ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 8, Nomor 4, Tahun 2019, Halaman 439 - 450
Online di: http://ejournal3.undip.ac.id/index.php/gaussian

REGRESI KOMPONEN UTAMA *ROBUST S-ESTIMATOR* UNTUK ANALISIS PENGARUH JUMLAH PENGANGGURAN DI JAWA TENGAH

Jeffri Nelwin J. O. Siburian[1], Rita Rahmawati[2], Abdul Hoyyi [3]
[1,2,3] Departemen Statistika FSM Universitas Diponegoro
ritarahmawati@gmail.com

ABSTRACT

Robust principal component regression s-estimator is principal component regression that applies robust approach method at principal component analysis and s-estimator at principal component regression analysis. The aim of robust principal component regression s-estimator is to overcome multicollinearity problems in multiple linier regression Ordinary Least Square (OLS) and to overcome outlier problems in principal component regression so get the most effective model. Minimum Volume Ellipsoid (MVE) is one of the robust approach methods that applied when doing principal component analysis and S-Estimator is one of the estimation methods that applied when doing principal component regression analysis. The case in this study is the factors that influence the Number of Unemployment in Central Java in 2017. The model that provides the most effective result to handling multicolliniearity and ouliers in the case study Number of Unemployment in Central Java in 2017 is using robust principal component regression MVE-(S-Estimator) with Adjusted R^2 value of 0.9615 and RSE value of 0.4073.

Keywords: Robust Principal Component Regression S-Estimator, Multicollinearity, Outliers, Minimum Volume Ellipsoid (MVE), Number of Unemployment.

1. PENDAHULUAN

Pengangguran merupakan salah satu masalah di dalam suatu negara yang selalu menjadi pusat perhatian pemerintah. Menurut BPS (2018), pengangguran terbuka merupakan penduduk yang telah masuk dalam angkatan kerja tetapi tidak memiliki pekerjaan dan sedang mencari pekerjaan, mempersiapkan usaha, serta sudah memiliki pekerjaan tetapi belum mulai bekerja. Faktor-faktor yang diduga mempengaruhi jumlah pengangguran di Jawa Tengah adalah jumlah penduduk miskin, Indeks Pembangunan Manusia (IPM), rata-rata lama sekolah, Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), dan jumlah penduduk.

Studi awal penelitian ini menggunakan regresi linier berganda dengan estimasi parameter Metode Kuadrat Terkecil (MKT), dalam proses analisis data jumlah pengangguran beserta faktor-faktornya terjadi pelanggaran asumsi non multikolinieritas. Pelanggaran asumsi tersebut menyebabkan adanya hubungan linear antar beberapa atau semua variabel independen dari model regresi (Gujarati, 2004). Menurut Montgomery dan Peck (2012), masalah multikolinieritas dapat diatasi dengan menggunakan regresi komponen utama. Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama. Pada tahap analisis komponen utama ditemukan kembali masalah baru yaitu terdapat data yang mengandung pencilan, sehingga pada tahap ini akan digunakan analisis komponen utama *robust* dengan pendekatan *Minimum Volume Ellipsoid* (MVE).

Terdapatnya o*utlier* mengakibatkan metode *Ordinary Least Square* (OLS) memiliki estimasi koefisien garis regresi yang diperoleh menjadi tidak tepat, artinya nilai *estimator* dari parameter menjadi bias. Menurut Draper dan Smith (1992) menggunaan regresi *robust* memiliki kelebihan dibandingkan dengan metode OLS, yaitu kurang peka terhadap penyimpangan-penyimpangan yang sering terjadi dari asumsi ideal. Menurut Chen (2002), regresi *robust* terdiri dari 5 metode estimasi, yaitu *M-estimator*, *Least Median Square (LMS)-estimator*, *Least Trimmed Square (LTS)-estimator*, *S-estimator*, dan *MM-estimator*. Kelima metode tersebut memiliki kelebihan dan kelemahan masing-masing. Dilihat dari nilai *breakdown point*-nya, *S-estimator* merupakan estimasi *robust* yang memiliki nilai *breakdown point* paling tinggi. Proses analisis Regresi Komponen Utama akan menghasilkan beberapa model dan akan ditentukan model terbaik.

2. TINJAUAN PUSTAKA

2.1. Analisis Regresi Linier Berganda

Menurut Montgomery dan Peck (2012), Model regresi berganda dinyatakan dalam bentuk persamaan :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \tag{1}$$

Jika ditulis dalam bentuk matriks, persamaan (1) dapat ditulis sebagai berikut :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \; ; \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \; ; \; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \; ; \; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.1.1 Estimasi Parameter Model Regresi Linier Berganda

Metode yang digunakan untuk mengestimasi parameter model regresi adalah dengan meminimumkan jumlah kuadrat residual atau yang sering dikenal dengan Metode Kuadrat Terkecil (MKT) (Gujarati, 2004).

$$\boldsymbol{J} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\varepsilon_i}^{\boldsymbol{2}} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$$

