BAB I
PENDAHULUAN

1.1 Latar Belakang Topik

Para ilmuan, ekonomi, psikolog, dan sosiolog selalu berkepentingan dengan masalah peramalan, karena dapat digunakan untuk menyelesaikan masalah dalam pengelolaan dan manajemen. Salah satu metode yang dapat digunakan untuk menyelesaikan masalah tersebut adalah dengan menggunakan metode statistik, yang salah satunya menggunakan analisis regresi linear.

Regresi linear adalah suatu metode yang digunakan untuk meramalkan nilai dari satu atau lebih variabel terikat apabila nilai dari variabel bebas berubah-ubah. Metode ini juga dapat digunakan untuk meramalkan pengaruh dari variabel bebas terhadap variabel terikat.

Pada metode regresi linier terbagi menjadi dua, yaitu analisis regresi linier sederhana dan analisis regresi linier berganda. Yang membedakan keduanya adalah terletak pada variabel bebas, untuk analisis regresi linier sederhana variabel bebasnya hanya satu sedangkan untuk analisis regresi linier berganda variabel bebasnya lebih dari satu.

Dalam analisis regresi linear, pembahasan yang menarik adalah saat beberapa asumsi seperti homoskedastisitas, tidak adanya multikolinearitas, jumlah pengamatan harus lebih besar dari jumlah variabel yang diamati, tidak adanya autokorelasi, dan linearitas tidak terpenuhi sehingga menimbulkan permasalahan yang harus diselesaikan.

Masalah yang sering ditemukan dalam banyaknya variabel bebas adalah saat variabel bebas yang satu dengan variabel bebas yang lainnya terjadi korelasi atau dinamakan dengan multikolinearitas. Masalah inilah yang akan menyebabkan model dari regresi linear sendiri tidak dapat ditentukan secara tepat karena tujuan dari regresi linear yaitu memperoleh nilai variansi dan standar error yang minimum tidak akan tercapai.

Myers (1990) dalam Nurhasanah (2006) memperkenalkan beberapa metode untuk mengatasi masalah multikolinearitas diantaranya seperti Regresi Komponen Utama dan Regresi Ridge, meskipun penggunaannya masih dalam perdebatan.

Beberapa peneliti lain menggunakan metode Regresi Komponen Utama untuk kasus dengan sampel besar (jumlah pengamatan > 30), dan menggunakan metode Regresi Ridge untuk kasus dengan sampel kecil (jumlah pengamatan < 30), dan Sarwoko (2005) dalam bukunya dasar-dasar ekonometrika memperkenalkan solusi yang paling sederhana untuk mengatasi multikolinearitas yaitu dengan metode penghilangan variabel-variabel yang menyebabkan multikolinearitas apabila variabel-variabel tersebut tidak relevan dalam regresi.

Untuk mengetahui kefektifitasan dalam mengatasi multikolinearitas pada regresi linear berganda harus dilakukan perbandingan dari ketiga metode tersebut dengan kasus yang berbeda. Harapannya dengan melakukan perbandingan tersebut dapat diketahui kekurangan dan kelebihan setiap metode, sehingga dapat pula ditentukan jenis kasus yang cocok dalam setiap penggunaan metode tersebut.

Dari latar belakang inilah penulis merasa tertarik untuk mengambil topik ” *Analisis Efektifitas Metode Perbaikan Model Regresi Linear Berganda yang Terdapat Multikolinearitas”* .

1.2 Rumusan Masalah

Berdasarkan pemaparan di atas maka permasalahan yang akan dibahas dalam penulisan ini adalah:

1. Bagaimana cara pembentukan model regresi komponen utama dari regresi linear berganda?
2. Bagaimana cara pembentukan model Regresi Ridge dari model regresi linear berganda?
3. Bagaimana cara pembentukan model regresi linear berganda setelah dilakukan penghilangan variabel bebas yang diduga mengandung multikolinearitas?
4. Apa kelebihan dan kekurangan masing-masing metode?
5. Jenis kasus seperti apa yang dapat digunakan oleh masing-masing metode dalam mengatasi multikolinearitas?

1.3 Batasan Masalah.

Dalam penelitian ini, ada beberapa batasannya yaitu:

1. Pembahasannya hanya mengenai masalah multikolinearitas saja. Karena masalah yang sering terjadi pada saat pemilihan banyaknya variabel bebas adalah terjadinya multikolinearitas.

2. Asumsikan bahwa beberapa asumsi seperti homoskedastisitas, jumlah pengamatan harus lebih besar dari jumlah variabel yang diamati, tidak adanya autokorelasi, dan linearitas tetap terpenuhi.

1.4 Tujuan dan Manfaat Penelitian

Tujuan Penelitian.

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah:

1. Untuk mengetahui cara pembentukan model regresi komponen utama dari regresi linear berganda.
2. Untuk mengetahui cara pembentukan model Regresi Ridge dari model regresi linear berganda
3. Untuk mengetahui cara pembentukan model regresi linear berganda setelah dilakukan penghilangan variabel bebas yang diduga mengandung multikolinearitas
4. Untuk mengetahui kelebihan dan kekurangan dari masing-masing metode.
5. Untuk mengetahui Jenis kasus seperti apa yang dapat digunakan oleh masing-masing metode dalam mengatasi multikolinearitas.

Manfaat Penelitian Secara Umum

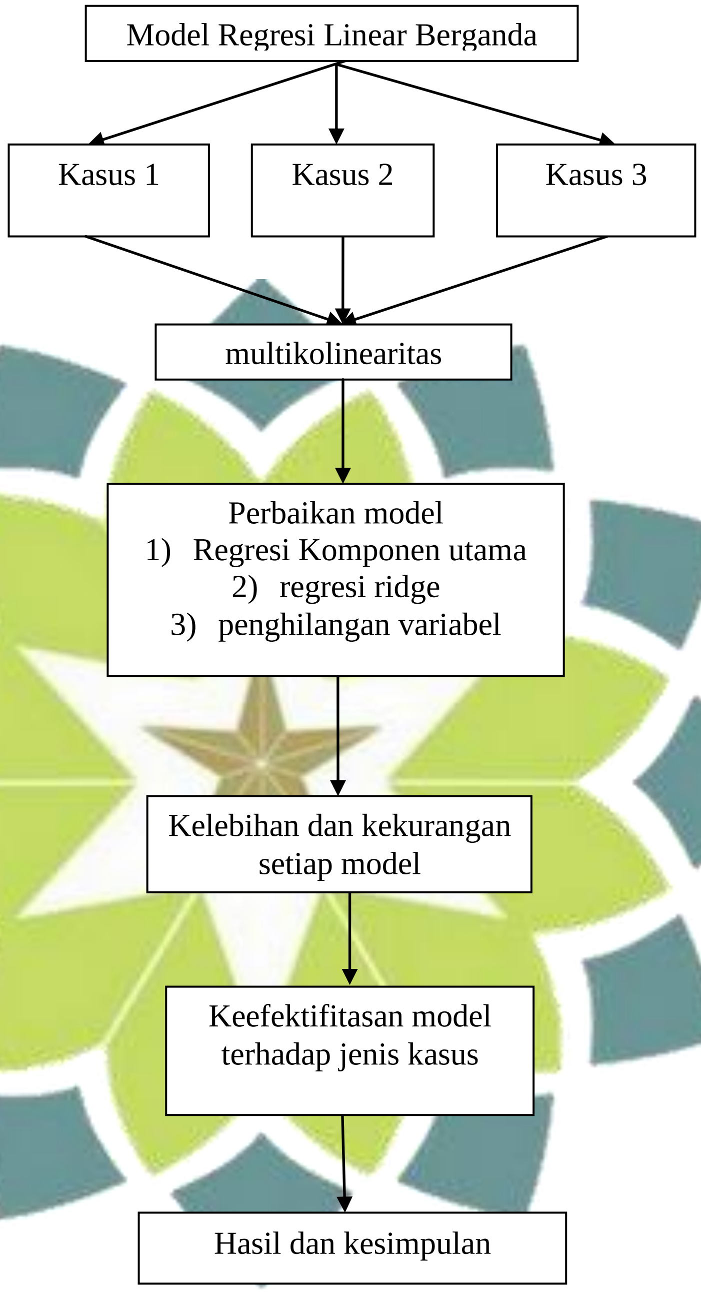
Dari penelitian yang akan dilakukan, penulis berharap dapat memberikan solusi yang paling tepat bagi pengguna regresi linear saat menemukan masalah multikolinearitas. Sehingga dalam melakukan peramalan terhadap suatu masalah dapat ditentukan model dari regresi linear berganda dengan tepat.

1.5 Kerangka Pemikiran

Regresi linear berganda adalah salah satu metode statistik yang digunakan untuk mengetahui pengaruh dari banyaknya variabel bebas terhadap satu variabel terikat. Masalah yang sering ditemukan dalam pemilihan banyaknya variabel bebas dalam model regresi linear berganda adalah terjadi multikolinearitas, yaitu adanya korelasi antara variabel bebas yang satu dengan variabel bebas yang lainnya.

Akibat dari adanya multikolinearitas pada model regresi sangat merugikan, karena harapan untuk model regresi linear berganda sendiri adalah memiliki standar error dan variansi yang minimum tidak akan tercapai. Cara untuk mengetahui adanya multikolinearitas dalam model tersebut adalah dengan melihat perolehan nilai *Variance Inflation Factor* yang melebihi sepuluh.

Diperkenalkan tiga metode untuk mengatasi multikolinearitas diantaranya yaitu 1) metode Regresi Komponen Utama, 2) metode Regresi Ridge dan 3) metode penghilangan variabel. Untuk mengetahui kefektifitasan dalam mengatasi multikolinearitas pada regresi linear berganda harus dilakukan perbandingan dari ketiga cara tersebut. Perbandingan dilakukan dengan menganalisis tiga kasus dengan kasus pertamanya dengan jumlah sampel besar, kasus kedua dengan jumlah sampel kecil, dan kasus ketiga dengan terdapatnya variabel yang tidak relevan dimasukan ke dalam persamaan regresi. Sehingga dengan melakukan perbandingan dari metode tersebut akan ditemukan jenis kasus seperti apa yang dapat digunakan oleh masing-masing metode.



Gambar 1.5.1. Kerangka Pemikiran untuk Penelitian

BAB II

LANDASAN TEORI

2.1 Matriks

Definisi 2.1.1: Matriks adalah suatu susunan bilangan berbentuk segiempat. Bilangan-bilangan dalam susunan itu disebut anggota dalam matriks tersebut **[1]**. Sebuah matriks yang berukuran *m* baris dan *n* kolom dengan a_{ij} dapat ditulis:

$$a_{mxn} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Atau dapat juga ditulis $A = \lfloor a_{ij} \rfloor$ *i = 1, 2, ..., m j = 1, 2, ..., n*

2.1.1 Jenis – Jenis Matriks

Matriks Bujur Sangkar

Sebuah matriks *A* dengan *n* baris dan *n* kolom, dapat ditulis $A = \lfloor a_{ij} \rfloor_{n \times n}$.

Misal

$$a_{n \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Dan anggota-anggota $a_{11}, a_{22}, \ldots, a_{nn}$ disebut sebagai anggota dari diagonal utamanya.

Matriks Diagonal

Matriks bujur sangkar $A = \lfloor a_{ij} \rfloor$ dinamakan matriksdiagonal jika semua elemen selain diagonal utama adalah nol, $a_{ij} = 0$ untuk $i \neq j$.

Matriks Identitas

Matriks bujur sangkar dengan nilai 1 pada diagonal utama dan nilai 0 pada anggota selain diagonal utamanya, dilambangkan dengan $A = \lfloor a_{ij} \rfloor_{m \times n} = I$ dan untuk $m = n$ maka

$a_{ij} = 1 \rightarrow i = j$

$a_{ij} = 0 \rightarrow i \neq j$

Matriks Singular

Matriks bujur sangkar A= $[a_{ij}\ ]$ dikatakan singular jika semua elemen pada salah satu baris atau kolom adalah nol atau jika semua kofaktor dari elemen suatu baris atau kolom sama dengan nol.

Definisi dari kofaktor sendiri yaitu: jika A adalah suatu matriks bujur sangkar, maka minor anggota a_{ij} dan dinyatakan oleh M_{ij} dan didefinisikan sebagai determinan sub-matriks yang masih tersisa setelah baris ke-i dan kolom ke-j dihilangkan dari A. Bilangan $(-1)^{i+j} M_{ij}$ dinyatakan oleh C_{ij} disebut kofaktor anggota a_{ij}

Untuk melihat kesingularan suatu matriks adalah dengan menghitung determinan matriks tersebut. Apabila determinannya sama dengan nol maka matriks tersebut singular.

Matriks Ortogonal

Matriks bujur sangkar A= $[a_{ij}\ ]$ dikatakan dapat didiagonalisasi secara orthogonal jika terdapat matriks orthogonal P sehingga berlaku $P^{-1}AP = P'AP$. Matriks orthogonal didefinisikan sebagai matriks bujur sangkar yang inversnya sama dengan transposenya,

sehingga :$P^{-1} = P'$, maka P adalah matriks orthogonal.

Matriks Topi

Sebuah matriks H dikatakan matriks topi atau *hat matrix* bila:

$$H = X(X'X)^{-1}X$$

Maka mudah terlihat bahwa $H' = H$ dan $HH = H^2 = H$. Jadi H merupakan suatu matriks yang simetri dan idempoten. Dengan jalan yang sama, akan diperlihatkan bahwa $1 - H$ memiliki sifat yang sama, yaitu:

1. $[1 - H]' = [1 - H]$ (simetri)
2. $[1 - H][1 - H] = [1 - H - H + H^2]$

   $= [1 - H]$ (idempoten)

2.1.2 Operasi Matriks

Penjumlahan Matriks dan Pengurangan Matriks

Definisi 2.1.2: jika *A* dan *B* adalah matriks-matriks berukuran sama, maka jumlah *A* + *B* adalah matriks yang diperoleh dengan menambahkan anggota-anggota *B* dengan anggota-anggota *A* yang berpadanan, dan selisih *A* – *B* adalah matriks yang diperoleh dengan mengurangkan anggota-anggota *A* dengan anggota-anggota *B* yang berpadanan. Matriks-Matriks berukuran berbeda tidak bisa dikurangkan.

Dalam notasi matriks, jika $A = \lfloor a_{ij} \rfloor$ dan $B = \lfloor b_{ij} \rfloor$ mempunyai ukuran yang sama, maka $(A + B) = (A) + (B) = \lfloor a_{ij} + b_{ij} \rfloor$, dan $(A - B) = (A) - (B) = \lfloor a_{ij} - b_{ij} \rfloor$.

Perkalian Matriks terhadap Skalar

Definisi 2.1.3: jika *A* adalah sebarang matriks dan c adalah sebarang skalar, maka hasil kali *cA* adalah matriks yang diperoleh dengan mengalikan setiap anggota *A* dengan *c*. Dalam notasi matriks, jika $A = \lfloor a_{ij} \rfloor$, maka $(cA) = c(A) = c\lfloor a_{ij} \rfloor$.

Perkalian Matriks terhadap Matriks

Definisi 2.1 4: jika *A* adalah sebuah matriks *m* × *r*, dan *B* adalah sebuah matriks *r* × *n*, maka hasil kali *AB* adalah matriks *m* × *n* yang anggota-anggotanya didefinisikan sebagai berikut: Untuk mencari anggota dalam baris *i* dan kolom *j* dari *AB*, pilih baris *i* dari matriks *A* dan kolom *j* dari matriks *B*, kalikan anggota-

anggota yang berpadanan dari baris dan kolom secara bersama-sama dan kemudian jumlahkan hasil kalinya.

Transpose Suatu Matriks

Definisi 2.1.4: jika A adalah sebarang matriks $m \times n$, maka transpos A, dinyatakan dengan A' didefinisikan sebagai matriks $n \times m$ yang didapatkan dengan mempertukarkan baris dan kolom dari A; yaitu, kolom pertama dari A' adalah baris pertama dari A, kolom kedua dari A' adalah baris kedua dari A, dan seterusnya.

Dalam notasi matriks $(A')_{ij} = (A)_{ji}$.

Trace suatu Matriks

Definisi 2.1.5: jika A adalah suatu matriks bujur sangkar, maka *trace* A dinyatakan dengan $tr(A)$, didefinisikan sebagai jumlah anggota-anggota pada diagonal utama A. *Trace* A tidak terdefinisi jika A bukan matriks bujur sangkar.

Invers Matriks

Definisi 2.1.6: jika A adalah sebuah matriks bujur sangkar, dan jika sebuah matriks B yang berukuran sama bisa didapatkan sedemikian rupa sehingga $AB = BA = I$, maka A disebut bisa dibalik dan B disebut invers dari A.

Determinan Matriks

Definisi 2.1.7: anggap *A* adalah suatu matriks bujur sangkar. Fungsi determinan dinyatakan dengan det, dan mendefinisikan *det(A)* sebagai jumlah semua hasil kali dasar bertanda dari *A*. Angka *det(A)* disebut determinan *A*.

2.2 Nilai Eigen dan Vektor Eigen

Definisi 2.2.1: jika *A* adalah sebuah matriks $n \times n$ maka sebuah vektor tak nol *X* pada R^n disebut vektor eigen (*eignvector*) dari *A* jika *AX* adalah sebuah kelipatan skalar dari *X*; yakni

$$AX = \lambda X$$

Untuk sebarang skalar λ. Skalar λ disebut nilai eigen (*eigenvalue*) dari *A*, dan *X* disebut sebagai vektor eigen dari *A* yang bersesuaian dengan λ.

Untuk menentukan nilai eigen matriks *A* yang berukuran *nxn*:

Misalkan $A_{n\times n} = \begin{bmatrix} a_{11} & . & . & . & a_{1n} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ a_{n1} & . & . & . & a_{nn} \end{bmatrix}$ $I_{n\times n} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ $X = \begin{bmatrix} x_i \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Sehingga $$AX = \lambda X, \quad X \neq 0$$

$$AX = \lambda IX$$

$$\lambda IX - AX = 0$$

$$(\lambda I - A)X = 0$$

karena $X \neq 0$ maka $|\lambda I - A| = 0$

Persamaan $|\lambda I - A| = 0$ disebut persamaan karakteristik. Dan nilai λ dapat diperoleh dari:

$$\begin{vmatrix} \lambda - a_{11} & . & . & . & -a_{n1} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ -a_{n1} & . & . & . & \lambda - a_{nxn} \end{vmatrix} = 0$$

$$f(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n = 0$$

Dari persamaan $a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n = 0$ memiliki sebanyak – banyaknya n solusi yang berbeda, sehingga sebuah matriks $n \times n$ memiliki sebanyak n solusi berbeda **[2]**.

