ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 10, Nomor 3, Tahun 2021, Halaman 402 - 412
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PEMODELAN REGRESI *RIDGE ROBUST S,M, MM-ESTIMATOR* DALAM PENANGANAN MULTIKOLINIERITAS DAN PENCILAN (Studi Kasus : Faktor-Faktor yang Mempengaruhi Kemiskinan di Jawa Tengah Tahun 2020)

Anggun Perdana Aji Pangesti[1], Sugito[2], Hasbi Yasin[3]
[1,2,3]Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

ABSTRACT

The Ordinary Least Squares (OLS) is one of the most commonly used method to estimate linier regression parameters. If there is a violation of assumptions such as multicolliniearity especially coupled with the outliers, then the regression with OLS is no longer used. One method can be used to solved the multicollinearity and outliers problem is Ridge Robust Regression. Ridge Robust Regression is a modification of ridge regression method used to solve the multicolliniearity and using some estimators of robust regression used to solve the outlier, the estimator including : Maximum likelihood estimator (M-estimator), Scale estimator (S-estimator), and Method of moment estimator (MM-estimator). The case study can be used with this method is data with multicollinearity and outlier, the case study in this research is poverty in Central Java 2020 influenced by life expentancy, unemployment number, GRDP rate, dependency ratio, human development index, the precentage of population over 15 years of age with the highest education in primary school, mean years school. The result of estimation using OLS show that there is a multicollinearity and presence an outliers. Applied the ridge robust regression to case study prove that ridge robust regression can improve parameter estimation. The best ridge robust regression model is Ridge Robust Regression S-Estimator. The influence value of predictor variabels to poverty is 73,08% and the MSE value is 0,00791.
Keywords : Ordinary Least Square (OLS), Multicolliniearity, Ridge Regression, Outliers, Robust Regression, Ridge Robust Regression, Poverty

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan yang dihadapi setiap negara. Kemiskinan adalah kondisi ketidakmampuan ekonomi seseorang atau sekelompok untuk memenuhi hak-hak dasarnya untuk mempertahankan hidup dan bekerja. Jumlah penduduk miskin di Provinsi Jawa Tengah pada Maret 2020 mencapai 3.98 juta orang atau sebesar 11.41 persen (BPS, 2019). Ada banyak hal yang dapat mempengaruhi angka kemiskinan, dengan mengetahui faktor-faktor yang mempengaruhi kemiskinan maka dapat dilakukan upaya untuk menurunkan angka kemiskinan.

Analisis regresi linier berganda dengan menggunakan Metode Kuadrat Terkecil (MKT) merupakan salah satu metode yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi kemiskinan. Sebuah model regresi yang baik apabila dapat memenuhi semua asumsi yaitu normalitas, non-autokorelasi, homoskedastisitas dan multikolinieritas (Montgomery *et al*, 2012)

Salah satu pelanggaran asumsi yang terjadi adalah multikolinieritas dimana terdapat hubungan antara beberapa atau semua variabel prediktor dalam model regresi. Regresi *ridge* merupakan salah satu metode yang digunakan mengatasi multikolinieritas. (Montgomery *et al*, 2012). Terjadinya pelanggaran asumsi dikarenakan terindikasi adanya pencilan. Pencilan merupakan kondisi dimana terdapat data yang nilainya berbeda jauh dengan data lainnya. Metode yang dapat menganalisis data yang terindikasi adanya pencilan adalah regresi *robust.* Regresi *robust* memiliki beberapa estimator antara lain *Maximum likelihood estimator (M-estimator), Scale estimator (S-estimator)*, dan *Method of moment estimator (MM-estimator)* (Chen, 2002).

Studi awal penelitian menggunakan regresi linier berganda dengan estimasi parameter Metode Kuadrat Terkecil, diketahui pada data kemiskinan terdapat pelanggaran asumsi berupa multikolinieritas dan heteroskedastisitas, dan terindikasi terdapat beberapa pencilan. Untuk menangani permasalahan tersebut diperlukan metode yang hasil estimasi parameternya stabil terhadap multikolinieritas dan pencilan. Metode tersebut adalah regresi *ridge robust* yang merupakan gabungan dari metode regresi *ridge* dan regresi *robust* (Lukman *et al*, 2014). Penelitian ini membahas analisis data kemiskinan tiap kabupaten/kota di Jawa Tengah dengan menggunakan metode regresi *ridge robust M-estimator*, *S-estimator* dan *MM-estimator*. Dari hasil analisis akan ditentukan model regresi *ridge robust* terbaik. *Software* yang digunakan dalam penelitian ini adalah *software R*.