Untuk mendapatkan estimator kuadrat terkecil ($\widehat{\boldsymbol{\beta}}$) yang meminimumkan $\boldsymbol{J}$ disyaratkan bahwa $\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0$. Turunan pertama dari $\boldsymbol{J}$ terhadap $\widehat{\boldsymbol{\beta}}$ adalah :

$$\frac{\partial J}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = -\boldsymbol{2X'y} + \boldsymbol{2X'X\widehat{\beta}} \tag{3}$$

karena, $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0$, maka

$$-\boldsymbol{2X'y} + \boldsymbol{2X'X\widehat{\beta}} = 0$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-\boldsymbol{1}} \boldsymbol{X'y} \tag{4}$$

Menurut Kurtner *et al*. (2005), dalam mengestimasi parameter model regresi perbedaan satuan akan menyebabkan koefisien tidak bisa dibandingkan. Oleh karena itu pada penelitian ini akan dilakukan pembakuan data dengan rumus berikut:

$$Y_i^* = \left(\frac{Y_i - \bar{Y}}{S_Y}\right) ; X_{ji}^* = \left(\frac{X_{ji} - \bar{X}_j}{S_j}\right) \tag{5}$$

dimana: $S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$; $S_j = \sqrt{\frac{\sum_{i=1}^{n}(X_{ji} - \bar{X}_j)^2}{n-1}}$

dengan:
- Y_i^* : variabel dependen dalam bentuk standar
- X_{ji}^* : variabel independen dalam bentuk standar
- $\bar{Y}$: rata-rata variabel dependen
- $\bar{X}_j$: rata-rata variabel independen

S_Y : standar deviasi variabel dependen
S_j : standar deviasi variabel independen

Sehingga diperoleh model regresi standar sebagai berikut:

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^*$$

2.1.2 Uji Hipotesis dalam Regresi Linier Berganda

a. Uji Signifikansi Kecocokan Model Regresi

Uji signifikansi regresi digunakan untuk menguji apakah ada hubungan linier antara variabel respon y dan variabel prediktor $x_1, x_2, \dots, x_k$ secara bersama-sama (Montgomery dan Runger, 2011). Berikut langkah-langkahnya :

1. Hipotesis
   $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$
   H_1 : terdapat $\beta_j \neq 0$, dengan j = 1,2,…,k
2. Statistik uji
   $$F_{hitung} = \frac{JKR/k}{JKS/(n-k-1)} = \frac{\text{MSR}}{\text{MSE}} \quad (6)$$
3. Kriteria uji
   H_0 ditolak jika $F_{hitung} > F_{(\alpha,k,n-k-1)}$ atau *p-value* $<$ α

b. Uji Koefisien Regresi secara Individual

Uji koefisien regresi secara individual digunakan untuk menguji ada tidaknya pengaruh yang signifikan antara masing – masing variabel prediktor terhadap model regresi linier, berikut langkah-langkah pengujiannya :

1. Hipotesis
   $H_0 : \beta_j = 0$
   $H_1 : \beta_j \neq 0$
2. Statistik uji
   $$t_{hitung} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \text{ ; dengan } Se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \quad (7)$$
3. Kriteria uji
   H_0 ditolak jika $|t_{hitung}| > t_{(\alpha/2, n-k-1)}$ atau *p-value* $<$ α

2.1.3 *Goodness Of Fit* (Ukuran Kecocokan Model)

a. Koefisien Determinasi yang disesuaikan (*Adjusted* R^2)

Menurut Montgomery dan Peck (2012), koefisien determinasi merupakan suatu nilai atau ukuran yang dapat digunakan untuk menilai kecocokan dari suatu model regresi.

$$R^2_{Adj,k} = 1 - \frac{JKS/(n-k-1)}{JKT/(n-1)} \quad (8)$$

b. RSE (Mean Square Error)

Residual Standard Error (RSE) adalah estimasi simpangan baku (*standard deviation*) dari ε .

$$RSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}} = \sqrt{MSE} \quad (9)$$

2.2 Multikolinieritas

Menurut Gujarati (2004), multikolinieritas merupakan adanya hubungan linear antara beberapa atau semua variabel independen dari model regresi (Gujarati, 2004). Ada beberapa cara yang digunakan untuk mendeteksi adanya multikolinieritas yaitu sebagai berikut:

a. Melihat nilai koefisien determinasi (R^2)
b. Melihat korelasi antar variabel independen
c. Melihat nilai *Variance Inflation Factor* (VIF)

Menurut Montgomery dan Peck (2012), selain menggunakan cara-cara tersebut multikolinieritas dapat juga ditangani dengan menggunakan analisis regresi *ridge* dan regresi komponen utama.

2.3 Pencilan

Menurut Neter *et al.* (1988), pencilan merupakan suatu pengamatan yang ekstrim. Salah satu metode yang dapat digunakan untuk mendeteksi pencilan adalah *DFFITS (Difference in Fit Standardized).* Rumus $DFFITS_i$ didefinisikan :

$$(DFFITS_i) = t_i \left(\frac{h_{ii}}{1-h_{ii}}\right)^{\frac{1}{2}} \quad (10)$$

dimana t_i adalah *studentized deleted residual* untuk kasus ke-*i* dengan rumus:

$$t_i = e_i\sqrt{\frac{n-p-1}{SS_E(1-h_{ii})-e_i{}^2}} \quad (11)$$

Suatu data disebut pencilan jika nilai $|DFFITS| > 2\sqrt{p/n}$ untuk gugus data yang berukuran besar, dengan $p = k+1$, dan n adalah banyaknya observasi (Montgomery dan Peck, 2012).