2.3 Diagonlisasi

Definisi 2.3.1: sebuah matriks bujur sangkar A dikatakan dapat didiagonalisasi jika terdapat sebuah matriks P yang dapat dibalik sedemikian rupa sehingga $P^{-1}AP$ adalah sebuah matriks diagonal sehingga matriks P dikatakan mendiagonalisasi A.

Terdapat bebrapa cara untuk mendiagonalisasikan sebuah matriks. Misal terdapat matriks A, maka langkah untuk mendiagonalisasikan adalah sebagai berikut:

1. tentukan n vektor eigen dari A yang bebas linear, misalkan $P_1, P_2, \ldots P_n$
2. bentuklah sebuah matriks P dengan $P_1, P_2, \ldots P_n$ sebagai vektor kolomnya

3. Matriks $P^{-1}AP$ kemudian akan menjadi diagonal dengan $\lambda_1, \lambda_2, \ldots \lambda_n$ sebagai entri-entri diagonalnya secara berurutan, di mana λ_i adalah nilai eigen yang terkait dengan P_i, untuk $i = 1, 2, \ldots, n$

Jika diberikan sebuah matriks A $n \times n$, dan apabila terdapat matriks ortogonal P sedemikian rupa sehingga matriks $P^{-1}AP = P'AP$ merupakan diagonal, maka matriks A dikatakan dapat didiagonalisasikan secara ortogonal dan P dikatakan mendiagonalisasi secara ortogonal matriks A.

2.4 Model Regresi Linear Berganda

Regresi linear adalah salah satu metode statistik yang digunakan untuk mengetahui pengaruh dari banyaknya variabel bebas terhadap satu variabel terikat. Menurut banyaknya variabel bebas, terdapat dua macam model regresi linear yaitu regresi linear sederhana dengan memiliki satu variabel bebas dan regresi linear berganda dengan memiliki lebih dari satu variable bebas.

Model regresi linear sederhana mempunyai bentuk:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.4.1)$$

Model regresi linear berganda mempunyai bentuk:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_r X_r + \varepsilon \qquad (2.4.2)$$

Dengan

Y : variabel terikat

$X_1, \ldots, X_r$: variabel bebas

$\beta_0, \beta_1, \ldots, \beta_r$: parameter yang tidak diketahui

ε : *error*.

Variabel terikat adalah variabel yang nilainya ditentukan oleh variabel lain, sedangkan variabel bebas adalah variabel yang digunakan untuk memprediksi nilai variabel lain **[14]**. Parameter yang tidak diketahui merupakan koefisien regresi yang menunjukan angka peningkatan ataupun penurunan variabel terikat yang didasarkan pada perubahan variabel bebas **[4]**. Dengan melakukan pengamatan sebanyak n pada Y maka model lengkap regresi linier berganda berbentuk:

$$\left.\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \ldots + \beta_r X_{1r} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \ldots + \beta_r X_{2r} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \ldots + \beta_r X_{nr} + \varepsilon_n \end{aligned}\right\} \quad (2.4.3)$$

Persamaan (2.4.3) dapat diperlihatkan dengan matriks berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1r} \\ 1 & X_{21} & X_{22} & \cdots & X_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Atau $\underset{(nx1)}{Y} = \underset{(nx(r+1))}{X} \underset{((r+1)x1)}{\beta} + \underset{(nx1)}{\varepsilon}$ (2.4.4)

Untuk memperoleh model regresi linear berganda yang tepat, maka harus memenuhi beberapa asumsi sebagai berikut **[15]**:

a. Nilai rata-rata error adalah nol, yaitu: $E(\varepsilon_i) = 0$ untuk $i = 1,2,\ldots,n$

b. $\text{var}(\varepsilon_i) = E(\varepsilon_i^{\,2}) = \sigma^2$,adalah konstan untuk semua error (asumsi homoskedastisitas). Variansi sendiri adalah bilangan yang menyatakan bervariasinya nilai suatu variabel terhadap nilai rata – rata hitungnya **[18]**.

c. Tidak ada korelasi antara error yang satu dengan error yang lainnya, berarti $kov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$ (asumsi non autokorelasi). kovariansi sendiri adalah bilangan yang menyatakan bervariasinya nilai suatu variabel dalam hubungan asosiatifnya dengan variabel lain, rumusan kovarian sama dengan variansi hanya saja penggunaan kovarian biasa digunakan untuk menyatakan hubungan antara dua variabel **[18]**.

d. Variabel bebas dengan error tidak berkorelasi (saling bebas)

e. Tidak ada multikolinearitas diantara variabel bebas

2.5 Penaksir Kuadrat Terkecil

Tujuan dari regresi adalah untuk mendapatkan nilai prediks ($\hat{Y}$) yang sedekat mungkin dengan data aktualnya (*Y*), maksudnya untuk mendapatkan *error* yang sekecil mungkin **[6]**. Salah satu metode yang dapat digunakan adalah metode penaksir kuadrat terkecil.

Misalkan b adalah taksiran untuk β, sehingga persamaan estimasi dapat ditulis $Y = Xb + \varepsilon$ atau $\varepsilon = Y - Xb$. Tujuan dari metode kuadrat terkecil adalah meminimumkan jumlah kuadrat error, yaitu

$\sum_{i=1}^{n} \varepsilon_i^{\ 2} =$ minimum, maka

$$\sum_{i=1}^{n} \varepsilon_i^{\ 2} = \varepsilon_1^{\ 2} + \varepsilon_2^{\ 2} + \ldots + \varepsilon_n^{\ 2} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Sehingga

$$\sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 X_{i1} - \ldots - b_r X_{jr}\right)^2 \qquad (2.5.1)$$

Dengan menurunkan $\sum_{i=1}^{n} \varepsilon_i^{\,2}$ terhadap $b_0, b_1, b_2, \ldots, b_r$ secara parsial kemudian samakan dengan nol maka akan diperoleh **[13]**:

$$\frac{\partial \varepsilon' \varepsilon}{\partial b_0} = -2\sum \left(Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - \ldots - b_r X_{ir}\right) = 0$$

$$\frac{\partial \varepsilon' \varepsilon}{\partial b_1} = -2\sum \left(Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - \ldots - b_r X_{ir}\right) X_{i1} = 0$$

$$\frac{\partial \varepsilon' \varepsilon}{\partial b_2} = -2\sum \left(Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - \ldots - b_r X_{ir}\right) X_{i2} = 0$$

$$\vdots$$

$$\frac{\partial \varepsilon' \varepsilon}{\partial b_r} = -2\sum \left(Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - \ldots - b_r X_{ir}\right) X_{ir} = 0$$

Setelah disusun kembali maka persamaan di atas dapat ditulis sebagai

$$\sum Y_i = nb_0 + b_1 \sum X_{i1} + b_2 \sum X_{i2} + \ldots + b_r \sum X_{ir}$$

$$\sum Y_i X_{i1} = b_0 \sum X_{i1} + b_1 \sum X_{i1}^{\,2} + b_2 \sum X_{i2} X_{i1} + \ldots + b_r \sum X_{ir} X_{i1}$$

$$\sum Y_i X_{i2} = b_0 \sum X_{i2} + b_1 \sum X_{i1} X_{i2} + b_2 \sum X_{i2}^{\,2} + \ldots + b_r \sum X_{ir} X_{i2}$$

$$\vdots$$

$$\sum Y_i X_{ir} = b_0 \sum X_{ir} + b_1 \sum X_{i1} X_{ir} + b_2 \sum X_{i2} X_{ir} + \ldots + b_r \sum X_{ir}^{\,2}$$

Bentuk persamaan matriks menjadi

$$\begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \dots & \sum X_{ir} \\ \sum X_{i1} & \sum X_{i1}{}^2 & \sum X_{i2}X_{i1} & \dots & \sum X_{ir}X_{i1} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}{}^2 & & \sum X_{ir}X_{i2} \\ \vdots & \dots & \dots & \dots & \vdots \\ \sum X_{ir} & \sum X_{i1}X_{ir} & \sum X_{i2}X_{ir} & \dots & \sum X_{ir}{}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & X_{31} & \dots & X_{r1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{r2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{1r} & X_{2r} & X_{33} & \dots & X_{ri} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_r \end{bmatrix}$$

Atau menjadi : $(X'X)b = X'Y$

$$b = (X'X)^{-1}X'Y \qquad (2.5.2)$$

Untuk taksiran dari β biasanya dilambangkan dengan $\hat{\beta}$ maka penaksir kuadrat terkecil pada regresi linear berganda adalah $\hat{\beta} = (X'X)^{-1}X'Y$

2.6 Jumlah Dekomposisi Kuadrat

Teorema 2.6.1: Misal X sebanyak $r+1 \le n$ dengan penaksir kuadrat terkecil dari β adalah $\hat{\beta} = (X'X)^{-1}X'Y$, dengan *error*nya adalah: $\hat{\varepsilon} = Y - \hat{Y} = [(X'X)^{-1}X'Y]$ memenuhi $Z'\hat{\varepsilon} = 0$ dan $y'\hat{\varepsilon} = 0$ dengan jumlah *error* kuadratnya adalah:

$$\hat{\varepsilon}'\hat{\varepsilon} = Y'[I - X(X'X)X']Y$$
$$= [Y' - Y'X(X'X)X']Y$$
$$= Y'Y - Y'X(X'X)X'Y$$
$$= Y'Y - Y'X\hat{\beta}$$

Pembuktian:

Misalkan dinyatakan bahwa $\hat{\beta} = (X'X)^{-1}X'Y$ maka

$\hat{\varepsilon} = Y - \hat{Y}$

$= Y - X\hat{\beta}$

$= Y - X(X'X)^{-1}X'Y$

$= [I - (X'X)^{-1}X']Y$ (2.6.1)

Dengan $[I - (X'X)^{-1}X'] = H$ (H merupakan *hat matrix*)

Maka $Y'\hat{\varepsilon} = Y'(Y - \hat{Y})$

$= Y'[1 - X(X'X)^{-1}X']Y$

$= 0$ (2.6.2)

dan $Y'\hat{\varepsilon} = \hat{\beta}X'\hat{\varepsilon}$.

$= \hat{\beta}X'Y'[1 - X(X'X)^{-1}X']Y$

$= 0$ (2.6.3)

Sehingga $\hat{\varepsilon}'\hat{\varepsilon} = (Y - \hat{Y})'(Y - \hat{Y})$

$= ([1 - X(X'X)^{-1}X']Y)'([1 - X(X'X)^{-1}X']Y)$

$= Y'[1 - H]'[1 - H]Y$

$= Y'[1 - X(X'X)^{-1}X']Y$

$= Y'Y - Y'X(X'X)^{-1}X'Y$

$= Y'Y - Y'X\hat{\beta}$ (2.6.4)

Terbukti bahwa $\hat{\varepsilon}'\hat{\varepsilon} = Y'Y - Y'X\hat{\beta}$

Dari persamaan 2.5.3 diperlihatkan bahwa $y'\hat{\varepsilon} = 0$, jadi jumlah variabel terikat total kuadrat $Y'Y = \sum_{j=1}^{n} y_j^{\,2}$ memenuhi

$$Y'Y = (\hat{Y} + Y - \hat{Y})'\,(\hat{Y} + Y - \hat{Y})$$

$$= (\hat{Y} + \hat{\varepsilon})'(\hat{Y} + \hat{\varepsilon})$$

$$= \hat{Y}'\hat{Y} + \hat{Y}'\hat{\varepsilon} + \hat{\varepsilon}'\hat{Y} + \hat{\varepsilon}'\hat{\varepsilon}$$

$$= \hat{Y}'\hat{Y} + 0 + 0 + \hat{\varepsilon}'\hat{\varepsilon}$$

$$= \hat{Y}'\hat{Y} + \hat{\varepsilon}'\hat{\varepsilon} \qquad (2.6.5)$$

Karena kolom pertama dari X adalah 1, kondisi $X'\hat{\varepsilon} = 0$ memenuhi persamaan

$$0 = 1'\hat{\varepsilon}$$

$$= \sum_{j=1}^{n} \hat{\varepsilon}_j$$

$$= \sum_{j=1}^{n} y_j - \sum_{j=1}^{n} \hat{y}_j \text{ sehingga } \bar{Y} = \bar{\hat{Y}} \qquad (2.6.6)$$

Jika kedua ruas dari persamaan (2.6.5) dikurangi $n\bar{Y}^2 = n\bar{\hat{Y}}^2$ diperoleh dekomposisi (pemisahan variabel) dasar dari jumlah rata-rata kuadrat

$$Y'Y - n\bar{Y}^2 = \hat{Y}'\hat{Y} - n(\bar{\hat{Y}})^2 + \hat{\varepsilon}'\hat{\varepsilon} \qquad (2.6.7)$$

Atau $\sum_{j=1}^{n} (y_j - \bar{y})^2 = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2 + \sum_{j=1}^{n} \hat{\varepsilon}^2{}_j$ (2.6.8)

Jumlah kuadrat tersebut menyarankan kualitas dari model yang tepat dapat diukur dengan menghitung koefisien determinasi yaitu

$$R^2 = \frac{\sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2} \qquad (2.6.9)$$

Dimana :

R^2 : koefisien determinasi

$\bar{Y}$: rata - rata nilai Y

Nilai dari R^2 merupakan koefisien determinasi yang menunjukan arah dan kuatnya hubungan antara variabel terikat dengan variabel bebas. R^2 juga merupakan fungsi yang memiliki sifat selalu menaik, yaitu semakin banyak variabel yang tercakup dalam suatu model regresi, makin besar juga nilai R^2 tersebut **[6]**.

2.7 Sifat Sampling dari Penaksir Kuadrat Terkecil

Sebelum variabel terikat $Y = X\beta + \varepsilon$ dilakukan pengamatan, maka Y merupakan vektor acak **[6]**. Maka untuk $\hat{\beta} = (X'X)^{-1}X'y$

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

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

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

$$= \beta + (X'X)^{-1}X'\varepsilon \qquad (2.7.1)$$

Dan $\hat{\varepsilon} = [I - X(X'X)^{-1}X']y$

$$= [I - X(X'X)^{-1}X'][X\beta + \varepsilon]$$

$$= [I - X(X'X)^{-1}X']X\beta + [I - X(X'X)^{-1}X']\varepsilon$$

$$= [(I - X(X'X)^{-1}X')X]\beta + [I - X(X'X)^{-1}X']\varepsilon$$

$$= [X - X(X'X)^{-1}X'X]\beta + [I - X(X'X)^{-1}X']\varepsilon$$

$$= [X - XI]\beta + [I - X(X'X)^{-1}X']\varepsilon$$

$$= [I - (X'X)^{-1}X']\varepsilon \qquad (2.7.2)$$

Dari sifat ekspektasi yaitu bila a dan b tetapan maka $E(aY+b)=aE(Y)+b$

Sehingga untuk $E(\hat{\beta})=E(\beta+(X'X)^{-1}X'\varepsilon)$

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

$$=\beta+(X'X)^{-1}X'0$$

$$=\beta \tag{2.7.3}$$

Dari sifat variansi yaitu bila a dan b tetapan maka $E(aY+b)=a^2E(Y)$

Sehingga untuk var($\hat{\beta}$) $=var(\beta+(X'X)^{-1}X'\varepsilon$

$$=(X'X)^{-1}X'\text{var}(\varepsilon)(X'X)^{-1}X$$

$$=\sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$=\sigma^2(X'X)^{-1} \tag{2.7.4}$$

Untuk variansinya yaitu $\sigma^2(X'X)^{-1}$

Untuk $\hat{\varepsilon}'\hat{\varepsilon}=\varepsilon'(I-H)'(I-H)\varepsilon$

$$=\varepsilon'(I-H)\varepsilon$$

$$=\text{tr}[\varepsilon'(I-H)\varepsilon]$$

$$=\text{tr}[(I-H)\varepsilon\varepsilon'] \tag{2.7.5}$$

Sekarang untuk perkalian *n x n* matriks acak *W* adalah

$$E(tr(W))=E(W_{11}+W_{12}+\ldots+W_{nm})$$

$$=E(W_{11}+W_{12}+\ldots+W_{nm})=tr[E(W)]$$

Maka E($\hat{\varepsilon}'\hat{\varepsilon}$) = tr([$I-H$)] E($\varepsilon\varepsilon'$))

$$=\sigma^2\ \text{tr}[I-H]$$

$$=\sigma^2\text{tr(I)}-\sigma^2\text{tr}[X(X'X)^{-1}X']$$

$$= \sigma^2 n - \sigma^2 \,\mathrm{tr}[(X'X)^{-1}X'X]$$

$$= n\sigma^2 - \sigma^2 \,\mathrm{tr}\left[\underset{(r+1)\times(r+1)}{I}\right]$$

$$= \sigma^2 (n-r-1) \qquad (2.7.6)$$

Karena pada umumnya σ^2 tidak diketahui maka σ^2 diduga dengan s^2. Maka hasil untuk $s^2 = \hat{\varepsilon}\hat{\varepsilon}'$/(n-r-1), dan untuk standar errornya yaitu

$$Se = \sqrt{s^2 (X'X)^{-1}} \qquad (2.7.7)$$

Dimana:

Se : standar error

s^2 : variansi untuk sampel

Standar error sendiri yaitu penyimpangan titik variabel dari garis regresi.

2.8 Analisis Variansi (ANAVA)

Pada persamaan (2.6.8) yaitu $\sum_{j=1}^{n}(y_j - \bar{y})^2 = \sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2 + \sum_{j=1}^{n}\hat{\varepsilon}^2{}_j$ merupakan teknik analisis variansi dengan memecah jumlah kuadrat total (JKT) yaitu $\sum_{j=1}^{n}(y_j - \bar{y})^2$ menjadi dua komponen yaitu $\sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2$ yang merupakan jumlah kuadrat regresi (JKR) dan $\sum_{j=1}^{n}\hat{\varepsilon}^2{}_j$ merupakan jumlah kuadrat error/sisa (JKS). Apabila dinyatakan dalam bentuk matriks, maka akan diperoleh:

$JKT = JKR + JKS$

$$(Y'Y - n\bar{Y}^2) = \left(\hat{\beta}'(X'Y) - n\bar{Y}^2\right) + Y'Y - \hat{\beta}'(X'Y)$$

Berikut ini terdapat tabel analisis variansi (Anava) dengan pendekatan matriks

Sumber variasi	Jumlah kuadrat	Rata-rata kuadrat	Derajat kebebasan
regresi	$JKR = \hat{\beta}'(X'Y) - n\bar{Y}^2$	$(\hat{\beta}'(X'Y) - n\bar{Y}^2)/(K-1)$	$k-1$
residu	$JKS = Y'Y - \hat{\beta}'(X'Y)$	$(Y'Y - \hat{\beta}'(X'Y))/(n-k)$	$n-k$
total	$JKT = JKR + JKS$		$n-1$

$$F_{hitung} = \frac{JKR/(k-1)}{JKS/(n-k)}$$

Distribusi F inilah yang digunakan untuk menguji kelinearan suatu regresi. Jika $F_{hitung} > F_{tabel}$ dengan taraf signifikan yang dipilih maka dapat disimpulkan bahwa regresi tersebut merupakan regresi linear. Alasan menggunakan distribusi F karena dapat digunakan untuk mengevaluasi pengaruh semua variabel bebas terhadap variabel terikat.