2. TINJAUAN PUSTAKA

2.1. Kemiskinan

Kemiskinan merupakan salah satu permasalahan yang banyak dihadapi negara. kemiskinan adalah kondisi ekonomi seseorang atau sekelompok orang yang tidak terpenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermatabat. Sedangkan penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan (BPS,2019)

2.2. Analisis Regresi

Menurut Montgomery *et al.* (2012), Model regresi berganda dinyatakan dalam bentuk persamaan :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \ ; \ i = 1, 2, \dots, n$$

2.2.1. Estimasi Parameter Model Regresi Linier Berganda

Menurut Montgomery *et al.* (2012), untuk mengestimasi parameter model regresi yaitu dengan meminimumkan jumlah kuadrat *error* yang dikenal dengan Metode Kuadrat Terkecil.

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

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

$$\left.\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}$$

karena, $\left.\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0$ maka

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

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

2.2.2. Uji Hipotesis Regresi Linear Berganda

a. Uji F

Uji F digunakan untuk menguji ada atau tidaknya hubungan linier antara variabel respon y dengan variabel prediktor $x_1, x_2, x_3, .., x_k$ secara bersama-sama (Montgomery *et al.*, 2012). Berikut langkah-langkahnya :

1. Hipotesis
   H_0 : $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ (tidak terdapat hubungan antara variabel respon y dengan variabel prediktor x_j secara bersama-sama)
   H_1 : terdapat $\beta_j \neq 0$, dengan j = 1,2,…,k (terdapat hubungan antara variabel respon y dengan variabel prediktor x_j secara bersama-sama)
2. Statistik uji

$$F_{hitung} = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{\text{MSR}}{\text{MSE}}$$

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 t)

Pengujian ini digunakan untuk menguji ada atau tidaknya pengaruh yang signifikan antara masing-masing variabel prediktor terhadap model regresi linier.. Langkahnya :

1. Hipotesis

H$_0$: $\beta_j = 0$ (koefisien parameter variabel $x_j$tidak signifikan terhadap y)

H$_1$: $\beta_j \neq 0$, (koefisien parameter variabel x_j signifikan terhadap y)

2. Statistik uji

$$t = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}$$

3. Kriteria uji

H$_0$ ditolak jika $|t_{hitung}| > t_{(\alpha/_2, n-k-1)}$ atau *p-value* < α

2.2.3. *Goodness of Fit* (Ukuran Kecocokan Model)

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

Menurut Gujarati (2003), R^2 merupakan suatu ukuran yang digunakan untuk mengetahui seberapa baik kecocokan dari suatu model regresi.

$$R^2{}_{Adj,k} = 1 - \frac{SSE/_{(n-k-1)}}{SST/_{(n-1)}} = 1 - \frac{(n-1)}{(n-k-1)}(1 - R^2)$$

b. MSE (*Mean Square Error*)

Menurut Montgomery *et al.* (2012), dalam analisis regresi rumus MSE adalah

$$\text{MSE} = \frac{SSE}{n-k-1}$$

2.2.3. Uji Asumsi dalam Regresi Linier Berganda

a. Uji Normalitas

Menurut Gujarati (2003), pada regresi linier klasik diasumsikan bahwa setiap ε_i didistribusikan normal dengan $\varepsilon_i \sim N(0, \sigma^2)$. Salah satu uji formal yang dapat digunakan adalah Kolmogorov-Smirnov

$$D = \sup_e |F_0(e) - S(e)|$$

b. Uji Autokorelasi

Autokorelasi dapat didefinisikan sebagai korelasi antar anggota serangkaian pengamatan yang diurutkan waktu, dilambangkan dengan $E(\varepsilon_i \varepsilon_j) = 0 \quad i \neq j$. Salah satu uji autokorelasi yang dapat dilakukan adalah dengan uji Durbin Watson.