2.4 Analisis Komponen Utama

Menurut Johnson dan Winchern (2007), komponen utama merupakan kombinasi linier yang didasarkan pada skala pengukuran variabel *random* k (X_1, X_2, …, X_k). Kemudian dari variabel random tersebut akan dibentuk suatu variabel yang baru U yang akan menjadi komponen utama. Bentuk dari kombinasi linier tersebut sabagai berikut:

$$\begin{aligned} U_1 &= \mathbf{a_1}'\,\mathbf{X} = a_{11} X_1 + a_{12} X_2 + \ldots + a_{1k} X_k \\ U_2 &= \mathbf{a_2}'\,\mathbf{X} = a_{21} X_1 + a_{22} X_2 + \ldots + a_{2k} X_k \\ &\vdots \\ U_k &= \mathbf{a_k}'\,\mathbf{X} = a_{k1} X_1 + a_{k2} X_2 + \ldots + a_{kk} X_k \end{aligned} \quad (12)$$

dengan : Cov (U_j,U_l) = $\mathbf{a_j}'\,\boldsymbol{\Sigma}\,\mathbf{a_l}$ dan Var (U_j) = $\mathbf{a_j}'\,\boldsymbol{\Sigma}\,\mathbf{a_j}$ di mana j = 1, 2,…, k

proporsi dari total varians ke-*j* = $\frac{\lambda_j}{\lambda_1+\lambda_2+\cdots+\lambda_k} \times 100\%$; *j* = 1, 2, …, *k* (13)

Komponen utama juga dapat diperoleh dari variabel yang distandarkan, sehingga akan diperoleh komponen utama yang dibentuk sebagai kombinasi linier dari variabel yang distandarkan yaitu Z_1, Z_2, …, Z_k adalah:

$$\begin{aligned} U_1 &= \mathbf{a_1}'\,\mathbf{Z} = a_{11} Z_1 + a_{12} Z_2 + \ldots + a_{1k} Z_k \\ U_2 &= \mathbf{a_2}'\,\mathbf{Z} = a_{21} Z_1 + a_{22} Z_2 + \ldots + a_{2k} Z_k \\ &\vdots \\ U_k &= \mathbf{a_k}'\,\mathbf{Z} = a_{k1} Z_1 + a_{k2} Z_2 + \ldots + a_{kk} Z_k \end{aligned} \quad (14)$$

Ada beberapa cara untuk menentukan seberapa banyak jumlah komponen yang digunakan yaitu:

a. Memilih komponen-komponen utama yang mempunyai kumulatif proporsi keragaman total 75%.
b. Memilih komponen utama yang mempunyai *eigenvalue* lebih besar dari satu.
c. Melihat *scree plot* di mana *scree plot* tersebut merupakan plot antara λ_j dan j.

2.5 Regresi Komponen Utama

Setelah melakukan analisis komponen utama, selanjutnya melakukan regresi komponen utama dengan cara meregresikan skor komponen utama terhadap variabel dependen menggunakan *Ordinary Least Square* (OLS). Untuk bentuk persamaan regresinya sebagai berikut:

$$Y_i = \alpha_0 + \alpha_1 U_{1i} + \alpha_2 U_{2i} + \cdots + \alpha_m U_{mi} + \gamma_i \tag{15}$$

Jika ditulis dalam bentuk matriks, persamaan (15) dapat ditulis sebagai berikut :

$$\mathbf{y} = \boldsymbol{U\alpha} + \boldsymbol{\gamma} \tag{16}$$

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix};\ \mathbf{U} = \begin{bmatrix} 1 & U_{11} & U_{21} & \cdots & U_{m1} \\ 1 & U_{12} & U_{22} & \cdots & U_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & U_{1n} & U_{2n} & \dots & U_{mn} \end{bmatrix};\ \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix};\ \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}$$

Parameter $\hat{\boldsymbol{\alpha}}$ dapat diestimasi menggunakan metode *Ordinary Least Squares* (OLS) menggunakan rumus:

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{U'U})^{-1}(\boldsymbol{U'y}) \tag{17}$$

Seperti halnya regresi pada umumnya, akan dilakukan uji signifikansi pada model regresi komponen utama baik itu signifikansi pada model ataupun signifikansi secara individu. Untuk hipotesis yang digunakan dalam signifikansi model yaitu:

a. Uji Signifikansi Kecocokan Model Regresi

Uji signifikansi regresi digunakan untuk menguji apakah ada hubungan linier antara variabel respon *y* dan variabel prediktor $x_1, x_2, \dots, x_k$ secara bersama-sama (Montgomery dan Runger, 2011). Berikut langkah-langkahnya :

