**PERBANDINGAN METODE REGRESI *RIDGE* DAN METODE REGRESI KOMPONEN UTAMA DALAM MENANGANI MULTIKOLINEARITAS**

**(Skripsi)**

**Oleh**

**SITI ULFA NABILA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2018**

# ABSTRACT

## COMPARISON STUDY OF *RIDGE* REGRESION METHOD AND PRINCIPAL COMPONENT REGRESSION METHOD IN HANDLING MULTICOLLINEARITY

**By**

**SITI ULFA NABILA**

*Ridge* regression and Principal Component Regression are methods that can solve multicollinearity. *Ridge* regression solves multicollinearity by adding a bias constant to diagonal $X^TX$ matrix while Principal Component Regression solves multicollinearity by reducing the dimension of independent variables without losing any important information in it. The purpose of this study is to estimate regression parameters by using *ridge* regression, Principal Component Regression and OLS and to estimate the best method for handling multicollinearity. The results show that Principal Component Regression gives better estimator in handling multicollinearity than *ridge* regression and OLS based on the value of regression coefficient, MSE and AMSE .



Keywords: *ridge* regression, Principal Component Regression, multicollinearity, MSE, AMSE

# ABSTRAK

## PERBANDINGAN METODE REGRESI *RIDGE* DAN REGRESI KOMPONEN UTAMA DALAM MENANGANI MULTIKOLINEARITAS

**Oleh**

**SITI ULFA NABILA**

Regresi *ridge* dan Regresi Komponen Utama merupakan metode - metode yang dapat mengatasi multikolinearitas. Regresi *ridge* mengatasi multikolinearitas dengan menambahkan konstanta bias pada diagonal matriks $X^T X$ sedangkan Regresi Komponen Utama mengatasi multikolinearitas dengan mereduksi dimensi variabel bebas menjadi lebih sederhana tanpa kehilangan informasi penting didalamnya. Tujuan dari penelitian ini adalah menduga parameter regresi dengan menggunakan metode regresi *ridge*, metode Regresi Komponen Utama dan MKT dan mengetahui metode terbaik dalam menangani multikolinearitas. Hasil dari penelitian ini menunjukkan bahwa Regresi Komponen Utama lebih baik dalam menangani multikolinearitas dibandingkan dengan metode regresi *ridge* dan MKT berdasarkan nilai dari koefisien regresi, MSE dan AMSE.



Kata Kunci: regresi *ridge*, Regresi Komponen Utama, multikolinearitas, MSE, AMSE

# PERBANDINGAN METODE REGRESI *RIDGE* DAN METODE REGRESI KOMPONEN UTAMA DALAM MENANGANI MULTIKOLINEARITAS

**Oleh**

**SITI ULFA NABILA**

**Skripsi**

Sebagai Salah Satu Syarat untuk Mencapai Gelar
**SARJANA SAINS**

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2018**

Judul Skripsi : **PERBANDINGAN METODE REGRESI *RIDGE* DAN METODE REGRESI KOMPONEN UTAMA DALAM MENANGANI MULTIKOLINEARITAS**

Nama Mahasiswa : Siti Ulfa Nabila

Nomor Pokok Mahasiswa : 1417031111

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

**MENYETUJUI**

1. Komisi Pembimbing

| | |
|---|---|
| **Ir. Netti Herawati, M.Sc., Ph.D.** | **Prof. Drs. Mustofa Usman, M.A., Ph.D.** |
| NIP 19650125 199003 2 001 | NIP 19570101 198403 1 020 |

2. Ketua Jurusan Matematika

**Prof. Dra. Wamiliana, M.A., Ph.D.**
NIP 19631108 198902 2 001

**MENGESAHKAN**

1. Tim Penguji

| | | |
|---|---|---|
| Ketua | : **Ir. Netti Herawati, M.Sc., Ph.D.** | ........................ |
| Sekretaris | : **Prof. Drs. Mustofa Usman, M.A., Ph.D.** | ........................ |
| Penguji Bukan Pembimbing | : **Widiarti, S.Si., M.Si.** | ........................ |

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

**Prof. Warsito, S.Si., D.E.A., Ph.D.**
NIP 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **18 April 2018**

## PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Siti Ulfa Nabila**

Nomor Pokok Mahasiswa : **1417031111**

Jurusan : **Matematika**

Judul Skripsi : **Perbandingan Metode Regresi *Ridge dan* Regresi Komponen Utama dalam Menangani Multikolinearitas**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 18 April 2018

Yang Menyatakan

METERAI TEMPEL

TGL. 20

17159AFF070364938

6000

ENAM RIBU RUPIAH

**Siti Ulfa Nabila**

## RIWAYAT HIDUP

Penulis bernama lengkap Siti Ulfa Nabila, lahir di Bandar Lampung pada 17 Juli 1996. Penulis merupakan anak pertama dari 3 bersaudara, pasangan bapak Wawan Hermawan dan ibu Sri Sulastri.

Penulis menempuh pendidikan dasar di SD Negeri 1 Kupang Teba dari tahun 2002-2008. Kemudian melanjutkan pendidikan di SMP Negeri 16 Bandar Lampung dan lulus pada tahun 2011. Kemudian menempuh pendidikan di SMA Negeri 8 Bandar Lampung dan lulus pada tahun 2014.

Pada tahun 2014, penulis diterima sebagai mahasiswi di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Selama menempuh pendidikan di Universitas Lampung penulis merupakan salah satu mahasiswa penerima beasiswa Bidik Misi. Pada tahun 2017 penulis melakukan Kerja Praktik di Kantor Badan Pusat Statistik Provinsi Lampung dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata di Desa Penantian, Kecamatan Ulu Belu, Kabupaten Tanggamus.

**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)

“*Hai orang-orang yang beriman, jadikanlah sabar dan shalat sebagai penolongmu, sesungguhnya Allah bersama orang-orang yang sabar*“.

(QS. Al-Baqarah : 153)

*"Dan janganlah kamu berputus asa daripada rahmat Allah, Sesungguhnya tiada berputus asa daripada rahmat Allah kecuali orang - orang yang kufur"*

(QS. Yusuf : 87)

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

(H.R. Muslim)

# PERSEMBAHAN

Karyaku yang sederhana ini kupersembahkan kepada:

**Bapak dan Ibu**

Terima kasih kepada Bapak dan Ibu yang selalu mendo’akan kesuksesanku, memberi semangat, nasihat, dukungan serta kasih sayang yang tiada henti.

**Adikku Nafisha dan Fatia**

Terima kasih kepada adik - adikku yang selalu memberikan semangat dan keceriaan dalam hidupku

**Alm. kakek dan Almh. Nenekku**

Terima kasih kepada kakek dan nenekku yang selalu memberikan perhatian, semangat, motivasi, doa dan kasih sayang selama ini.

**Sahabat-sahabatku Wahyu, Amel, Yunita**

Terima kasih kepada para sahabatku yang selalu memberikan semangat, do’a, dan motivasi, serta kenangan indah selama ini.

**Almamater dan Negeriku**

## SANWACANA

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Metode Regresi *Ridge* dan Metode Regresi Komponen Utama dalam Menangani Multikolinearitas” dengan baik dan tepat pada waktunya.