$$d = \frac{\sum_{i=2}^{i=n} (e_i - e_{i-1})^2}{\sum_{i=2}^{i=n} e_i^2}$$

c. Uji Heteroskedastisitas

Uji heteroskedastisitas digunakan untuk mengetahui pada model regresi terjadi ketidaksamaan varian dan residual atau tidak. Lambang homoskedastisitas adalah $E(\varepsilon_i^2) = \sigma^2 \; i \neq j$. Menurut Gujarati (2003), untuk mengetahui ada tidaknya heteroskedastisitas dapat diperiksa dengan uji Glejser. Uji Glejser dilakukan dengan meregresikan nilai absolut residual $|e_i|$ terhadap variabel-variabel prediktor yang mempunyai hubungan erat dengan σ^2

$$|e_i| = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + v$$

2.3. Ukuran Pemusatan dan Penskalaan (*Centering and Scaling*)

Menurut Kutner *et al.* (2005), perbedaan unit satuan pada model regresi perlu dilakukan standarisasi dengan rumus berikut:

$Y_i^* = \frac{y_i}{\sqrt{n-1}\,(S_Y)} = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\bar{Y}}{S_Y}\right)$ dimana $S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}}$

$Z_{ij} = \frac{x_{ij}}{\sqrt{n-1}\,(S_j)} = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ij}-\bar{X}_j}{S_j}\right)$ $S_j = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij}-\bar{X}_j\right)^2}{n-1}}$

untuk $i = 1, 2, \dots, n \,;\, j = 1, 2, \dots, k$

sehingga diperoleh model regres standar sebagai berikut:

$$Y_i^* = \beta_1^* Z_{i1} + \beta_2^* Z_{i2} + \cdots + \beta_k^* Z_{ik} + \varepsilon_i$$

hubungan estimator regresi bentuk standar dengan bentuk asli adalah:

$$\beta_j = \left(\frac{S_Y}{S_j}\right)\beta_j^*,\; j = 1, 2, \dots, k \text{ dan } \beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 \dots - \beta_k\bar{X}_k$$

2.4. Multikolinieritas

Multikolinieritas adalah kondisi dimana terdapat hubungan linier dari variabel prediktor. Model regresi yang baik adalah yang tidak terdapat multikolinieritas, untuk mengukur besarnya multikolinieritas menggunakan *Variance Inflation Factors* (VIF)

$$VIF\left(x_j\right) = \frac{1}{(1-R_j^2)}, \text{ i = 1,2,...,n}$$

Nilai VIF >10 menunjukkan multikolinearitas yang kuat (Montgomery *et al.*, 2012).

2.5. Pencilan

Pencilan merupakan pengamatan yang nilainya ekstrim. Keberadaan pencilan dapat mengganggu akan tetapi menghapusnya bukan keputusan yang baik (Draper dan Smith, 1998). Metode yang dapat mendeteksi pencilan adalah *Difference in Fit Standarized* ($DFFITS_i$)

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

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

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

Data disebut pencilan jika $|DFFITS| > 1$ untuk data yang berukuran kecil (n < 15) dan nilai $|DFFITS| > 2\sqrt{p/n}$ untuk gugus data yang berukuran besar, dengan $p = k + 1$, dan n adalah banyaknya observasi (Kutner *et al.*, 2005)

2.6. Regresi *Ridge*

Menurut Dereny dan Rashwan (2011), model persamaan *ridge* didasarkan pada penambahan tetapan K pada diagonal utama matriks $\boldsymbol{X'X}$ sehingga model persamaan *ridge* menjadi:

$$\boldsymbol{Y} = \boldsymbol{X\beta_R} + \boldsymbol{\varepsilon}$$

Estimasi regresi *ridge* diperoleh dengan meminimumkan jumlah kuadrat *error*:

$$\sum_{i=1}^{n}\varepsilon_i^2 = \boldsymbol{\varepsilon'\varepsilon} = (\boldsymbol{Y}-\boldsymbol{X\beta_R})'(\boldsymbol{Y}-\boldsymbol{X\beta_R})$$

dengan kendala $\boldsymbol{\beta'_R\beta_R} = \boldsymbol{c^2}$. Bila $\boldsymbol{c}$ tetapan positif yang berhingga, dengan menggunakan metode pengali *langrange* maka diperoleh:

$$L(\boldsymbol{\beta_R}, K) = (\boldsymbol{Y}-\boldsymbol{X\beta_R})'(\boldsymbol{Y}-\boldsymbol{X\beta_R}) + K\left(\boldsymbol{\beta'_R\beta_R} - \boldsymbol{c^2}\right)$$

sehingga diperoleh estimator regresi *ridge* yaitu:

$$\boldsymbol{\beta_R} = (\boldsymbol{X'X} + K\boldsymbol{I})^{-1}\boldsymbol{X'Y}$$

Dengan *K* merupakan tetapan bias.

