PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK PADA KASUS MULTIKOLINEARITAS DENGAN REGRESI KOMPONEN UTAMA DAN *STEPWISE REGRESSION*

(Skripsi)

Oleh

WIDIYAWATI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

ABSTRACT

SELECTION OF THE BEST MULTIPLE LINEAR REGRESSION MODEL IN MULTICOLINEARITY CASE WITH PRINCIPAL COMPONENT REGRESSION AND STEPWISE REGRESSION

By

Widiyawati

One of assumption violation in regression analysis is be found linear relationship between some or all independent variables in regression model which is called multicollinearity. The consequence of the multicollinearity problem that is variance regression coefficients becomes very large, even though the least squares method still meets the requirement of the Best Linear Unbiased Estimator (BLUE) with high multicollinearity but the estimator obtained is unstable. This study aims to determine the performance of the Principal Component Regression and stepwise regression method in overcoming multicollinearity in multiple linear regression based on simulation data and real data on the percentage of poor people in the province of Indonesia. The result of this study indicate that the Stepwise Regression method is better in selecting multiple linear regression models in the case multicollinearity compared to the Principal Component Regression method based on the Adjusted R^2 dan MSE criteria values. The percentage of Poor Population data is indluenced by the human development index, population density, and old school hope on Stepwise Regression method while the Principal Component Regression method is the human development index, per capita expenditure, and population density.

Key Word : *Principal Component Regression, Stepwise Regression*, Multiple Linear Regression, Multicollinearity.

ABSTRAK

PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK PADA KASUS MULTIKOLINEARITAS DENGAN REGRESI KOMPONEN UTAMA DAN *STEPWISE REGRESSION*

Oleh

Widiyawati

Salah satu pelanggaran asumsi pada analisis regresi adalah terdapatnya hubungan linear diantara beberapa atau semua variabel bebas dalam model regresi yang disebut multikolinearitas. Konsekuensi dari masalah multikolinearitas yaitu variansi dari koefisien regresi menjadi sangat besar, meskipun metode kuadrat terkecil tetap memenuhi syarat *Best Linear Unbiased Estimator* (BLUE) dengan multikolinearitas yang tinggi namun penduga yang didapatkan tidak stabil. Penelitian ini bertujuan untuk mengetahui performa metode Regresi Komponen Utama dan *Stepwise Regression* dalam mengatasi masalah multikolinearitas pada regresi linear berganda berdasarkan data simulasi dan data real Persentase Penduduk Miskin Provinsi di Indonesia. Hasil penelitian ini menunjukkan bahwa metode *Stepwise Regression* lebih baik dalam pemilihan model regresi linear berganda pada kasus multikolinearitas dibandingkan dengan metode Regresi Komponen Utama berdasarkan nilai kriteria *Adjusted* R^2 dan MSE. Data Persentase Penduduk Miskin dipengaruhi oleh indeks pembangunan manusia, kepadatan penduduk, dan harapan lama sekolah pada metode *Stepwise Regression* sedangkan pada metode Regresi Komponen Utama adalah indeks pembangunan manusia, pengeluaran per-kapita disesusaikan, dan kepadatan penduduk.

Kata Kunci : Regresi Komponen Utama, *Stepwise Regression*, Regresi Linear Berganda, Multikolinearitas.

PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK PADA KASUS MULTIKOLINEARITAS DENGAN REGRESI KOMPONEN UTAMA DAN *STEPWISE REGRESSION*

Oleh

Widiyawati

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi	: **PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK PADA KASUS MULTIKOLINEARITAS DENGAN REGRESI KOMPONEN UTAMA DAN *STEPWISE REGRESSION***
Nama Mahasiswa	: **Widiyawati**
Nomor Pokok Mahasiswa	: **1717031003**
Jurusan	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Drs. Eri Setiawan, M.Si.	**Amanto, S.Si., M.Si.**
NIP. 19581101 198803 1 002	NIP. 19730314 200012 1 002

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua Penguji : **Drs. Eri Setiawan, M.Si.** ....

Sekretaris : **Amanto, S.Si., M.Si.** ....

Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **04 Agustus 2021**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Widiyawati**
Nomor Pokok Mahasiswa	: **1717031003**
Jurusan	: **Matematika**

Dengan ini menyatakan bahwa skripsi saya yang berjudul " **PEMILIHAN MODEL REGRESI LINEAR BERGANDA TERBAIK PADA KASUS MULTIKOLINEARITAS DENGAN REGRESI KOMPONEN UTAMA DAN *STEPWISE REGRESSION*"** adalah hasil pekerjaan saya sendiri. Semua hasil tulisan dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil Salinan atau telah dibuat orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 05 Agustus 2021
Penulis,

METERAI TEMPEL
E3164AJX391941595

Widiyawati
NPM. 1717031003

RIWAYAT HIDUP

Penulis bernama lengkap Widiyawati, lahir di Bandar Sukabumi pada tanggal 7 September 1999. Penulis merupakan anak pertama dari tiga bersaudara, pasangan Bapak Mukhlisin dan Ibu Kardena Wati.

Penulis telah menempuh pendidikan dasar di SD Negeri 1 Bandar Sukabumi Tahun 2005-2011, Sekolah Menengah Pertama Negeri (SMPN) 1 Bandar Negeri Semuong Tahun 2011-2014, dan Sekolah Menengah Atas Negeri (SMAN) 1 Kotaagung pada Tahun 2014-2017.

Tahun 2017 penulis terdaftar sebagai mahasiswa S1 Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMMPTN. Selama menempuh pendidikan di Universitas Lampung penulis merupakan salah satu mahasiswa penerima beasiswa Bidik Misi. Pada tahun 2020 penulis melakukan Kerja Praktik di Knator Badan Pusat Statistik Kabupaten Tanggamus dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata Dari Rumah di Kelurahan Kuripan, Kecamatan Kotaagung Pusat, Kabupaten Tanggamus. Selama menjadi mahasiswa penulis aktif di organisasi Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila sebagai anggota Bidang Eksternal periode 2018, serta Redaktur Media Dalam Jaringan UKMF Natural periode 2019-2020.

KATA INSPIRASI

“Sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari suatu urusan), kerjakanlah dengan sungguh-sungguh (urusan yang lainnya). Dan hanya kepada tuhanmu lah hendaknya kamu berharap.”

(QS. AL-Insyirah: 6-8)

“Wahai orang-orang yang beriman! Mohonlah pertolongan (kepada Allah) dengan sabar dan shalat. Sungguh Allah bersama orang-orang yang sabar”

(QS. Al-Baqarah : 153)

“Barang siapa yang berbuat kebaikan seberat zarrah, niscaya dia akan melihat (balasan) Nya.”