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena dukungan, bimbingan, saran, serta do’a dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing satu yang telah memberikan bimbingan, pengarahan, dan saran kepada penulis dalam mengerjakan skripsi.
2. Bapak Prof., Drs. Mustofa Usman, M.A., Ph.D., selaku pembimbing dua yang telah memberikan saran serta pembelajaran yang sangat bermanfaat dalam menyelesaikan skripsi.
3. Ibu Widiarti, S.Si., M.Si., selaku pembahas dan penguji skripsi yang telah memberikan evaluasi, arahan, dan saran demi perbaikan skripsi.
4. Bapak Drs.Suharsono S., M.Sc., Ph.D., selaku dosen pembimbing akademik.
5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Bapak dan Ibu tercinta yang selalu mendo'akan kesuksesan dunia dan akhirat.
9. Adik - adik yang telah mendo'akan dan memberi saran serta keceriaan.
10. Sahabat-sahabat tersayang, Wahyu, Amel, Yunita, Wayan, Lidya, Kasandra, Fitrotin, Septi, Restika, Riyana, Rahmad, Agus yang telah mendo'akan, memberi dukungan dan kenangan indah kepada penulis.
11. Fietra, Mona, Putri, Jelli, Rium, dan teman-teman satu bimbingan lainnya, terima kasih atas semangat dan saran selama penyelesaian skripsi.
12. HIMATIKA yang telah memberikan pengalaman berharga.
13. Teman-teman mahasiswa Jurusan Matematika angkatan 2014.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, April 2018

Penulis

**Siti Ulfa Nabila**

## DAFTAR ISI

Halaman

# DAFTAR TABEL

## DAFTAR GAMBAR

Gambar Halaman

# I. PENDAHULUAN

## 1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistika yang dapat digunakan untuk menyelidiki atau mengetahui hubungan serta membangun hubungan antara dua variabel atau lebih. Variabel tersebut terdiri dari variabel yang dijelaskan disebut dengan variabel terikat (dilambangkan dengan Y) dan variabel penjelas yang disebut variabel bebas (dilambangkan dengan X). Analisis regresi sering diterapkan dalam beberapa bidang keilmuan, diantaranya mencakup bidang ekonomi, ilmu-ilmu sosial, ilmu biologi, bidang pendidikan, serta dalam bidang teknik.

Sebuah model regresi 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 autokolerasi, dan tidak terjadi multikolinearitas antar variabel bebas. Salah satu metode penduga terbaik dalam analisis regresi adalah Metode Kuadrat Terkecil (MKT). Pendugaan parameter dengan menggunakan metode kuadrat terkecil akan menghasilkan penduga yang tak bias, akan tetapi metode ini peka terhadap adanya penyimpangan asumsi. Apabila terdapat salah satu asumsi regresi yang tidak terpenuhi maka penduga dengan metode kuadrat terkecil tidak lagi efisien untuk digunakan.

Asumsi dalam analisis regresi berganda yang memungkinkan tidak terpenuhi adalah terjadinya multikolinearitas. Multikolinearitas merupakan keadaan dimana terjadinya korelasi atau hubungan linear antarvariabel bebas. Adanya multikolinearitas dapat menyebabkan variansi koefisien regresi menjadi sangat besar, meskipun metode kuadrat terkecil tetap memenuhi syarat *Best Linear Unbiast Estimator* (BLUE) dengan multikolinearitas yang tinggi namun penaksir yang didapatkan tidak stabil.

Salah satu metode analisis yang dapat digunakan untuk mengatasi masalah multikolinearitas adalah dengan menggunakan regresi *ridge*. Prinsip dari regresi *ridge* ini adalah menambahkan konstanta bias pada diagonal matriks $X^T X$ sehingga dapat melemahkan multikolinearitas. Terdapat beberapa cara dalam memilih sebuah konstanta bias dalam regresi *ridge* diantaranya adalah dengan menggunakan rumus HKB, LW, KM dan iterasi GCV. Pemilihan konstanta bias dengan iterasi GCV dalam regresi *ridge* adalah dengan memilih konstanta bias yang memiliki statistik GCV minimum dan menghasilkan penduga parameter yang stabil. Selain itu metode lain yang dapat digunakan untuk mengatasi multikolinearitas adalah dengan menggunakan regresi komponen utama.

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 yang mengatasi multikolinearitas yang terdapat antarvariabel bebas. Suatu pemodelan atau peramalan dikatakan akurat apabila selisih antara nilai aktual dengan nilai ramalan adalah mendekati nol. Salah satu kriteria yang dapat digunakan untuk melihat keakuratan suatu pemodelan adalah dengan menggunakan kriteria MSE.

Dalam penelitian ini akan dikaji tentang perbandingan metode regresi *ridge* dan metode regresi komponen utama dalam menangani multikolinearitas dimana metode yang memiliki nilai berdasarkan kriteria MSE yang terkecil adalah metode yang terbaik.

## 1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menduga parameter regresi dengan menggunakan metode regresi *ridge*, metode regresi komponen utama dan metode kuadrat terkecil.
2. Membandingkan nilai dugaan antara metode regresi ridge, metode regresi komponen utama dan metode kuadrat terkecil.
3. Menentukan metode terbaik dalam menangani multikolinearitas dengan menggunakan kriteria MSE.

## 1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah menambah pengetahuan serta diharapkan dapat menjadi masukkan bagi para peneliti, mahasiswa, dan pembaca tentang metode regresi *ridge* dan metode regresi komponen utama dalam mengatasi multikolinearitas.

# II. TINJAUAN PUSTAKA

## 2.1 Analisis Regresi

Analisis regresi merupakan salah satu metode statistik yang dapat digunakan untuk menyelidiki atau meramalkan atau membangun model hubungan antara dua peubah (variabel) atau lebih. Dalam analisis regresi dibedakan menjadi dua jenis variabel 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 variasi variabel terikat dan biasanya diplot pada sumbu datar (sumbu-X).

Misalkan diasumsikan model hubungan antara variabel X dan Y adalah linier dan ingin menentukan garis dugaan terbaiknya, maka harus menyadari bahwa garis dugaan dari masalah yang sebenarnya dan tidak dapat 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).

## 2.2 Analisis Regresi Linear Berganda

Analisis regresi linier berganda merupakan analisis hubungan secara linier antara dua atau lebih variabel bebas dengan satu variabel terikat. Persamaan umum garis regresi untuk regresi linear berganda adalah :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \ , i = 1,2,3,\ldots,n \, , j = 1,2,3,\ldots,k \quad (2.1)$$

dengan,

$Y_i$ = variabel tak bebas pengamatan ke-i

$X_{ki}$ = variabel bebas pengamatan ke-i

$\beta_0$ = konstanta (parameter)

$\beta_k$ = koefisien regresi atau *slope* (parameter) ke-k

$\varepsilon_i$ = sisaan (galat) pengamatan ke-i

Dalam regresi linear berganda yang akan diduga adalah $\beta_0$ dan $\beta_k$ 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 ŷ bila x sama dengan nol

$b_k$ = gradien garis regresi (perubahan nilai dugaan ŷ per satuan perubahan nilai x) ke-k

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

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.3}$$

dengan,

$Y$ = vektor pengamatan berukuran n x 1

$X$ = matriks variabel bebas berukuran n x k

$\beta$ = vektor parameter yang akan ditaksir berukuran k x 1

$\varepsilon$ = vektor galat berukuran n x 1

## 2.3 Asumsi Analisis Regresi Linear

Menurut Drapper dan Smith (1992), agar mampu memiliki kesimpulan yang benar tentang parameter $\beta_0$ dan $\beta_k$, pemenuhan asumsi-asumsi model regresi harus terpenuhi. Asumsi-asumsi tersebut adalah :

1. Nilai $\varepsilon_i$ adalah bebas satu dengan yang lainnya atau korelasi $(\varepsilon_i , \varepsilon_j) = 0$. Untuk asumsi pertama yang menyatakan *independent*, artinya $\varepsilon_i$ merupakan

variabel acak dengan nilai tengah nol dan $\sigma^2$ yang tidak diketahui. Jadi, $E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma^2$, $\varepsilon_i$ dan $\varepsilon_j$ tidak berkolerasi, $i \neq j$, sehingga $covv(\varepsilon_i , \varepsilon_j) = 0$. Jadi, $\varepsilon_i$ merupakan variabel acak normal, dengan nilai tengah nol dan ragam $\sigma^2$ dengan kata lain $\varepsilon_i$~ $N\ (0, \sigma^2)$.