Adapun tujuan dari analisis variansi sendiri yaitu **[15]**:

1. Menguji secara bersama-sama seluruh koefisien regresi yaitu menguji hipotesis nol bahwa koefisien regresi yang sebenarnya nol, dengan alternatif bahwa paling tidak ada satu yang tidak sama dengan nol.

   $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$

   $H_1 : \beta_j \neq 0$

   Hipotesis ini berarti bahwa seluruh variabel bebas tidak mempengaruhi variabel terikat, sehingga apabila H_0 diterima dengan kriteria $F_{hitung} < F_{tabel}$, regresi linear tidak boleh digunakan untuk meramalkan variabel terikat Y.

2. Memperkirakan/memperhitungkan kontribusi dari beberapa variabel bebas terhadap variabel terikat. Hal ini digunakan untuk menguji apakah penambahan satu variabel bebas ke dalam model regresi dapat menambah atau memperbesar R^2 yang berarti meningkatkan ketelitian hasil perkiraan variabel terikat Y.

2.9 Matriks Korelasi

Matriks X didefinisikan sebagai berikut:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

Jika pengamatan sebuah sampel sebanyak n, maka rata-ratanya didefinisikan sebagai berikut:

$$\bar{X}_j = \frac{1}{n}\sum X_{ji} \quad ; j = 1,2,3,\dots n \text{ dan } i = 1, 2, 3, \dots, k \qquad (2.9.1)$$

Variansi sampel didefinisikan sebagai berikut:

$$S_j^2 = \frac{1}{n-1}\sum\left(X_{ji} - \bar{X}_j\right)^2 \quad j = 1,2,3,\dots,n \qquad (2.9.2)$$

$S_j^2 = S_{jj}$ = variansi sampel ke-j

Variansi sampel yang menunjukan tingkat hubungan antara dua sampel didefinisikan sebagai berikut:

$$S_{jh} = cov\left(X_j\ X_h\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ji} - \bar{X}_j\right)\left(X_{hi} - \bar{X}_h\right) \qquad (2.9.3)$$

Dengan $j = 1,2,3,\dots,k$ dan $h = 1, 2, 3, \dots, k$

S_{jh} = kovariansi antara X_j dan X_h

Untuk analisis dengan satuan variabel yang berbeda maka dilakukan pembakuan dengan pemusatan dan penskalaan sehingga variabel terikat Y dan variabel bebas X didefinisikan sebagai berikut:

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{(n-1)S_{yy}}} \text{ dan } X_i^* = \frac{X_i - \bar{X}}{\sqrt{(n-1)S_{jj}}} \tag{2.9.4}$$

Model regresi untuk model yang dibakukan diatas dapat dibuat dalam bentuk matriks seperti ini:

$$\begin{bmatrix} Y_1^* \\ Y_2^* \\ \vdots \\ Y_n^* \end{bmatrix} = \begin{bmatrix} X_{11}^* & X_{12}^* & \cdots & X_{1r}^* \\ X_{21}^* & X_{22}^* & \cdots & X_{2r}^* \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1}^* & X_{n2}^* & \cdots & X_{nr}^* \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_r^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{bmatrix}$$

Dari persamaan di atas diperoleh matriks $X^{*'}X^*$ dan $X^{*'}Y^*$, yaitu

$$X^{*'}X^* = \begin{bmatrix} \sum_{i=1}^{n} X_{1i}^{*2} & \sum_{i=1}^{n} X_{1i}^* X_{2i}^* & \cdots & \sum_{i=1}^{n} X_{1i}^* X_{ni}^* \\ \sum_{i=1}^{n} X_{1i}^* X_{2i}^* & \sum_{i=1}^{n} X_{2i}^{*2} & \cdots & \sum_{i=1}^{n} X_{i2}^* X_{ni}^* \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{1i}^* X_{ni}^* & \sum_{i=1}^{n} X_{ik}^* X_{i2}^* & \cdots & \sum_{i=1}^{n} X_{ni}^{*2} \end{bmatrix}$$

Dan

$$X^{*'}Y^* = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^* Y_i^* \\ \sum_{i=1}^{n} X_{i2}^* Y_i^* \\ \vdots \\ \sum_{i=1}^{n} X_{in}^* Y_i^* \end{bmatrix}$$

Matriks $X^{*'}X^*$ dan $X^{*'}Y^*$ dapat juga ditulis dalam bentuk matriks korelasi yaitu sebagai berikut:

$$\sum_{i=1}^{n} X_{1i}^{*2} = \sum \left\{ \frac{X_{1i}{}^{2} - \bar{X}_1}{\sqrt{(n-1)}S_1} \right\}^2$$

$$= \frac{\sum \left(X_{1i}{}^{2} - \bar{X}_1 \right)}{(n-1)S_1{}^{2}}$$

$$= \frac{(n-1)S_1{}^{2}}{(n-1)S_1{}^{2}} = 1$$

Sama halnya untuk $\sum_{i=1}^{n} X_{2i}^{*2}$ dan $\sum_{i=1}^{n} X_{ni}^{*2}$

Untuk

$$\sum_{i=1}^{n} X_{1i}^{*} X_{2i}^{*} = \sum \left\{ \frac{X_{1i} - \bar{X}_1}{\sqrt{(n-1)}S_1} \right\} \left\{ \frac{X_{2i} - \bar{X}_2}{\sqrt{(n-1)}S_2} \right\}$$

$$= \sum_{i=1}^{n} \left\{ \frac{\left(X_{1i} - \bar{X}_1 \right)(X_{2i} - \bar{X}_2)}{(n-1)S_1 S_2} \right\}$$

$$= \frac{\sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)(X_{2i} - \bar{X}_2)}{(n-1)\sqrt{\frac{\sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)^2}{(n-1)}} \sqrt{\frac{\sum_{i=1}^{n} (X_{2i} - \bar{X}_2)^2}{(n-1)}}}$$

$$= \frac{\sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)(X_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)^2} \sqrt{\sum_{i=1}^{n} (X_{2i} - \bar{X}_2)^2}}$$

$$= r_{12} = r_{21}$$

sehingga $r_{xjx\,h=} = \frac{\sum_{i=1}^{n} \left(X_{ji} - \bar{X}_j \right)(X_{hi} - \bar{X}_h)}{\sqrt{\sum_{i=1}^{n} (X_{ji} - \bar{X}_j)^2} \sqrt{\sum_{i=1}^{n} (X_{hi} - \bar{X}_h)^2}}$

Maka matriks korelasi antar variabel bebasnya adalah

$$r_{XX} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix}$$

2.10 *Variance Inflation Factor*

Variance Inflation Factor adalah faktor yang mempengaruhi kenaikan variansi berdasarkan nilai koefisien determinasinya **[5]**. VIF didefinisikan sebagai berikut:

$$(VIF)_j = \frac{1}{1 - R^2{}_j} \tag{2.10.1}$$

Terdapat persamaan regresi linear berganda:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_1 + \ldots + \beta_r X_r + \varepsilon \tag{2.10.2}$$

Maka langkah-langkah menghitung VIF pada tiap variabel adalah sebagai berikut **[12]**:

1. Menjalankan regresi dengan menggunakan metode penaksir kuadrat terkecil dimana variabel bebas X_i merupakan fungsi dari semua variabel bebas lainnya di dalam persamaan itu. Jika i=1 maka persamaan adalah

   $$X_1 = a_1 + a_2 X_2 + a_3 X_3 + \ldots + a_r X_r + u \tag{2.10.3}$$

   Persamaan (2.10.3) disebut regresi *Auxilary*, dengan demikian terdapat r regresi *auxilary* apabila satu per satu dari variabel-variabel dalam persamaan (2.10.2) menjadi variabel bebas
2. Menghitung VIF dengan menggunakan:

$$(VIF)_j = \frac{1}{1 - R^2{}_j}$$

Dimana $R^2{}_j$ adalah koefisien determinasi pada regresi *auxilary* pada langkah pertama.

2.11 Multikolinearitas

2.11.1 Pengertian multikolinearitas

Istilah multikolinearitas atau kolinearitas ganda diciptakan oleh Ragner Frish yang berarti, adanya hubungan linear yang sempurna diantara variabel-variabel bebas dalam model regresi **[15]**.

Menurut tinggi rendahnya masalah multikolinearitas dibedakan menjadi dua yaitu **[18]**:

a. Multikolinearitas sempurna, adalah hubungan antara dua atau lebih variabel bebas yang sifatnya deterministik yaitu mengakibatkan nilai menjadi nol. Nilai dari satu variabel bebasnya dapat dinyatakan dengan perkalian nilai bebas yang lain dengan suatu bilangan tertentu.

b. Multikolinearitas hampir sempurna, adalah hubungan antara dua atau lebih variabel bebas yang korelasinya kuat meskipun tidak deterministik.

Suatu hubungan linear (hubungan antar variabel tidak bebas linear) dikatakan ada apabila kondisi berikut dipenuhi:

$$k_1X_1 + k_2X_2 + \ldots + k_nX_n = 0 \qquad (2.11.1)$$

dimana $k_1, k_2, \ldots, k_n$ adalah konstanta yang sedemikian rupa tidak semuanya sama dengan nol **[5]**. Saat ini, istilah multikolinearitas digunakan dalam pengertian yang lebih luas untuk memasukan kasus multikolinearitas sempurna maupun kasus dimana variabel X berkorelasi tetapi tidak secara sempurna. Persamaan (2.11.1) merupakan persamaan untuk multikolinearitas sempurna, dan untuk multikolinearitas tidak sempurna memiliki persamaan **[5]**:

$$k_1 X_1 + k_2 X_2 + \ldots + k_n X_n + \varepsilon_i = 0 \qquad (2.11.2)$$

Dimana ε_i adalah *error*nya. Untuk melihat perbedaan antara keduanya adalah misal asumsikan bahwa $k_2 \neq 0$ maka persamaan (2.11.1) dapat ditulis sebagai:

$$X_2 = -\frac{k_1}{k_2} X_1 - \frac{k_3}{k_2} X_3 - \ldots - \frac{k_n}{k_2} X_n \qquad (2.11.3)$$

Persamaan di atas menunjukan bagaimana X_2 dapat diperoleh dari kombinasi linear variabel X lain. Untuk multikolinearitas tidak sempurna, dengan asumsi bahwa $k_2 \neq 0$ persamaan (2.11.3) dapat ditulis sebagai:

$$X_2 = -\frac{k_1}{k_2} X_1 - \frac{k_3}{k_2} X_3 - \ldots - \frac{k_n}{k_2} X_n - \frac{1}{k_2} \varepsilon_i \qquad (2.11.4)$$

Persamaan (2.11.4) menunjukan bahwa X_2 bukan merupakan kombinasi linear yang pasti dari X lainnya karena ditentukan pula oleh error ε_i.

2.11.2 Akibat dari multikolinearitas

Beberapa akibat yang ditimbulkan karena adanya multikolinearitas adalah sebagai berikut:

a. Untuk multikolinearitas yang sempurna, perkiraan koefisien regresi untuk β tidak dapat ditentukan dan variansi serta standar errornya tidak terhingga **[18]**. Hal ini diperlihatkan pada saat penentuan

$\hat{\beta} = (X'X)^{-1} X'Y$, untuk $(X'X)^{-1}$ nilai determinannya adalah tidak terdefinisi. Begitupun untuk variansinya yaitu var $(\hat{\beta})$ = $\sigma^2 (X'X)^{-1}$ tidak terdefinisi dan standar errornya yaitu $Se = \sqrt{s^2 (X'X)^{-1}}$ tak terdefinisi juga. Ini artinya model untuk regrei linear klasik tidak dapat ditentukan.

b. Untuk multikolinearitas yang kurang sempurna, masih mungkin untuk menghitung perkiraan koefisien regresi, tetapi nilai variansi dan standar errornya besar **[12]**. Misalkan $X_{n2} = kX_{n1} + \varepsilon_n$ maka matriks untuk persamaan $(X'X)^{-1}$ pada regresi linear klasik adalah

$$\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ kX_{11} + v_1 & kX_{21} + v_2 & \cdots & kX_{n1} + v_n \\ \vdots & \vdots & \ddots & \vdots \\ X_{1r} & X_{2r} & \cdots & X_{nr} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & kX_{11} + v_1 & \cdots & X_{1r} \\ 1 & X_{21} & kX_{21} + v_2 & \cdots & X_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & kX_{n1} + v_n & \cdots & X_{nr} \end{bmatrix} \right)^{-1}$$

Untuk $k \neq 0$ dan ε_n adalah *error*.

Setelah persamaan matriks di atas diselesaikan, dapat terlihat bahwa determinan dari matriks $X'X$ dapat diperkirakan namun tergantung pada ε_n. Apabila ε_n sangat kecil, maka akan sangat mendekati nol yang tentu saja akan mendekati multikolinearitas sempurna **[5]**. Maka standar errornya akan cenderung membesar nilainya sewaktu tingkat multikolinearitas antara variabel bebas juga meningkat **[15].**

2.11.3 Deteksi Multikolinearitas

Salah satu cara mengukur multikolinearitas adalah menggunakan nilai *variance inflation factor* (*VIF*) yaitu merupakan cara untuk mendeteksi multikolinearitas dengan melihat sejauh mana sebuah variabel bebas mempengaruhi variabel bebas lainnya di dalam persamaan regresi **[12]**. Dimana $VIF = \left(1 - R_j^{\,2}\right)^{-1}$, dan dikatakan terdapat multikolinearitas apabila nilai *VIF* lebih 10. Penggunaan *VIF* merupakan perkiraan seberapa besarnya multikolinearitas dapat meningkatkan variansi pada suatu koefisien estimasi sebuah variabel bebas, sehingga *VIF* yang tinggi menunjukan bahwa multikolinearitas telah menaikan sedikit variansi pada perkiraan koefisien **[12].**

BAB III

METODE UNTUK MENGATASI MULTIKOLINEARITAS

Multikolinearitas pada regresi linear berganda menyebabkan matriks $X'X$ nya hampir singular, sehingga menghasilkan nilai penaksir koefisien model regresi tidak stabil. Karena itulah diperkenalkan beberapa metode untuk mengatasi multikolinearitas, diantaranya yaitu metode Regresi Komponen Utama, metode Regresi Ridge, dan metode Penghilangan Variabel.

3.1 Regresi Komponen Utama

Regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan teknik analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap sebagai analisis antara. Regresi komponen utama merupakan metode untuk mengatasi masalah multikolinearitas dengan mengeliminasi dimensi variabel bebas yang merupakan penyebab adanya korelasi antar variabel bebas itu sendiri **[10]**. Dalam hal ini akan dicari beberapa variabel baru yang saling bebas dan merupakan kombinasi linear dari variabel asal. Variabel-variabel inilah yang dinamakan komponen utama.

Cara pembentukan regresi komponen utama (RKU) melalui analisis komponen utama terdapat dua cara yaitu menggunakan matriks kovarian pada saat skala pengukuran variabel-variabelnya sama dan menggunakan matriks korelasi pada saat skala pengukuran variabel-variabelnya berbeda **[4]**.

3.1.1 Pembentukan RKU yang dibentuk oleh matriks kovarian

Terdapat matriks kovarian Σ dari vektor acak $X' = [X_1, X_2, \dots, X_p]$ dengan pasangan nilai eigen dan vektor eigen adalah $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$. Dimana $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, maka komponen utama ke-i didefinisikan sebagai berikut **[6]**:

$W_i = e_i'X = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p \quad i = 1,2,\dots,p$

Secara lengkapnya yaitu:

$$W_1 = e_1'X = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p$$

$$W_2 = e_2'X = e_{12}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p$$

$$\vdots$$

$$W_p = e_p'X = e_{1p}X_1 + e_{2p}X_2 + \cdots + e_{pp}X_p \qquad (3.1.1)$$

Dimana W_1 adalah komponen utama pertama yang memenuhi maksimum nilai $e_1'\Sigma e_1 = \lambda_1$. W_2 adalah komponen kedua yang memenuhi sisa keragaman selain komponen pertama dengan memaksimumkan nilai $e_2'\Sigma e_2 = \lambda_2$. W_p adalah komponen ke-p yang memenuhi sisa keragaman selain komponen utama $W_1, W_2, \dots W_{p-1}$ dengan memaksimumkan nilai $e_p'\Sigma e_p = \lambda_p$. Urutan $W_1, W_2, \dots W_p$ harus memenuhi persyaratan $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

Pada persamaan 3.1.1 apabila ditulis dalam notasi matriks yaitu $W = XP$, dimana P adalah matriks orthogonal dengan memenuhi persamaan $P'P = PP' = I$. Maka proses persamaan regresi linear berganda menjadi regresi komponen utama yaitu **[10]**:

$Y = X\beta + \varepsilon$

$= XPP'\beta + \varepsilon$ dengan $W = XP$ dan $a = P'\beta$

$$= Wa + \varepsilon \qquad (3.1.2)$$

Model regresi komponen utama yang telah direduksi menjadi k komponen adalah

$$Y = \beta_0 1 + W_k a_k + \varepsilon \qquad (3.1.3)$$

Y : variabel terikat
β_0 : kemiringan
1 : vektor yang elemen-elemennya satu berukuran $n \times 1$
W_k : matriks brukuran $n \times k$ yang elemennya merupakan komponen utama
a_k : vektor koefisien komponen utama berukura $k \times 1$
ε : vektor sisa (error) berukuran $n \times 1$

3.1.2 Pembentukan RKU yang dibentuk oleh matriks korelasi

Selain berdasarkan matriks kovariansi, komponen utama juga dapat dibentuk berdasarkan matriks korelasi, hal ini dilakukan apabila skala pengukuran variabel-variabelnya berbeda. Persamaan regresi komponen utama berdasarkan matriks korelasi pada dasarnya hampir sama, perbedaannya variabel $X_1, X_2, \dots, X_p$ berdasarkan variabel-variabel yang telah dibakukan $Z' = [Z_1, Z_2, \dots, Z_p]$ dengan $cov(Z) = \rho$. Maka persamaannya didefinisikan sebagai berikut **[7]**:

$$W_p = e_{1p} Z_1 + e_{2p} Z_2 + \cdots + e_{pp} Z_p \qquad (3.1.4)$$

Proses persamaan regresi linear berganda menjadi regresi komponen utamanya pun secara umum hampir sama yaitu

$Y = Z\beta + \varepsilon$

$= ZPP'\beta + \varepsilon$ dengan $W = ZP$ dan $a = P'\beta$

$$= Wa + \varepsilon \qquad (3.1.5)$$

Model regresi komponen utama yang telah direduksi menjadi k komponen adalah

$$Y = \beta_0 1 + W_k a_k + \varepsilon \qquad (3.1.6)$$

Y : variabel terikat
β_0 : kemiringan
1 : vektor yang elemen-elemennya satu berukuran $n \times 1$
W_k : matriks brukuran $n \times k$ yang elemennya merupakan komponen utama, dimana $W_k = ZE_k$
a_k : vektor koefisien komponen utama berukura $k \times 1$
ε : vektor sisa (error) berukuran $n \times 1$

3.1.3 Penaksir koefisien regresi komponen utama

Pendugaan koefisien regresi komponen utama dapat dilakukan dengan menggunakan metode penaksir kuadrat terkecil, yaitu pada persamaan matriks $Y = Wa + \varepsilon$, dengan a merupakan koefisien regresi komponen utama.