(QS. Az-Zalzalah : 7)

“Barang siapa yang menempuh jalan untuk menuntut ilmu, niscaya Allah subhanahu wata’ala akan memudahkan baginya jalan menuju surga”

(H.R. Muslim)

PERSEMBAHAN

Dengan mengucapkan segala puji dan syukur kehadirat Allah SWT yang telah memberikan petunjuk dan kemudahan untuk menyelesaikan skripsi ini. Kupersembahkan karyaku yang sederhana ini kepada:

Bapak dan Ibu

sebagai tanda bakti, hormat, dan rasa terimakasih kepada bapak dan ibu yang tidak pernah lelah untuk selalu mendoakan, yang telah mendidik dan membesarkanku dengan penuh kasih saying, serta sebagai motivasi terbesar dalam hidupku. Semoga ini menjadi langkah awal untuk membuat bapak dan ibu bangga kepadaku.

Adik-adikku Selvi dan Tirta

Yang telah memberikan semangat, do'a, dan keceriaan dalam hidupku.

Dosen Pembimbing dan Penguji

Yang telah senantiasa mengarahkan dan memberi motivasi kepada penulis

Almamater Tercinta, Universitas Lampung

SANWACANA

Alhamdulillah puji syukur kepada Allah SWT yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga skripsi ini dapat diselesaikan. Shalawat serta salam semoga selalu tercurahkan kepada junjungan alam Nabi Muhammad SAW, penuntun jalan bagi seluruh umat manusia. Skripsi saya yang berjudul "Pemilihan Model Regresi Linear Berganda Terbaik Pada Kasus Multikolinearitas dengan Regresi Komponen Utama dan *Stepwise Regression*" adalah salah satu syarat untuk memperoleh gelar Sarjana Matematika di Universitas Lampung

Penulis menyadari bahwa terselesainya skripsi ini tidak akan terwujud tanpa bantuan dan doa dari mereka yang senantiasa mendukung penulis. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada:

1. Bapak Drs. Eri Setiawan, M.Si., selaku dosen pembimbing I yang telah memberikan bimbingan, pengarahan dan edukasi kepada penulis dalam menyelesaikan skripsi.
2. Bapak Amanto, S.Si., M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan dan saran dalam proses penyusunan skripsi ini.
3. Bapak Prof. Drs. Mustofa Usman, M.A.,Ph.D., selaku dosen pembahas dan penguji skripsi yang telah memberikan kritik dan arahan demi perbaikan skripsi.
4. Bapak Ir. Warsono, M.S., Ph.D., selaku dosen pembimbing akademik.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.

7. Seluruh Dosen dan Staf Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Kedua orang tuaku dan adik yang selalu memberikan do’a dan dukungan tak terhingga serta menjadi penyemangat tersendiri bagi penulis.
9. Kakek dan nenek yang selalu mendoakan dan memberikan dukungan yang tak terhingga bagi penulis.
10. Sahabat-sahabat tersayang (Mayda, Yoga, Mutiara, Deski, Yesi, Novia,Zelva, Umroh, Meta, Suaini, Ginda) dan orang-orang yang hadir dalam kehidupan perkuliahan penulis yang telah membantu, saling menguatkan, memberikan semangat, do’a dan berbagi keceriaan kepada penulis.
11. Yulica, Dewi, Krisdiana, Siti dan teman-teman satu bimbingan yang telah membantu dan saling menyemangati dalam proses penulisan skripsi.
12. Teman-teman Jurusan Matematika angkatan 2017.
13. Seluruh pihak yang telah memotivasi, membantu dan mendoakan penulis yang tidak dapat disebutkan satu per satu.

Bandar Lampung, 05 Agustus 2021
Penulis,

Widiyawati
NPM. 1717031003

DAFTAR ISI

Halaman

IV. HASIL DAN PEMBAHASAN

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis Regresi adalah salah satu metode statistik yang digunakan untuk mengetahui atau menyelidiki hubungan serta membangun hubungan antara dua variabel atau lebih. Variabel tersebut terdiri dari variabel terikat dan variabel bebas. Variabel terikat adalah variabel yang akan diestimasi nilainya dan biasanya diplot pada sumbu tegak (sumbu-Y). Sedangkan variabel bebas adalah variabel yang diasumsikan memberikan pengaruh terhadap variasi variabel terikat dan biasa di plot pada sumbu datar (sumbu-X). Bila dalam analisisnya hanya melibatkan sebuah variabel bebas maka analisis yang digunakan adalah analisis regresi linear sederhana, sedangkan jika dalam analisisnya melibatkan lebih dari satu atau beberapa variabel bebas, maka analisis yang digunakan adalah analisis regresi linear berganda.

Sebuah model regresi linear berganda dikatakan baik atau cocok apabila memenuhi asumsi-asumsi klasik, yaitu galat menyebar normal dengan rata-rata nol, ragam dari galat bersifat homogen, galat tidak mengalami autokorelasi dan tidak terjadi multikolinearitas antar variabel bebas. Apabila terdapat salah satu asumsi regresi yang tidak terpenuhi maka penduga dengan metode kuadrat terkecil tidak lagi efisien untuk digunakan.

Asumsi dalam analisis regresi linear berganda yang memungkinkan tidak terpenuhi adalah terjadinya multikolinearitas. Multikolinearitas merupakan suatu keadaaan dimana terjadinya korelasi antar variabel-variabel bebas dalam analisis regresi linear berganda. Adanya multikolinearitas akan menyebabkan kesimpulan yang didapat dari hasil pengujian untuk model regresi linear berganda maupun untuk masing-masing peubah yang ada dalam model seringkali tidak tepat. Oleh karena itu, masalah multikolinearitas harus diatasi agar estimasi koefisien parameter regresi optimal.

Terdapat beberapa cara yang dapat dilakukan untuk mengatasi masalah multikolinearitas yaitu dengan metode regresi komponen utama dan *stepwise regression*. Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis. Prinsip yang mendasari regresi komponen utama ini adalah dengan mereduksi dimensi variabel yang diamati sehingga menjadi lebih sederhana tanpa kehilangan informasi penting didalamnya. Hasil pereduksian ini dapat mengatasi multikolinearitas yang terdapat antar variabel bebas.

Stepwise regression merupakan prosedur statistik untuk menentukan peubah mana yang akan dimasukkan kedalam persamaan regresi, dengan memperhitungkan korelasi parsial sebagai prosedur dalam analisis. Selain itu, metode *stepwise regression* juga merupakan kombinasi dari metode seleksi langkah maju (*Forward regression)* dan metode eliminasi langkah mundur (*Backward regression)*. Suatu pemodelan atau peramalan dikatan baik apabila selisih antara nilai aktual dengan nilai ramalan mendekati nol. Kriteria yang digunakan dalam pemilihan model terbaik adalah dengan menggunkan kriteria *Mean Square Error* (MSE) dan kriteria R^2 yang disesuaikan (*Adjusted* R^2).