2. Nilai tengah dari Y adalah fungsi linear dari X, yaitu jika dihubungkan titik-titik dari nilai tengah yang berbeda, maka akan diperoleh garis lurus $\mu(y|x) = \beta_0 + \beta_k X_{ki}$.
3. Ragam galat homogen (homoskedasitas) yaitu galat memiliki nilai ragam yang sama antara galat ke-i dan galat ke-j. Secara matematis ditulis $Var(\varepsilon_i) = \sigma^2$.
4. Ragam galat menyebar normal dengan rata-rata nol dan suatu ragam tertentu. Asumsi ke empat menyatakan untuk sembarang kombinasi tetap dari variabel bebas X, variabel tak bebas Y berdistribusi normal atau yang biasa disebut asumsi kenormalan. Dengan kata lain $\varepsilon_i$~ $N(0, \sigma^2)$.

## 2.4 Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch pada tahun 1934 yang artinya terdapat hubungan linear diantara beberapa atau semua variabel bebas dalam model regresi. Masalah multikolinearitas hanya ditemukan dalam 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).

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 (Gujarati, 1995).
Uji untuk mengetahui gejala multikolinearitas dengan melihat nilai VIF dan TOL tersebut dilakukan dengan menggunakan langkah-langkah:

1. Menghitung VIF dari $X_1$.
2. Meregresikan variabel bebas selain $X_1$ terhadap $X_1$.
3. Menghitung koefisien determinasi dari regresi variabel bebas selain $X_1$ terhadap $X_1$ dan diperoleh $R_j^2$.
4. Menghitung nilai TOL dengan rumus $TOL = (1 - R_j^2)$.
5. Menghitung nilai VIF dengan rumus $VIF = \frac{1}{TOL}$.

Masalah multikolinearitas bisa timbul karena berbagai sebab. Pertama, karena sifat-sifat yang terkandung dalam kebanyakan variabel ekonomi berubah bersama-sama sepanjang waktu. Besaran-besaran ekonomi dipengaruhi oleh faktor-faktor yang sama. Oleh karena itu, sekali faktor-faktor yang mempengaruhi itu menjadi operatif, maka seluruh variabel akan cenderung berubah dalam satu arah. Dalam data *time series*, pertumbuhan dan faktor-faktor kecenderungan merupakan penyebab utama adanya multikolinearitas. Kedua, penggunaan nilai lag (*lagget values*) dari variabel-variabel bebas tertentu dalam model regresi. Mengingat sifat yang sangat mendasar dari data, multikolinearitas diperkirakan terdapat pada sebagian besar hubungan-hubungan ekonomi. Oleh karena itu, perhatian sesungguhnya bukan lagi terletak pada ada atau tidaknya multikolinearitas, tetapi lebih pada akibat-akibat yang ditimbulkan oleh adanya multikolinearitas dalam sampel (Sumodiningrat, 1998).

## 2.5 Konsekuensi Multikolinearitas

Jika asumsi pada model regresi linear klasik terpenuhi, maka penaksir kuadrat terkecil/ *Ordinary Least Square* (OLS) dari koefisien regresi linear adalah linear, tak bias dan mempunyai varian minimum dalam arti penaksir tersebut adalah penaksir tak bias kolinear terbaik/ *Best Linear Unbiased Estimator* (BLUE), meskipun multikolinearitas sangat tinggi, penaksir kuadrat terkecil biasa masih tetap memenuhi syarat BLUE, tetapi penaksir tersebut tidak stabil (Gujarati, 1995).

Multikolinearitas berpengaruh terhadap estimasi kuadrat terkecil dari koefisien regresi. Akan diperlihatkan bagaimana $\hat{\beta}$, variansi $\left(\hat{\beta}_j\right)$ dan kovariansi $\left(\hat{\beta}_j, \hat{\beta}_h\right)$ dengan $j, h = 1,2, \dots, k$. Misalkan ada dua variabel bebas $(X_1, X_2)$ dan $Y$ variabel terikat sehingga diperoleh model,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{2.4}$$

Persamaan matriks dengan metode kuadrat terkecil adalah

$$(\boldsymbol{X^T X})\boldsymbol{\hat{\beta}} = \boldsymbol{X^T Y} \tag{2.5}$$

$$\boldsymbol{X^T X} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}, \qquad \boldsymbol{X^T Y} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

diperoleh

$$[\boldsymbol{X^T X}]^{-1} = \begin{bmatrix} \dfrac{1}{1 - r_{12}^2} & \dfrac{-r_{12}}{1 - r_{12}^2} \\ \dfrac{-r_{21}}{1 - r_{12}^2} & \dfrac{1}{1 - r_{12}^2} \end{bmatrix}$$

Elemen diagonal utama dari matriks $[\boldsymbol{X^T X}]^{-1}$ merupakan *varians inflasion vactor* (VIF) yaitu :

$$C_{jj} = \frac{1}{1-R_j^2} \; ; j = 1,2, \dots, k \tag{2.6}$$

dengan,

$R_j^2$ = koefisien determinansi dari regresi $X_j$

$C_{jj}$ = *varians inflasion vactor* (VIF)

$r_{12}$ = korelasi antara $X_1 \; dan \; X_2$

$r_{X_jY}$ = korelasi antara $X_j$ dan $Y$

$$(\boldsymbol{X^T X})\boldsymbol{\hat{\beta}} = \boldsymbol{X^T Y}$$
$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

sehingga,

$$\hat{\beta}_1 = \frac{r_{1y} - r_{21}r_{2y}}{(1 - r_{12}^2)}$$
$$\hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1 - r_{12}^2)}$$

Jika ada multikolinieritas antara $X_1$ dan $X_2$ yang sangat erat dan $r_{12} \to 1$. Variansi dan kovariansi koefisien regresi menjadi sangat besar karena $V(\hat{\beta}_j) = C_{jj}r_2 \to \infty$ seperti $|r_{12}| \to 1$, galat $Cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\delta^2 \to \pm\infty$, variansi yang besar untuk $\hat{\beta}_j$ menyatakan bahwa koefisien regresi adalah perkiraan yang sangat lemah. Pengaruh multikolinieritas adalah untuk memperkenalkan sebuah ketergantungan linier yang dekat dalam kolom matriks. Selanjutnya jika kita mengasumsikan $X_1^T Y \to X_2^T Y$, seperti $|r_{12}| \rightarrow 1$, perkiraan koefisien regresi menjadi sama besarnya, tetapi berlawanan tanda, yaitu $\hat{\beta}_1 = -\hat{\beta}_2$.

Masalah yang sama terjadi bila masalah multikolinieritas disajikan dan ada lebih dari dua variabel bebas. Umumnya elemen diagonal matriks $C = [X^T X]^{-1}$ dapat ditulis sebagai berikut :

$$C_{jj} = \frac{1}{1 - R_j^2}$$

$R_j^2$ dihasilkan dari meregresikan $X_1$ pada variabel bebas lainnya. Konsekuensinya kita biasa menyebut

$$VIF(\hat{\beta}_j) = \frac{1}{1-R_j^2} \tag{2.7}$$

*Varians inflasion factor* (VIF) untuk $\hat{\beta}_j$ ini adalah ukuran penting dalam perkiraan multikolinieritas.

Menurut Sumodiningrat (1998), dalam hal terdapat multikolinearitas sempurna, penaksir dengan kuadrat terkecil bisa menjadi tak tentu dan variansi serta standar deviasinya menjadi tak terhingga. Sedangkan jika multikolinearitas tinggi, tetapi tidak sempurna maka konsekuensinya
adalah sebagai berikut :

a. Meskipun penaksir melalui kuadrat terkecil bisa didapatkan, standar deviasinya cenderung besar jika derajat kolinearitas antara peubah bertambah.

b. Karena standar deviasi besar, interval kepercayaan bagi parameter populasi yang relevan akan menjadi besar.
c. Taksiran-taksiran parameter kuadrat terkecil didapatkan tetapi standar deviasi akan menjadi sangat sensitif terhadap perubahan.
d. Jika multikolinearitas tinggi, mungkin bisa tinggi namun tidak satu pun (sangat sedikit) taksiran koefisien regresi yang signifikan secara statistic.

