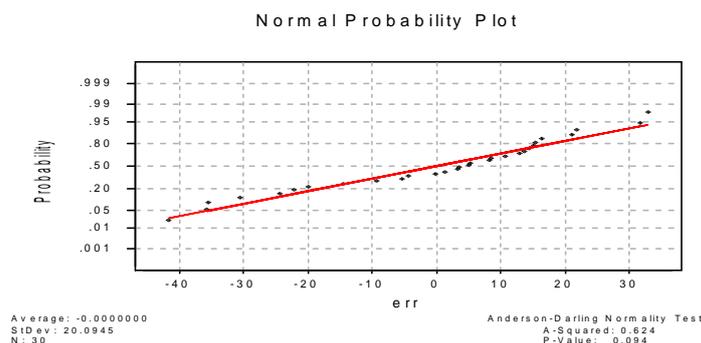


Grafik 5

Berdasarkan hasil plot antara sisaan dengan lag-1 pada grafik 5, menunjukkan bahwa tidak ada pola atau kecenderungan sebaran antara sisaan dengan lag-1.

Berdasarkan Uji Durbin-Watson dengan hipotesis  $H_0: \rho_s = 0$  dan  $H_1: \rho_s \neq 0$ , diperoleh Statistik Durbin-Watson ( $d$ ) sebesar  $1,45 > d_L = 1,12$  dan  $(4-d) > 1,54$ , sehingga tidak cukup alasan untuk menolak  $H_0$ . Hal ini dapat disimpulkan, bahwa tidak terjadi korelasi serial sisaan pada lag-1 atau dengan kata lain asumsi independen dapat terpenuhi.

### *Asumsi Kenormalan*



Grafik 6

Berdasarkan plot normal sisaan grafik 6, menunjukkan pola plot sisaan berupa garis lurus yang merupakan ciri bila sebaran data berdistribusi normal. Adapun hipotesis dari Anderson-Darling Normality Test adalah

$H_0$ : Data sisaan berdistribusi normal

$H_1$ : Data sisaan tidak berdistribusi normal

Pada plot tersebut, juga ditunjukkan bahwa p-value sebesar 0,094 dari Anderson-Darling Normality Test, masih lebih besar dari taraf  $\alpha = 0.05$ , sehingga menjadi alasan untuk tidak menolak  $H_0$ . Oleh karena itu, dapat disimpulkan bahwa data sisaan berdistribusi normal.

### *Multikolinearitas*

Setelah asumsi yang digunakan dalam regresi linear terpenuhi, maka langkah selanjutnya adalah memeriksa apakah ada korelasi yang kuat diantara variabel prediktor. Jika ada korelasi yang kuat di antara variabel prediktor akan muncul masalah multikolinearitas.

Tabel 1: Matrik korelasi Dari Data Limbah Padat

	X1	X2	X3	X4	X5
X1	1.00000				
X2	0.57060	1.00000			
X3	0.58863	0.27063	1.00000		
X4	0.89999	0.63858	0.54832	1.00000	
X5	0.59413	0.69166	0.34460	0.55650	1.00000

Merujuk kepada tabel 1 di atas, bahwa korelasi antara  $X_1$  dengan  $X_4$  cukup kuat yakni sebesar 0,8999, sedang yang berkorelasi lemah, antara  $X_3$  dengan  $X_5$  sebesar 0,345, korelasi  $X_2$  dan  $X_3$  sebesar 0,271, sedang korelasi antara variabel prediktor lainnya lebih besar dari 0,5. Hal ini memberikan indikasi bahwa data yang digunakan dapat menyebabkan munculnya masalah multikolinearitas.

Untuk mengatasi hal tersebut, dilakukan transformasi data dengan menggunakan analisis komponen utama, sehingga diperoleh variabel prediktor yang tidak saling berkorelasi, sebagaimana yang ditunjukkan pada tabel 1.