Berdasarkan pemaparan diatas, peneliti tertarik untuk membandingkan metode regresi komponen utama dan metode *stepwise regression* sebagai penyelesaian masalah multikolinearitas. Oleh karena itu peneliti mengangkat judul untuk penelitian ini yaitu Pemilihan Model Regresi Linier Berganda Terbaik Pada Kasus Multikolinearitas Berdasarkan Metode Regresi Komponen Utama dan Metode *Stepwise Regression*.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Mengetahui performa metode regresi komponen utama dan *Stepwise regression* dalam mengatasi masalah multikolinearitas pada regresi linear berganda.
2. Membandingkan metode Regresi Komponen Utama dan *Stepwise Regression* dalam pemilihan model regresi terbaik yang mengandung multikolinearitas dengan kriteria MSE dan *Adjusted* R^2.
3. Menentukan variabel yang mempengaruhi Persentase Penduduk Miskin di Provinsi Indonesia.

1.3 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah :

1. Mendapatkan model regresi terbaik pada kasus multikolinearitass dengan metode regresi komponen utama dan metode s*tepwise regression.*
2. Menambah wawasan bagi penulis tentang metode regresi komponen utama dan metode *stepwise regression*.
3. Penelitian ini dapat dijadikan referensi bagi pembaca dalam melakukan penelitian pada kasus multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu metode dalam statistika yang dapat digunakan untuk memodelkan hubungan antara dua variabel (peubah) atau lebih. Dalam analisis regresi variabel dibedakan menjadi dua jenis yaitu variabel bebas atau variabel prediktor dan juga variabel tak bebas atau terikat atau peubah respon. Variabel terikat adalah variabel yang akan diestimasi nilainya dan biasa diplot pada sumbu tegak (sumbu-Y). Sedangkan variabel bebas adalah variabel yang diasumsikan memberikan pengaruh terhadap variansi variabel terikat dan biasa diplot pada sumbu datar (sumbu-X).

Misalkan diasumsikan model hubungan antara variabel X dan Y adalah linear dan ingin menentukan garis dugaan terbaiknya, maka harus menyadari bahwa garis dugaan dari masalah yang sebenarnya diharapkan mampu memprediksi dengan tepat setiap individu Y oleh setiap individu X. Aspek yang sangat penting dari analisis regresi adalah pengumpulan data karena kesimpulan dari analisis sangat tergantung pada data yang dikumpulkan. Pengumpulan data yang baik akan memberikan banyak manfaat, termasuk penyederhanaan analisis dan membangun model yang secara umum dapat dipergunakan dan dipertanggungjawabkan (Usman dan Warsono, 2000).

Berdasarkan banyak variabel bebas yang digunakan, analisis regresi linear dibagi menjadi dua yaitu, analisis regresi linear sederhana dan analisis regresi linear berganda. Menurut Myers dan Milton (1991), analisis regresi yang hanya melibatkan satu peubah bebas disebut analisis regresi linear sederhana, sedangkan analisis regresi linear dengan peubah respon dipengaruhi oleh lebih dari satu peubah bebas disebut analisis regresi linier berganda.

2.2 Analisis Regresi Linear Berganda

Analisis regresi linear berganda merupakan analisis regresi yang memiliki variabel bebas lebih dari satu. Analisis regresi linear berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel bebas ($X_1, X_2, \ldots, X_k$) terhadap variabel tak bebas (Y). Secara umum model regresi linear berganda dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad (2.1)$$

dengan,

i	= banyaknya pengamatan dengan	i = 1,2,3,…,n.
k	= banyaknya variabel bebas dengan	k = 1,2,3,…,k.
Y_i	= variabel terikat pengamatan ke-i	
X_{ki}	= variabel bebas pengamatan ke-i.	
β_0	= intersep.	
β_k	= koefisien regresi atau *slope* (parameter) ke-k.	
ε_i	= galat pengamatan ke-i.	

Dalam regresi linear berganda yang akan diduga adalah β_0 dan β_k artinya $(\beta_1, \beta_2, \ldots, \beta_k)$. Persamaan linear untuk pendugaan garis regresi linear ditulis dalam bentuk.

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} \quad (2.2)$$

dengan,

$\hat{y}_i$ = nilai dugaan variabel terikat pengamatan ke-i

x_{ki} = nilai variabel bebas pengamatan ke-i

b_0 = titik potong garis regresi pada sumbu-y atau nilai dugaan $\hat{y}$ bila x sama dengan nol

b_k = gradien garis regresi (perubahan nilai dugaan $\hat{y}$ per satuan perubahan nilai x) ke-k

Model regresi linear berganda juga dapat ditulis dalam bentuk matriks yaitu:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (2.3)$$

dengan,

Y = vektor pengamatan n x 1

X = matriks variabel bebas berukuran n x k

β = vektor parameter yang akan ditaksir berukuran k x 1

ε = vektor galat berukuran n x 1

2.3 Asumsi Regresi Linear Berganda

Uji asumsi klasik terhadap model regresi linear berganda dilakukan agar dapat mengetahui bahwa model tersebut baik atau tidak untuk digunakan. Tujuan pengujian asumsi ini adalah untuk memberikan kepastian bahwa persamaan regresi yang diperoleh memiliki ketepatan dalam estimasi, tidak bias, dan konsisten. Sehingga, sebelum melakukan analisis regresi berganda maka terlebih dahulu dilakukan pengujian asumsi.

Menurut Chatterjee & Hadi (2006), untuk melakukan analisis regresi linear berganda harus memenuhi asumsi-asumsi berikut:

1. Normalitas

   Asumsi normalitas harus ter penuhi untuk mengetahui apakah residual atau galat dari data berdistribusi normal atau untuk mengetahui apakah data sampel berasal dari populasi berdistribusi normal. Uji statistik yang sering digunakan adalah Kolmogorov-Smirnov.
2. Autokorelasi

   Autokorelasi terjadi karena galat antar pengamatan tidak saling bebas dan berkaitan antara satu dengan lainnya pada model regresi. Persyaratan yang harus dipenuhi adalah tidak adanya autokorelasi dalam model regresi.
3. Homoskedastisitas

   Homoskedastisitas dapat diartikan sebagai distribusi dari galat yang memiliki ragam yang konstan (homogen). Apabila varian galat dalam model tidak konstan disebut heteroskedatisitas. Hesteroskedatisitas disebabkan karena variabel yang digunakan memiliki nilai yang sangat beragam, sehingga menghasilkan nilai galat yang tidak konstan.
4. Multikolinearitas

   Salah satu syarat yang harus dipenuhi dalam pembentukan model regresi dengan beberapa variabel bebas adalah tidak adanya kasus multikolinearitas, yaitu tidak terdapat korelasi antara satu variabel bebas dengan variabel bebas lainnya. Dalam model regresi, adanya korelasi antar variabel bebas menyebabkan taksiran parameter regresi yang dihasilkan akan memiliki error yang sangat besar. Pendeteksian kasus multikolinearitas dilakukan menggunakan kriteria VIF (*Varians Inflation Factor*) lebih besar dari 10 menunjukkan adanya multikolinearitas antar variabel bebas.