## 2.6 Ukuran Pemusatan dan Penskalaan (*centering and scaling*)

Pemusatan dan penskalaan data merupakan bagian dari membakukan (*standardized*) variabel. Modifikasi sederhana dari pembakuan atau standarisasi variabel ini adalah transformasi korelasi (*correlation transformation*).

Pemusatan merupakan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Sedangkan penskalaan meliputi gambaran pengamatan pada kesatuan (unit) standar deviasi dari pengamatan untuk variabel (Kutner, *et al*., 2005).

Dalam hal ini yang akan dibakukan (distandarisasi) adalah model regresi linear berganda yang ditujukkan pada model di bawah ini.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \tag{2.8}$$

Berikut ini merupakan pembakuan variabel terikat $Y$ dan variabel bebas $X_1, X_2, \dots, X_k$

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \; dengan \; S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \tag{2.9}$$

$$X_j^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \; dengan \; S_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2}{n-1}} \tag{2.10}$$

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

dengan :

$\bar{Y}$ = rata-rata dari Y

$\bar{X}_j$ = rata-rata dari pengamatan $X_j$

$S_Y$ = standar deviasi dari Y

$S_{X_j}$ = standar deviasi dari $X_j$

Model regresi berganda terstandarisasi adalah tranformasi dari model regresi berganda (didefinisikan sebagai transformasi korelasi)

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^* \tag{2.11}$$

Model di atas disebut sebagai model regresi yang baku (*standardized regression model*). Diantara parameter $\beta_1^*, \; \beta_2^*, \dots, \beta_k^*$ pada model regresi baku, dengan parameter $\beta_1, \beta_2, \dots, \beta_k$ dengan model regresi linear berganda terdapat suatu hubungan linear. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan seperti di bawah ini (Kutner, *et al.*, 2005).

$$\beta_j = \left(\frac{S_Y}{S_{X_j}}\right)\beta_j^* \;, \qquad j = 1,2, \dots, k \tag{2.12}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \cdots - \beta_k \bar{X}_k \tag{2.13}$$

$$\beta_0 = \bar{Y} - \sum_{j=1}^{k} \beta_j \bar{X}_j \tag{2.14}$$

Prosedur ini disebut dengan prosedur penskalaan. Dari persamaan (2.8) di atas dapat dibentuk menjadi :

$$Y_i = \beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_1\bar{X}_1 + \beta_2(X_{2i} - \bar{X}_2) + \beta_2\bar{X}_2 + \cdots + \beta_k(X_{ki} - \bar{X}_i)$$
$$+ \beta_k\bar{X}_k + \varepsilon_i$$
$$Y_i = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + \beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i \tag{2.15}$$

Berdasarkan persamaan (2.13) Maka berlaku :

$$\bar{Y} = \beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k$$

Sehingga

$$Y_i - \bar{Y} = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + (\beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i)$$
$$- (\beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k)$$
$$Y_i - \bar{Y} = \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i \tag{2.16}$$

Jika

$$y_i = Y_i - \bar{Y}$$
$$x_{1i} = X_{1i} - \bar{X}_1$$
$$x_{2i} = X_{2i} - \bar{X}_2$$
$$x_{ki} = X_{ki} - \bar{X}_k$$

Maka didapat model baru yaitu:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \tag{2.17}$$

Prosedur untuk membentuk model pertama menjadi model terakhir disebut dengan prosedur pemusatan. Prosedur ini mengakibatkan hilangnya $\beta_0$ (*intercept*) yang membuat perhitungan untuk mencari model regresi menjadi lebih sederhana. Keseluruhan dari prosedur di atas disebut prosedur pemusatan dan penskalaan.

## 2.7 Metode Kuadrat Terkecil (MKT)

Metode kuadrat terkecil merupakan suatu metode yang digunakan untuk mengestimasi $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ dengan cara meminimumkan jumlah kuadrat galat. Parameter $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ tidak diketahui dan perlu dicari nilai estimasinya (Montgomery, 2006).

Secara umum persamaan untuk mendapatkan nilai estimasi dengan metode kuadrat terkecil dapat ditulis :

$$\hat{\beta}^{MKT} = \arg\min_{\beta} \|y - X\beta\|_2^2 = \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} \beta_j \tilde{x}_j \right\|_2^2 \tag{2.18}$$

Pada notasi matriks jumlah kuadrat galat $e_i^2$ dapat ditulis sebagai berikut

$$\sum_{i=1}^{n} e_i^2 = e_i^T e_i = \begin{bmatrix} e_1 & e_2 & \cdots & e_i \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \end{bmatrix} = e_1^2 + e_2^2 + \ldots + e_i^2 \tag{2.19}$$

Dari persamaan umum regresi linear berganda dapat ditulis sebagai berikut

$$Q(\hat{\beta}_j) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2 \tag{2.20}$$

Berdasarkan persamaan umum regresi linear berganda dengan matriks diperoleh

$$\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X\beta} \tag{2.21}$$

Oleh Karena itu, perkaliaan matriks galat menjadi :

$$\boldsymbol{e_i^T e_i} = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{Y} - \boldsymbol{X\beta})$$

$$\boldsymbol{e_i^T e_i} = \boldsymbol{Y^T Y} - \boldsymbol{Y^T X\beta} - \boldsymbol{\beta^T X^T Y} + \boldsymbol{\beta^T X^T X\beta}$$

(karena $\boldsymbol{X^T \beta^T Y} = \boldsymbol{Y^T X\beta}$ )

$$\boldsymbol{e_i^T e_i} = \boldsymbol{Y^T Y} - \boldsymbol{2Y^T X\beta} + \boldsymbol{\beta^T X^T X\beta} \tag{2.22}$$

Untuk mencari nilai-nilai $\beta$ yaitu dengan meminimumkan jumlah kuadrat galat, kemudian dicari turunan dari $Q(\beta_j)$ secara parsial terhadap $\beta_j$ ; $j = 1,2,\ldots,k$ dan disamakan dengan nol.

Jika ditulis dalam bentuk matriks maka bentuknya menjadi

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \cdots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \cdots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Atau secara lengkap jika ditulis kedalam bentuk matriks menjadi

$$[\boldsymbol{X^T X}]\widehat{\boldsymbol{\beta}} = [\boldsymbol{X^T Y}] \tag{2.23}$$

Pada persamaan diatas kedua ruasnya dikalikan invers dari matriks $(\boldsymbol{X^T X})^{-1}$, sehingga diperoleh:

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

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

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

Sehingga diperoleh estimator untuk MKT adalah

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} \qquad (2.24)$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut

1. $\hat{\beta}$ linear

   $\hat{\beta}$ linear jika $\hat{\beta}$ merupakan fungsi linear dari $\beta$

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

$$= (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})$$

$$= (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon}$$

$$= \boldsymbol{I\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon}$$

2. $\hat{\beta}$ tak bias

   $\hat{\beta}$ adalah penduga tak bias jika $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

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

$$= \boldsymbol{E}((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + \boldsymbol{\varepsilon})$$

$$= \boldsymbol{E}((\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \varepsilon})$$

$$= (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta} + (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T E}(\boldsymbol{\varepsilon})$$

$$= (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X\beta}$$

$$= \boldsymbol{\beta}$$

Sehingga $\hat{\boldsymbol{\beta}}$ merupakan penduga tak bias dari $\boldsymbol{\beta}$