Terdapat persamaan regresi komponen utama yaitu $Y = Wa + \varepsilon$, maka koefisien regresi komponen utama dapat dicari dengan:

$$\varepsilon = Y - Wa$$

$$\varepsilon'\varepsilon = (Y - Wa)'(Y - Wa)$$

$$= Y'Y - 2a'W'Y + a'W'Wa$$

Berdasarkan sifat transpose matriks yaitu $(Wa)' = a'W'$ dan oleh karena $a'W'Y$ suatu skalar maka sama dengan transposnya yaitu $Y'Wa$. Turunan pertama $\varepsilon'\varepsilon$ terhadap a adalah

$$\frac{\partial \varepsilon'\varepsilon}{\partial a} = -2W'Y + 2W'Wa$$

Kemudian jika turunan pertama disamakan dengan nol maka diperoleh:

$W'Y = W'Wa$ sehingga diperoleh

$$a = (W'W)^{-1}W'Y \tag{3.1.7}$$

Maka koefisien regresi komponen utama yaitu $a = (W'W)^{-1}W'Y$

Matriks W merupakan matriks komponen utama dengan sifat orthogonal satu sama lain dalam elemen, maka dengan $W = XP$ maka

$$W'W = (XP)'(XP) = P'X'XP = diag(\lambda_1, \lambda_2, \dots, \lambda_k)$$

Nilai variansi dan ekspektasi koefisien regresi komponen utama dinyatakan dalam bentuk :

$$var(\hat{\alpha}) = \hat{\sigma}^2 diag(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_k^{-1})$$

$$E[\hat{\alpha}] = \sum_{m=k+1}^{p} ((n-1)\lambda_m)^{-1} e_m e_m' X'Y$$

Kovarian antara regresi komponen utama dengan komponen utama yang lainnya adalah saling bebas, dinyatakan dalam bentuk :

$cov(W_i, W_j) = 0; \quad i, j = 1,2, \dots p$ dan $i \neq j$

Dengan kovarian antar regresi komponen utama saling bebas, maka nilai VIF adalah satu karena tidak berkorelasinya antara komponen utama yang satu dengan komponen utama yang lainnya. Dengan demikian terlihat bahwa analisis Regresi Komponen Utama tidak lain adalah meregresikan variabel tak bebas terhadap komponen-komponen utama yang saling bebas, maka jelas tidak ada masalah multikolinearitas lagi **[4]**.

3.1.4 Tahapan pembentukan model Regresi komponen utama

Secara umum tahapan pembentukan regresi komponen utama yaitu **[4]**:

1. Tentukan matriks X yang merupakan matriks berisi data variabel bebas untuk variabel dengan skala pengukuran yang sama, dan tentukan matriks

Z yang merupakan matriks berisi data dari variabel bebas X yang telah dibakukan untuk variabel dengan skala pengukuran berbeda

2. Tentukan matriks kovarian dari matriks X, atau tentukan matriks korelasi dari matriks Z
3. Menghitung nilai eigen dan vektor eigen dari matriks kovarian untuk variabel dengan pengukuran skala yang sama, dan menghitung nilai eigen dan vektor eigen dari matriks korelasi untuk variabel dengan skala pengukuran yang berbeda
4. Membuat komponen utama. Nilai eigen disusun secara terurut menurun kemudian vektor eigen disusun sesuai dengan nilai eigennya. Vektor eigen yang tersusun itulah disebut sebagai komponen utama
5. Pemilihan komponen utama dengan menggunakan kriteria persen varians, dimana jumlah komponen utama yang digunakan memiliki persentasi kumulatif varians minimal 85%. Rumus yang digunakan untuk menghitung persentasi kumulatif varians adalah: $\frac{jumlah\ nilai\ eigen}{jumlah\ variabel\ bebas} \geq 85\%$
6. Pembentukan koefisien Regresi Komponen Utama yang dibentuk dengan menggunakan persamaan (3.1.7) terhadap komponen utama yang telah terpilih.
7. Pembentukan model Regresi Komponen Utama dengan mengalikan vektor transpose komponen utama terpilih dengan koefisiennya.
8. Pentransformasian model Regresi Komponen Utama menjadi model regresi untuk variabel bebas X.

3.2 Regresi Ridge

Regresi ridge merupakan salah satu metode yang digunakan untuk mengatasi kondisi buruk (*ill conditioned*) yang diakibatkan oleh korelasi tinggi antara beberapa variabel bebas didalam regresi sehingga menyebabkan matriks *X'X* nya hampir singular **[3]**. Metode ini juga merupakan metode yang dapat menstabilkan parameter regresi karena adanya multikolinearitas yang dilakukan melalui modifikasi terhadap metode kuadrat terkecil. Modifikasi tersebut dilakukan dengan cara menambahkan tetapan bias *c* yang relatif kecil pada diagonal utama matriks *X'X*. Sehingga penduga koefisien regresi ridge adalah

$$\hat{\beta}^*(c) = (X'X + cI)^{-1}X'Y \qquad (3.2.1)$$

Metode regresi ridge ini meninggalkan metode kuadrat terkecil yang biasa digunakan dan terlihat menggunakan cara penaksiran yang bias. Dalam penggunaannya, metode ini bersedia menerima sejumlah bias tertentu dalam taksiran agar variansi penaksir koefisien regresinya dapat diperkecil.

Sifat dari penduga koefisien regresi ridge yaitu **[11]**:

1. Bias

$$\begin{aligned}
\hat{\beta}^* &= (X'X + cI)^{-1}X'Y \\
&= (X'X + cI)^{-1}X'X(X'X)^{-1}X'Y \text{ dengan} Y = X\hat{\beta} \\
&= (X'X + cI)^{-1}X'X(X'X)^{-1}X'X\hat{\beta} \\
&= (cI)^{-1}X'X(X'X)^{-1}X'X\hat{\beta} + (X'X)^{-1}X'X(X'X)^{-1}X'X\hat{\beta} \\
&= (cI)^{-1}X'X\hat{\beta} + \hat{\beta} \\
&= (I + (cI)^{-1}X'X)\hat{\beta}
\end{aligned}$$

$= [I + c(X'X)^{-1}]^{-1}\hat{\beta}$

$= Z\hat{\beta}$ dengan $Z = [I + c(X'X)^{-1}]^{-1}$

Sehingga $E[\hat{\beta}^*] = E[Z\hat{\beta}]$

$= ZE[\hat{\beta}]$

$= Z\beta$

Sehingga penduga koefisien regresi ridge memiliki sifat bias.

2. Variansi minimum

$$var(\hat{\beta}) = (X'X + cI)^{-1}X' \sum [(X'X + cI)X']'$$

$$= (X'X + cI)^{-1}X'\sigma^2 IX(X'X + cI)^{-1}$$

$$= \sigma^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1}$$

Dari sifat penduga koefisien regresi ridge yang minimum, nilai VIF merupakan diagonal utama dari matriks $(X'X + cI)^{-1}X'X(X'X + cI)^{-1}$**[7]**.

Dalam pemilihan konstanta bias c merupakan hal yang perlu diperhatikan, karena konstanta tersebut mencerminkan jumlah bias dalam penduga $\hat{\beta}(c)$. Tetapan bias yang diinginkan adalah tetapan bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien yang relatif stabil. Untuk pemilihan tetapan bias c tersebut digunakan ridge trace, yaitu plot dari penduga regresi ridge secara keseluruhan bersama dengan semua kemungkinan tetapan bias c yang biasanya terdapat pada interval 0 – 1 **[8]**. Disamping cara tersebut, tetapan bias c dapat ditentukan berdasarkan nilai VIF bagi setiap koefisien regresi ridge. Nilai c yang terpilih yaitu pada saat nilai-nilai VIF cukup kecil dengan nilai mendekati 1.

3.3. Metode Penghilangan Variabel bebas

Salah satu metode yang paling mudah dilakukan untuk mengatasi masalah multikolinearitas adalah dengan menghilangkan salah satu variabel bebas yang mempunyai hubungan linear kuat **[17]**. Ketika dihadapkan dengan multikolinearitas yang parah sekalipun, salah satu cara yang paling sederhana adalah dengan menghilangkan satu dari variabel yang berkorelasi **[5]**.

Akan tetapi, dengan mengeluarkan suatu variabel dari model regresi akan berakibat adanya kesalahan spesifikasi **[5]**. Kesalahan spesifikasi terjadi karena melakukan kesalahan dalam menentukan spesifikasi model yang dipergunakan dalam analisa, maksudnya salah dalam menentukan variabel yang tepat dalam suatu model regresi **[15]**.

Untuk melihat konsekuensi dari kesalahan spesifikasi, misalkan model yang tepat dalam regresi linear adalah

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad (3.3.1)$$

Tetapi misalkan menggunakan model yang dispesifikasikan secara salah dengan merumuskan model sebagai berikut:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \upsilon_i \qquad (3.3.2)$$

Diketahui bahwa

$$\hat{\alpha}_1 = \frac{\sum y_i x_{1i}}{\sum x^2{}_{1i}} \qquad (3.3.4)$$

$$\hat{\beta}_1 = \frac{(\sum y_i x_{1i})(\sum x^2{}_{2i}) - (\sum y_i x_{2i})(\sum x_{1i} x_{2i})}{(\sum x^2{}_{1i})(\sum x^2{}_{2i}) - (\sum x_{1i} x_{2i})^2} \qquad (3.3.5)$$

Sekarang dapat ditunjukan dari (3.3.4) dan (3.3.5) bahwa

$$E(\hat{\alpha}_1) = \beta_1 + b_{21}\beta_2$$

Dimana b_{21} merupakan koefisien kemiringan dalam regresi X_2 atas X_1. Sehingga dari $\hat{\alpha}_1$ merupakan taksiran bias dari β_2 selama b_{21} berbeda dengan nol (diasumsikan bahwa β_2 bebeda dengan nol; kalau tidak, maka tidak ada artinya untuk memasukan X_2 ke dalam model semula). Tentu saja apabila b_{21} adalah nol, maka tidak mempunyai masalah multikolinearitas dari awal.

Dari uraian diatas jelas bahwa mengeluarkan satu variabel dari model untuk mengurangi masalah multikolinearitas bisa mengakibatkan kesalahan spesifikasi. Dalam beberapa situasi penyembuhan model yang dicapai akan lebih buruk dari model sebelumnya, karena perkiraan parameter yang diperoleh bukan parameter yang dimaksudkan **[15]**.

Situasi yang tepat menggunakan metode ini apabila multikolinearitas mempengaruhi variabel-variabel yang tidak penting **[18]**. Kadang-kadang, solusi sederhana dengan menghapus variabel-variabel bebas yang berkorelasi merupakan tindakan bagus apabila memasukan begitu banyak variabel bebas di dalam persamaan yang pada dasarnya variabel tersebut mengukur kondisi yang sama **[12]**.

3.4 Kelebihan dan Kekurangan Setiap Metode

Terdapat multikolinearitas pada model regresi linear berganda merupakan masalah serius, maka harus dilakukan penghilangan multikolinearitas. Sehingga terdapat banyak cara untuk mengatasi masalah ini, diantaranya menggunakan metode Regresi Komponen Utama, Regresi Ridge, dan penghilangan variabel.

Diantara ketiga metode ini terdapat beberapa kelebihan dan kekurangan, sehingga dapat terlihat keefektifitasan masing-masing metode.

Setelah dilakukan pembahasan pada setiap metode, dapat diketahui keefektifitasan metode dari tingkat kesulitan pembuatan model regresi, sifat dari pembentukan model, nilai bias, nilai variansi dan dari jenis kasus yang memungkinkan menggunakan salah satu metode dari ketiga metode yang telah dibahas.

Dilihat dari tingkat kesulitan pembuatan model, Regresi Komponen Utama memiliki tingkatan yang cukup sulit karena harus dilakukan banyak langkah untuk menghilangkan multikolinearitas dan diperlukan pemahaman yang kuat dalam memahami teorinya untuk menentukan langkah-langkah dalam pembuatan model regresinya. Metode Regresi Ridge dikatakan memiliki tingkatan kesulitan sedang karena secara umum, dengan dilakukannya pemilhan tetapan bias kemudian dilihat pola Ridge trace dan nilai VIF diharapkan bisa menangani masalah multikolinearitas. Sedangkan utuk metode penghilangan variabel dikatakan memiliki tingkatan paling sederhana, karena hanya dengan melihat variabel bebas berkorelasi maka salah satu variabel bebas itulah yang di hilangkan.

Dari ketiga metode yaitu regresi komponen utama, regresi ridge, dan penghilangan variabel dapat dilihat kekurangan dan kelebihannya dari sifat pembuatan model. Pada Regresi Ridge dikatakan bersifat subjektif karena pada saat pemilihan tetapan bias c, yang dilihat dari pola RidgeTrace dan dari menurunnya nilai VIF diserahkan pada analisisnya sendiri **[8]**. Pada metode penghilangan variabel bersifat subjektif karena pada saat menentukan salah satu variabel bebas yang harus dihilangkan dari banyaknya variabel bebas yang berkorelasi diserahkan kepada analis sendiri. Untuk Regresi Komponen Utama tidak bersifat subjektif karena setiap langkah pembentukan model Regresinya memilihi langkah-langkah tertentu menggunakan perhitungan sistematis.

Kriteria dengan sifat penaksir koefisien bias atau tak bias dapat dijadikan kriteria untuk menentukan tingkat kefektifitasan model. Sifat penaksir koefisien Regresi Ridge adalah bias, karena pembentukan model Regresi Ridgenya sendiri menggunakan penambahan tetapan bias c. Hal ini bisa dilihat dari $\hat{\beta}^*(c) = (X'X + cI)^{-1}X'Y$ sehingga menghasilkan $E[\hat{\beta}^*] \neq \beta$ **[11]**. Untuk metode penghilangan variabel bersifat bias karena mengeluarkan suatu variabel dari model regresi akan berakibat adanya kesalahan spesifikasi (bias spesifikasi) sehingga adanya kesalahan dalam menentukan model regresi **[5]**.

Dilihat dari nilai variansi model Regresi Komponen Utama dan Regresi Ridge memiliki nilai variansi minimum, hal ini dapat dilihat dari $var(\hat{\beta}) = \sigma^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1}$ untuk model Regresi Ridge, dan $var(\hat{\alpha}) = \hat{\sigma}^2 diag(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_k^{-1})$ untuk Regresi Komponen Utama.

Dengan melihat VIF pada metode, Regresi komponen Utama dapat menghilangkan korelasi antar variabel bebas dengan bersih, hal ini disebabkan karena $cov(W_i, W_j) = 0;\quad i, j = 1,2, \ldots p$ dan $i \neq j$ artinya komponen utama yang satu dengan yang lainnya saling bebas sehingga menjadikan nilai VIF adalah satu maka masalah multikolinearitas benar-benar teratasi. Pada Metode Regresi Ridge, pemilihan tetapan bias c dengan melihat nilai VIF menurun menuju ke nilai satu maka dikatakan metode ini dapat mengurangi dampak multikolinearitas saja. Pada metode penghilangan variabel hanya dapat mengurangi dampak multikolinearitas, dan tidak dapat menghilangkan multikolinearitas pada model regresi linear berganda.

Dalam berbagai penelitian, peneliti sering dihadapkan pada permasalahan yang melibatkan data yang besar dengan variabel yang banyak. Sehingga dikembangkan analisis Regresi Komponen Utama untuk mereduksi data yang besar menjadi lebih sederhana. Analisis komponen utamanya dapat dijadikan tahap antara untuk penelitian yang bersipat lebih besar **[18]**.

Apabila pada regresi linear berganda terdapat multikolinearitas dengan jenis kasusnya memiliki data besar dan variabel banyak, maka metode yang paling efektif untuk digunakan adalah menggunakan metode Regresi Komponen Utama.

Dalam penelitian lain dengan data kecil dan variabel yang sedikit, terdapatnya multikolinearitas pada regresi linear berganda dapat diatasi dengan efektif menggunakan metode Regresi Ridge. Hal ini disebabkan karena dengan jumlah variabel yang sedikit akan menghasilkan jumlah variabel bebas yang berkorelasi akan sedikit pula sehingga dapat diatasi dengan pemilihan tetapan bias yag relatif kecil dengan koefisien regresi yang stabil.

Berbeda dengan metode Regresi Komponen Utama dan metode Regresi Ridge, metode penghilangan variabel tidak bergantung pada besar kecilnya data ataupun banyak sedikitnya variabel yang digunakan, tetapi solusi sederhana ini dapat digunakan apabila terdapat variabel bebas penyebab multikolinearitas yang tidak relevan/tidak penting masuk ke dalam persamaan regresi. Memasukan variabel yang tidak penting ini biasanya terjadi pada peneliti yang memasukan variabel-variabel untuk mengukur barang/kondisi yang sama. Dalam kasus seperti ini, variabel multikolinear tidak relevan **[12]**.