2.4 Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frich pada tahun 1934 yang artinya terdapat hubungan linear diantara beberapa atau semua variabel bebas dalam model regresi. Masalah multikolinearitas hanya ditemukan pada regresi linear berganda. Model yang baik adalah model yang bebas dari multikolinearitas. Suatu model yang bebas dari multikolinearitas adalah model yang memiliki nilai faktor *Variance Inflation Factor* (*VIF*) < 10. Nilai *VIF* >10 mengindikasi terdapat multikolinearitas (Myers, 1990).

Selain melihat nilai *Variance Inflation Factor* (VIF) multikolinearitas dapat dianalisis dengan melihat koefisien korelasi sederhana antar variabel bebasnya. Multikolinearitas dapat diduga dari tingginya nilai korelasi antar variabel bebasnya yang memiliki nilai korelasi $0{,}8 \leq r \leq 1$.

Menururt Gujarati (1995), salah satu cara untuk menguji gejala multikolinearitas dalam model regresi adalah dengan melihat nilai TOL (*tolerance*) dan VIF (*Variance Inflation Factor*) dari masing-masing variabel bebas terhadap variabel terikatnya. Nilai VIF dapat dicari menggunakan rumus sebagai berikut:

$$\text{TOL} = (1 - R_j^2) \tag{2.4}$$

$$VIF = \frac{1}{TOL}. \tag{2.5}$$

R_j^2 merupakan koefisien determinasi yang didapat dari variabel bebas X_j yang diregresikan dengan variabel bebas lainnya. Jika X_j tidak berkorelasi dengan peubah bebas lainnya, maka nilai R_j^2 akan bernilai kecil dan nilai VIF akan mendekati 1.

Dampak multikolinearitas yaitu:

a. Koefisien partial regresi tidak terukur secara presisi. Oleh karena itu, nilai standar errornya besar.
b. Perubahan kecil pada data dari sampel ke sampel akan menyebabkan perubahan vektor pada nilai koefisien regresi partial.
c. Perubahan pada satu variabel dapat menyebabkan perubahan besar pada nilai koefisien regresi parsial variabel lainnya.
d. Nilai selang kepercayaan sangat lebar, sehingga akan menjadi sangat sulit untuk menolak hipotesis nol pada sebuah penelitian jika dalam penelitian tersebut terdapat multikolinearitas.

2.5 Analisis Komponen Utama

Metode analisis komponen utama (AKU) merupakan teknik statistika yang digunakan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel melalui bebarapa variabel baru dimana variabel baru ini saling bebas dan merupakan kombinasi linear dari variabel asal. Selanjutnya variabel baru ini disebut sebagai komponen utama. Analisis komponen utama bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan dimensinya. Hal ini dilakukan dengan menghilangkan korelasi variabel melalui transformasi variabel asal ke varaibel baru yang tidak berkorelasi.

Nilai *eigen* merupakan keragaman dari setiap komponen utama. Ketika nilai *eigen* dari komponen utama mendekati nol hal tersebut juga mengindentifikasikan adanya multikolinearitas pada data asli.

Variabel baru (Q) disebut komponen utama yang merupakan hasil transformasi dari variabel asal X yang modelnya dalam matriks adalah sebagai berikut:

$$\boldsymbol{Q = VX} \tag{2.6}$$

dengan,

Q = vektor variabel baru komponen utama yang merupakan hasil transformasi dari variabel asal X

V = matriks yang melakukan transformasi terhadap varaibel asal X sehingga diperoleh komponen vektor **Q**

komponen utama pertama adalah kombinasi linear terbobot variabel asal yang dapat menerangkan keragaman terbesar (Gasperz, 1991). Komponen utama pertama dapat dituliskan sebagai

$$Q_1 = v_{11}X_1 + v_{21}X_2 + \cdots + v_{p1}X_p = v_1^T X$$

Komponen utama kedua dapat dituliskan sebagai

$$Q_2 = v_{12}X_1 + v_{22}X_2 + \cdots + v_{p2}X_p = v_2^T X$$

Secara umum komponen utama ke-j dapat dituliskan sebagai

$$Q_j = v_{1j}X_1 + v_{2j}X_2 + \cdots + v_{pj}X_p = v_j^T X\beta \tag{2.7}$$

Setelah beberapa komponen hasil AKU yang bebas multikolinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisis pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi. Keunggulan metode AKU diantaranya adalah dapat menghilangkan korelasi secara bersih tanpa harus mengurangi jumlah variabel asal

2.6 Regresi Komponen Utama

Regresi komponen utama merupakan salah satu analisis regresi yang menggunakan komponen utama untuk mengatasi masalah multikolinearitas pada regresi berganda. Analisis komponen utama merupakan analisis yang memperkecil dimensi variabel tanpa kehilangan banyak informasi didalamnya, dengan tujuan menyederhanakan variabel yang diamati dengan mereduksi dimensinya. Regresi komponen utama menggunakan analisis komponen utama sebagai variabel bebas dan meregresikan menggunakan *scores* komponen utama tersebut sebagai variabel bebas terhadap variabel terikat.

Menurut Montgomery and Peck (2006) bentuk umum persamaan regresi komponen utama adalah sebagai berikut:

$$\boldsymbol{Y} = \boldsymbol{Q}\boldsymbol{a} + \boldsymbol{\varepsilon} \tag{2.8}$$

dimana,

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{V} \qquad \boldsymbol{a} = \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{\beta}$$

dengan,

$\boldsymbol{Y}$ = vektor variabel tak bebas

$\boldsymbol{X}$ = matriks yang elemen-elemennya dikurang dengan rataannya yang mensyaratkan rataan 0 dan standar deviasi 1.