2.7. Regresi *Robust*

Regresi *robust* merupakan metode regresi yang digunakan jika terdapat beberapa pencilan yang berpengaruh pada model sehingga hasil menjadi resistan terhadap pencilan (Draper dan Smith, 1998)

Menurut Chen (2002), regresi *robust* memiliki beberapa metode estimasi, diantaranya *M-estimator, S-estimator* dan *MM-estimator*.

2.7.1. *Robust M-Estimator*

Menurut Montgomery *et al.* (2012). *M-estimator* meminimumkan fungsi sisaan :

$$\min_{\beta} \sum_{i=1}^{n} \rho(u_i) = \min_{\beta} \sum_{i=1}^{n} \rho(\frac{e_i}{s}) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)$$

dimana s adalah skala estimasi robust. Estimasi s yang digunakan adalah:

$$s = \frac{MAD}{0{,}6475} = \frac{median|e_i - median(e_i)|}{0{,}6475} \text{ dan } u_i = \frac{e_i}{s}$$

Untuk mendapatkan estimasi parameter dengan mencari turunan parsial pertama dari ρ terhadap β_j $(j = 0,1, \dots, k)$ disamakan dengan 0

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

Draper dan Smith (1998) memberikan solusi dengan mendefinisikan fungsi pembobot :

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

dan $w_i = w(u_i)$. Kemudian estimasi persamaan dapat ditulis:

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

Kemudian diselesaikan dengan *Iteratively Reweighted Least Square* (IRLS). Iterasi akan berhenti jika $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ konvergen yaitu selisih nilai $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{m+1})}$ dan $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}^{(\boldsymbol{m})}$ mendekati 0.

2.7.2. *Robust S-Estimator*

Salah satu cara untuk mengukur kerobustan estimator adalah dengan *Breakdown point*, *S-estimator* merupakan salah satu estimator yang memiliki nilai *breakdown point* yang tinggi. *S-estimator* didefinisikan:

$$\hat{\beta} = \min_{\beta} s(\, e_1(\boldsymbol{\beta}), e_2(\boldsymbol{\beta}) \dots, e_n(\boldsymbol{\beta}))$$

e_i merupakan residual ke-i dari β dan $s(e_1, \dots, e_n)$ didefinisikan sebagai solusi dari:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \frac{1}{1}\sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = \delta$$

Agar *breakdown point* 50%, maka $\delta = E_{\Phi}\rho(u_i)$ = 0.1995, dengan skala *robust* (s) yang digunakan untuk iterasi pertama :

$$s = \frac{MAD}{0{,}6475} = \frac{median|e_i - median(e_i)|}{0{,}6475}$$

dan untuk iterasi berikutnya

$$s = \sqrt{\frac{1}{n\delta}\sum_{i=1}^{n} w_i {e_i}^2}$$

Dengan pembobot $w_i = w(u_i) = \frac{\rho(u_i)}{{u_i}^2}$

Dengan menurunkan parsial pertama ρ terhadap β_j seperti pada *M-Estimator* sehingga didapatkan estimasi persamaan:

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

Kemudian diselesaikan dengan *Iteratively Reweighted Least Square* (IRLS). Iterasi akan berhenti jika $\widehat{\boldsymbol{\beta}}_j$ konvergen yaitu selisih nilai $\widehat{\boldsymbol{\beta}}_j^{(m+1)}$ dan $\widehat{\boldsymbol{\beta}}_j^{(m)}$ mendekati 0.

2.7.3. *Robust MM-Estimator*

Menurut Chen (2002), *MM-estimator* merupakan kombinasi antara estimator yang memiliki *breakdownt point* tinggi dan *M-estimator* yang memiliki efisiensi tinggi. Yohai dalam Stuart (2011) mendiskripsikan tiga tahapan *MM-estimator* :

1. Menentukan estimasi awal yang ditunjukkan $\widehat{\boldsymbol{\beta}}$ dengan estimator yang memiliki *breakdown point* tinggi, biasanya *S-estimator*
2. Menghitung residual dan skala *M-estimator* dengan 50% *breakdown point*. $s(e_1(\widehat{\boldsymbol{\beta}}), e_2(\widehat{\boldsymbol{\beta}}) \dots, e_n(\widehat{\boldsymbol{\beta}}))$ dinotasikan s_n.
3. Menghitung estimasi parameter akhir dengan M-*estimator*, menggunakan turunan fungsi pengaruh $\psi(u_i)$ dan skala s_n didapatkan dari langkah kedua.