3. $\hat{\boldsymbol{\beta}}$ memiliki variansi minimum

$$\begin{aligned}
var(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\boldsymbol{T}}] \\
&= E[(\,(\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TY}-\boldsymbol{\beta})((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TY}-\boldsymbol{\beta})^{\mathrm{T}}] \\
&= E[\,(\,(\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T}(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})-\boldsymbol{\beta})\,((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T}(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})-\boldsymbol{\beta})^{\mathrm{T}}] \\
&= E[((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TX\beta}+(\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})-\boldsymbol{\beta})((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TX\beta}+ \\
&\quad (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})-\boldsymbol{\beta})^{\mathrm{T}}] \\
&= \mathrm{E}\,[(\boldsymbol{\beta}+(\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon}-\boldsymbol{\beta})\,(\boldsymbol{\beta}+(\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon}-\boldsymbol{\beta})^{\mathrm{T}}] \\
&= \mathrm{E}\,[((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})\,((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})^{\mathrm{T}}] \\
&= \mathrm{E}\,[((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})\,((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon})^{\mathrm{T}}] \\
&= \mathrm{E}\,[((\boldsymbol{X^TX})^{-\mathbf{1}}\,\boldsymbol{X^T\varepsilon^T\varepsilon}\,\boldsymbol{X}\,(\boldsymbol{X^TX})^{-\mathbf{1}}] \\
&= \mathrm{E}\,[((\boldsymbol{X^TX})^{-\mathbf{1}}\,\boldsymbol{X^T}\,\boldsymbol{E}(\boldsymbol{\varepsilon\varepsilon^T})\,\boldsymbol{X}\,(\boldsymbol{X^TX})^{-\mathbf{1}}] \\
&= (\boldsymbol{X^TX})^{-\mathbf{1}}\,\boldsymbol{X^TX}\,(\boldsymbol{X^TX})^{-\mathbf{1}}\,E[\boldsymbol{\varepsilon\varepsilon^T}] \\
&= (\boldsymbol{X^TX})^{-\mathbf{1}}\sigma^2
\end{aligned}$$

$Var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X^TX})^{-\mathbf{1}}\,\sigma^2$ merupakan varians terkecil dari semua penaksir linear tak bias, hal ini dijamin dengan teorema Gauss-Markov.

Untuk menunjukkan bahwa $Var(\hat{\boldsymbol{\beta}})$ adalah varians yang paling minimum. Maka akan diasumsikan penduga lain yang linear dan tak bias, kemudian dibuktikan bahwa variansinya lebih besar dari $Var(\hat{\boldsymbol{\beta}})$.

Misalkan $\hat{\boldsymbol{\beta}}^*$ adalah penduga yang linier dan tak bias bagi $\boldsymbol{\beta}$. Asumsikan bahwa :

$$\hat{\boldsymbol{\beta}}^* = ((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T}+\boldsymbol{Z})\boldsymbol{y}$$

dimana **Z** adalah matriks konstanta $(k \times n)$ yang diketahui

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^* &= ((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{y} \\
&= ((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) \\
&= \left((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TX\beta} + (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})\right) \\
&= ((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TX\beta} + (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}) \\
&= (\boldsymbol{I\beta} + (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon})
\end{aligned}
$$

sehingga

$$
\begin{aligned}
E\left(\widehat{\boldsymbol{\beta}}^*\right) &= E(\boldsymbol{\beta} + (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}) \\
&= \boldsymbol{\beta} + \boldsymbol{ZX\beta} \qquad (karena\ E(\boldsymbol{\varepsilon}) = \mathbf{0}) \\
&= (\mathbf{1} + \boldsymbol{ZX})\boldsymbol{\beta}
\end{aligned}
$$

Agar $\widehat{\boldsymbol{\beta}}^*$ estimasi tak bias dari $\boldsymbol{\beta}$ maka $\boldsymbol{ZX} = \mathbf{0}$, sehingga :

$$
\boldsymbol{Var}\left(\widehat{\boldsymbol{\beta}}^*\right) = E\left(\left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right)^{\boldsymbol{T}}\right)
$$

dengan $\left(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right) = (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{ZX\beta} + \boldsymbol{Z\varepsilon}$. Dan diasumsikan bahwa

$\boldsymbol{E}(\boldsymbol{\varepsilon\varepsilon^T}) = \sigma_u^2\boldsymbol{I}_u$. Karena $\boldsymbol{ZX\beta} = \mathbf{0}$ sehingga

$$
\begin{aligned}
\boldsymbol{Var}\left(\widehat{\boldsymbol{\beta}}^*\right) &= E\left(((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})^{\boldsymbol{T}}\right) \\
&= E\left(((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T\varepsilon} + \boldsymbol{Z\varepsilon})(\boldsymbol{\varepsilon^TZ^T} + \boldsymbol{\varepsilon^TX}(\boldsymbol{X^TX})^{-\mathbf{1}})\right) \\
&= E\left(((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})\boldsymbol{\varepsilon\varepsilon^T}(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-\mathbf{1}})\right) \\
&= \left(((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})\sigma_u^2\boldsymbol{I}_u(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-\mathbf{1}})\right) \\
&= \sigma^2\left(((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^T} + \boldsymbol{Z})(\boldsymbol{Z^T} + \boldsymbol{X}(\boldsymbol{X^TX})^{-\mathbf{1}})\right) \\
&= \sigma^2((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TZ^T} + (\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TX}(\boldsymbol{X^TX})^{-\mathbf{1}} + \boldsymbol{ZZ^T} + \boldsymbol{ZX}(\boldsymbol{X^TX})^{-\mathbf{1}}) \\
&= \sigma^2((\boldsymbol{X^TX})^{-\mathbf{1}}\boldsymbol{X^TZ^T} + \boldsymbol{I}(\boldsymbol{X^TX})^{-\mathbf{1}} + \boldsymbol{ZZ^T} + \boldsymbol{ZX}(\boldsymbol{X^TX})^{-\mathbf{1}})
\end{aligned}
$$

Karena $\boldsymbol{ZX} = \boldsymbol{X^TZ^T} = \mathbf{0}$ maka

$$
\boldsymbol{Var}\left(\widehat{\boldsymbol{\beta}}^*\right) = \sigma^2((\boldsymbol{X^TX})^{-\mathbf{1}} + \boldsymbol{ZZ^T})
$$

Matriks $\boldsymbol{ZZ^T}$ adalah definit positif, karena semua unsur diagonalnya berbentuk kuadrat. Jadi terbukti bahwa variansi dari setiap unsur dari vektor $\widehat{\boldsymbol{\beta}}^*$ selalu lebih besar, atau paling kecil sama dengan variansi unsur $\widehat{\boldsymbol{\beta}}$ yang sesuai.

Estimator kuadrat terkecil yang memenuhi sifat linear, tak bias, dan mempunyai variansi minimum ini bersifat *Best Linear Unbiased Estimator* (*BLUE)*

## 2.8 Regresi *Ridge*

Metode regresi ridge pertama kali dikemukakan oleh Hoerl (1962) dan dikembangkan oleh Hoerl dan Kennard (1970). Penaksir koefisien regresi *ridge* adalah sebagai berikut

$$\widehat{\boldsymbol{\beta}}_R = (\boldsymbol{X^T X} + k\boldsymbol{I})^{-1}\boldsymbol{X^T y} \tag{2.25}$$

dengan,

$I$ = matriks identitas berukuran $pxp$

$k$ = konstanta bias $0 \le k \le 1$

Regresi ridge memberikan estimasi koefisien regresi yang bias dengan memodifikasi metode kuadrat terkecil untuk mendapatkan pengurangan varian dengan menambahkan suatu tetapan k dalam menstabilkan koefisien (Mardikyan dan Cetin, 2008).

Menurut Dereny dan Rashwan (2011), teknik *Ridge* didasarkan pada penambahan konstanta bias k pada diagonal matrik $\boldsymbol{X^T X}$, sehingga koefisien penduga *ridge* dipengaruhi oleh besarnya tetapan bias k, dimana nilai k bernilai antara 0 sampai 1. Dalam regresi *ridge* variabel bebas X dan variabel terikat Y ditransformasikan kedalam bentuk baku (standarisasi) kedalam $X^*$ dan $Y^*$. Ditransformasikan dengan rumus sebagai berikut:

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\sqrt{n-1}\, S_j} \; dan \; Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{n-1}\, S_Y}$$

Pada regresi *ridge* diubah menjadi persamaan tanpa koefisien intersep variabel bebas X dan variabel terikat Y dalam bentuk baku maka diperoleh model sebagai berikut:

$$Y_i^* = \beta_1^* X_1 + \beta_2^* X_2 + \cdots + \beta_p^* X_p + \varepsilon$$

Maka persamaan diatas dalam bentuk matriks sebagai berikut:

$$Y^* = X^* \beta^* + \varepsilon^* \tag{2.26}$$

Penduga regresi baku *ridge* diperoleh dengan memasukan konstanta bias k kedalam persamaan normal MKT dan matriks korelasi variabel bebas X dan variabel terikat Y.