$\boldsymbol{Q}$ = matriks berukuran n x k yang elemennya terdapat komponen utama dimana $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{V}$

$\boldsymbol{a}$ = vektor koefisien komponen utama berukuran k x 1

$\boldsymbol{\varepsilon}$ = vektor galat berukuran n x 1

Regresi komponen utama menggunakan estimasi kuadrat terkecil namun menggunakan skor (nilai) komponen utama sebagai penjelas (variabel bebas). Secara umum estimasi dengan regresi komponen utama dengan fungsi objektif adalah sebagai berikut:

$$\hat{a} = \arg\min \|y - Qa\|_2^2 = \arg\min \left\| y - \sum_{i=1}^{n} a_i \tilde{q}_i \right\|_2^2$$

sehingga,

$$Q(\hat{a}_i) = \sum_{i=1}^{n} (y_i - \hat{a}_0 - \hat{a}_1 q_{1i} - \hat{a}_2 q_{2i} - \cdots - \hat{a}_k q_{ki})^2$$

Apabila disusun kembali dalam bentuk matriks adalah sebagai berikut:

$$[\boldsymbol{Q^T Q}]\boldsymbol{\hat{a}} = \boldsymbol{Q^T y}$$

$$[\boldsymbol{Q^T Q}]^{-\boldsymbol{1}}\boldsymbol{\hat{a}} = [\boldsymbol{Q^T Q}]^{-\boldsymbol{1}}\boldsymbol{Q^T y}$$

$$\boldsymbol{\hat{a}} = [\boldsymbol{Q^T Q}]^{-\boldsymbol{1}}\boldsymbol{Q^T y} \qquad (2.9)$$

$$\boldsymbol{\hat{\beta}} = \boldsymbol{v\hat{a}} \qquad (2.10)$$

Apabila dibentuk dalam matriks

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = [v_1 \quad v_2 \quad \cdots \quad v_k] \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_k \end{bmatrix}$$

Dalam analisis regresi komponen utama, tidak digunakan semua komponen Q melainkan memilih komponen utama berdasarkan kriteria tertentu. Menurut D.F. Marison tahun 1976 dalam edisi kedua *Multivariat Statistical Methoda*, komponen-komponen dapat dihitung sampai sejumlah tertentu proporsi keragaman (>75%) yang telah terjelaskan (Draper and Smith, 1992).

Dengan semua variabel terstandarisasi sehingga $\boldsymbol{X^T X}$ merupakan matriks korelasi dari X dan $\boldsymbol{X^T Y}$ merupakan vektor korelasi antara X dan Y. Misalkan $\boldsymbol{V} = [\boldsymbol{V_1}, \cdots, \boldsymbol{V_p}]$ merupakan matriks berukuran p x p yang kolom-kolomnya eigen vektor ternormalisasi dari $\boldsymbol{X^T X}$ dan misalkan $\lambda_1 \quad \cdots \quad \lambda_p$ merupakan nilai eigen yang bersesuaian. Misalkan $\boldsymbol{Q} = [\boldsymbol{Q_1}, \cdots, \boldsymbol{Q_p}] = \boldsymbol{XV}$. Sehingga $\boldsymbol{Q_j} = \boldsymbol{XV_j}$ adalah komponen utama ke-j dari X.

Beberapa hal penting dari regresi komponen utama adalah:

1. $\boldsymbol{V^T V} = \boldsymbol{V V^T} = \mathbf{I_p}$, dimana matriks **V** ortonormal
2. $\boldsymbol{Q^T Q} = \boldsymbol{\Lambda}$ dimana $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$, **Q** orthogonal dan $|\boldsymbol{Q_j}| = \sqrt{\boldsymbol{\lambda_j}}$
3. $\boldsymbol{X} = \boldsymbol{Q V^T}$ dan $X_i = \sum_{k=1}^{p} v_{jk} Q_k$

Model regresi dapat ditulis sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \boldsymbol{XVV^T\beta} + \boldsymbol{\epsilon} = \boldsymbol{Qa} + \boldsymbol{\epsilon} \qquad \text{dimana, } \boldsymbol{a} = \boldsymbol{V^T\beta}$$

Berdasarkan persamaan ini, estimasi kuadrat terkecil dari **a** adalah

$$\boldsymbol{\hat{a}} = (\boldsymbol{Q^T Q})^{-1}\boldsymbol{Q^T Y} = \boldsymbol{\Lambda^{-1} Q^T Y}$$

Dan estimator regresi komponen utama adalah sebagai berikut.

$$\boldsymbol{\hat{\beta}} = \boldsymbol{V\hat{a}}$$

$$\boldsymbol{\hat{\beta}} = \boldsymbol{V\Lambda^{-1}Q^T Y}$$

Apabila semua komponen digunakan maka penduga dengan menggunakan regresi komponen utama sama dengan penduga yang menggunakan regresi kuadrat terkecil, namun dalam praktiknya hanya sebuah subset $\boldsymbol{Q}_{(s)} = [\boldsymbol{Q_1}, \cdots, \boldsymbol{Q_p}]$ dari komponen utama yang digunakan dalam menduga a sehingga

$$\boldsymbol{\hat{a}} = \left(\boldsymbol{Q}_{(s)}^{\boldsymbol{T}}\boldsymbol{Q}_{(s)}\right)^{-1}\boldsymbol{Q}_{(s)}^{\boldsymbol{T}}\boldsymbol{Y} = \boldsymbol{\Lambda}_{s}^{-1}\boldsymbol{Q}_{(s)}^{\boldsymbol{T}}\boldsymbol{Y}$$

Sehingga estimator regresi komponen utama dari subset ini adalah

$$\boldsymbol{\hat{\beta}} = \boldsymbol{V}_{(s)}\boldsymbol{\Lambda}_{s}^{-1}\boldsymbol{Q}_{(s)}^{\boldsymbol{T}}\boldsymbol{Y} \tag{2.11}$$

Adapun sifat dari regresi komponen utama adalah

1. Bersifat Bias

$$
\begin{aligned}
E(\widehat{\boldsymbol{\beta}}) &= E(\boldsymbol{V\Lambda^{-1}Q^{T}Y}) \\
&= E(\boldsymbol{V}(\boldsymbol{Q^{T}Q})^{-1}\boldsymbol{Q^{T}Y}) \\
&= E(\boldsymbol{V}(\boldsymbol{Q^{T}Q})^{-1}\boldsymbol{Q^{T}Q\alpha} + \boldsymbol{\varepsilon}) \\
&= E(\boldsymbol{V}(\boldsymbol{Q^{T}Q})^{-1}\boldsymbol{Q^{T}QV^{T}\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{V^{T}\beta}
\end{aligned}
$$