2.8. Regresi *Ridge Robust*

Regresi *ridge robust* merupakan penggabungan dari metode regresi ridge dan regresi ro*bust*, perbedaan dengan regresi *ridge* biasa adalah pada penduga parameter yang digunakan. Penduga yang dihasilkan akan resisten terhadap pencilan.

Rumus penduga parameter regresi *ridge robust* adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{RR}} = (\boldsymbol{X}'\boldsymbol{X} + K\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Menurut Dereny dan Rashwan (2011), salah satu cara menentukan *K* dengan metode yang dikenalkan oleh Hoerl, Kennard, dan Balwin (HKB)

$$K_{HKB} = \frac{k(MSE^{Robust})}{\widehat{\boldsymbol{\beta}}^{Robust\prime}\widehat{\boldsymbol{\beta}}^{Robust}}$$

3. METODE PENELITIAN

3.1. Sumber Data dan Variabel Penelitian

Data yang digunakan adalah data sekunder, bersumber dari Provinsi Jawa Tnegah dalam Angka tahun 2021 dan Statistik Kesejahteraan Rakyat Jawa Tengah tahun 2020 yang dapat diakses https://jateng.bps.go.id/

3.2. Variabel Penelitian

Y : Kemiskinan
X_1 : Angka Harapan Hidup
X_2 : Jumlah Pengangguran
X_3 : Laju PDRB
X_4 : Rasio Ketergantungan
X_5 : Indeks Pembangunan Manusia (IPM)
X_6 : Persentase Penduduk Berusia lebih dari 15 Tahun dengan Pendidikan Tertinggi SD
X_7 : Rata-rata Lama Sekolah

3.3. Tahapan Analisis Data

1. Mengumpulkan data.
2. Menentukan variabel terikat dan variabel bebas.
3. Melakukan standarisasi data dengan Transfomasi Korelasi
4. Mengestimasi koefisien regresi $\widehat{\boldsymbol{\beta}}$ menggunakan Metode Kuadrat Terkecil.
5. Melakukan Uji F dan Uji t
6. Menghitung *Adjusted* R^2 dan MSE.
7. Melakukan uji multikolinearitas dengan melihat nilai VIF.

8. Melakukan pendeteksian pencilan dengan $DFFITS_i$.
9. Mengestimasi nilai $\widehat{\boldsymbol{\beta}}^{(robust)}$) dengan menggunakan metode Regresi Robust dengan pembobot tukey's-bisquare
10. Menghitung nilai $\widehat{\boldsymbol{\beta}}_{RR}$ pada metode Regresi *Ridge Robust*
11. Melakukan uji F dan uji t untuk model regresi *ridge robust*
12. Menghitung *Adjusted* R^2 dan MSE untuk model regresi *ridge robust.*
13. Membandingkan nilai *Adjusted* R^2 dan MSE model Regresi MKT dan Regresi *Ridge Robust* untuk mendapatkan model terbaik

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:

$$\hat{y} = -0{,}157531Z_1 + 0{,}478338Z_2 + 0{,}099946Z_3 + 0{,}250985Z_4 - 0{,}095399Z_5 + 0{,}049871Z_6 - 0{,}117738Z_7$$

4.1.2. Uji Signifikansi Parameter

a. Uji F

1. Hipotesis
   $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_{6=}\ 0$
   H_1 : terdapat $\beta_j \neq 0$, j = 1,2,3,4,5,6,7
2. Taraf signifikansi
   α = 5% = 0,05
3. Statistik uji
   F_{hitung} =9,81 dan *p-value* = 0,000005
4. Kriteria uji
   H_0 ditolak jika $F_{hitung} > (F_{(5\%,;6;28)} = 2{,}45)$ atau *p-value* < α (0,05)
5. Keputusan
   H_0 ditolak karena $F_{hitung} = 5{,}655063 > (F_{(5\%;6;27)} = 2{,}37)$ atau *p-value* (0,000005) < α (0,05)
6. Kesimpulan
   Pada taraf signifikansi 5% dapat disimpulkan bahwa terdapat hubungan antara variabel respon kemiskinan dengan variabel prediktor (AHH, pengangguran, laju PDRB, rasio ketergantungan, IPM, presentase penduduk berusia lebih dari 15 tahun dengan pendidikan tertinggi SD, rata-rata lama sekolah) secara bersama-sama.