Untuk lebih jelasnya, dapat dilihat pada tabel di bawah ini mengenai kekurangan dan kelebihan dari setiap metode ini adalah:

Tabel 3.4.1 kekurangan dan kelebihan setiap metode

Kekurangan/kelebihan metode yang dilihat dari:	Metode Penghilangan Multikolinearitas		
	RKU	RR	Penghilangan Variabel
Tingkat kesulitan pembuatan model	sulit	sedang	sederhana
Sifat pembuatan model	objektif	subjektif	subjektif
Sifat penaksir koefisien regresi	Bias dan variansi minimum	Bias dan variansi minimum	Bias dan variansi minimum
Dampak multikolinearitas	Mengurangi multikolinearitas	Mengurangi multikolinearitas	Mengurangi multikolinearitas
Jenis kasus yang seuai	Kasus dengan data besar (n > 30) dan variabel banyak (variabel bebas > 3)	Kasus dengan data kecil dan variabel sedikit	Kasus dengan data terdapat variabel multikolinear yang tidak relevan

BAB IV
APLIKASI METODE PERBAIKAN DATA YANG TERDAPAT MULTIKOLINEARITAS

Untuk mengetahui aplikasi dari setiap metode yang dapat memperbaiki model regresi linear berganda yang terdapat multikolinearitas maka akan dikembangkan beberapa contoh kasus pertama dengan jumlah sampel kecil (n<30), dan kasus kedua dengan jumlah sampel besar (n>30) sehingga dapat diperlihatkan keefektifitasan setiap metode Regresi Komponen Utama, metode Regresi Ridge,dan metode penghilangan variabel.

4.1 Contoh Kasus Pertama

Terdapat contoh kasus dimana data variabel terikat Bodyfat dipengaruhi oleh data variabel bebas Triceps, Thigh, dan Midarm, dengan data ditunjukan pada tabel 4.1.1

Tabel 4.1.1 data kasus pertama

no	Bodyfat	Triceps	Thigh	Midarm
1	11,9	19,5	43,1	29,1
2	22,8	24,7	49,8	28,2
3	18,7	30,7	51,9	37,0
4	20,1	29,8	54,3	31,1
5	12,9	19,1	42,2	30,9
6	21,7	25,6	53,9	23,7
7	27,1	31,4	58,5	27,6
8	25,4	27,9	52,1	30,6
9	21,3	22,1	49,9	23,2
10	19,3	25,5	53,5	24,8
11	25,4	31,1	56,6	30,0

12	27,2	30,4	56,7	28,3
13	11,7	18,7	46,5	23,0
14	17,8	19,7	44,2	28,6
15	12,8	14,6	42,7	21,3
16	23,9	29,5	54,4	30,1
17	22,6	27,7	55,3	25,7
18	25,4	30,2	58,6	24,6
19	14,8	22,7	48,2	27,1
20	21,1	25,2	51,0	27,5

Sumber: Regresi dan Korelasi dalam Genggaman Anda 2011.

Tabel 4.1.2 koefisien regresi linear berganda

variabel	Koefisien regresi
Konstan	117.0847
X_1	4.3341
X_2	-2.8568
X_3	-2.1861

Dengan dicari koefisien regresi menggunakan persamaan 2.5.2 maka model regresi linear bergandanya adalah:

Y = 117.0847+4.3341$X_1 - 2.8568X_2 - 2.1861X_3$

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	396,985	3	132,328	21,52
Residual	98,4049	16	6,15031	
Total (Corr.)	495,389	19		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian regresi linear dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1: \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(3,16,0.05)} = 3.24$, karena $F_{hit} > F_{tab}$ maka tolak H_0 dan dinyatakan bahwa regresi signifikan.

Untuk mengetahui model dari regresi linear berganda di atas terdapat multikolinearitas, dapat dideteksi dengan menggunakan nilai VIF. Apabila nilai VIF > 10 maka diindikasikan bahwa model regresi linear berganda terdapat multikolinearitas

Tabel 4.1.3 nilai VIF setiap variabel bebas

Variabel bebas	VIF
X_1	708.843
X_2	564.343
X_3	104.606

Dari tabel di atas, dilihat dari masng-masing nilai VIF variabel bebas adalah lebih dari sepuluh maka dapat diindikasikan bahwa model regresi ini terdapat multikolinearitas.

Karena model tersebut diindikasikan memiliki multikolinearitas, maka akan dilakukan penghilangan multikolinearitas dengan beberapa metode yaitu metode Regresi Komponen Utama, metode Regresi Ridge, dan metode penghilangan variabel. Sebelum dilakukan penghilangan multikolinearitas maka setiap variabel dilakukan standarisasi terlebih dahulu dengan tujuan untuk meminimumkan kesalahan pembulatan dan untuk menganggap regresi sudah dipenuhi kenormalannya. Berikut tabel hasil standarisasi yang didapat dari persamaan 2.9.2:

Tabel 4.1.4 merupakan tabel dengan variabel yang sudah distandarisasi

Y^*	Z1	Z2	Z3
-0.3727	-0.2651	-0.3537	0.0931
0.1170	-0.0276	-0.0600	0.0365
-0.0672	0.2464	0.0320	0.5900
-0.0043	0.2053	0.1372	0.2189
-0.3278	-0.2834	-0.3931	0.2063
0.0676	0.0135	0.1196	-0.2466
0.3102	0.2784	0.3212	-0.0013
0.2339	0.1185	0.0408	0.1875
0.0496	-0.1464	-0.0557	-0.2780
-0.0402	0.0089	0.1021	-0.1774
0.2339	0.2647	0.2380	0.1497
0.3147	0.2327	0.2424	0.0428
-0.3817	-0.3017	-0.2047	-0.2906
-0.1076	-0.2560	-0.3055	0.0616
-0.3322	-0.4889	-0.3712	-0.3975
0.1665	0.1916	0.1416	0.1560
0.1081	0.1094	0.1810	-0.1208
0.2339	0.2236	0.3256	-0.1900
-0.2424	-0.1190	-0.1302	-0.0327
0.0407	-0.0048	-0.0075	-0.0075

- Metode Regresi Komponen Utama

Terdapat langkah-langkah untuk membentuk model Regresi komponen utama yaitu:

1. Menentukan matriks X yang telah dibakukan karena variabel merupakan variabel dengan skala pengukuran yang berbeda. Data tersebut terdapat pada tabel 4.1.4
2. Menentukan matriks korelasi dari variabel yang telah distandarisasikan

1.0000	0.9095	0.3772
0.9095	1.0000	0.0848
0.3772	0.0848	1.0000

3. Menentukan nilai eigen dan vektor eigen dari matriks korelasi, kemudian vektor eigen disusun berdasarkan nilai eigen yang terurut mulai dari nilai eigen terkecil ke nilai eigen terbesar

   Susunan nilai eigen

   0.0007 0.9328 2.0665

   Maka susunan vektor eigennya:

0.7176	0.0501	0.6947
-0.6401	0.4405	0.6294
-0.2745	-0.8963	0.3482

4. Pembentukan Komponen Utama

Dari susunan nilai eigen dan vektor eigen, maka komponen utama yang terbentuk yaitu

$$W_1 = 0.7176Z_1 - 0.6401Z_2 - 0.2745Z_3$$
$$W_2 = 0.0501Z_1 + 0.4405Z_2 - 0.8963Z_3$$
$$W_3 = 0.6947Z_1 + 0.6294Z_2 + 0.3482Z_3$$

5. Pemilihan komponen utama

   Komponen utama yang digunakan adalah komponen utama dengan persentasi kumulatif varians minimal 85%. Rumus yang digunakan yaitu:

$$\frac{jumlah\ nilai\ eigen}{jumlah\ variabel\ bebas} \geq 85\%$$

Persentasi kumulatif nilai eigen

Component Number	*Eigenvalue*	*Percent of Variance*	*Cumulative Percentage*
1	2,0665	68,883	68,883
2	0,932774	31,092	99,976
3	0,000725769	0,024	100,000

Dari nilai kumulatif tersebut, akan digunakan dua komponen utama karena hanya dengan dua komponen utama dengan nilai kumulatif variansi sebesar 0.9998 dapat menerangkan keragaman sekitar 99.98% ≥ 85%. Jadi komponen utama yang dipilih yaitu

$$W_2 = 0.0501Z_1 + 0.4405Z_2 - 0.8963Z_3$$

$$W_3 = 0.6947Z_1 + 0.6294Z_2 + 0.3482Z_3$$

6. Pembentukan model Regresi Komponen Utama

Taksiran koefisien regresi komponen utamanya yaitu

0.3231

0.5749

Sehingga

$$Y = 0.3231W_2 + 0.5749W_3$$

Maka persamaan menjadi $Y^* = 0.4156Z_1 + 0.5042Z_2 - 0.0894Z_3$

Dikembalikan ke variabel semula, didapatkan

$$Y = -12.2046 + 0.4225X_1 + 0.4918X_2 - 0.1252X_3$$

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	386.6323	3	128.8774	18.9600
Residual	108.7572	16	6.7973	
Total (Corr.)	495.3895	19		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\ (3,16,0.05)} = 3.24$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Dengan nilai VIF dan variansinya:

Tabel 4.1.5 tabel VIF dan Variansi model RKU

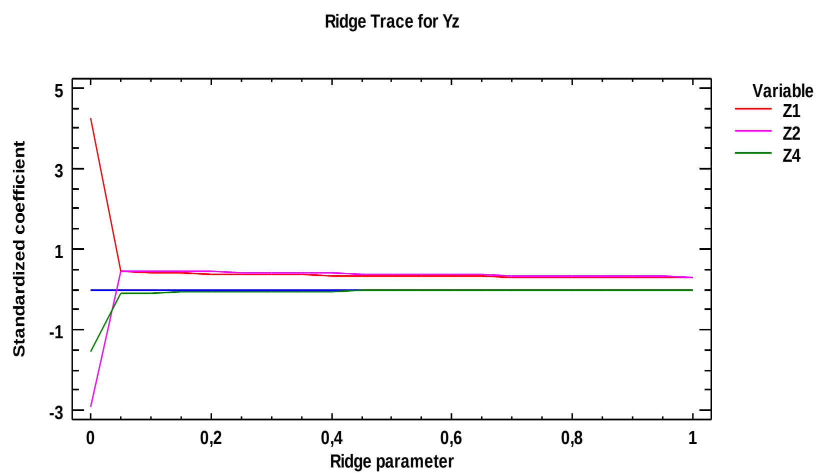
Variabel bebas	VIF	variansi
konstan		1.1004
X_1	2.5210	0.0010
X_2	0.5108	0.0007
X_3	0.8889	0.0003

.

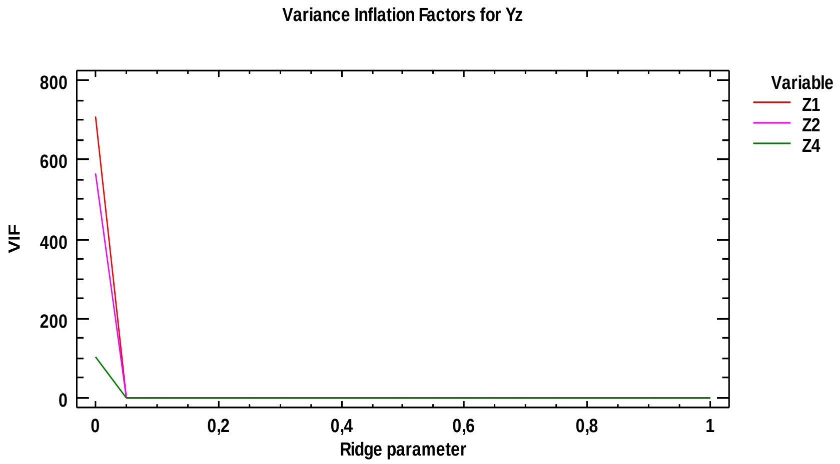
- Metode Regresi Ridge

Terdapat beberapa langkah untuk memodelkan menggunakan Regresi Ridge yaitu

1. Standarisasi variabel X dan Y, hasilnya terdapat pada tabel 4.1.4
2. Dengan data yang telah ditransformasi, maka akan dilakukan pemilihan nilai c dengan melihat nilai VIF dan ridge trace



Dari grafik Ridge Trace diatas terlihat bahwa pada ridge parameter dari 0 sampai 1 yang merupakan nilai c, yang mana koefisien standar terlihat stabil pada saat nilai c sekitar 0 sampai 0.05.



Dari grafik VIF di atas, terlihat mulai tampak ada penurunan pada saat nilai c di sekitar 0 sampai 0.05. Hal inipun menunjukan bahwa dengan c pada ridge parameter tersebut, koefisien dari Regresi lebih stabil dengan nilai VIF nya kurang dari 10 yang menandakan berkurangnya multikolinearitas.

Tabel 4.1.6 nilai VIF dengan berbagai nilai c

Ridge Parameter	Xstandar1	Xstandar2	Xstandar3	R-Squared
0,0	709,68	565,037	104,688	80,13
0,00333333	22,9161	18,4488	4,23115	78,27
0,00666667	7,07285	5,83795	1,90768	77,96
0,01	3,48224	2,97877	1,37642	77,76
0,0133333	2,12377	1,89615	1,17172	77,58
0,0166667	1,46777	1,37263	1,06982	77,42
0,02	1,10161	1,0798	1,01038	77,26
0,0233333	0,8765	0,899245	0,971635	77,11
0,0266667	0,728194	0,779833	0,944183	76,96
0,03	0,625256	0,696543	0,923425	76,82
0,0333333	0,550831	0,63596	0,906896	76,67
0,0366667	0,495223	0,590367	0,893178	76,53
0,04	0,452533	0,555068	0,881406	76,39
0,0433333	0,419004	0,527075	0,871029	76,25
0,0466667	0,392154	0,50441	0,861685	76,11
0,05	0,370286	0,485724	0,853125	75,97

0,0533333	0,352213	0,470071	0,845173	75,83
0,0566667	0,337078	0,456769	0,837705	75,69
0,06	0,324256	0,44532	0,830627	75,56
0,0633333	0,313278	0,435352	0,82387	75,42
0,0666667	0,30379	0,426581	0,817383	75,29
0,07	0,295518	0,41879	0,811124	75,15
0,0733333	0,288249	0,411809	0,805061	75,02
0,0766667	0,281814	0,405504	0,799169	74,88
0,08	0,276079	0,399768	0,793428	74,75
0,0833333	0,270936	0,394514	0,787822	74,62
0,0866667	0,266296	0,389672	0,782336	74,49
0,09	0,262087	0,385184	0,77696	74,36
0,0933333	0,258251	0,381003	0,771684	74,23
0,0966667	0,254736	0,37709	0,766501	74,10
0,1	0,251502	0,37341	0,761403	73,97

Dari berbagai nilai c yang ada, terlihat adanya penurunan nilai VIF sedikit demi sedikit, nilai c yang akan diambil adalah pada saat nilai VIF relatif dekat dengan 1 yaitu c = 0.02.

Tabel 4.1.7 nilai koefisien regresi ridge dengan nilai tetapan bias c = 0.02

variabel	Koefisien Regresi Ridge
Z_1	0.545879
Z_2	0.377816
Z_3	-0.136748

Maka dapat dibentuk model regresi ridgenya yaitu:

$$Y^* = 0.545879Z_1 + 0.377816Z_2 - 0.136748Z_3$$

Apabila model di atas dikembalikan ke variabel-variabel asal maka diperoleh:

$$Y = -7.4171 + 0.5549X_1 + 0.3685X_2 - 0.1915X_3$$

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	382.7371	3	127.5790	18.1200
Residual	112.6524	16	7.0408	
Total (Corr.)	495.3895	19		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(3,16,0.05)} = 3.24$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Tabel 4.1.8 nilai VIF dan variansi model Regresi Ridge

Variabel bebas	VIF	variansi
konstan		1.1398
X_1	1.10161	0.0010
X_2	1.0798	0.0008
X_3	1.01038	0.0003

- Metode Penghilangan Variabel

Untuk mengetahui variabel bebas mana yang akan dihilangkan, yaitu dengan melihat korelasi antar variabel bebas yang hampir sempurna atau mendekati nilai satu. Berikut matriks korelasi antar variabel bebas:

$$\begin{bmatrix} 1 & 0,924 & 0,458 \\ 0,924 & 1 & 0,085 \\ 0,458 & 0,085 & 1 \end{bmatrix}$$

Dari matriks tersebut dapat dilihat bahwa ada korelasi antara variabel bebas Z_1 dengan variabel bebas Z_2 sebesar 0, 924, maka variabel yang akan dihilangkan adalah salah satu diantara variabel tersebut. Untuk mengetahui variabel mana yag akan dihilangkan, yaitu dengan melihat masing-masing konsekuensi yang dihasilkan apabila variabel bebas Z_1 dihilangkan atau variabel bebas Z_2 yang dihilangkan. Konsekuensinya seperti melihat berkurangnya nilai VIF, dan nilai variansi.

- ✓ Konsekuensi apabila variabel bebas Z_1 yang dihilangkan

Tabel 4.1.9 Tabel koefisien regresi saat penghilangan variabel Z_1

variabel	Koefisien regresi
Z_2	0.872
Z_3	0.069

Tabel 4.1.10 Tabel konsekuensi setelah penghilangan variabel X_1

variabel	Koefisien regresi	VIF	variansi
Konstan	-25.997		48.9625
X_2	0.851	1.007	0.0126
X_3	0.096	1.007	0.0260

- ✓ Konsekuensi apabila variabel bebas Z_2 yang dihilangkan

Tabel 4.1.11 Tabel koefisien regresi saat penghilangan variabel Z_2

variabel	Koefisien regresi
Z_1	0.984
Z_3	-0.308

Tabel 4.1.12 Tabel konsekuensi setelah penghilangan variabel X_2

variabel	Koefisien regresi	VIF	variansi
Konstan	6.792		46.6786
X_1	1.001	1.265	0.0121
X_3	-0.431	1.265	0.0248

Disini, dengan menghilangkan variabel bebas X_1 ataupun X_2 nilai VIF nya menandakan multikolinearitas sudah teratasi sehingga akan dilihat dari nilai variansinya dimana dengan menghilangkan variabel bebas X_2 variansinya lebih kecil dibandingkan dengan menghilangkan variabel bebas X_1. Sehingga untuk kasus ini akan dihilangkan variabel bebas X_2.