1. Hipotesis
   $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m$
   H_1 : terdapat $\alpha_j \neq 0$, dengan j = 1,2,…,m
2. Statistik uji
   $$F_{hitung} = \frac{JKR/k}{JKS/(n-m-1)} = \frac{\text{MSR}}{\text{MSE}} \tag{18}$$
3. Kriteria uji
   H_0 ditolak jika $F_{hitung} > F_{(\alpha,k,n-m-1)}$ atau *p-value* $< \alpha$

b. Uji Koefisien Regresi secara Individual

Uji koefisien regresi secara individual digunakan untuk menguji ada tidaknya pengaruh yang signifikan antara masing – masing variabel prediktor terhadap model regresi linier, berikut langkah-langkah pengujiannya :

1. Hipotesis
   $H_0 : \alpha_j = 0$
   $H_1 : \alpha_j \neq 0$
2. Statistik uji
   $$t_{hitung} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}\ ;\ \text{dengan } Se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} \tag{19}$$
3. Kriteria uji
   H_0 ditolak jika $|t_{hitung}| > t_{(\alpha/2, n-m-1)}$ atau *p-value* $< \alpha$

2.6 Metode *Minimum Volume Ellipsoid* (MVE)

Minimum Volume Ellipsoid (MVE) pertama kali diperkenalkan oleh Rousseuw merupakan penaksir *robust*. Karena dalam mencari solusi eksak dalam menyelesaikan masalah pada MVE ini dapat dikatakan sulit, maka alternatif lain adalah melalui pendekatan algoritma *resampling* yang disebut juga algoritma $(k + 1)$ subset. Berikut merupakan langkah-langkah algoritma *resampling* untuk mendapatkan elipsoid dengan volume minimum:

1. Bentuk subsampel yang mengandung $k + 1$ pengamatan yang disimbolkan dengan indeks $j = \{i_1, i_{2,} \dots . i_{k+1}\} \subset \{1,2, \dots n\}$
2. Menentukan rata-rata (*mean*) sampel dengan rumus:
$$\boldsymbol{T_j} = \boldsymbol{\bar{x}_J} = \frac{\mathbf{1}}{\boldsymbol{k+1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k+1}} \boldsymbol{x_{ij}} \tag{20}$$
Selanjutnya ditentukan matriks kovarian dari sampel dengan rumus:
$$\boldsymbol{S_j} = \frac{\mathbf{1}}{\boldsymbol{k}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k+1}} (\boldsymbol{X_i} - \boldsymbol{\bar{X}})'(\boldsymbol{X_i} - \boldsymbol{\bar{X}}) \tag{21}$$
3. Menghitung jarak Mahalanobis untuk semua pengamatan dengan rumus:
$$D_j{}^2 = [(\boldsymbol{X_i} - \boldsymbol{\bar{X}})'\boldsymbol{S^{-1}}(\boldsymbol{X_i} - \boldsymbol{\bar{X}})]_{\boldsymbol{h:n}} \tag{22}$$
4. Sebagai penentu yang sangat penting dalam MVE adalah volume ellipsoid yang proporsional dengan nilai:
$$V_J = \left(\frac{D_J}{C}\right)^k \sqrt{\det(S_J)} \tag{23}$$
5. Setelah selesai langkah keempat, maka ulangi langkah pertama sampai keempat di atas untuk subsampel kedua. Proses diulang sampai sebanyak $\binom{n}{k+1}$ subsampel.
6. Pilih subsampel yang elipsoidnya memiliki volume paling minimum.
7. Hitung $\boldsymbol{T}(\boldsymbol{X})$ dan $\boldsymbol{S}(\boldsymbol{X}) = c^2(n,k)(X_{k,\alpha}^2)^{-1} \, D_j^2 \, \boldsymbol{S_j}$ dari V_j yang akan digunakan sebagai estimasi mean dan matriks varians dan kovarians, dengan $c^2(n,k) = \left[1 + \frac{15}{n-k}\right]^2$ yang disebut *correction term.*

2.7 Metode *S-Estimator*

Menurut Stuart (2011), salah satu cara yang digunakan untuk mengukur kekekaran suatu *estimator* adalah dengan *breakdown point*. *Breakdown point* merupakan proporsi minimal dari banyaknya *outlier* dibandingkan seluruh data pengamatan. *S-estimator* didefinisikan dengan:
$$\widehat{\boldsymbol{\beta}} = \min_{\beta} s(\, e_1(\boldsymbol{\beta}), e_2(\boldsymbol{\beta}) \dots, e_n(\boldsymbol{\beta})) \tag{24}$$
e_i merupakan residual ke-i dari $\boldsymbol{\beta}$ dan $s(e_1, e_2 \dots, e_n)$ didefinisikan sebagai solusi dari:
$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = K \tag{25}$$
dengan K merupakan suatu konstanta yang didefinisikan sebagai K = 0,1995, untuk mencapai *breakdown point* 50% (Rousseeuw dan Yohai, 1984), sedangkan s merupakan skala estimasi *robust*. Menurut Maronna, Martin dan Yohai (2006), nilai s untuk iterasi kedua dan seterusnya dapat diperoleh dengan:
$$s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i e_i{}^2} \tag{26}$$
Pembobot $w_i = w(u_i) = \frac{\rho(u_i)}{u_i{}^2}$ untuk iterasi kedua dan seterusnya. Untuk iterasi pertama menggunakan:
$$s = \frac{MAD}{0{,}6745} = \frac{median|e_i - median(e_i)|}{0{,}6745} \tag{27}$$

dengan menurunkan parsial pertama dari ρ terhadap β_j $(j = 0,1,\dots,k)$ disamakan dengan 0 seperti berikut:

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i-\sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = 0, \qquad j = 0,1,\dots,k \tag{28}$$

dimana $\psi = \rho'$ dan x_{ij} adalah pengamatan ke-*i* pada parameter ke-*j* dan $x_{i0} = 1$.