b. Uji Koefisien Regresi secara Individual (Uji t)

1. Hipotesis
   $H_0 : \beta_j = 0$, dimana j=1,2,3,4,5,6,7
   $H_1 : \beta_j \neq 0$, dimana j=1,2,3,4,5,6,7
2. Taraf signifikansi
   α = 5% = 0,05
3. Statistik uji
   Disajikan pada Tabel 1.
4. Kriteria uji
   H_0 ditolak jika $|t_{hitung}| > (t_{(2,5\%\ ;27)} = 2{,}052)$ atau *p-value* < α (0,05)
5. Keputusan

Tabel 1. Tabel Uji t Regresi Linier Berganda dengan MKT

Variabel	t_{hitung}	*p-value*	Keputusan
$\boldsymbol{Z_1}$	-0,891753	0,380404	H_0 diterima
$\boldsymbol{Z_2}$	3,977133	0,000470	H_0 ditolak
$\boldsymbol{Z_3}$	0,925263	0,363026	H_0 diterima
$\boldsymbol{Z_4}$	1,657337	0,109028	H_0 diterima
$\boldsymbol{Z_5}$	-0,203321	0,840409	H_0 diterima
$\boldsymbol{Z_6}$	0,217616	0,829364	H_0 diterima
$\boldsymbol{Z_7}$	-0,260652	0,796339	H_0 diterima

6. Kesimpulan

   Pada taraf signifikansi 5% hanya koefisien parameter jumlah pengangguran (Z_2) yang berpengaruh signifikan terhadap kemiskinan (*y*). Sedangkan koefisien parameter AHH (Z_1), laju PDRB (Z_3), rasio ketergantungan (Z_4), IPM (Z_5), presentase penduduk berusia lebih dari 15 tahun dengan pendidikan tertinggi SD (Z_6), rata-rata lama sekolah (Z_7) tidak berpengaruh signifikan terhadap kemiskinan (*y*). Karena ada beberapa variabel bebas yang tidak berpengaruh, maka dengan menggunakan metode *backward* diperoleh model baru yaitu :

   $\hat{y} = 0{,}497273Z_2 - 0{,}544549Z_7$

4.1.3. *Goodness Of Fit*

a. ***Adjusted*** $\boldsymbol{R^2}$

$R^2{}_{Adj,k} = 0{,}638253$

Artinya Berdasarkan nilai tersebut maka berarti kemiskinan dipengaruhi oleh AHH dan rata-rata lama sekolah sebesar 63,83% dan sisanya 36,17% dipengaruhi oleh faktor lain.

b. **MSE**

MSE = 0,010639

4.1.4. Uji Asumsi dalam Regresi Linier Berganda

a. **Uji Normalitas**

Nilai D = 0,074594 dan *p-value* = 0,9889

Kesimpulan : Pada taraf signifikansi 5% residual data berdistribusi normal

b. **Uji Autokorelasi**

Nilai d = 1,76392

$d_U\ (1{,}5838) \leq d(1{,}76392\ ) \leq 4 - d_U(2{,}4162)$

Kesimpulan : Pada taraf signifikansi 5%, tidak ada autokorelasi positif atau negatif antar residual.

c. **Uji Heteroskedastisitas**

Tabel 2. Tabel Uji Glejser

Variabel	t_{hitung}	$p - value$	Keputusan
$\boldsymbol{Z}_2$	2,593929	0,014952	H_0 ditolak
$\boldsymbol{Z}_7$	-0,3871589	0,701201	H_0 diterima

Kesimpulan : Pada taraf signifikansi 5% terdapat heteroskedastisitas

4.2 Pendeteksian Multikolinieritas dan Pencilan

a. Pendeteksian Multikolinieritas

Tabel 2. Nilai VIF

Variabel	VIF	Keterangan
$\boldsymbol{Z_1}$	2,984893	VIF < 10
$\boldsymbol{Z_2}$	1,383600	VIF < 10
$\boldsymbol{Z_3}$	1,116037	VIF < 10
$\boldsymbol{Z_4}$	2,193583	VIF < 10
$\boldsymbol{Z_5}$	21,057710	VIF > 10
$\boldsymbol{Z_6}$	5,023295	VIF < 10
$\boldsymbol{Z_7}$	19,516113	VIF > 10

Kesimpulan : Pada taraf signifikansi 5% terjadi multikolinieritas pada variabel IPM (Z_5) dan rata-rata lama sekolah (Z_7)

b. Pendeteksian Pencilan

Dikarenakan jumlah data (n) = 35 termasuk gugus data besar, maka batas suatu data dikatakan sebagai pencilan jika nilai $|DFFITS| > 2\sqrt{(k+1)/n}$ dengan k = 7, sehingga $|DFFITS| >$ 0,9561829. Sehingga diperoleh 4 pencilan terdeteksi, yaitu data ke 1, 28, 29 dan 33.