Penduga regresi komponen utama merupakan penduga yang bias karena

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{V^{T}\beta} \neq \boldsymbol{\beta}$$

2. Menurut Gaspersz (1991), ragam dugaan dari koefisien regresi komponen utama

$$V(\hat{a}_j) = \frac{1}{\lambda_j}s^2; j = 1,2,\cdots,m$$

dengan,

$V(\hat{a}_j)$ = *varians* dugaan koefisien regresi komponen utama

λ_j = Akar ciri ke-j

s^2 = Ragam dugaan dari bentuk galat

Sehingga penduga *varians* dari koefisien b dengan persamaan regresi asli adalah sebagai berikut:

$$V(b_i) = s^2 \sum_{j=1}^{m} v_{ij}^2/\lambda_j \tag{2.12}$$

dengan,

$V(b_i)$ = *varians* dugaan koefisien regresi komponen utama dengan persamaan asli

λ_j = Akar ciri ke-j

s^2 = *varians* dugaan dari bentuk galat

v_{ij} = komponen dari vektor ciri ke-I yang berhubungan dengan akar ciri ke-j

2.7 *Stepwise Regression*

Metode *stepwise regression* merupakan sebuah metode yang dapat digunakan dalam menentukan model regresi linear berganda. Prosedur yang dilakukan yaitu dengan menyusupkan satu demi satu variabel bebas sampai diperoleh persamaan regresi yang dibutuhkan. Urutan penyisipan ditentukan dari koefisien korelasi parsial (r) (Draper & Smith, 1992). Jadi, variabel bebas yang mempunyai korelasi terbesar dengan variabel terikat akan masuk pertama kali dalam model.

Apabila salah satu peubah telah dimasukkan ke dalam model regresi, maka peubah lainnya tidak perlu dimasukkan lagi ke dalam model regresi, karena pengaruhnya telah diwakili oleh peubah yang sudah masuk ke dalam model regresi. Sehingga tidak terdapat multikolinearitas pada model regresi yang dihasilkan (Sembiring, 1995).

Tahap penyelesaian pada metode *stepwise* yaitu membentuk matriks koefisien korelasi. Dimana koefisien regresi yang dibutuhkan adalah koefisien korelasi linear sederhana Y dan X_i dengan rumus:

$$r_{yx_i} = \frac{\sum(X_{ij}-\bar{X}_i)(Y_j-\bar{Y})}{\sqrt{\sum(X_{ij}-\bar{X}_i)^2\sum(Y_j-\bar{Y})^2}} \qquad (2.12)$$

Oleh karena itu, dapat dibentuk matriks koefisien korelasi linier sederhana antara Y dan X_i.

$$\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Kemudian membentuk persamaan regresi linear yang pertama. Variabel yang pertama diregresikan adalah variabel yang mempunyai harga mutlak koefisien korelasi yang terbesar antara Y dan X_i, misalnya X_i.

$$X = \begin{pmatrix} 1 & X_{h1} \\ 1 & X_{h2} \\ \vdots & \vdots \\ 1 & X_{hn} \end{pmatrix} (X^T X)^{-1} = \begin{pmatrix} n & \sum X_h \\ \sum X_h & \sum X_h^2 \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} X^T Y = \begin{pmatrix} \sum Y \\ \sum YX_i \end{pmatrix}$$

Langkah selanjutnya yaitu seleksi variabel kedua diregresikan. Cara menyeleksi variabel yang diregresikan adalah memilih parsial korelasi variabel sisa terbesar. Untuk menghitung harga masing-masing parsial korelasi sisa digunakan rumus:

$$r_{YX_h.X_k} = \frac{r_{YX_h} - r_{YX_k} r_{X_{h.}X_k}}{\sqrt{\left(1 - r_{YX_k}^2\right)\left(1 - r_{X_h X_k}^2\right)}} \qquad (2.13)$$

Tahap selanjutnya adalah memasukkan variabel bebas yang memiliki korelasi terbesar kedua dengan variabel terikat. Tahap ini terus dilakukan sampai semua variabel masuk dalam model, kemudian membuat persamaan regresi kedua :

$$Y = b_0 + b_1 X_1 + \cdots + b_k X_k \qquad (2.14)$$

Setelah diperoleh satu variabel bebas yang masuk dalam model perlu dilakukan pengujian keberartian variabel dalam model.

Adapun algoritma metode *stepwise*:

1. Menetukan matriks koefisien korelasi antara variabel terikat (Y) terhadap variabel bebas (X).
2. Pemilihan variabel yang pertama diregresikan yaitu variabel yang mempunyai harga mutlak koefisien korelasi terbesar terhadap reponden (Y).
3. Pembentukan regresi pertama yaitu regresi sederhana untuk variabel terpilih pada langkah kedua dan menguji keberartian regresi.

4. Pemilihan variabel kedua diregresikan. Bila pada langkah ketiga ternyata tidak tolak H_0 maka dilakukan pemilihan variabel kedua untuk diregresikan selanjutnya. Variabel terpilih adalah variabel sisa (di luar regresi) yang mempunyai parsial korelasi terbesar.
5. Pembentukan regresi kedua yaitu merupakan regresi ganda dan menguji keberartian regresi. Bila tidak signifikan maka proses dihentikan, sedangkan apabila signifikan maka seluruh variabel tetap.
   Pembentukan penduga apabila proses pemasukan variabel terhadap regresi sudah selesai, maka ditetapkan persamaan regresi yang menjadi penduga linear yang diinginkan yaitu merupakan persamaan regresi yang diperoleh terakhir.

2.8 Pendeteksi Kelayakan Model Regresi Terbaik

Salah satu tujuan di dalam analisis regresi adalah untuk mendapatkan model terbaik yang menjelaskan hubungan antara variabel terikat dengan variabel bebas. Model terbaik adalah model yang seluruh koefisien regresinya berarti (signifikan) dan mempunyai kriteria model terbaik optimum.

Suatu model regresi dikatakan layak dan terbaik apabila model tersebut memenuhi kriteria sebagai berikut:

2.8.1 Nilai Adjusted R-Square

Karena adanya kelemahan dalam perhitungan R^2, banyak peneliti yang menyarankan untuk menggunakan *Adjusted* R-Square. Interpretasinya sama dengan R-Square, akan tetapi nilai *adjusted* r-square dapat naik atau turun dengan adanya penambahan variabel baru, tergantung dari korelasi antara variabel bebas tambahan tersebut dengan variabel terikatnya. Nilai *adjusted* R-square dapat bernilai negatif, sehingga jika nilainya negatif maka nilai tersebut dianggap nol, atau variabel bebas sama sekali tidak mampu menjelaskan varians dari variabel terikatnya.