### **Regresi Komponen Utama**

Jika pemilihan komponen utama didasarkan pada nilai eigen yang kurang dari 0,01,<sup>11</sup> maka model yang dihasilkan oleh Regresi komponen utama adalah:

$$Y = 714 - 46.9 Z_1 + 22.0 Z_2 + 53.1 Z_3 + 8.59 Z_4 - 8.9 Z_5.$$

Tabel 2: Hasil Uji Koefisien Komponen Utama Regresi Secara Parsial

Predictor	Coef	StDev	T	P
<b>Constant</b>	713.800	3.682	193.88	0.000
<b>z1</b>	-46.880	2.055	-22.82	0.000
<b>z2</b>	21.980	4.069	5.40	0.000
<b>z3</b>	53.078	5.594	9.49	0.000
<b>z4</b>	8.591	6.876	<b>1.25</b>	<b>0.224</b>
<b>z5</b>	-8.91	12.72	<b>-0.70</b>	<b>0.490</b>

Berdasarkan model tersebut di atas dengan merujuk pada tabel 2, diperoleh hasil bahwa koefisien dari komponen utama  $Z_4$  dan  $Z_5$  tidak signifikan berdasarkan p-value dari kedua komponen utama tersebut yaitu masing-masing 0,224 dan 0,490. Oleh karena itu, mungkin ada model yang lebih baik dapat diperoleh dari model tersebut di atas.

Dengan berdasar pada jurnal yang ditulis oleh Mansfield, E.R. at. al. (1977), bahwa metode pengurangan jumlah variabel di dalam analisis regresi komponen utama dapat menggunakan *backward elimination procedure*.<sup>12</sup> Hal tersebut didasarkan pada pertambahan jumlah kuadrat erornya yang kecil bila menggunakan prosedur tersebut di atas, dan didahului dengan menghapus komponen utama yang mempunyai nilai eigen yang terkecil.

Tabel 3: Hasil Backward Elimination Pada Tabel Anova

ANOVA <sup>d</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	261009.4	5	52201.889	128.374	.000 <sup>a</sup>
	Residual	9759.353	24	406.640		
	Total	270768.8	29			
2	Regression	260809.7	4	65202.435	163.676	.000 <sup>b</sup>
	Residual	9959.061	25	398.362		
	Total	270768.8	29			
3	Regression	260175.0	3	86724.989	212.845	.000 <sup>c</sup>
	Residual	10593.834	26	407.455		
	Total	270768.8	29			

a. Predictors: (Constant), Z5, Z1, Z2, Z4, Z3

b. Predictors: (Constant), Z1, Z2, Z4, Z3

c. Predictors: (Constant), Z1, Z2, Z3

d. Dependent Variable: Y

Dari hasil prosedur *backward elimination*, pada table 3, maka model yang akan digunakan dalam regresi komponen utama adalah model-3 dengan nilai F terbesar yaitu 212,845, sebagaimana berikut ini:

$$Y = 714 - 46.9 Z_1 + 22.0 Z_2 + 53.1 Z_3$$

Untuk memberikan interpretasi pada model, maka model tersebut di atas komponen utamanya harus dikembalikan ke variabel asli. Sebagaimana model diperoleh sebagai berikut:

$$Y = 53.052 + 0.081415 X_1 + 0.090587 X_2 + 1.7337 X_3 + 0.018712 X_4 - 7.0272 X_5$$

Dari model tersebut di atas, menunjukkan tidak ada variabel prediktor yang dikeluarkan dari model, yang berarti tidak ada informasi yang hilang.

Model tersebut mampu menjelaskan 0,962849 keragaman yang ada di dalam data, yang dinyatakan oleh besarnya nilai  $R^2$ . Dengan memerhatikan besarnya  $R^2_{adj}$  adalah 0,955110, yang berarti keragaman yang mampu dijelaskan oleh model sebesar 95,5%. Adapun nilai MSE yang diperoleh dari model sebesar 419,134, artinya semakin kecil nilai dari MSE, maka taksiran yang dihasilkan oleh model akan semakin mendekati nilai parameter.

---

### Metode *New stepwise*

Berdasarkan prosedur pemilihan variabel dengan menggunakan metode *New stepwise*, diperoleh model sebagai berikut:

$$Y = 110 + 0.0657 X_1 - 0.0836 X_2 + 0.0221 X_4$$

Tabel 4 : Hasil Uji Parameter Regresi Secara Parsial

Predictor	Coef	StDev	T	P
Constant	110.16	28.80	3.82	0.001
x1	0.06567	0.01535	4.28	0.000
x2	-0.08364	0.06037	-1.39	0.178
x4	0.022068	0.003405	6.48	0.000

$$S = 21.22 \quad R\text{-Sq} = 95.7\% \quad R\text{-Sq(adj)} = 95.2\%$$

Berdasarkan tabel 4, keragaman yang mampu dijelaskan oleh model dengan 3 variabel prediktor yang ditunjukkan oleh  $R^2$  sebesar 95,7 %, sedangkan nilai  $R^2_{\text{adj}}$  sebesar 95,2 %. Hal ini menunjukkan bahwa model tersebut cukup baik digunakan untuk data limbah padat. Adapun nilai MSE yang dihasilkan dari model tersebut di atas sebesar 450.