4.3. Regresi *Ridge Robust M-estimator*

4.3.1. Estimasi Parameter Regresi *Robust M-estimator*

Model regresi *robust M-estimator* yang didapatkan setelah proses 27 kali iterasi adalah sebagai berikut :

$$\hat{y} = 0{,}008566 - 0{,}205064Z_1 + 0{,}585185Z_2 + 0{,}136933Z_3 + 0{,}241089Z_4 - 0{,}264145Z_5 - 0{,}040626Z_6 - 0{,}004172Z_7$$

4.3.2 Estimasi Parameter Regresi *Ridge Robust-MM*

Pada *output* regresi *ridge robust M-estimator* dengan nilai *K* = 0,077646 didapatkan model Regresi *Ridge Robust M-estimator* adalah sebagai berikut:

$$\hat{y} = 0{,}461210Z_2 + 0{,}256313Z_4 - 0{,}345947Z_7$$

4.4. Regresi *Ridge Robust S-estimator*

4.4.1. Estimasi Parameter Regresi *Robust S-estimator*

Model regresi *robust S-estimator* yang didapatkan setelah proses 34 kali iterasi adalah sebagai berikut :

$$\hat{y} = 0{,}017040 - 0{,}207593Z_1 + 0{,}677475Z_2 + 0{,}141624Z_3 + 0{,}069428Z_4 - 0{,}264423Z_5 - 0{,}060280Z_6 - 0{,}064634Z_7$$

4.4.2. Estimasi Parameter Regresi *Ridge Robust S-estimator*

Pada *output* regresi *ridge robust S-estimator* dengan nilai *K* = 0,0076887 didapatkan model Regresi *Ridge Robust S-estimator* adalah sebagai berikut:

$$\hat{y} = -0{,}180450Z_1 + 0{,}461065Z_2 + 0{,}104639Z_3 + 0{,}2551741Z_4 - 0{,}2360846Z_7$$

4.5. Regresi *Ridge Robust MM-estimator*

4.5.1. Estimasi Parameter Regresi *Robust MM-estimator*

Model regresi *robust MM-estimator* yang didapatkan setelah proses 34 kali iterasi dengan *S-estimator* dan 12 kali iterasi dengan *M-estimator* adalah sebagai berikut :

$$\hat{y} = 0{,}023327 - 0{,}108132Z_1 + 0{,}826108Z_2 + 0{,}176386Z_3 + 0{,}173314Z_4 - 0{,}254300Z_5 + 0{,}071576Z_6 + 0{,}112614Z_7$$

4.5.2. Estimasi Parameter Regrei *Ridge Robust MM-estimator*

Pada *output* regresi *ridge robust MM-estimator* dengan nilai K = 0,032588 didapatkan model Regresi *Ridge Robust MM-estimator* adalah sebagai berikut:

$$\hat{y} = 0{,}461706Z_2 + 0{,}256303Z_4 - 0{,}346274Z_7$$

4.6. Pemilihan Model Regresi *Ridge Robust* Terbaik

Sebelumnya model regresi harus dikembalikan ke dalam bentuk variabel asli, didapatkan empat model regresi sebagai berikut:

a. Model regresi linier berganda dengan estimasi parameter menggunakan Metode Kuadrat Terkecil yaitu:
$\hat{y} = 295{,}212895 + 0{,}001488X_2 - 29{,}636952X_7$

b. Model regresi linier berganda dengan estimasi parameter menggunakan Regresi *Ridge Robust M-estimator* yaitu:
$\hat{y} = 14{,}472737 + 0{,}001380X_2 + 4{,}234321X_4 - 18{,}827703X_7$

c. Model regresi linier berganda dengan estimasi parameter menggunakan Regresi *Ridge Robust S-estimator* yaitu:
$\hat{y} = 473.216057 - 6{,}635838X_1 + 0{,}001380X_2 + 4{,}126659X_3 + 4{,}215507X_4 - 12{,}848849X_7$

d. Model regresi linier berganda dengan estimasi parameter menggunakan Regresi *Ridge Robust MM-estimator* yaitu:
$\hat{y} = 14{,}571856 + 0{,}001382X_2 + 4{,}234156X_4 - 18.845881X_7$

Pemilihan metode terbaik didasarkan pada besarnya nilai $Adjusted\ R^2$ dan MSE, disajikan dalam Tabel 3.