Suatu sifat penting R^2 adalah nilainya merupakan fungsi yang tidak pernah menurun dari banyaknya variabel bebas yang ada dalam model. Oleh karena itu, untuk membandingkan dua R^2 dari dua model harus memperhitungkan banyaknya variabel bebas yang ada dalam model tersebut. Ini dapat dilakukan dengan menggunakan "*adjusted* R-square". Istilah penyesuaian berarti nilai R^2 sudah disesuaikan dengan banyaknya variabel (derajat bebas) dalam model. Memang, R^2 yang disesuaikan ini juga akan meningkat bersamaan dengan meningkatnya jumlah variabel, tetapi peningkatan relatif kecil. Seringkali juga disarankan, jika variabel bebas lebih dari dua, sebaiknya menggunakan adjusted R square.

Menurut Sembiring (1995), statistik R^2 yang sering disebut dengan koefisien determinasi. Sedangkan *adjusted* R^2 memperhitungkan banyaknya variabel bebas dalam model serta telah disesuaikan terhadap derajat bebas masing-masing jumlah kuadrat. *Adjusted* R^2 dirumuskan dengan:

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{(N-1)}{(N-p)} \qquad (2.15)$$

dengan,

R^2 = Sampel R-square

p = banyaknya parameter termasuk Y

N = banyaknya pengamatan

2.8.2 *Mean Square Error* (MSE)

Menurut Ghazali(2006), MSE merupakan salah satu pengukuran kesalahan yang populer dan mudah digunakan. Umumnya, semakin kecil MSE semakin akurat nilai suatu peramalan atau suatu pemodelan. Selain itu dalam kasus multikolinearitas metode terbaik diartikan sebagai metode yang dapat melakukan perbaikan masalah multikolinearitas lebih baik dari metode yang lain. Efisiensi dari metode ini untuk menangani multikolinearitas akan dievaluasi berdasarkan rata-rata dari *Mean Square Error* (MSE) dari hasil estimasi parameter $\hat{\beta}$, yang didefinisikan sebagai berikut:

$$MSE(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}(\hat{\beta}_j - \beta)^2 \; ; \;\; j = 1,2,\dots,m \qquad (2.16)$$

dengan,

$\hat{\beta}_j$ = Penduga parameter regresi

β = Parameter regresi

m = Banyaknya pengulangan

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2020/2021 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah.simulasi yang mengandung multikolinearitas dan data real tentang persentase penduduk miskin tahun 2019. Untuk data simulasi yakni data yang dibangkitkan dengan variabel bebas sebanyak p=6 dan n observasi yaitu n = 25, 50, dan 75 yang diulang sebanyak 1000 kali.

Untuk mendapatkan data multikolinearitas pada setiap himpunan data, X_p dibangkitkan menggunakan simulasi Monte Carlo dengan persamaan sebagai berikut:

$$X_p = \sqrt{(1-\rho^2)z_{ij}} + \rho z_{i(p+1)} \qquad ; i = 1,2, \ldots, n \quad ; j = 1,2, \ldots, p$$

dengan $z_{i1}, z_{i2}, \dots, z_{i(p+1)}$ merupakan data yang dibangkitkan dalam bentuk normal standar atau distribusi normal N(0,1) dan ρ ditentukan sehingga korelasi antarvariabel bebas diberikan oleh ρ^2. Dua himpunan dari variabel yang saling berkorelasi dalam penelitian ini dibuat berdasarkan nilai ρ = 0.99 dan ρ = 0.3. Untuk lebih jelasnya dapat dilihat pada tabel berikut.

Tabel 1. Simulasi Data X_{ip} untuk 6 Variabel Bebas

Jumlah Variabel Bebas	Simulasi Data X_{ip}
6	$Z_{ij} = N(0,1)$ j = 1, 2,…, 7 $X_{i1} = (1-\rho^2)^{1/2} z_{i1} + \rho z_{i7}$ $X_{i2} = (1-\rho^2)^{1/2} z_{i2} + \rho z_{i7}$ $X_{i3} = (1-\rho^2)^{1/2} z_{i3} + \rho z_{i7}$ $X_{i4} = (1-\rho^2)^{1/2} z_{i4} + \rho z_{i7}$ $X_{i5} = (1-\rho^2)^{1/2} z_{i5} + \rho z_{i7}$ $X_{i6} = (1-\rho^2)^{1/2} z_{i6} + \rho z_{i7}$

Variabel terikat (Y) untuk setiap p variabel bebas diperoleh berdasarkan model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ dimana β adalah $\beta_{i,j} = 1$; untuk i = j, dan 0 selainnya. Dengan ε dibangkitkan berdasarkan distribusi normal N(0,1) sehingga Y merupakan kombinasi linear dari p variabel bebas ditambah galat yang ditampilkan pada tabel berikut.

Tabel 2. Variabel Terikat (Y) untuk 6 Variabel Bebas

P	Y
6	$Y = X_1 + X_2 + \cdots + X_p + N(0,1)$ p = 1,2, …,6

Sedangkan data real yang digunakan untuk menduga adanya multikolinearitas adalah data Persentase Penduduk Miskin beberapa provinsi di Indonesia untuk variabel terikat Y yang diperoleh dari Badan Pusat Statistik (BPS). Faktor-faktor yang mempengaruhi Persentase Penduduk Miskin adalah sebagai berikut:

X1 = Indeks Pembangunan Manusia(IPM)