Secara umum rumus yang digunakan untuk mendapatkan penduga regresi *ridge* yang memiliki fungsi objektif dan kendala sebagai berikut

$$\hat{\beta}^{RIDGE} = \arg\min_{\beta} \|y - X\beta\|_2^2 + k\|\beta\|_2^2 \tag{2.27}$$

Misalkan

$$Q(\hat{\beta}_{Rj}) = \sum_{i=1}^{n} (y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2 + \sum_{j=1}^{k} k\beta_j^2$$

Langkah selanjutnya adalah meminimumkan fungsi $Q(\hat{\beta}_{Rj})$ dan menyamadengankanya dengan nol .

Apabila disusun kembali dalam bentuk matriks adalah sebagai berikut

$$[\boldsymbol{X^T X}]\begin{bmatrix}\beta_1\\ \vdots\\ \beta_k\end{bmatrix}+\begin{bmatrix}k & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & k\end{bmatrix}\begin{bmatrix}\beta_1\\ \vdots\\ \beta_k\end{bmatrix}=[\boldsymbol{X^T Y}]$$

$$[\boldsymbol{X^T X}]\boldsymbol{\beta}+[k\boldsymbol{I}]\boldsymbol{\beta}=[\boldsymbol{X}^T\boldsymbol{Y}]$$

$$[\boldsymbol{X^T X}+k\boldsymbol{I}]\boldsymbol{\beta}=[\boldsymbol{X^T Y}] \tag{2.28}$$

Dalam bentuk matriks Langkah awal untuk menentukan penduga regresi baku ridge adalah meminimumkan jumlah kuadrat galat untuk model pada persamaan dengan menggunakan metode pengali Langrange yang meminimumkan fungsi :

$$\boldsymbol{\varepsilon^T\varepsilon}=(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{T}}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})$$

dengan syarat pembatas $\boldsymbol{\beta_R^T\beta_R-k^2=0}$

$$\boldsymbol{G}=\left(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)^{\boldsymbol{T}}\left(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)+k\boldsymbol{I}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}-\boldsymbol{k^2}\right)$$

$$\boldsymbol{G}=\boldsymbol{Y^TY}-\boldsymbol{2X^TY}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}+\boldsymbol{X^TX}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)^{\boldsymbol{2}}+k\boldsymbol{I}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}-\boldsymbol{k^2}\right)$$

Dengan menggunakan syarat minimum persamaan diatas didefinisikan terhadap $\widehat{\boldsymbol{\beta}}$ dan estimasi regresi *Ridge* diperoleh sebagai berikut:

$$\frac{\partial\boldsymbol{G}}{\partial\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}}=-2\boldsymbol{X^TY}-2\boldsymbol{X^TX}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}+2k\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}=\boldsymbol{0}$$

$$\boldsymbol{X^TX}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}+k\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}=\boldsymbol{X^TY}$$

$$(\boldsymbol{X^TX}+k\boldsymbol{I})\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}=\boldsymbol{X^TY}$$

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}(k)=(\boldsymbol{X^TX}+k\boldsymbol{I})^{-1}\boldsymbol{X^TY}$$

Sehingga diperoleh penduga regresi *ridge* yaitu:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = (\boldsymbol{X^T X} + k\boldsymbol{I})^{-1}\boldsymbol{X^T Y} \tag{2.29}$$

Adapun sifat-sifat dari regresi *ridge* antara lain:

1. Menurut Montgomery dan Peck (2006) estimator regresi *ridge* merupakan transformasi linier dari penduga metode kuadrat terkecil oleh karena itu

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) &= E[(\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}\boldsymbol{X^T Y}] \\ &= E[(\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}\boldsymbol{X^T}(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})] \\ &= (\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}(\boldsymbol{X^T X})E(\boldsymbol{\beta}) + (\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}\boldsymbol{X}E(\boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X^T X} + \boldsymbol{kI})^{-\mathbf{1}}(\boldsymbol{X^T X})\boldsymbol{\beta} \\ &= \boldsymbol{H\beta} \end{aligned}$$

   Karena $E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) \neq \boldsymbol{\beta}$ maka $\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}$ merupakan penduga yang bias.

2. Menurut Montgomery dan Peck (2006) varian dari $\boldsymbol{\beta}_{\boldsymbol{R}}$ dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\boldsymbol{V}(\boldsymbol{\beta}_{\boldsymbol{R}}) = \sigma^2(\boldsymbol{X^T X} + \boldsymbol{kI})^{-\mathbf{1}}(\boldsymbol{X^T X})(\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}} \tag{2.30}$$

   Sehingga nilai *varians inflasion factor* (VIF) merupakan diagonal utama dari matriks $(\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}(\boldsymbol{X^T X})(\boldsymbol{X^T X} + k\boldsymbol{I})^{-\mathbf{1}}$.

3. Menurut Hoerl dan Kennard (1970) jumlah kuadrat galat untuk nilai dugaan regresi *ridge* adalah

$$SSE = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) \tag{2.31}$$

Yang dapat ditulis dalam bentuk:

$$SSE = \boldsymbol{y^T y} - \boldsymbol{\beta_R^T X^T y} - \boldsymbol{k\beta_R^T \beta_R}$$

4. Menurut Montgomery dan Peck (2006) rata-rata jumlah kuadrat dari regresi *ridge* adalah sebagai berikut.

$$\boldsymbol{MSE}(\boldsymbol{\beta_R}) = \boldsymbol{Var}(\boldsymbol{\beta_R}) + \left(\boldsymbol{bias}(\boldsymbol{\beta_R})\right)^{\mathbf{2}}$$

$$= \sigma^2 Tr\left[\left(\boldsymbol{X^T X} + k\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X^T X}\right)\left(\boldsymbol{X^T X} + k\boldsymbol{I}\right)^{-1}\right] + [-\boldsymbol{\beta}k\ [\boldsymbol{X^T X} + k\boldsymbol{I}]^{-1}]^2$$

$$= \sigma^2 Tr\left[\left(\boldsymbol{X^T X} + k\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X^T X}\right)\left(\boldsymbol{X^T X} + k\boldsymbol{I}\right)^{-1}\right] + [k^2\boldsymbol{\beta}'\ \boldsymbol{X^T X} + k\boldsymbol{I}^{-1}]^{-2}\boldsymbol{\beta}$$

$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{\lambda_j + k^2} + k^2(\boldsymbol{X^T X} + kI)^{-2}\boldsymbol{\beta} \qquad (2.32)$$

Hoerl dan Kennard (1970) menyarankan metode grafik yang disebut *ridge* trace untuk memilih nilai parameter k. Grafik plot berdasarkan nilai komponen individu $\beta(k)$ dengan barisan dari $k$ $(0 < k < 1)$. Konstanta k mencerminkan jumlah bias dalam estimator $\hat{\beta}_R$ bila $k = 0$ maka estimator $\hat{\beta}_R$ akan bernilai sama dengan estimator kuadrat terkecil $\beta$. Bila $k > 0$ estimator *ridge* akan bias terhadap estimator $\beta$ tetapi cenderung lebih stabil dibandingkan estimator kuadrat terkecil. Pemilihan besarnya tetapan bias k merupakan masalah yang perlu diperhatikan. Tetapan bias yang diinginkan adalah tetapan bias yang menghasilkan galat yang relatif kecil dan menghasilkan koefisien estimator yang relatif stabil.