Draper dan Smith (1998) memberikan solusi pada persamaan di atas dengan mendefinisikan fungsi pembobot:

$$w(u_i) = \frac{\psi\left(\frac{y_i-\sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)}{\left(\frac{y_i-\sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)} \tag{29}$$

dan $w_i = w(u_i)$. Kemudian estimasi pada turunan parsial pertama dari ρ terhadap β_j dapat ditulis:

$$\sum_{i=1}^{n} x_{ij}w_i\left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) = 0 \qquad , j = 0, 1, \dots, k \tag{30}$$

Persamaan (30) diselesaikan dengan *Iteratively Reweighted Least Square* (IRLS). Pada notasi matrik, persamaan (30) dapat ditulis:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\mathbf{y} \tag{31}$$

Iterasi akan berhenti jika $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ konvergen yaitu selisih nilai $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{m}+\boldsymbol{1})}$ dan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{m})}$ mendekati 0

2.8 Pengangguran

Menurut BPS (2018), pengangguran terbuka merupakan jumlah orang yang masuk dalam angkatan kerja (usia 15 tahun ke atas) yang sedang mencari pekerjaan dan belum mendapatkannya. Faktor-faktor yang mungkin mempengaruhi jumlah pengangguran yaitu jumlah penduduk miskin, Indeks Pembangunan Manusia (IPM), rata-rata lama sekolah, Angka Partisipasi Sekolah (APS), Angka Partisipasi Murni (APM), dan jumlah penduduk.

3. METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah melalui publikasi Jawa Tengah Dalam Angka 2018. Objek penelitiannya adalah jumlah pengangguran, jumlah penduduk miskin, Indeks Pembangunan Manusia (IPM), rata-rata lama sekolah, Angka Partisipasi Sekolah (APS) untuk umur 16-18 tahun, Angka Partisipasi Murni (APM) untuk umur 16-18 tahun, dan jumlah penduduk di setiap kabupaten/kota di Provinsi Jawa Tengah tahun 2017.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah:

y : Jumlah pengangguran
x_1 : Jumlah penduduk miskin
x_2 : Indeks Pembangunan Manusia
x_3 : Rata-rata lama sekolah
x_4 : Angka Partisipasi Sekolah untuk umur 16-18 tahun
x_5 : Angka Partisipasi Murni untuk umur 16-18 tahun
x_6 : Jumlah penduduk

3.3 Langkah-Langkah Penelitian

1. Mengumpulkan data dan menentukan variabel dependen dan variabel independen.
2. Mempersiapkan *software R 3.4.3.*, dan menginstal beberapa *package* dalam *software R 3.4.3* diantaranya yaitu *lmtest, car,* MASS, dan *robustbase*.
3. Memasukkan data ke dalam *software R 3.4.3*.
4. Melakukan analisis deskriptif dan standarisasi terhadap data penelitian.
5. Melakukan analisis regresi linier berganda dengan metode OLS.
6. Melakukan pemeriksaan asumsi non-multikolinieritas dengan cara melihat nilai VIF dan nilai koefisien korelasi *pearson* antar variabel independen.
7. Mendeteksi adanya *outlier* pada variabel independen menggunakan jarak mahalanobis.
8. Melakukan analisis komponen utama klasik dan analisis komponen utama *robust* terhadap variabel independen.
9. Meregresikan komponen-komponen utama terpilih dari analisis komponen utama klasik dan analisis komponen utama *robust* dengan variabel dependen menggunakan metode OLS.
10. Melakukan uji signifikansi model regresi melalui uji F dan uji koefisien regresi secara individu melalui uji t.
11. Melakukan deteksi multikolinieritas dan diperoleh hasil bahwa masalah multikolinieritas sudah teratasi.
12. Menghitung nilai *Adjusted* R^2 dan nilai RSE.
13. Mendeteksi adanya *outlier* pada dua buah model yang diperoleh dari langkah 12 menggunakan metode *DFFITS*.
14. Jika ada *outlier*, maka lakukan analisis regresi komponen utama *robust*, yaitu meregresikan komponen-komponen utama terpilih dengan variabel dependen menggunakan metode *S-estimator*.
15. Melakukan uji signifikansi model regresi melalui uji F dan uji koefisien regresi secara individu melalui uji t.
16. Menghitung nilai *Adjusted* R^2 dan nilai RSE.
17. Mengembalikan persamaan regresi ke bentuk variabel asal.
18. Menarik kesimpulan

4 HASIL DAN PEMBAHASAN

4.1. Regresi Linier Berganda

Pada awal penelitian dilakukan analisis regresi linier berganda dengan estimasi parameter menggunakan Metode Kuadrat Terkecil.