X2 = Pengeluaran PerKapita Disesuaikan

X3 = Angka Partisipasi Sekolah

X4 = Kepadatan Penduduk

X5 = Harapan Lama Sekolah

X6 = Rata-rata lama sekolah

Tabel 3. Data Persentase Penduduk Miskin Tahun 2019

No	Provinsi	Y	X1	X2	X3	X4	X5	X6
1	ACEH	15,01	71,9	9603	32,54	93	14,3	9,18
2	SUMATERA UTARA	8,63	71,74	10649	25,75	200	13,15	9,45
3	RIAU	6,9	73	11255	28,16	80	13,14	9,03
4	JAMBI	7,51	71,26	10592	23,32	72	12,93	8,45
5	BENGKULU	14,91	71,21	10409	30,71	100	13,59	8,73
6	LAMPUNG	12,3	69,57	10114	20,69	244	12,63	7,92
7	KEP. BANGKA BELITUNG	4,5	71,3	12959	17,01	91	11,94	7,98
8	DKI JAKARTA	3,42	80,76	18527	24,52	15900	12,97	11,06
9	JAWA BARAT	6,82	72,03	11152	22,71	1394	12,48	8,37
10	JAWA TENGAH	10,58	71,73	11102	22,41	1058	12,68	7,53
11	DI YOGYAKARTA	11,44	79,99	14394	51,85	1227	15,58	9,38
12	JAWA TIMUR	10,2	71,5	11739	24,8	831	13,16	7,59
13	BANTEN	4,94	72,44	12267	21,43	1338	12,88	8,74
14	BALI	3,61	75,38	14146	27,86	750	13,27	8,84
15	NUSA TENGGARA BARAT	13,88	68,14	10640	25,59	273	13,48	7,27
16	NUSA TENGGARA TIMUR	20,62	65,23	7769	29,27	112	13,15	7,55
17	KALIMANTAN TENGAH	4,81	70,91	11236	23,98	18	12,57	8,51
18	KALIMANTAN SELATAN	4,47	70,72	12253	24,34	110	12,52	8,2
19	KALIMANTAN TIMUR	5,91	76,61	12359	29,89	29	13,69	9,7
20	KALIMANTAN UTARA	6,49	71,15	9343	23,11	10	12,84	8,94
21	SULAWESI UTARA	7,51	72,99	11115	22,55	181	12,73	9,43
22	SULAWESI SELATAN	8,56	71,66	11118	34,44	189	13,36	8,26
23	SULAWESI TENGGARA	11,04	71,2	9436	31,27	71	13,55	8,91
24	GORONTALO	15,31	68,49	10075	30,97	107	13,06	7,69
25	MALUKU	17,65	69,45	8887	38,58	38	13,94	9,81

3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka yaitu mempelajari buku-buku penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dan mendapatkan hasil yang akurat menggunakan *software* R-Studio. Adapun langkah-langkah yang dilakukan dalam penelitian ini yaitu :

a. Melakukan Simulasi Data
b. Mengidentifikasi multikolinearitas dengan melihat nilai korelasi dan nilai VIF (*Variance Inflation Factor)*
c. Melakukan analisis dengan menggunakan regresi komponen utama
   Dengan langkah-langkah sebagai berikut:
   a. Melakukan *center and scale* data
   b. Membuat matriks korelasi dari *center and scale* data
   c. Menghitung nilai eigen $\lambda_i$dan vector eigen V dari matriks korelasi
   d. Menghitung nilai komponen utama Q
   e. Memilih komponen utama yang memiliki nilai eigen lebih dari 1
   f. Menghitung nilai duga regresi komponen utama berdasarkan komponen yang terpilih
   g. Mentransformasi nilai duga β regresi komponen utama kedalam bentuk variabel asli terstandarisasi dengan mensubstitusi nilai komponen utama Q
   h. Menstransformasi kembali nilai duga β regresi komponen utama kedalam bentuk variabel asli
d. Mencatat nilai duga β metode regresi komponen utama
e. Melakukan analisis dengan menggunakan regresi *stepwise*
   Dengan langkah-langkah sebagai berikut:
   a. Hitung koefisien korelasi antara variabel independen dengan variabel dependen
   b. Pilih variabel independen yang mempunyai korelasi tertinggi dengan variabel dependen
   c. Kemudian, regresikan antar variabel independen sehingga mendapatkan nilai korelasi parsial.

d. Lakukan pengujian parameter, apabila hasil pengujian menyimpulkan bahwa variabel independen berpengaruh, maka tambahkan variabel independen tertinggi berikutnya. Apabila sebaliknya, maka hilangkan variabel independen tersebut dari model.

f. Mencatat nilai MSE dengan metode regresi *stepwise*

g. Mencatat nilai *adjusted R-square* metode regresi komponen utama dan regresi *stepwise*

h. Menghitung MSE dengan metode regresi komponen utama dengan rumus:

$$MSE(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta\right)^2 ; \quad j = 1,2,\dots,m \tag{3.1}$$

i. Membandingkan nilai MSE dan *adjusted R-square* metode regresi komponen utama dan regresi *stepwise*.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan maka diperoleh kesimpulan sebagai berikut:

1. Metode Regresi Komponen Utama dan *Stepwise Regression* dapat menangani multikolinearitas pada model regresi linear berganda yang mengalai multikolinearitas
2. Berdasarkan nilai MSE, *Adjusted* R^2, dan Rataan Kuadrat Sisa (S^2), model regresi linear berganda yang diperoleh dari *Stepwise Regression* lebi baik dibandingkan Regresi Komponen Utama ada data yang mengandung multikolinearitas. Semakin besar ukuran sampelyang digunakan semakin kecil nilai MSE sehingga baik dalam menentukan nilai suatu pemodelan.
3. Berdasarkan nilai uji t untuk melihat signifikansi parameter dilakukan dengan menggunakan data real pada metode *stepwise regression* diperoleh variabel indeks pembangunan manusia,kepadatan penduduk , dan harapan lama sekolah mempengaruhi persentase penduduk miskin, sedangkan metode regresi komponen utama diperoleh variabel indeks pembangunan manusia, pengeluaran per-kapita disesuaikan, dan kepadatan penduduk yang mempengaruhi persentase penduduk miskin.

DAFTAR PUSTAKA

Badan Pusat Statistik. 2018. *Tabel Dinamis Tahun 2018*. Badan Pusat Statistik, Jakarta.

Chatterjee, S. & Hadi, A.S. 2006. *Regression Analysis by Example*. Ed. Ke4. John Wiley & Sons Inc., New Jersey.

Draper, N.R. & Smith, H. 1992. *Analisis Regresi Terapan*. Ed. Ke-2. Diterjemahkan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.

Gaspersz, V. 1991. *Ekonometri Terapan.* Ed. Ke-2. Tarsito, Bandung.

Ghozali, I. 2006. *Aplikasi Analisis Multivariate dengan Program SPSS*. Ed. Ke-4. Badan Penerbit Universitas Diponegoro, Semarang.

Montgomery, D.C. & Peck, A..E. 2006. *Introduction to Linear Regression Analysis.* Wiley-Intersection Publication, New York.

Morrison, D.F. 1978. *Multivariate Statistical Methods Series in Probability and Statistics*. Mc Graw Hill, Singapore.

Myers, R.H. 1990. *Clasical and Modern Regression With Application.* PWSKENT publishing Company, Boston.

Pujilestari, S., *et al.* 2017. Pemilihan Model Regresi Linear Berganda Terbaik Pada Kasus Multikolinearitas Berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise. UNNES Journal of Mathematics. **6**(1): 70-81.

Sembiring, R.K. 2003. *Analisis Regresi.* Ed. Ke-2. Institut Teknologi Bandung, Bandung.

Usman, M. dan Warsono. 2000. *Teori Model Linear dan Aplikasinya*. C.V. Darmajaya, Bandar Lampung.