Dari analisis kedua metode tersebut di atas diperoleh hasil bahwa keragaman yang dapat dijelaskan oleh model yang dihasilkan dari Regresi komponen utama, yang dinyatakan dengan nilai  $R^2$  sebesar 96,3%, hampir sama dengan keragaman yang dapat dijelaskan oleh model yang dihasilkan dengan metode *New stepwise* yang dinyatakan dengan nilai  $R^2$  sebesar 95,7%.

Demikian pula, jika dilihat dari keragaman yang ditunjukkan oleh model yang diperoleh dari Regresi komponen utama, dengan berdasar pada nilai  $R^2_{\text{adj}}$  sebesar 95,5%, juga hampir sama dengan keragaman yang ditunjukkan oleh model yang diperoleh dari metode *New stepwise* dengan nilai  $R^2_{\text{adj}}$  sebesar 95,2%. Jika memerhatikan nilai MSE dari Regresi komponen utama sebesar 419,134, perbedaannya tidak terlalu besar, jika dibandingkan dengan MSE yang diperoleh dari metode *New stepwise* sebesar 450.

Berdasarkan analisis di atas, dapat disimpulkan bahwa Regresi komponen utama tidak lebih baik dalam memilih variabel yang masuk ke dalam model, dibandingkan dengan menggunakan metode *New stepwise* pada kasus data limbah padat pabrik tebu yang digunakan dalam penelitian ini.

## SIMPULAN DAN IMPLIKASI PENELITIAN

### Simpulan

Berdasarkan analisis dan pembahasan dengan memerhatikan tujuan penelitian, maka dapat disimpulkan:

- Perbandingan  $MSE(\hat{b}_z)$  dari Regresi komponen utama dengan  $MSE(\hat{\phantom{b}})$  dari metode *New stepwise* diperoleh hasil, bahwa untuk  $n$  yang besar,

---

$V \hat{a}r(\hat{b}_z) / V \hat{a}r(\hat{u}) \cong 1$ , yang berarti bahwa variansi penaksir dari penaksir parameter Regresi komponen utama hampir sama dengan variansi penaksir dari penaksir parameter *New stepwise*.

- b. Dengan menggunakan Regresi komponen utama, semua variabel prediktor masuk dalam model, sebagaimana berikut ini:  $Y = 53.052 + 0.081415 X_1 + 0.090587 X_2 + 1.7337 X_3 + 0.018712 X_4 - 7.0272 X_5$

Sedangkan besarnya keragaman yang dapat dijelaskan oleh model tersebut di atas, yang dinyatakan dengan nilai  $R^2$  sebesar 96,3% dan nilai  $R^2_{adj}$  sebesar 95,5%. Sedangkan MSE yang dihasilkan model adalah sebesar 419,134.

Bila menggunakan metode *New stepwise*, maka variabel yang terpilih masuk model adalah  $X_1$  (berat kapur tohor),  $X_2$  (berat sulfur) dan  $X_4$  (berat tebu) yang dinyatakan sebagai berikut:  $Y = 110 + 0.0657 X_1 - 0.0836 X_2 + 0.0221 X_4$

Adapun keragaman yang dapat dijelaskan oleh model tersebut di atas, ditunjukkan oleh nilai  $R^2$  sebesar 95,7% dan nilai  $R^2_{adj}$  sebesar 95,2%. Sedangkan MSE yang diperoleh dari model tersebut sebesar 450.

Berdasarkan nilai  $R^2$  yang hampir sama dari model yang dihasilkan oleh kedua metode yang dibandingkan dalam penelitian ini, menunjukkan bahwa metode *New stepwise* cenderung lebih baik daripada Regresi komponen utama dalam memilih variabel dengan memerhatikan banyaknya variabel prediktor yang masuk ke dalam model.