4.1.1. Estimasi Parameter Regresi

Dari hasil *output* didapatkan model regresi linier berganda sebagai berikut:

$$\widehat{\mathrm{Q}} = 0{,}18606\ \mathrm{Z}_1 - 0{,}32160\ \mathrm{Z}_2 + 0{,}62781\ \mathrm{Z}_3 - 0{,}22919\ \mathrm{Z}_4 - 0{,}00012\ \mathrm{Z}_5 + 0{,}75035\ \mathrm{Z}_6$$

Keterangan:

$\widehat{\mathrm{Q}}$: Data Jumlah Pengangguran yang Distandarisasi

Z_1 : Data Jumlah Penduduk Miskin yang Distandarisasi

Z_2 : Data Indeks Pembangunan Manusia (IPM) yang Distandarisasi

Z_3 : Data Rata-rata Lama Sekolah yang Distandarisasi

Z_4 : Data Angka Partisipasi Sekolah (APS) yang Distandarisasi

Z_5 : Data Angka Partisipasi Murni (APM) yang Distandarisasi

Z_6 : Data Jumlah Penduduk yang Distandarisasi

4.1.2. Uji Signifikansi dalam Regresi Linier Berganda

a. Uji F

1. Hipotesis
   $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$
   H_1 : terdapat $\beta_j \neq 0$, j = 1,2,3,4,5,6
2. Taraf signifikansi
   $\alpha = 5\%$
3. Statistik uji
   $F_{hitung} = 21,628240$
4. Kriteria uji
   H_0 ditolak jika $F_{hitung} > (F_{(5\%,6,29)} = 2,432)$
5. Keputusan
   H_0 ditolak karena nilai $F_{hitung} = 17,73 > F_{(5\%,6,29)} = 2,432$
6. Kesimpulan
   Pada taraf signifikansi 5% paling sedikit ada satu variabel independen (jumlah penduduk miskin, IPM, rata-rata lama sekolah, APS, APM, dan jumlah penduduk) yang ada hubungan dengan variabel dependen jumlah penganguran secara bersama-sama.

b. Uji Koefisien Regresi secara Individual (Uji t)

1. Hipotesis
   $H_0 : \beta_j = 0$, dimana j=1,2,3,4,5,6
   $H_1 : \beta_j \neq 0$, dimana j=1,2,3,4,5 ,6
2. Taraf signifikansi
   $\alpha = 5\%$
3. Statistik uji
   Disajikan pada Tabel 2.
4. Kriteria uji
   H_0 ditolak jika $|t_{hitung}| > (t_{(2,5\%,29)} = 2,045)$ atau *p-value* < α (0,05)
5. Keputusan

Tabel 1. Tabel Uji t Regresi Linier Berganda

Variabel	$\mathbf{\lvert t_0 \rvert}$	***p-value***	**Keputusan**	**Kesimpulan**
Z_1	1,006	0,322698	H_0 diterima	Tidak Signifikan
Z_2	0,908	0,371548	H_0 diterima	Tidak Signifikan
Z_3	1,786	0,084643	H_0 diterima	Tidak Signifikan
Z_4	0,926	0,362209	H_0 diterima	Tidak Signifikan
Z_5	0,001	0,999505	H_0 diterima	Tidak Signifikan
Z_6	4,263	0,000195	H_0 ditolak	Signifikan

6. Kesimpulan
   Pada taraf signifikansi 5% koefisien parameter variabel yang berpengaruh signifikan terhadap jumlah pengangguran (Q) adalah jumlah penduduk (Z_6).

4.1.3. *Goodness Of Fit*

Dalam penelitian ini, uji kecocokan model yang digunakan adalah nilai koefisien determinasi yang disesuaikan dan nilai RSE. Diperoleh nilai koefisien determinasi yang disesuaikan sebagai berikut:

$R^2_{Adj,k} = 0,7415$

Hal ini mengindikasikan besar pengaruh jumlah penduduk miskin, Indeks Pembangunan Manusia, rata-rata lama sekolah, Angka Partisipasi Sekolah untuk umur 16-18 tahun, Angka

Partisipasi Murni untuk umur 16-18 tahun, dan jumlah penduduk terhadap jumlah pengangguran adalah 74,15%. Sedangkan sisanya sebesar 25,85% jumlah pengangguran dipengaruhi model lain

Nilai RSE untuk regresi linier berganda menggunakan estimasi Metode Kuadrat Terkecil adalah sebagai berikut:

RSE = 0,5011

4.2. Deteksi Multikolinieritas

Tabel 2. Nilai VIF

Variabel	VIF	Keterangan
Z_1	4,63012	VIF < 10
Z_2	16,99603	VIF > 10
Z_3	16,73693	VIF > 10
Z_4	8,29703	VIF < 10
Z_5	5,48171	VIF < 10
Z_6	4,19507	VIF < 10

Dilihat dari Tabel 3, variabel IPM (Z_2) dan rata-rata lama sekolah (Z_3) mempunyai nilai VIF lebih besar dari 10 yang mengindikasikan adanya multikolinieritas dalam model regresi.