Tabel 3. Tabel Pembanding Regresi (MKT) dan Regresi *Ridge Robust*

	$Adjusted\ R^2$	MSE
Metode Kuadrat Terkecil	0,638253	0,010639
Ridge Robust M-estimator	0,645562	0,010425
Ridge Robust S-estimator	0,730798	0,007918
Ridge Robust MM-estimator	0,642049	0,010528

Kriteria pemilihan model regresi *robust* terbaik yaitu mempunyai $R^2{}_{Adj,k}$ terbesar dan nilai MSE terkecil. Dari Tabel 3, disimpulkan bahwa model terbaik adalah Regresi *Ridge Robust S-estimator*

$$\hat{y} = 473.216057 - 6{,}635838X_1 + 0{,}001380X_2 + 4{,}126659X_3 + 4{,}215507X_4 - 12{,}848849X_7$$

dengan nilai *Adjusted* R^2 terbesar yaitu 0,730798 dan MSE terkecil yaitu 0,007918. Selain itu, terdapat lebih banyak variabel prediktor yang signifikan terhadap variabel respon.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan dalam pemodelan kemiskinan di Jawa Tengah tahun 2020, diperoleh kesimpulan sebagai berikut :

1. Estimasi parameter pertama menggunakan MKT, di mana pada model yang diperoleh asumsi homoskedastisitas dan multikolinieritas tidak terpenuhi. Terdapat 2 variabel dengan masalah multikolinieritas dan terdapat 4 buah pencilan. Oleh karena itu dilakukkan penanganan dengan metode regresi *ridge robust.*. Setelah dilakukan estimasi parameter menggunakan Regresi *Ridge Robust S, M, MM-estimator,* lalu dicari model terbaik dengan membandingkan nilai *Adjusted* R^2 dan MSE. Model terbaik yang diperoleh yaitu dengan menggunakan Regresi *Ridge Robust S-estimator.* Model yang diperoleh sebagai berikut :

$$\hat{y} = 473.216057 - 6{,}635838X_1 + 0{,}001380X_2 + 4{,}126659X_3 + 4{,}215507X_4 - 12{,}848849X_7$$

2. Kemiskinan di Jawa Tengah tahun 2020 yang dipengaruhi oleh AHH, pengangguran, laju PDRB, rasio ketergantungan, dan rata-rata lama sekolah sebesar 73,08% dengan nilai MSE yaitu 0,007081.

DAFTAR PUSTAKA

BPS. (2019). *Provinsi Jawa Tengah dalam Angka tahun 2019* . Semarang : Badan Pusat Statistik Jawa Tengah .

Chen, C. (2002). Robust Regression and Outlier Detection with ROBUSTREG Procedure . *Sugi 27,* Paper 265-27.

Dereny, M., & Rashwan, N. I. (2011). Solving Multicolliniearity Problem Using Ridge Regression Models. *Int. J Contemp. Math Sciences, Vol. 6*(No. 12), 585-600.

Draper, N. R., & Smith, H. (1998). *Aplied Regression Analysis Third Edition.* USA: John Wiley and Sons, Inc.

Gujarati, D. N. (2003). *Basic Econometrics Fourth Edition.* New York: The McGraw-Hill Companies.

Kutner, M. H., Nachtsheim , C. J., Neter , J., & Li , W. (2005). *Applied Linier Statistical Models* . New York : The McGraw-Hill Comapanies .

Lukman, A., Arowolo, O., & Ayinde , K. (2014). Some Robust Ridge Regression for Handling Multicollinearity and Outlier. *International Journal of Sciences : Basic and Applied Research, Vol. 16*(No. 2), 192-202.

Montgomery , D. C., Peck, E. A., & Vining , G. G. (2012). *Introduction To Linier Regression Analysis Fifth Edition.* New Jersey: John Wiley and Sons, Inc.