Sehingga modelnya didapatkan

$$Yz = 0.984Z_1 - 0.308Z_3$$

Dikembalikan ke variabel semula, didapatkan

$$Y = 6.792 + 1.001X_1 - 0.431X_3$$

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	389.455	2	194.728	31.249
Residual	105.934	17	6.231	
Total (Corr.)	495.3895	19		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasil, dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(2,17,0.05)} = 3.59$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Tabel 4.1.13 Tabel nilai VIF dan variansi pada saat penghilangan variabel X_2

variabel	VIF	variansi
Konstan		46.6786
X_1	1.265	0.0121
X_3	1.265	0.0248

- Perbandingan Setiap Metode

Persamaan yang dihasilkan dari setiap metode yaitu:

Persamaan RLB : $Y = 117.0847 + 4.3341X_1 - 2.8568X_2 - 2.1861X_3$

Persamaan metode RKU : $Y = -12.2046 + 0.4225X_1 + 0.4918X_2 - 0.1252X_3$

Persamaan metode RR : $Y = -7.4171 + 0.5549X_1 + 0.3685X_2 - 0.1915X_3$

Persamaan metode PV : $Y = 6.792 + 1.001X_1 - 0.431X_3$

Perbandingan metode Regresi Komponen Utama, Regresi Ridge, dan penghilangan variabel dilihat dari nilai VIF, variansi, dan MSE

Tabel 4.1.14 Tabel perbandingan nilai VIF kasus pertama

Variabel bebas	Nilai VIF dari metode		
	RKU	RR	PV
X_1	0.0014	1.10161	1.265
X_2	0.0018	1.0798	
X_3	0.0860	1.01038	1.265

Tabel 4.1.15 Tabel perbandingan nilai variansi kasus pertama

Variabel bebas	Nilai variansi dari metode		
	RKU	RR	PV
konstan	1.1004	1.1398	48.9625
X_1	0.0010	0.0010	
X_2	0.0007	0.0008	0.0126
X_3	0.0003	0.0003	0.0260

Ket:
RLB:regresi linear berganda
RKU: regresi komponen utama
PV: penghilangan variabel

Apabila dilihat dari nilai variansinya, metode Regresi Komponen Utama memiliki variansi yang paling kecil. Namun metode ini tidak dapat dikatakan efektif untuk mengatasi multikolinearitas karena, apabila dilihat dari dampak multikolinearitas yang hampir bersih dengan melihat nilai VIF mendekati nilai satu dan melihat perbedaan variansi yang cukup kecil diantara regresi komponen utama dengan regresi ridge, penggunaan metode Regresi Ridge dengan jumlah sampel kecil dan variabel sedikit akan lebih efektif mengatasi multikolinearitas.

4.2 Contoh kasus kedua

Terdapat contoh kasus dimana data variabel terikat Y dipengaruhi variabel bebas X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 dan X_9 dengan jumlah pengamatan > 30.

Tabel 4.2.1 data kasus kedua

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
30.4	1.1398	1.0569	0.9093	0.8779	0.7183	0.581	0.6195	0.6472	0.3779
41.8	1.4455	1.3258	1.1355	1.1068	0.9213	0.7835	0.8363	0.8662	0.5016
44.1	1.4607	1.3413	1.152	1.1216	0.9333	0.7976	0.8503	0.8807	0.51
42.7	1.5637	1.4364	1.2334	1.2019	0.9942	0.8534	0.9117	0.944	0.537
38.7	1.3598	1.2561	1.0819	1.0488	0.8614	0.7275	0.7761	0.806	0.5644
39.9	1.44	1.3299	1.1479	1.1155	0.9225	0.7791	0.8303	0.8614	0.5036
35.9	1.4534	1.3434	1.1604	1.1257	0.9225	0.7677	0.8197	0.8529	0.4878
40.8	1.5675	1.44	1.2434	1.2108	1.0183	0.8651	0.9203	0.9518	0.5644
38.6	1.3571	1.2475	1.0723	1.041	0.8595	0.718	0.7663	0.7965	0.4619
41.6	1.4391	1.3166	1.1262	1.098	0.9298	0.7696	0.8215	0.8508	0.5016
44.8	1.5791	1.4431	1.2339	1.205	1.0027	0.8691	0.928	0.9589	0.5505
44.8	1.7892	1.6594	1.4527	1.4197	11.9675	1.039	0.1104	1.1379	0.6604
43.6	1.6179	1.4921	1.2911	1.2603	1.057	0.9143	0.974	1.0056	0.5815
43.1	1.5615	1.4343	1.2343	1.2027	1.0001	0.8658	0.9233	0.9543	0.5504
39.6	1.4028	1.2825	1.0943	1.0621	0.8749	0.735	0.7865	0.8174	0.4687
45.2	1.5439	1.4282	1.2417	1.2151	1.029	0.8967	0.952	0.9811	0.5823
41.8	1.5455	1.4256	1.2309	1.2018	0.9986	0.8554	0.9121	0.9433	0.5462
43.3	1.6108	1.4852	1.2872	1.2569	1.0546	0.9113	0.8679	0.9992	0.5888
41.6	1.4499	1.3411	1.165	1.1362	0.9573	0.8129	0.866	0.8961	0.5194

31.6	1.2834	1.1894	1.0246	0.9895	0.8034	0.6483	0.6926	0.724	0.4122
43.0	1.4015	1.2831	1.0962	1.0678	0.8793	0.7597	0.811	0.8395	0.4809
35.9	1.3636	1.2631	1.0923	1.0619	0.8868	0.7314	0.779	0.8089	0.4774
36.0	1.3922	1.2863	1.1081	1.0752	0.8876	0.7385	0.7868	0.8174	0.4782
42.3	1.4416	1.3212	1.126	1.0939	0.8945	0.7651	0.8182	0.8492	0.4808
43.3	1.4938	1.3744	1.1893	1.1612	0.9882	0.8735	0.8912	0.9206	0.5466
45.4	1.4985	1.367	1.1671	1.1399	0.95004	0.8215	0.8766	0.906	0.5466
40.7	1.6116	1.4886	1.2867	1.2518	1.0368	0.8858	0.944	0.9781	0.5602
40.4	1.4788	1.3565	1.168	1.1379	0.9583	0.8048	0.8571	0.8871	0.5277
44.5	1.6615	1.5273	1.3244	1.2934	1.0961	0.9508	0.1011	1.0413	0.6116
46.3	1.5601	1.4344	1.2445	1.2187	1.044	0.9056	0.9621	0.9896	0.5922
39.1	1.5353	1.4165	1.2242	1.1923	0.997	0.843	0.8984	0.9303	0.5419
37.6	1.3876	1.2804	1.1013	1.068	0.8758	0.7356	0.7851	0.8157	0.4707
37.1	1.284	1.1826	1.0132	0.9829	0.8001	0.6708	0.7161	0.745	0.4277
39.4	1.4004	1.2862	1.102	1.0697	0.8826	0.7444	0.7951	0.8252	0.4734
41.6	1.3603	1.256	1.0893	1.0601	0.8858	0.7578	0.8068	0.8354	0.49
36.4	1.3842	1.2807	1.1059	1.0725	0.8842	0.7384	0.7869	0.8175	0.4767

Sumber: Naes T. 1985. *Multivariate Calibration When the Error Covariance Matrix is Structured. Technometrics.* V-27, no.3:301-311. Dikutip dari Nurhasanah, perbandingan Regresi Komponen Utama Terkoreksi dengan Regresi Ridge dalam Mengatasi Multikolinearitas, 2006.

Tabel 4.2.2 koefisien regresi linear berganda

variabel	Koefisien regresi
Konstan	36.1339
X_1	7.1126
X_2	128.0696
X_3	-180.4833
X_4	-109.8327
X_5	0.2433
X_6	11.5915
X_7	1.4004
X_8	161.4825
X_9	4.0443

Model regresinya:

$$Y = 36.1339 + 7.1126X_1 + 128.0696X_2 - 180.4833X_3 - 109.8327X_4 + 0.2433X_5 + 11.5915X_6 + 1.4004X_7 + 161.4825X_8 + 4.0443X_9$$

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
Model	469.0195	9	52.1133	63.6616
Residual	21.2835	26	0.8186	
Total (Corr.)	490.3031	35		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(9,26,0.05)} = 2.27$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Mendeteksi adanya multikolinearitas yaitu dengan nilai VIF > 10 maka diindikasikan terdapat multikolinearitas

Tabel 4.2.3 nilai VIF setiap variabel bebas

Variabel bebas	VIF
X_1	110.611
X_2	1.237E4
X_3	129.546
X_4	8.942E-5
X_5	2.511
X_6	228.235
X_7	1.758
X_8	404.971
X_9	11.772

Dari tabel di atas diindikasikan bahwa model regresi ini terdapat multikolinearitas.

Karena model tersebut memiliki multikolinearitas, maka akan dilakukan penghilangan multikolinearitas dengan beberapa metode. Sebelum dilakukan penghilangan multikolinearitas maka setiapa variabel dilakukan standarisasi terlebih dahulu dengan tujuan untuk meminimumkan kesalahan pembulatan dan untuk menganggap regresi sudah dipenuhi kenormalannya. Berikut tabel hasil standarisasinya:

Tabel 4.2.4 merupakan tabel dengan variabel yang sudah distandarisasi

Yz	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
-0.4623	-0.4520	-0.4423	-0.4279	-0.4277	-0.0483	-0.4123	-0.1621	-0.4194	-0.4076
0.0526	-0.0316	-0.0395	-0.0500	-0.0463	-0.0296	-0.0367	0.0332	-0.0341	-0.0481
0.1564	-0.0107	-0.0163	-0.0225	-0.0216	-0.0285	-0.0105	0.0458	-0.0086	-0.0237
0.0932	0.1309	0.1261	0.1135	0.1122	-0.0229	0.0930	0.1011	0.1027	0.0548
-0.0874	-0.1495	-0.1439	-0.1396	-0.1429	-0.0351	-0.1405	-0.0210	-0.1400	0.1345
-0.0332	-0.0392	-0.0334	-0.0293	-0.0318	-0.0295	-0.0448	0.0278	-0.0426	-0.0423
-0.2139	-0.0208	-0.0132	-0.0084	-0.0148	-0.0295	-0.0660	0.0182	-0.0575	-0.0882
0.0074	0.1362	0.1315	0.1302	0.1270	-0.0207	0.1147	0.1089	0.1164	0.1345
-0.0920	-0.1532	-0.1568	-0.1556	-0.1559	-0.0353	-0.1581	-0.0299	-0.1567	-0.1635
0.0435	-0.0404	-0.0533	-0.0656	-0.0609	-0.0288	-0.0624	0.0199	-0.0612	-0.0481
0.1880	0.1521	0.1362	0.1143	0.1174	-0.0221	0.1221	0.1158	0.1289	0.0941
0.1880	0.4411	0.4602	0.4798	0.4752	0.9850	0.4373	-0.6208	0.4438	0.4135
0.1339	0.2055	0.2096	0.2099	0.2095	-0.0172	0.2060	0.1573	0.2111	0.1842
0.1113	0.1279	0.1230	0.1150	0.1135	-0.0224	0.1160	0.1116	0.1208	0.0938
-0.0468	-0.0903	-0.1044	-0.1189	-0.1207	-0.0339	-0.1266	-0.0117	-0.1200	-0.1437
0.2061	0.1037	0.1138	0.1274	0.1342	-0.0197	0.1733	0.1374	0.1680	0.1865
0.0526	0.1059	0.1099	0.1093	0.1120	-0.0225	0.0967	0.1015	0.1015	0.0816
0.1203	0.1957	0.1992	0.2034	0.2039	-0.0174	0.2004	0.0617	0.1998	0.2054
0.0435	-0.0256	-0.0166	-0.0008	0.0027	-0.0263	0.0179	0.0600	0.0185	0.0037
-0.4081	-0.2545	-0.2438	-0.2353	-0.2417	-0.0404	-0.2874	-0.0963	-0.2843	-0.3079
0.1068	-0.0921	-0.1035	-0.1157	-0.1112	-0.0335	-0.0808	0.0104	-0.0811	-0.1082
-0.2139	-0.1442	-0.1335	-0.1222	-0.1211	-0.0328	-0.1333	-0.0184	-0.1349	-0.1184
-0.2094	-0.1049	-0.0987	-0.0958	-0.0989	-0.0327	-0.1201	-0.0114	-0.1200	-0.1161
0.0751	-0.0370	-0.0464	-0.0659	-0.0678	-0.0321	-0.0708	0.0169	-0.0640	-0.1085
0.1203	0.0348	0.0333	0.0398	0.0444	-0.0235	0.1303	0.0827	0.0615	0.0827
0.2151	0.0413	0.0222	0.0027	0.0089	-0.0270	0.0338	0.0695	0.0359	0.0827
0.0029	0.1968	0.2043	0.2025	0.1954	-0.0190	0.1531	0.1302	0.1627	0.1222
-0.0107	0.0142	0.0064	0.0043	0.0056	-0.0262	0.0029	0.0519	0.0026	0.0278

0.1745	0.2654	0.2623	0.2655	0.2647	-0.0136	0.2737	-0.6292	0.2739	0.2716
0.2558	0.1260	0.1231	0.1320	0.1402	-0.0183	0.1898	0.1465	0.1829	0.2153
-0.0694	0.0919	0.0963	0.0981	0.0962	-0.0227	0.0737	0.0892	0.0786	0.0691
-0.1371	-0.1112	-0.1075	-0.1072	-0.1109	-0.0338	-0.1255	-0.0129	-0.1230	-0.1379
-0.1597	-0.2537	-0.2540	-0.2543	-0.2527	-0.0407	-0.2457	-0.0751	-0.2473	-0.2629
-0.0558	-0.0936	-0.0989	-0.1060	-0.1081	-0.0332	-0.1092	-0.0039	-0.1063	-0.1300
0.0435	-0.1488	-0.1441	-0.1272	-0.1241	-0.0329	-0.0843	0.0066	-0.0883	-0.0818
-0.1913	-0.1159	-0.1071	-0.0995	-0.1034	-0.0330	-0.1203	-0.0113	-0.1198	-0.1205

- Metode Regresi Komponen Utama

Terdapat langkah-langkah untuk membentuk model Regresi komponen utama yaitu:

1. Menentukan matriks X yang telah dibakukan karena variabel merupakan variabel dengan skala pengukuran yang berbeda. Data tersebut terdapat pada tabel 4.2.4
2. Menentukan matriks korelasi dari variabel yang telah distandarisasikan

1.0000	0.9986	0.9937	0.9941	0.4861	0.9808	-0.1281	0.9883	0.9277
0.9986	1.0000	0.9980	0.9979	0.5051	0.9817	-0.1426	0.9892	0.9298
0.9937	0.9980	1.0000	0.9997	0.5247	0.9838	-0.1627	0.9901	0.9343
0.9941	0.9979	0.9997	1.0000	0.5202	0.9871	-0.1565	0.9927	0.9375
0.4861	0.5051	0.5247	0.5202	1.0000	0.4827	-0.6218	0.4891	0.4568
0.9808	0.9817	0.9838	0.9871	0.4827	1.0000	-0.1263	0.9973	0.9538
-0.1281	-0.1426	-0.1627	-0.1565	-0.6218	-0.1263	1.0000	-0.1310	-0.1262
0.9883	0.9892	0.9901	0.9927	0.4891	0.9973	-0.1310	1.0000	0.9522
0.9277	0.9298	0.9343	0.9375	0.4568	0.9538	-0.1262	0.9522	1.0000

3. Menentukan nilai eigen dan vektor eigen dari matriks korelasi, kemudian vektor eigen disusun berdasarkan nilai eigen yang terurut mulai dari nilai eigen terkecil ke nilai eigen terbesar

   Susunan nilai eigen

   0.0000 0.0001 0.0016 0.0066 0.0226 0.1048 0.2987 1.3861 7.1795

   Maka susunan vektor eigennya:

4.2.5 tabel susunan eigen Vektor

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
0.2770	-0.3750	-0.1092	-0.6861	-0.2761	0.2806	0.0423	-0.0847	0.3688
-0.4953	0.6595	-0.0796	-0.1104	-0.2939	0.2692	0.0250	-0.0689	0.3701
0.6663	0.1979	-0.0500	0.5400	-0.1934	0.2119	0.0179	-0.0494	0.3712
-0.4815	-0.6127	0.0365	0.4553	-0.0933	0.1845	0.0185	-0.0554	0.3714
0.0021	0.0000	-0.0011	-0.0383	0.0136	-0.0683	-0.7696	0.5961	0.2144
-0.0088	0.0453	-0.5834	-0.0393	0.7095	-0.0756	0.0566	-0.0879	0.3687
0.0037	-0.0007	0.0006	0.0083	-0.0073	-0.0277	-0.6227	-0.7778	-0.0796
0.0459	0.0879	0.7983	-0.1211	0.4376	-0.0020	0.0513	-0.0836	0.3702
-0.0033	-0.0027	-0.0154	-0.0209	-0.3100	-0.8710	0.1048	-0.0865	0.3550

4. Pembentukan Komponen Utama

Dari susunan nilai eigen dan vektor eigen, maka komponen utama yang terbentuk yaitu

$$W_1 = 0.2766Z_1 - 0.4932Z_2 + 0.6664Z_3 - 0.4837Z_4 + 0.0020Z_5 - 0.0084Z_6 + 0.0036Z_7 + 0.0460Z_8 - 0.0033Z_9$$

$$W_2 = -0.3755Z_1 + 0.6612Z_2 + 0.1960Z_3 - 0.6114Z_4 - 0.0000Z_5 + 0.0453Z_6 - 0.0008Z_7 + 0.0870Z_8 - 0.0025Z_9$$

$$W_3 = -0.1104Z_1 - 0.0795Z_2 - 0.0489Z_3 + 0.0366Z_4 - 0.0009Z_5 - 0.5834Z_6 + 0.0006Z_7 + 0.7982Z_8 - 0.0154Z_9$$

$$W_4 = -0.6859Z_1 - 0.1107Z_2 + 0.5406Z_3 + 0.4546Z_4 - 0.0382Z_5 - 0.0384Z_6 + 0.0083Z_7 - 0.1221Z_8 - 0.0208Z_9$$

$$W_5 = -0.2762Z_1 - 0.2935Z_2 - 0.1934Z_3 - 0.0936Z_4 + 0.0136Z_5 + 0.7096Z_6 - 0.0073Z_7 + 0.4376Z_8 - 0.3100Z_9$$

$$W_6 = 0.2804Z_1 + 0.2692Z_2 + 0.2120Z_3 + 0.1845Z_4 - 0.0683Z_5 - 0.0756Z_6 - 0.0274Z_7 - 0.0020Z_8 - 0.8711Z_9$$

$$W_7 = 0.0424Z_1 + 0.0250Z_2 + 0.0179Z_3 + 0.0187Z_4 - 0.7697Z_5 + 0.0565Z_6 - 0.6227Z_7 + 0.0514Z_8 + 0.1046Z_9$$

$$W_8 = -0.0847Z_1 - 0.0688Z_2 - 0.0493Z_3 - 0.0554Z_4 + 0.5961Z_5 - 0.0878Z_6 - 0.7779Z_7 - 0.0836Z_8 - 0.0867Z_9$$

$$W_9 = 0.3688Z_1 + 0.3701Z_2 + 0.3712Z_3 + 0.3714Z_4 + 0.2144Z_5 + 0.3687Z_6 - 0.0796Z_7 + 0.3702Z_8 + 0.3550Z_9$$

5. Pemilihan komponen utama

Komponen utama yang digunakan adalah komponen utama dengan persentasi kumulatif varians minimal 85%. Rumus yang digunakan yaitu: $\frac{jumla\ h\ nilai\ eigen}{jumla\ h\ variabel\ bebas} \geq 85\%$