4.3. Analisis Komponen Utama

4.3.1 Deteksi *Outlier*

Dengan nilai $\chi^2_{(6;0,05)}$ = 12,592 terdapat ada tiga *outlier* yang terdeteksi, yaitu data ke-30, 32 dan 33 dengan nilai d^2_{MD} berturut-turut adalah 13,304; 16,789 dan 19,526. Karena terdapat *outlier* pada variabel independen, maka pada penelitian ini, analisis komponen utama dilakukan dengan dua cara, yaitu analisis komponen utama klasik dan analisis komponen utama *robust*.

4.3.2 Analisis Komponen Utama Klasik

Menghitung skor komponen utama dihitung dengan mengalikan variabel independen yang dalam hal ini adalah Z_1, Z_2, Z_3, Z_4, Z_5, dan Z_6 dengan *eigenvector* dari komponen-komponen utama yang terpilih. Sehingga diperoleh persamaan sebagai berikut:

$$\mathrm{K}_1 = 0{,}39474\,\mathrm{Z}_1 - 0{,}47476\,\mathrm{Z}_2 - 0{,}47137\,\mathrm{Z}_3 - 0{,}44252\,\mathrm{Z}_4 - 0{,}32179\,\mathrm{Z}_5 + 0{,}31181\,\mathrm{Z}_6$$

$$\mathrm{K}_2 = -0{,}47445\,\mathrm{Z}_1 - 0{,}04226\,\mathrm{Z}_2 - 0{,}03211\,\mathrm{Z}_3 - 0{,}32629\,\mathrm{Z}_4 - 0{,}57994\,\mathrm{Z}_5 - 0{,}57383\,\mathrm{Z}_6$$

4.3.3 Analisis Komponen Utama *Robust*

Skor komponen utama dihitung dengan mengalikan variabel independen yang dalam hal ini adalah Z_1, Z_2, Z_3, Z_4, Z_5, dan Z_6 dengan *eigenvector* dari komponen-komponen utama yang terpilih. Sehingga diperoleh persamaan sebagai berikut:

$$\mathrm{W}_1 = 0{,}42006\,\mathrm{Z}_1 - 0{,}47816\,\mathrm{Z}_2 - 0{,}41548\,\mathrm{Z}_3 - 0{,}47326\,\mathrm{Z}_4 - 0{,}31462\,\mathrm{Z}_5 + 0{,}31514\,\mathrm{Z}_6$$

$$\mathrm{W}_2 = -0{,}51854\,\mathrm{Z}_1 - 0{,}12876\,\mathrm{Z}_2 - 0{,}13492\,\mathrm{Z}_3 - 0{,}25135\,\mathrm{Z}_4 - 0{,}53256\,\mathrm{Z}_5 - 0{,}59121\,\mathrm{Z}_6$$

4.3.4 Hasil Analisis Komponen Utama Klasik dan Analisis Komponen Utama *Robust*

Tabel 3. Proporsi Kumulatif Varian AKU Klasik dan AKU *Robust*

Metode Analisis	Proporsi Kumulatif Varian yang Dapat Dijelaskan	Proporsi Kumulatif Varian yang Tidak Dapat Dijelaskan
AKU Klasik	84,63%	15,37%
AKU *Robust*	90,52%	9,48%

4.4. Regresi Komponen Utama

4.4.1 Regresi Komponen Utama Klasik-OLS

RKU Klasik-OLS adalah metode regresi yang mengkombinasikan AKU Klasik dengan analisis regresi menggunakan OLS, model RKU Klasik-OLS yang diperoleh adalah:

$$\widehat{Q} = 0{,}26562\ K_1 - 0{,}45057\ K_2$$

4.4.2 Regresi Komponen Utama *Robust*-OLS

RKU *Robust*–OLS adalah metode regresi yang mengkombinasikan AKU *Robust* dengan analisis regresi menggunakan OLS, model RKU *Robust*-OLS yang diperoleh adalah:

$$\widehat{Q} = 0{,}27839\ W_1 - 0{,}48910\ W_2$$

4.4.3 Regresi Komponen Utama Klasik-(*S-Estimator*)

RKU Klasik-(*S-Estimator*) adalah metode regresi yang mengkombinasikan AKU Klasik dengan analisis regresi menggunakan (*S-Estimator*), model RKU Klasik- (*S-Estimator*) yang diperoleh adalah: $\widehat{Q} = 0{,}22176\ K_1 - 0{,}29447\ K_2$

4.4.4 Regresi Komponen Utama *Robust*-(*S-Estimator*)

RKU *Robust*-(*S-estimator*) adalah metode regresi yang mengkombinasikan AKU *Robust* dengan analisis regresi menggunakan *S-estimator*, model RKU *Robust*-(*S-estimator*) yang diperoleh adalah:

$$\widehat{Q} = 0{,}23491\ W_1 - 0{,}29548\ W_2$$

4.5. Perbandingan Efektivitas Model Regresi Komponen Utama

Perbandingan efektivitas keempat model regresi komponen utama (RKU) berdasarkan nilai *Adjusted* R^2 dan *Residual Standard Error* (RSE) disajikan dalam Tabel 5 berikut:

Tabel 4. Perbandingan Efektivitas Model RKU Klasik-OLS, RKU *Robust*-OLS, RKU Klasik-(*S-estimator*), dan RKU *Robust*-(*S-estimator*)

	Adjusted $\mathbf{R^2}$	**RSE**
RKU Klasik-OLS	0,5143	0,6869
RKU *Robust*-OLS	0,5730	0,6440
RKU Klasik-(*S-estimator*)	0,9135	0,4141
RKU *Robust*-(*S-estimator*)	0,9165	0,4073

Dari Tabel 4 dapat dilihat bahwa model regresi komponen utama yang menerapkan metode *robust* MVE dan *S-estimator*, yaitu RKU *Robust*-(*S-estimator*) merupakan model yang paling efektif jika dibandingkan dengan ketiga model lainnya. Hal ini dikarenakan model tersebut mempunyai nilai *Adjusted* R^2 yang paling besar yaitu 0,9165 dan nilai RSE yang paling kecil yaitu 0,4073 dengan model sebagai berikut:

$$\widehat{Q} = 0{,}23491\ W_1 - 0{,}29548\ W_2$$

Persamaan tersebut akan dikembalikan ke persamaan awal, sehingga diperoleh model RKU *Robust-(S-estimator)* sebagai berikut:

$$\widehat{Y} = 8544{,}63619 + 53{,}79346\ X_1 - 262{,}91108\ X_2 - 750{,}33763\ X_3 - 59{,}64477\ X_4 + 463{,}61536\ X_5 + 0{,}00936\ X_6$$

5. KESIMPULAN

Dalam analisis regresi linier berganda dengan parameter OLS terdapat pelanggaran asumsi non multikolinieritas sehingga digunakan regresi komponen utama untuk penanganannya. Dalam penentuan komponen utama pada tahap analisis komponen utama digunakan pendekatan *robust* MVE untuk mengatasi pencilan. Sedangkan dalam tahap regresi komponen utama digunakan *S-estimator* untuk mengatasi pencilan dalam data. Setelah didapatkan beberapa model dibandingkan berdasarkan nilai *Adjusted R-square* dan RSE, diperoleh model terbaik yaitu Regresi Komponen Utama *Robust*-(*S-Estimator*):

$$\hat{Y} = 8544{,}63619 + 53{,}79346\, X_1 - 262{,}91108\, X_2 - 750{,}33763\, X_3 - 59{,}64477\, X_4 + 463{,}61536\, X_5 + 0{,}00936\, X_6$$

Regresi Komponen Utama *Robust* S-*Estimator* menunjukan nilai Jumlah Pengangguran di Jawa Tengah dipengaruhi oleh jumlah penduduk Miskin(X_1), IPM (X_2), rata-rata lama sekolah (X_3), APM(X_4), APS (X_5) dan jumlah penduduk(X_6), sebesar 91,65% dan nilai RSE nya 0,4073.

DAFTAR PUSTAKA

[BPS] Badan Pusat Statistik. 2018. *Katalog Provinsi Jawa Tengah dalam Angka 2018*. Jakarta : BPS.

Chen, C. 2002. *Robust Regression and Outlier Detection with The ROBUSTREG Procedure,* paper 265-27. SAS Instutute Inc., Lary, NC.

Draper, N.R. dan Smith, H. 1992. *Analisis Regresi Terapan*. Edisi Kedua. Diterjemahkan oleh : Bambang Sumantri. Jakarta : Gramedia Pustaka Utama. Terjemahan dari : *Applied Regression Analysis.*

Gujarati, D. N. 2004. *Ekonometri Dasar*. Diterjemahkan oleh : Sumarno Zain. Jakarta : Erlangga. Terjemahan dari : Basic Econometrics.

Johnson, R. A. dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Pearson Prentice Hall.

Kutner M. H., Nachtsheim, C. J., Neter, J., dan Li, W. 2005. *Applied Linear Statistical Models.* Fifth Edition. New York: The McGraw-Hill Companies, Inc.

Montgomery, D. C. dan Peck, E. A. 2012. *Introduction To Linier Regression Analysis. Fifth Edition.* New York : John Wiley and Sons, Inc.

Montgomery, D. C and Runger, G. C. 2011.Applied Statistics and Probability for Engineers :John Wiley & Sons, Inc, USA

Neter, J., Wasserman, W., dan Kutner, M. H. 1988. *Applied Linier Regression Model.* USA : Richard D. Irwin. Inc.

Rousseeuw P. J. dan Aelst, S. V. 2009. *Minimum Volume Ellipsoid.* New York: John Wiley & Sons, Inc.

Rousseeuw, P. J. dan Yohai, V. 1984. “Robust Regression by Means of S Estimators”. *Robust and Nonlinier Time Series Analysis*, pp.256-272. New York : Springer Verlag.

Stuart, C. 2011. *Robust Regression.* England : Durham University.