Myers (1990) dalam Montgomery dan Peck (2006) menyarankan nilai k dengan menggunakan metode kriteria validasi silang tergeneralisasi (*generalized cross validation*). Secara sederhana manfaat dari prosedur ini adalah untuk pemilihan k guna mendapatkan model terbaik dan koefisien dugaan yang lebih stabil dengan meminimumkan *GCV* yang dapat dilihat melalui plot sederhana antara validasi silang tergeneralisasi ini dengan k. Pemilihan ini menggunakan rumus sebagai berikut

$$GCV = \frac{SSE\ k}{\{n - [1 - trace\boldsymbol{H}_k]\}^2} \quad (2.33)$$

dimana,

$$\boldsymbol{X}\ (\boldsymbol{X^T X} + k\boldsymbol{I})^{-1}\ \boldsymbol{X}^T \equiv \boldsymbol{H}_k$$

$$H_k = \sum_{j=1}^{p} \frac{\lambda_j}{\lambda_j + k}$$

dengan,

$SSE_k$ = jumlah kuadrat residual dengan regresi *ridge*

$\lambda_j$ = nilai eigen ke- j

k = konstanta antara 0 sampai 1

n = banyaknya data

## 2.9 Analisis Komponen Utama

Metode analisis komponen utama (AKU) merupakan teknik statistik yang digunakan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel melalui beberapa variabel baru dimana variable 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 variabel baru yang tidak berkorelasi.

Nilai *eigen* merupakan keragaman dari setiap komponen utama. Ketika nilai *eigen* dari komponen utama mendekati nol hal tersebut juga mengindikasikan adanya multikolinearitas pada data asli (Rougoor, e*t al*., 2000).

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

$$\boldsymbol{Q} = \boldsymbol{VX} \tag{2.34}$$

dengan,

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

**V** = matriks yang melakukan transformasi terhadap variabel 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

$$Q_j = v_{1j}X_1 + v_{22}X_2 + \cdots + v_{pj}X_p = v_j^T X\beta \tag{2.35}$$

Setelah beberapa komponen hasil PCA 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 PCA diantaranya adalah dapat menghilangkan korelasi secara bersih tanpa harus mengurangi jumlah variabel asal.

## 2.10 Regresi Komponen Utama

Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis. Analisis komponen utama merupakan analisis yang memperkecil dimensi variabel tanpa kehilangan banyak informasi, dengan tujuan menyederhanakan variabel yang diamati dengan cara 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 :

$$\boldsymbol{Y} = \boldsymbol{Q\alpha} + \boldsymbol{\varepsilon} \quad \text{dimana,} \quad \mathbf{Q} = \mathrm{XV} \quad \boldsymbol{\alpha} = \boldsymbol{V^T\beta}$$

dengan,

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

**X** = matriks yang elemen-elemennya dikurang dengan rataannya yang mensyaratkan rataan 0 dan standar deviasi 1

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

$\boldsymbol{a}$= 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 variabel penjelas (variabel bebas). Secara umum estimasi dengan regresi komponen utama dengan fungsi objektif adalah sebagai berikut :

$$\hat{a} = \arg\min_{\alpha} \|y - Qa\|_2^2 = \arg\min_{\alpha} \left\| y - \sum_{i=1}^{n} a_i \tilde{q}_i \right\|_2^2 \qquad (2.36)$$

Sehingga

$$Q(\hat{a}_i) = \sum_{i=1}^{n} (y_i - \hat{a}_0 - \hat{a}_1 q_{1i} - \hat{a}_2 q_{21} - \cdots - \hat{a}_k q_{ki})^2$$

Langkah selanjutnya adalah meminimumkan fungsi $Q(\hat{\beta}_{Rj})$ dan menyamadengankanya dengan nol.

Apabila disusun kembali dalam bentuk matriks adalah sebagai berikut

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

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

$$\hat{\boldsymbol{a}} = [\boldsymbol{Q}^T\boldsymbol{Q}]^{-1}\boldsymbol{Q}^T\boldsymbol{y} \tag{2.37}$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{v}\hat{\boldsymbol{a}} \tag{2.38}$$

Apabila dibentuk dalam bentuk matrik

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

Dalam analisis regresi komponen utama ,tidak digunakan semua komponen Q melainkan memilih komponen utama berdasarkan kriteria tertentu.

Menurut D.F Marrison tahun 1976 dalam edisi kedua *Multivariat Statistical Methods,* komponen-komponen dapat dihitung sampai sejumlah tertentu proporsi keragaman (> 75%) yang telah terjelaskan (Draper and Smith, 1992 ).

Dengan semua variabel yang terstandarisasi sehingga $\boldsymbol{X}^T\boldsymbol{X}$ merupakan matriks korelasi dari X dan $\boldsymbol{X}^T\boldsymbol{Y}$ merupakan vektor korelasi antara X dan Y .

Misalkan **V** = [$\mathbf{V}_1$ , . . .,$\mathbf{V}_p$] merupakan matriks berukuran pxp yang kolom kolomnya eigenvector ternormalisasi dari $\boldsymbol{X}^T\boldsymbol{X}$ dan misalkan $\lambda_1, \dots, \lambda_p$ merupakan nilai eigen yang bersesuaian.

Misalkan **Q** = [$\mathbf{Q}_1$, . . . ,$\mathbf{Q}_p$ ] = **XV**. Sehingga $\mathbf{Q}_j$ =$\mathbf{XV}_j$ adalah komponen utama ke-j dari X.

Beberapa hal penting dari regresi komponen utama adalah :

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

Model regresi dapat ditulis sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \boldsymbol{XVV^T\beta} + \boldsymbol{\varepsilon} = \boldsymbol{Qa} + \boldsymbol{\varepsilon} \quad dimana\ \boldsymbol{a} = \boldsymbol{V^T\beta}$$

Berdasarkan persamaan ini, estimasi kuadrat terkecil dari $\boldsymbol{a}$ adalah

$$\hat{\boldsymbol{a}} = (\boldsymbol{Q^TQ})^{-1}\boldsymbol{Q^tY} = \boldsymbol{\Lambda^{-1}Q^TY}$$

Dan estimator regresi komponen utama adalah sebagai berikut

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

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

Apabila semua komponen digunakan maka penduga dengan menggunakan regresi komponen utama sama dengan penduga dengan menggunakan regresi kuadrat terkecil, namun dalam praktiknya hanya sebuah subset $\mathbf{Q}_{(s)} = [\mathbf{Q}_1, \ldots, \mathbf{Q}_p]$ dari komponen utama yang digunakan dalam menduga $a$ sehingga

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

Sehingga estimator regresi komponen utama dari subset ini adalah

$$\hat{\boldsymbol{\beta}} = \boldsymbol{V}_{(s)}\boldsymbol{\Lambda}_{(s)}{}^{-1}\boldsymbol{Q}_{(s)}^{\boldsymbol{T}}\boldsymbol{Y} \tag{2.39}$$

Adapun sifat dari regresi komponen utama adalah

1. Bersifat Bias

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\boldsymbol{V\Lambda^{-1}Q^TY}) \\
&= \mathrm{E}(\boldsymbol{V}(\boldsymbol{Q^TQ})^{-1}\boldsymbol{Q^TY}) \\
&= \mathrm{E}(\boldsymbol{V}(\boldsymbol{Q^TQ})^{-1}\boldsymbol{Q^TQa} + \boldsymbol{\varepsilon}) \\
&= \mathrm{E}(\boldsymbol{V}(\boldsymbol{Q^TQ})^{-1}\boldsymbol{Q^TQV^T\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{V^T\beta}
\end{aligned}$$

Penduga regresi komponen utama merupakan penduga yang bias karena

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

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

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

dengan,

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

$\lambda_j$ = Akar ciri ke-j

$s^2$ = Ragam dugaan dari bentuk galat

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

$$V(b_i) = s^2 \sum_{j=1}^{m} v_{ij}^2/\lambda_j \qquad (2.40)$$

dengan,

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

$\lambda_j$ = 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

## 2.11 *Mean of Squares Error (MSE)*

Menurut Ghozali (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 yang metode yang lain. Efisiensi dari metode untuk menangani multikolinearitas ini akan dievalusi berdasarkan rata-rata dari *Mean Square Error* (MSE) dari hasil estimasi parameter $\hat{\beta}$, yang didefinisikan sebagai berikut :

$$\mathrm{MSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta\right)^2 ; \quad j = 1,2,\ldots,m \tag{2.41}$$

$$\mathrm{AMSE}\left(\hat{\beta}\right) = \frac{1}{m}\sum_{j=1}^{m}\left\|\hat{\boldsymbol{\beta}}_{\boldsymbol{j}} - \boldsymbol{\beta}\right\|^2 \tag{2.42}$$

dengan,

$\hat{\beta}_j$ = Penduga parameter regresi simulasi ke – $j$

$\beta$ = Parameter regresi

m = Banyaknya ulangan

# III. METODOLOGI PENELITIAN

## 3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2017/2018. Bertempat di jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

## 3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi. Untuk data simulasi yakni data yang dibangkitkan dengan variabel bebas sebanyak p = 6 dan banyaknya data yaitu n = 25, 50, dan 75 yang diulang sebanyak 100 kali dengan $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \ldots = \beta_p = 1$.