Maka:

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	7,17943	79,771	79,771
2	1,38606	15,401	95,172
3	0,298746	3,319	98,491
4	0,104847	1,165	99,656
5	0,0225926	0,251	99,907
6	0,00657353	0,073	99,981
7	0,00163878	0,018	99,999
8	0,0000790216	0,001	100,000
9	0,0000352112	0,000	100,000

Dari nilai kumulatif tersebut, akan digunakan dua komponen utama karena hanya dengan dua komponen utama dengan nilai kumulatif variansi sebesar 0.952 dapat menerangkan keragaman sekitar 95.2% ≥ 85%. Jadi komponen utama yang dipilih yaitu

$$W_8 = -0.0847Z_1 - 0.0688Z_2 - 0.0493Z_3 - 0.0554Z_4 + 0.5961Z_5 - 0.0878Z_6 - 0.7779Z_7 - 0.0836Z_8 - 0.0867Z_9$$

$$W_9 = 0.3688Z_1 + 0.3701Z_2 + 0.3712Z_3 + 0.3714Z_4 + 0.2144Z_5 + 0.3687Z_6 - 0.0796Z_7 + 0.3702Z_8 + 0.3550Z_9$$

6. Pembentukan model Regresi Komponen Utama

Taksiran koefisien regresi komponen utamanya yaitu

-0.2400

0.2800

Sehingga persamaannya yaitu

$$Y = -0.2400W_8 + 0.2800W_9$$

Sehingga menjadi

$$Y^* = 0.1236Z_1 + 0.1202Z_2 + 0.1158Z_3 + 0.1173Z_4 - 0.0830Z_5 + 0.1243Z_6 + 0.1644Z_7 + 0.1237Z_8 + 0.1202Z_9$$

Dikembalikan ke variabel semula, didapatkan

$$Y = 5.0256 + 3.7642X_1 + 3.9862X_2 + 4.2829X_3 + 4.3285X_4 - 0.1689X_5 + 5.1068X_6 + 3.2803X_7 + 4.8194X_8 + 7.7348X_9$$

Untuk tabel analisis variansinya yaitu:

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	315.2146	9	35.0238	5.2009
Residual	175.0885	26	6.7342	
Total (Corr.)	490.3031	35		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\ (9,26,0.05)} = 2.27$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan

4.2.6 tabel nilai VIF dan variansi Regresi komponen Utama

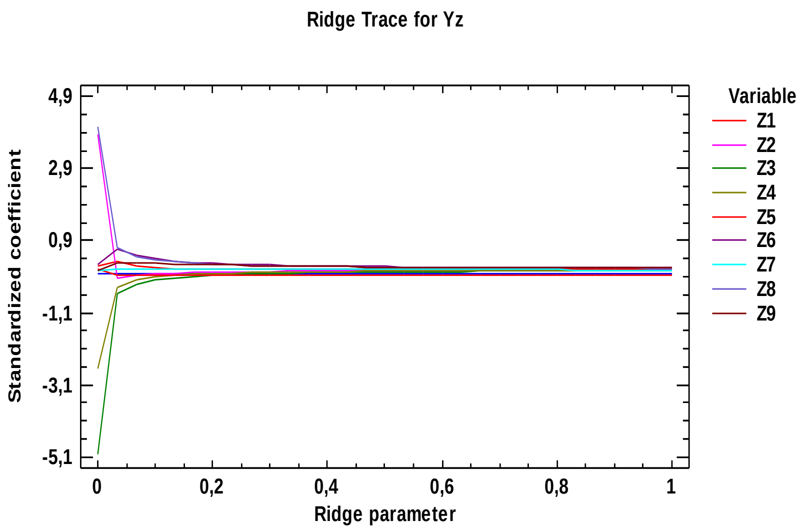
Variabel bebas	VIF	variansi
konstan		0.0011
X_1	0.0186	0.5188
X_2	0.0134	1.9007
X_3	0.0103	2.5000
X_4	0.0072	2.1429
X_5	0.4385	0.0000
X_6	0.0184	0.0600

X_7	0.9664	0.0001
X_8	0.0086	0.1165
X_9	0.1129	0.0069

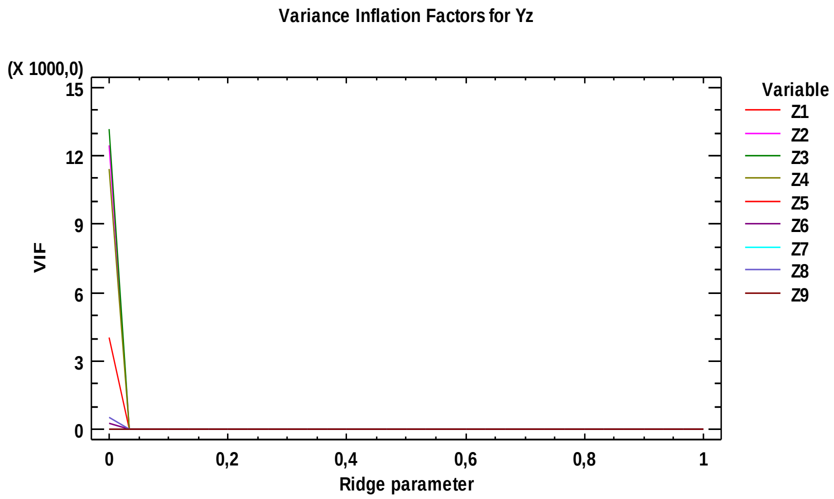
- Metode Regresi Ridge

Terdapat beberapa langkah untuk memodelkan menggunakan Regresi Ridge yaitu

1. Standarisasi variabel X dan Y, hasilnya terdapat pada tabel 4.2.4
2. Dengan data yang telah ditransformasi, maka akan dilakukan pemilihan nilai c dengan melihat nilai VIF dan ridge trace



Dari grafik Ridge Trace diatas terlihat bahwa pada ridge parameter dari 0 sampai 1 yang merupakan nilai c, yang mana koefisien standar terlihat stabil pada saat nilai c sekitar 0.25



Dari grafik VIF di atas, terlihat mulai tampak ada penurunan pada saat nilai c di sekitar 0.05. akan tetapi penurunan nilai VIF pada saat 0.05 tidak disertai koefisien regresi yang stabil, sehingga pemilihan nilai c akan memiliki bias cukup basar karena mengikuti koefisien regresi yang stabil. Maka berikut pemilihan nilai c dengan melihat kestabilan koefisien regresi

Tabel 4.2.7 nilai VIF dengan berbagai nilai c

Ridge Parameter	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
0,0	4040,09	12451,5	13144,8	11408,2	2,63704	258,246	2,13194	555,442	12,1544
0,0333333	2,9819	1,14191	1,76503	1,16116	1,88863	4,16356	1,47232	2,33018	4,91487
0,0666667	1,10776	0,552788	0,650655	0,430124	1,584	1,60043	1,26865	0,814564	3,02412
0,1	0,613009	0,345707	0,360742	0,243143	1,35509	0,85535	1,11113	0,425612	2,08295
0,133333	0,401492	0,243345	0,237721	0,163659	1,17675	0,537582	0,986033	0,267167	1,53074
0,166667	0,288867	0,183826	0,172515	0,121266	1,03449	0,372672	0,88456	0,186858	1,17605
0,2	0,220828	0,14569	0,133241	0,0955511	0,918878	0,276006	0,800759	0,140428	0,933847
0,233333	0,176188	0,119598	0,107527	0,0785966	0,82343	0,214407	0,730483	0,111088	0,760804
0,266667	0,145139	0,100871	0,0896666	0,066741	0,743559	0,172685	0,670763	0,0913126	0,632739
0,3	0,122575	0,0869252	0,0766985	0,0580776	0,675929	0,143075	0,619422	0,0773124	0,535236
0,333333	0,105608	0,0762306	0,0669507	0,051525	0,618067	0,121272	0,574833	0,0670097	0,459246
0,366667	0,0924921	0,067829	0,0594158	0,0464291	0,568104	0,10473	0,535761	0,0591858	0,398845
0,4	0,0821205	0,0610941	0,0534553	0,0423734	0,524604	0,0918633	0,501248	0,0530877	0,350025
0,433333	0,0737605	0,0556012	0,0486474	0,0390822	0,486449	0,0816445	0,470545	0,0482291	0,309989
0,466667	0,0669108	0,0510541	0,0447038	0,0363662	0,452758	0,073382	0,443057	0,0442844	0,276739
0,5	0,0612184	0,0472403	0,0414218	0,034092	0,422828	0,0665968	0,418307	0,0410291	0,248814
0,533333	0,0564287	0,0440044	0,0386555	0,032163	0,39609	0,0609487	0,395905	0,0383039	0,225129
0,566667	0,052354	0,0412305	0,0362971	0,030508	0,372084	0,0561906	0,375533	0,0359936	0,204861
0,6	0,0488534	0,0388304	0,0342662	0,0290736	0,350431	0,0521391	0,35693	0,0340128	0,187379
0,633333	0,0458194	0,0367364	0,0325012	0,0278189	0,330815	0,0486562	0,339875	0,0322972	0,17219
0,666667	0,0431689	0,0348955	0,0309545	0,026712	0,312975	0,0456361	0,324183	0,0307978	0,158906
0,7	0,0408366	0,0332658	0,029589	0,0257282	0,296692	0,0429969	0,309698	0,0294763	0,14722

0,733333	0,0387706	0,0318138	0,0283751	0,0248477	0,281779	0,0406738	0,296288	0,028303		0,136881
0,766667	0,0369295	0,0305126	0,027289	0,0240545	0,268077	0,0386157	0,283838	0,0272539		0,127688
0,8	0,0352795	0,0293401	0,0263117	0,0233358	0,255453	0,0367814	0,27225	0,0263101		0,119477
0,833333	0,0337933	0,0282783	0,0254274	0,022681	0,243788	0,0351374	0,261438	0,025456		0,112109
0,866667	0,0324481	0,0273122	0,0246234	0,0220815	0,232984	0,0336564	0,251329	0,0246789		0,105472
0,9	0,0312251	0,0264295	0,0238888	0,02153	0,222951	0,032316	0,241856	0,0239684		0,0994713
0,933333	0,0301086	0,0256195	0,0232149	0,0210205	0,213615	0,0310973	0,232964	0,0233158		0,094026
0,966667	0,0290855	0,0248735	0,022594	0,0205479	0,204908	0,0299848	0,2246	0,0227138		0,0890687
1,0	0,0281444	0,024184	0,0220198	0,0201078	0,196773	0,0289653	0,216721	0,0221563		0,0845417

Dari berbagai nilai c yang ada, terlihat adanya penurunan nilai VIF sedikit demi sedikit, nilai c yang akan diambil adalah pada saat nilai VIF dengan disertai kestabilan koefisien regresi ridge dengan c = 0.267

Tabel 4.2.8 nilai koefisien regresi ridge dengan nilai tetapan bias c = 0.267

variabel	Koefisien Regresi Ridge
Z_1	0.1105
Z_2	0.0379
Z_3	-0.0212
Z_4	0.0120
Z_5	-0.0701
Z_6	0.2376
Z_7	0.1165
Z_8	0.2172
Z_9	0.2111

Maka dapat dibentuk model regresi ridgenya yaitu:

$$Y^* = \quad 0.1105Z_1 + 0.0379Z_2 - 0.0212Z_3 + 0.0120Z_4 - 0.0701Z_5 + 0.2376Z_6 + 0.1165Z_7 + 0.2172Z_8 + 0.2111Z_9$$

Apabila model di atas dikembalikan ke variabel-variabel asal maka diperoleh:

$$Y = 10.3533 + 3.3658X_1 + 1.2568X_2 - 0.7825X_3 + 0.4427X_4 - 0.1427X_5 + 9.7587X_6 + 2.3250X_7 + 8.4598X_8 + 13.5837X_9$$

Tabel Anavanya

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	309.6672	9	34.4075	4.9525
Residual	180.6358	26	6.9475	
Total (Corr.)	490.3031	35		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(9,26,0.05)} = 2.27$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Tabel 4.2.9 tabel nilai VIF Regresi Ridge

Variabel bebas	VIF	variansi
konstan		0.0012
X_1	0,145136	0.5352
X_2	0,100869	1.9609
X_3	0,0896651	2.5792
X_4	0,06674	2.2108
X_5	0,743551	0.0000
X_6	0,172681	0.0619
X_7	0,670757	0.0001
X_8	0,0913109	0.1202
X_9	0,632728	0.0071

➢ Metode Penghilangan Variabel

Untuk mengetahui variabel bebas mana yang akan dihilangkan, yaitu dengan melihat korelasi antar variabel bebas yang hampir sempurna atau mendekati nilai satu. Berikut matriks korelasi antar variabel bebas:

1.0000	0.9986	0.9937	0.9941	0.4860	0.9808	-0.1267	0.9883	0.9278
0.9986	1.0000	0.9980	0.9979	0.5051	0.9817	-0.1411	0.9892	0.9298
0.9937	0.9980	1.0000	0.9997	0.5248	0.9838	-0.1615	0.9901	0.9343
0.9941	0.9979	0.9997	1.0000	0.5202	0.9871	-0.1552	0.9927	0.9375
0.4860	0.5051	0.5248	0.5202	1.0000	0.4827	-0.6182	0.4891	0.4568
0.9808	0.9817	0.9838	0.9871	0.4827	1.0000	-0.1260	0.9973	0.9538
-0.1267	-0.1411	-0.1615	-0.1552	-0.6182	-0.1260	1.0000	-0.1301	-0.1243
0.9883	0.9892	0.9901	0.9927	0.4891	0.9973	-0.1301	1.0000	0.9522
0.9278	0.9298	0.9343	0.9375	0.4568	0.9538	-0.1243	0.9522	1.0000

Dari matriks tersebut dapat dilihat bahwa ada korelasi antara variabel bebas $Z_1, Z_2,\ Z_3, Z_4, Z_6, Z_8, Z_9$ dengan nilai korelasi hamper menuju nilai satu. Maka dalam kasus ini akan dicari salah satu variabel yang akan dihilangkan dengan melihat masing-masing konsekuensi yang dihasilkan apabila masing-masing variable bebas yang berkorelasi dihilangkan. Konsekuensinya dilihat dari berkurangnya nilai VIF pada variable bebas.

✓ Konsekuensi apabila variabel bebas Z_1 yang dihilangkan

Tabel 4.2.10 Tabel Konsekuensi apabila variabel bebas Z_1 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.957E-5	
Z_2	4.254	340.989
Z_3	-7.621	361.767
Z_4	-3.042	2.466
Z_5	0.111	1.137E4

Z_6	0.216	234.939
Z_7	0.059	1.765
Z_8	3.801	397.913
Z_9	0.079	11.795

✓ Konsekuensi apabila variabel bebas Z_2 yang dihilangkan

Tabel 4.2.11 Tabel Konsekuensi apabila variabel bebas Z_2 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.123E-5	
Z_1	2.388	110.519
Z_3	-5.475	129.636
Z_4	-3.606	1.137E4
Z_5	.120	2.493
Z_6	.022	228.009
Z_7	.077	1.743
Z_8	3.722	404.286
Z_9	.072	11.765

✓ Konsekuensi apabila variabel bebas Z_3 yang dihilangkan

Tabel 4.2.12 Tabel Konsekuensi apabila variabel bebas Z_3 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	1.639E-5	
Z_1	1.750	2.703E3
Z_2	1.098	2.846E3
Z_4	-6.963	2.530
Z_5	0.132	258.384
Z_6	0.274	1.782
Z_7	0.095	475.188

Z_8	4.537	11.778
Z_9	0.037	8.056E3

✓ Konsekuensi apabila variabel bebas Z_4 yang dihilangkan

Tabel 4.2.13 Konsekuensi apabila variabel bebas Z_4 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.123E-5	
Z_1	2.388	110.519
Z_2	4.350	1.242E4
Z_3	-5.475	129.636
Z_5	0.120	2.493
Z_6	0.022	228.009
Z_7	0.077	1.743
Z_8	3.722	404.286
Z_9	0.072	11.765

✓ Konsekuensi apabila variabel bebas Z_6 yang dihilangkan

Tabel 4.2.14 Tabel konsekuensi apabila variabel bebas Z_6 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.195E-5	
Z_1	2.383	102.539
Z_2	3.870	1.118E4
Z_3	-5.475	129.563
Z_4	-3.433	1.095E4
Z_5	0.120	2.490
Z_7	0.077	1.742
Z_8	3.749	83.137
Z_9	0.073	11.707

✓ Konsekuensi apabila variabel bebas Z_8 yang dihilangkan

Tabel 4.2.15 Tabel konsekuensi apabila variabel bebas Z_8 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.912E-5	
Z_1	2.954	100.224
Z_2	3.587	1.267E4
Z_3	-10.389	6.345E3
Z_4	6.112	8.164E3
Z_5	0.105	2.535
Z_6	1.860	140.099
Z_7	0.059	1.830
Z_9	0.204	11.355

✓ Konsekuensi apabila variabel bebas Z_9 yang dihilangkan

Tabel 4.2.16 Tabel konsekuensi apabila variabel bebas Z_9 yang dihilangkan

variabel	Koefisien regresi	VIF
Konstan	-7.057E-5	
Z_1	2.347	106.836
Z_2	4.216	1.238E4
Z_3	-5.469	129.561
Z_4	-3.867	9.061E-5
Z_5	0.118	2.489
Z_6	0.044	226.885
Z_7	0.076	1.742
Z_8	3.804	389.372

Dilihat dari setiap table konsekuensi apabila salah satu variabel bebas yang diindikasikan berkorelasi dengan variabel lain dihilangkan, model regresi tetap memiliki masalah multikolinearitas, hal ini terlihat dari nilai VIF setiap tabel konsekuensi masih terdapat nilai yang melebihi sepuluh. Maka untuk contoh kasus kedua ini tidak dapat menggunakan metode penghilangan variabel.