### Implikasi Penelitian

Dari hasil penelitian ini, diharapkan:

- Data yang digunakan sebaiknya data yang memiliki variabel prediktor yang lebih banyak, sehingga kemungkinan terjadinya korelasi yang tinggi di antara variabel prediktor akan lebih besar. Pada kondisi tersebut, diharapkan metode *New stepwise* akan lebih nampak manfaatnya.
- Sungguhpun pada kasus data limbah padat ini, nilai  $R^2$  dan  $R^2_{adj}$  untuk metode *New stepwise* tidak lebih besar dari nilai  $R^2$  dan  $R^2_{adj}$  untuk metode Regresi komponen utama, namun model yang dihasilkan metode *New stepwise* memuat variabel prediktor lebih sedikit daripada model yang dihasilkan oleh Regresi komponen utama, sehingga dapat menjadi metode alternatif dalam membangun model, khususnya data yang mengandung variabel prediktor yang banyak, amatan yang lebih besar dan berkorelasi tinggi.
- Prosedur pemilihan variabel dengan metode *New stepwise*, sebaiknya dibuat dalam bentuk susunan program komputer (makro), sehingga kemungkinan terjadinya kesalahan prosedur dapat diminimalkan.

### CATATAN AKHIR

- S. Boneh and G. R. Mendieta, Variable selection in regression models using principal components, *Commun. Statist. - Theory & Meth.*, 23(1), 197-213, 1994.
- R. H. Myers, *Classical and Modern Regression with Applications*. Boston: PWS-KENT, 1989.

- 
3. D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis*. New York: John Wiley & Son, Inc., 1991.
  4. N. R. Draper and H. Smith, *Applied Regression Analysis*, 2<sup>nd</sup> ed. New York: John Wiley & Son, Inc., 1981.
  5. D. C. Montgomery and E. A. Peck, *op. cit.*
  6. *Ibid.*
  7. J. E. Jackson, *A User's Guide To Principal components*. New York: John Wiley & Son, Inc., 1991.
  8. S. Boneh and G. R. Mendieta, *op. cit.*
  9. J. E. Jackson, *op. cit.*
  10. S. Weisberg, *Applied Linear Regression, second edition*. York: John Wiley & Son, Inc., 1985.
  11. N. R. Draper & H. Smith, *op. cit.*
  12. E. R. Mansfield, J. T. Webster, & R. F. Gunst, An analytic variable selection technique for principal component analysis, *Appl. Statistl.*, 36, 34-40., 1977.

#### DAFTAR PUSTAKA

- Anton, H., *Aljabar Linear Elementer*, Penerbit Erlangga, Jakarta: Penerbit Erlangga, 1997.
- Boneh, S. and G. R. Mendieta, Variable selection in regression models using principal components, *Commun. Statist. – Theory & Meth.*, 23(1), 197-213., 1994
- Draper, N. R. and H. Smith, *Applied Regression Analysis*, 2<sup>nd</sup> ed. New York: John Wiley & Son, Inc., 1981.
- Jackson, J. E., *A User's Guide To Principal components*. New York: John Wiley & Son, Inc., 1991.
- Mahmud, M. E., *Perbandingan Pemilihan Variabel Independen Dalam Model Regresi Linear Dengan Pendekatan Cp Mallow, Tp Dan RTp Pada Data Limbah Padat Pabrik Gula Asempagus Situbondo*, TA, Surabaya, 1999.
- Mansfield, E. R., J. T. Webster, & R. F. Gunst, An analytic variable selection technique for principal component analysis, *Appl. Statistl.*, 36, 34-40., 1977.
- Mason, R. L. and R. F. Gunst, Selecting principal components in regression, *Stat. & Prob. Letters*, No.3, 299-301., 1985.
- Montgomery, D. C. and E. A. Peck, *Introduction to Linear Regression Analysis*. New York: John Wiley & Son, Inc., 1991.
- Myers, R. H., *Classical and Modern Regression with Applications*. Boston: PWS-KENT, 1989.
- Schott, R. J., *Matrix Analysis for Statistic*. New York: John Wiley & Son, Inc., 1997.
- Walpole, R. E. and R. H. Myers, *Probability and Statistics for Engineers and Scientists, fourth edition*. New York: Macmillan, 1989.
- Weisberg, S., *Applied Linear Regression, second edition*. York: John Wiley & Son, Inc., 1985.
- Wetherill, G. B., P. Duncombe, M. Kenward, J. Kollerstrom, S. R. Paul, and B. J. Vowden, *Regression Analysis With Application*, London: Chapman and Hall, 1986.