Untuk mendapatkan data kolinearitas pada setiap himpunan data, $X_{ip}$ dibangkitkan menggunakan simulasi Monte Carlo berdasarkan McDonald dan Galarneau (1975) dengan persamaan sebagai berikut :

$$X_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{i(p+1)} \qquad ; i = 1, 2, \ldots, n$$
$$j = 1, 2, \ldots, p \qquad (3.1)$$

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

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<br>$X_{i1} = (1-\rho^2)^{1/2}z_{i1} + \rho z_{i7}$<br>$X_{i2} = (1-\rho^2)^{1/2}z_{i2} + \rho z_{i7}$<br>$X_{i3} = (1-\rho^2)^{1/2}z_{i3} + \rho z_{i7}$<br>$X_{i4} = (1-\rho^2)^{1/2}z_{i4} + \rho z_{i7}$<br>$X_{i5} = (1-\rho^2)^{1/2}z_{i5} + \rho z_{i7}$<br>$X_{i6} = (1-\rho^2)^{1/2}z_{i6} + \rho z_{i7}$ |

Variabel terikat $(Y)$ untuk setiap $p$ variabel bebas diperoleh berdasarkan model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ dimana $\beta$ adalah $\beta_{i,j} = 1$ ; untuk i = j , dan 0 selainnya. Dengan $\varepsilon$ 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 | $Y = X_1 + X_2 + \cdots + X_p + N(0,1)$ $\quad p = 1, 2, \ldots, 6$ |

## 3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka, yaitu dengan mempelajari buku-buku teks penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dengan hasil yang akurat, penulis menggunakan software R versi 3.4.2. Adapun langkah-langkah yang dilakukan adalah sebagai berikut :

1. Melakukan simulasi data.
2. Mengidentifikasi multikolinearitas dengan melihat nilai korelasi dan nilai VIF (*Variance Inflation Factor*).
3. Melakukan analisis dengan menggunakan regresi *ridge.*

   Dengan langkah-langkah sebagai berikut :

   - Melakukan *center and scale* data.
   - Membuat matriks korelasi dari *center and scale* data.
   - Menghitung nilai duga β regresi *ridge* untuk semua nilai k, dengan $0 < k < 1$.
   - Menghitung statistik GCV.
   - Membuat plot validasi GCV untuk memilih nilai k
   - Mendapatkan nilai duga $\beta$ regresi *ridge* dengan nilai k yang telah dipilih.

- Mentransformasi nilai duga $\beta$ regresi *ridge* kedalam variabel asli.

4. Mencatat nilai duga $\beta$ metode regresi *ridge*.
5. Melakukan analisis dengan menggunakan regresi komponen utama.

   Dengan langkah-langkah sebagai berikut :

   - Melakukan *center and scale* data.
   - Membuat matriks korelasi dari *center and scale* data.
   - Menghitung nilai eigen $\lambda_i$ dan vector eigen $V$ dari matriks korelasi.
   - Menghitung nilai komponen utama $Q$.
   - Memilih komponen utama yang memiliki nilai eigen lebih dari 1.
   - Menghitung nilai duga regresi komponen utama berdasarkan komponen yang terpilih.
   - Mentransformasi nilai duga $\beta$ regresi komponen utama kedalam bentuk variabel asli terstandarisasi dengan mensubstitusi nilai komponen utama Q.
   - Mentransformasi kembali nilai duga $\beta$ regresi komponen utama kedalam bentuk variabel asli.
6. Mencatat nilai duga β metode regresi komponen utama.
7. Mencari nilai duga β dengan metode kuadrat terkecil.
8. Menghitung MSE dan AMSE dari nilai duga $\beta$ dengan metode regresi *ridge*, regresi komponen utama dan metode kuadrat terkecil dengan rumus

$$\mathrm{MSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta\right)^2 \tag{3.2}$$

$$\mathrm{AMSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left\|\hat{\boldsymbol{\beta}}_{\boldsymbol{j}} - \boldsymbol{\beta}\right\|^2 \tag{3.3}$$

dengan $m$ adalah banyaknya pengulangan, j= 1, 2, …, m.

$\hat{\beta}_j$ = nilai duga parameter regresi simulasi ke – *j*

9. Membandingkan MSE *(Mean Square Error)* dan AMSE (*Average Mean Square Error*) dari nilai duga β metode regresi *ridge*, regresi komponen utama dan metode kuadrat terkecil.

# V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan sebagai berikut :

1. Nilai penduga parameter dengan menggunakan metode regresi komponen utama lebih baik dibandingkan nilai penduga parameter dengan menggunakan metode regresi *ridge* dan MKT karena paling mendekati nilai parameter sesungguhnya.
2. Nilai MSE dan AMSE regresi komponen utama lebih kecil dibandingkan analisis regresi *ridge.*
3. Metode regresi komponen utama dan metode regresi *ridge* lebih baik dibandingkan metode MKT ketika terdapat multikolinieritas antar variabel bebas dan metode regresi komponen utama lebih baik dibandingkan dengan metode regresi *ridge* dalam menangani multikolinearitas berdasarkan kriteria MSE dan AMSE.
4. Semakin banyak data yang digunakan dalam analisis semakin baik nilai penduga parameter yang dihasilkan.

.

## DAFTAR PUSTAKA


Dereny, M. El. and Rashwan, N.I. 2011. Solving Multicollinearity Problem Using Ridge Regression Models. *Int. J. Contemp. Math. Sciences*. **6**(12):585-600.

Draper, N.R. and 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 Pogram SPSS.* Ed. Ke-4. Badan Penerbit Universitas Diponegoro, Semarang.

Gujarati, D. 1995. *Ekonometri Dasar.* Diterjemahkan oleh Sumarno Zain. Erlangga, Jakarta.

Hoerl, A.E. and Kennard, R.W. 1970. Ridge Regression: Biased Estimator to Nonorthogonal Problems. *Technometrics.* **12**(1):68-82.

Kutner, M.H., *et al*. 2005. *Applied linear Statistic Model*. Ed. Ke-5. Mc-Graw-hill, New York.

Mardikyan, S. and Cetin, E. 2008. Efficient Choice of Biasing Constant for Ridge Regression. *Int. J. Contemp. Math. Sciences*. **3**(11):527-536.

McDonald, G.C. and Galarneau, D.I. 1975. AMonte Carlo Evaluation of some Ridge-type Estimators. *J. Amer. Statist. Asoc.* **70**:407-416.

Montgomery, D.C., *et al*. 2006. *Introduction to Linear Regression Analysis.* Wiley and Sons, Inc., New York.

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

Rougoor, C.W., *et al*. 2000. The Relation Between Breeding Management and 305-day Milk Production, Determined via Principal Components Regression and Partial Least Squares. *Livestock Product Science*. **66**:71-83.

Sumodiningrat, G. 1998. *Ekonometrika Pengantar*. BPFE, Yogyakarta.

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