✓ Perbandingan Setiap Metode

Persamaan regresi linear berganda:

$$Y = 36.1339 + 7.1126X_1 + 128.0696X_2 - 180.4833X_3 - 109.8327X_4 + 0.2433X_5 + 11.5915X_6 + 1.4004X_7 + 161.4825X_8 + 4.0443X_9$$

Persamaan metode Regresi Komponen Utama :

$$Y = 5.0256 + 3.7642X_1 + 3.9862X_2 + 4.2829X_3 + 4.3285X_4 - 0.1689X_5 + 5.1068X_6 + 3.2803X_7 + 4.8194X_8 + 7.7348X_9$$

Persamaan metode Regresi Ridge :

$$Y = 10.3533 + 3.3658X_1 + 1.2568X_2 - 0.7825X_3 + 0.4427X_4 - 0.1427X_5 + 9.7587X_6 + 2.3250X_7 + 8.4598X_8 + 13.5837X_9$$

Persamaan metode PV : tidak digunakan

Perbandingan metode Regresi Komponen Utama dan metode Regresi Ridge dilihat dari nilai VIF dan nilai variansi

Tabel 4.2.17 tabel perbandingan nilai VIF kasus kedua

Variabel bebas	Nilai VIF dari metode	
	RKU	RR
X_1	0.0186	0,145136
X_2	0.0134	0,100869
X_3	0.0103	0,0896651
X_4	0.0072	0,06674
X_5	0.4385	0,743551
X_6	0.0184	0,172681
X_7	0.9664	0,670757
X_8	0.0086	0,0913109
X_9	0.1129	0,632728

Tabel 4.2.18 tabel perbandingan nilai variansi kasus kedua

Variabel bebas	Nilai variansi dari metode	
	RKU	RR
konstan	0.0011	0.0012
X_1	0.5188	0.5352
X_2	1.9007	1.9609
X_3	2.5000	2.5792
X_4	2.1429	2.2108
X_5	0.0000	0.0000
X_6	0.0600	0.0619
X_7	0.0001	0.0001
X_8	0.1165	0.1202
X_9	0.0069	0.0071

Ket:
RLB: regresi linear berganda
RKU: regresi komponen utama
RR: regresi ridge

Dilihat dari nilai VIF setiap variabel bebas antara metode Regresi Komponen Utama dengan Regresi Ridge terlihat kedua metode ini dapat mengatasi multikolinearitas meskipun nilai VIFnya beragam dan tidak stabil menuju nilai VIF satu , sehingga dilihat dari nilai variansinya, yang paling kecil adalah variansi pada metode Regresi Komponen Utama , maka untuk kasus dengan data besar variabel banyak ini akan lebih efektif menggunakan metode Regresi komponen Utama. Untuk metode penghilangan variabel tidak dapat digunakan karena pada analisis penghilangan variabelnya sama sekali tidak mengurangi dampak multikolinearitas.

4.3 Contoh Kasus Ketiga

Terdapat contoh kasus mengenai konsumsi daging ayam (Y) di Amerika yang dipengaruhi oleh harga per kg daging ayam (X_1), harga per kg daging babi (X_2), pendapatan siap pakai perkapita (X_3) , dan total pembelian barang dan jasa oleh pemerintah (X_4). Data secara lengkapnya adalah sebagai berikut

Tabel 4.3.1 Tabel data kasus ketiga

Y	X_1	X_2	X_3	X_4
44.8	12.4	48.5	71.24	677.0
48.3	13.9	66.1	78.9	689.3
48.4	11.0	62.4	86.97	704.2
50.04	11.1	58.6	96.03	713.2
51.5	10.3	56.7	101.33	723.6
52.6	12.7	55.5	107.77	743.8
54.5	15.9	57.3	119.14	766.9
56.3	14.8	53.7	125.94	813.4
58.1	12.5	52.6	132.13	855.4
61.9	11.0	61.1	138.53	881.5
63.8	9.2	66.6	148.84	886.8
67.5	14.9	69.5	157.74	904.4
70.4	9.3	74.6	166.89	932.6
73.5	7.1	72.7	171.82	944.0
76.8	8.6	71.3	180.32	936.9
78.9	10.0	72.6	185.64	929.8
80.5	7.6	66.7	192.59	922.5

Sumber: U.S Department of Agriculture Statistics. Dikutip dari Drs. Sarwoko, Dasar-Dasar Ekonometrika,2005

Tabel 4.3.2 koefisien regresi linear berganda

variabel	Koefisien regresi
Konstan	32.689
X_1	-0.2716
X_2	0.1535
X_3	0.3314
X_4	-0.0270

Maka didapat model regresinya:

$$Y = 32.6893 - 0.2716X_1 + 0.1535X_2 + 0.3314X_3 - 0.0270X_4$$

Tabel Anavanya

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	2.1558e+003	4	538.9461	247.2639
Residual	26.1557	12	2.1796	
Total (Corr.)	2.1819e+003	16		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(4,12,0.05)} = 3.26$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan. .

Mendeteksi adanya multikolinearitas yaitu dengan nilai VIF > 10 maka diindikasikan terdapat multikolinearitas

Tabel 4.3.3 nilai VIF setiap variabel bebas

Variabel bebas	VIF
X_1	1.526
X_2	2.196
X_3	17.389
X_4	16.330

Dari tabel di atas diindikasikan bahwa model regresi ini terdapat multikolinearitas.

Karena model tersebut memiliki multikolinearitas, maka akan dilakukan penghilangan multikolinearitas dengan beberapa metode. Sebelum dilakukan penghilangan multikolinearitas maka setiapa variabel dilakukan standarisasi terlebih dahulu dengan tujuan untuk meminimumkan kesalahan pembulatan dan untuk menganggap regresi sudah dipenuhi kenormalannya. Berikut tabel hasil standarisasinya:

Tabel 4.3.4 hasil standarisasi

Yz	Z_1	Z_2	Z_3	Z_4
-0.3479	0.1049	-0.4449	-0.3972	-0.3691
-0.2729	0.2494	0.1052	-0.3479	-0.3385
-0.2708	-0.0300	-0.0105	-0.2961	-0.3013
-0.2357	-0.0204	-0.1292	-0.2379	-0.2789
-0.2044	-0.0975	-0.1886	-0.2038	-0.2529
-0.1809	0.1338	-0.2261	-0.1624	-0.2025
-0.1402	0.4422	-0.1699	-0.0894	-0.1449
-0.1017	0.3362	-0.2824	-0.0457	-0.0290
-0.0631	0.1145	-0.3168	-0.0059	0.0758
0.0182	-0.0300	-0.0511	0.0352	0.1409
0.0589	-0.2035	0.1208	0.1015	0.1541
0.1381	0.3458	0.2114	0.1587	0.1980
0.2002	-0.1939	0.3708	0.2175	0.2683
0.2665	-0.4059	0.3114	0.2491	0.2967
0.3372	-0.2613	0.2677	0.3038	0.2790
0.3821	-0.1264	0.3083	0.3379	0.2613
0.4164	-0.3577	0.1239	0.3826	0.2431

- Metode Regresi Komponen Utama

Terdapat langkah-langkah untuk membentuk model Regresi komponen utama yaitu:

1. Menentukan matriks X yang telah dibakukan karena variabel merupakan variabel dengan skala pengukuran yang berbeda. Data tersebut terdapat pada tabel 4.3.4
2. Menentukan matriks korelasi dari variabel yang telah distandarisasikan

1.0000	-0.5372	-0.5412	-0.5084
-0.5372	1.0000	0.7142	0.7006
-0.5412	0.7142	1.0000	0.9685
-0.5084	0.7006	0.9685	1.0000

3. Menentukan nilai eigen dan vektor eigen dari matriks korelasi, kemudian vektor eigen disusun berdasarkan nilai eigen yang terurut mulai dari nilai eigen terbesar ke nilai eigen terkecil

   Susunan nilai eigen

   0.0306 0.3589 0.5988 3.0116

   Maka susunan vektor eigennya:

4.3.5 tabel susunan eigen Vektor

e_1	e_2	e_3	e_4
0.0314	0.2216	0.8823	-0.4140
-0.0104	0.8707	0.0124	0.4916
0.7189	-0.3035	0.3064	0.5451
-0.6943	-0.3173	0.3570	0.5383

4. Pembentukan Komponen Utama

   Dari susunan nilai eigen dan vektor eigen, maka komponen utama yang terbentuk yaitu

$$W_1 = 0.0314Z_1 - 0.0105Z_2 + 0.7190Z_3 + 0.0786Z_4$$

$$W_2 = 0.2218Z_1 + 0.8707Z_2 - 0.3034Z_3 - 0.3173Z_4$$

$$W_3 = 0.8823Z_1 + 0.0122Z_2 + 0.3065Z_3 + 0.3571Z_4$$

$$W_4 = -0.4140Z_1 + 0.4916Z_2 + 0.5451Z_3 + 0.5383Z_4$$

5. Pemilihan komponen utama

   Komponen utama yang digunakan adalah komponen utama dengan persentasi

Component		*Percent of*	*Cumulative*
Number	*Eigenvalue*	*Variance*	*Percentage*
1	3,0116	75,290	75,290
2	0,598841	14,971	90,261
3	0,358926	8,973	99,234
4	0,030628	0,766	100,000

Dari nilai kumulatif tersebut, akan digunakan dua komponen utama karena hanya dengan dua komponen utama dengan nilai kumulatif variansi sebesar 0.902 dapat menerangkan keragaman sekitar 90.26% ≥ 85%. Jadi komponen utama yang dipilih yaitu

$$W_3 = 0.8823Z_1 + 0.0122Z_2 + 0.3065Z_3 + 0.3571Z_4$$

$$W_4 = -0.4140Z_1 + 0.4916Z_2 + 0.5451Z_3 + 0.5383Z_4$$

6. Pembentukan model regresi komponen utama

   Taksiran koefisien regresi komponen utamanya yaitu

   0.2036

   0.5537

   Sehingga persamaannya yaitu

$$Y = 0.2036W_3 + 0.5537W_4$$

   Sehingga menjadi

$$Y^* = -0.0496Z_1 + 0.2747Z_2 + 0.3642Z_3 + 0.3708Z_4$$

Dikembalikan ke variabel semula, didapatkan

$$Y = -11.7685 - 0.2231X_1 + 0.4011X_2 + 0.1093X_3 + 0.0432X_4$$

Tabel anava

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	2.0689e+003	4	689.6309	79.3048
Residual	113.0474	12	8.6960	
Total (Corr.)	2.1819e+003	16		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1: \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\ (4,12,0.05)} = 3.26,$ karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

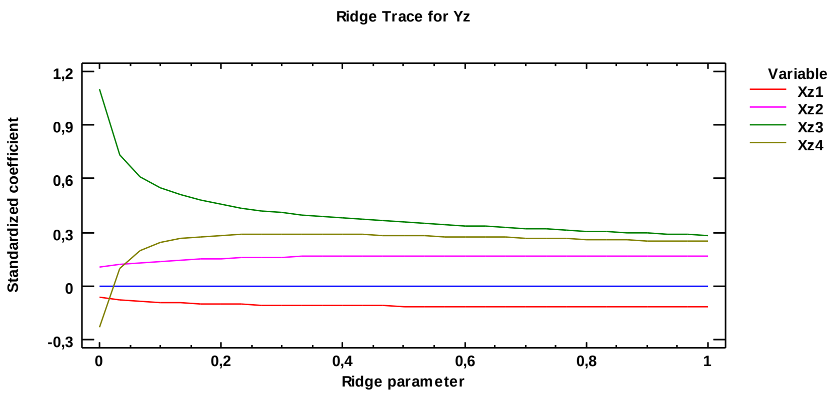
Tabel 4.3.6 nilai VIF dan variansi pada regresi komponen utama

Variabel bebas	VIF	variansi
konstan		277.9720
X_1	0.6780	0.1232
X_2	0.4559	0.0187
X_3	0.1589	0.0062
X_4	0.1882	0.0009

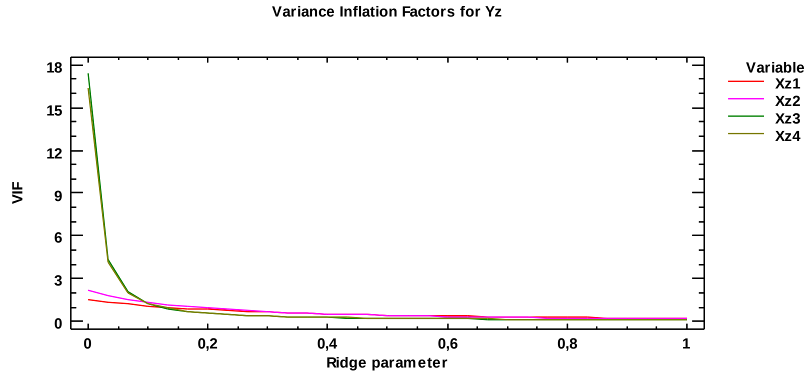
- Metode Regresi Ridge

Terdapat beberapa langkah untuk memodelkan menggunakan Regresi Ridge yaitu

1. Standarisasi variabel X dan Y, hasilnya terdapat pada tabel 4.3.4
2. Dengan data yang telah ditransformasi, maka akan dilakukan pemilihan nilai c dengan melihat nilai VIF dan ridge trace



Dari grafik Ridge Trace diatas terlihat bahwa pada ridge parameter dari 0 sampai 1 yang merupakan nilai c, yang mana koefisien standar terlihat stabil pada saat nilai c sekitar 0 sampai 0.4



Dari grafik VIF di atas, terlihat mulai tampak ada penurunan pada saat nilai c di sekitar 0.2. akan tetapi penurunan nilai VIF pada saat 0.2 tidak disertai koefisien regresi yang stabil, sehingga pemilihan nilai c akan memiliki bias cukup besar karena mengikuti koefisien regresi yang stabil. Maka berikut pemilihan nilai c dengan melihat kestabilan koefisien regresi

Tabel 4.3.7 nilai VIF dengan berbagai nilai c

Ridge Parameter	Xz1	Xz2	Xz3	Xz4
0,0	1,52601	2,19618	17,3887	16,3271
0,0333333	1,34423	1,8479	4,32175	4,12865
0,0666667	1,20765	1,57956	2,07622	2,02357
0,1	1,09342	1,36745	1,29222	1,28326
0,133333	0,995744	1,19674	0,920611	0,928975
0,166667	0,911278	1,05724	0,710861	0,726797
0,2	0,837617	0,941741	0,57828	0,597516
0,233333	0,772928	0,844992	0,487593	0,508063
0,266667	0,715772	0,763113	0,421874	0,442519
0,3	0,664997	0,69318	0,37212	0,392384
0,333333	0,619666	0,632953	0,33315	0,352739
0,366667	0,579014	0,580696	0,301786	0,320558
0,4	0,542406	0,535046	0,275985	0,293879
0,433333	0,509312	0,494918	0,254371	0,271374
0,466667	0,479291	0,459443	0,235988	0,252115
0,5	0,451965	0,427917	0,220152	0,235433
0,533333	0,427016	0,399765	0,206358	0,220833
0,566667	0,404171	0,374513	0,194228	0,207939
0,6	0,383196	0,351768	0,183472	0,196463
0,633333	0,363888	0,331201	0,173865	0,186179
0,666667	0,346072	0,312537	0,165227	0,176907
0,7	0,329596	0,295543	0,157416	0,168501
0,733333	0,314325	0,280019	0,150315	0,160844
0,766667	0,300143	0,265796	0,143829	0,153837
0,8	0,286947	0,25273	0,13788	0,1474
0,833333	0,274645	0,240693	0,132402	0,141465
0,866667	0,263157	0,229578	0,127339	0,135974
0,9	0,252411	0,219289	0,122645	0,130878

0,933333	0,242343	0,209744	0,11828	0,126136
0,966667	0,232897	0,20087	0,114208	0,12171
1,0	0,22402	0,192604	0,110402	0,117571

Dari berbagai nilai c yang ada, terlihat adanya penurunan nilai VIF sedikit demi sedikit, nilai c yang akan diambil adalah pada saat nilai VIF dengan disertai kestabilan koefisien regresi ridge dengan c = 0.267.

Tabel 4.3.8 nilai koefisien regresi ridge dengan nilai tetapan bias c = 0.3667

variabel	Koefisien Regresi Ridge
Z_1	-0.1107
Z_2	0.1645
Z_3	0.3880
Z_4	0.2887

Nilai c yang akan diambil adalah pada saat nilai c = 0.3667menghasilkan persamaan regresi Ridgenya yaitu

$$Y^* = -0.1107Z_1 + 0.1645Z_2 + 0.3880Z_3 + 0.2887Z_4$$

Dikembalikan ke variabel semula, didapatkan

$$Y = 8.3832 - 0.4983X_1 + 0.2401X_2 + 0.1165X_3 + 0.0336X_4$$

Tabel anava

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	1.8473e+003	4	615.7711	23.9222
Residual	334.6268	12	25.7405	
Total (Corr.)	2.1819e+003	16		

Setelah model diperoleh maka akan diuji signifikan dari model tersebut, untuk melakukan pengujian tersebut dilakukan sebagai berikut:

Hipotesa: $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; regresi tidak signifikan

$H_1 : \beta_j \neq 0$; regresi berarti signifikan

Kriteria: tolak H_0 bila $F_{hitung} > F_{tabel}$; dalam hal lain terima H_0

Hasilnya: dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\,(4,12,0.05)} = 3.26,$ karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Tabel 4.3.9 nilai VIF dan variansi pada regresi ridge

Variabel bebas	VIF	variansi
konstan		309.0472
X_1	0,578975	0.1370
X_2	0,580647	0.0207
X_3	0,301758	0.0069
X_4	0,320529	0.0010

- Metode Penghilangan Variabel

Untuk mengetahui variabel bebas mana yang akan dihilangkan, yaitu dengan melihat korelasi antar variabel bebas yang hampir sempurna atau mendekati nilai satu. Berikut matriks korelasi antar variabel bebas dari variabel yang telah distandarisasikan:

$$\begin{matrix} 1.0000 & -0.5372 & -0.5412 & -0.5084 \\ -0.5372 & 1.0000 & 0.7142 & 0.7006 \\ -0.5412 & 0.7142 & 1.0000 & 0.9685 \\ -0.5084 & 0.7006 & 0.9685 & 1.0000 \end{matrix}$$

Dari matriks tersebut dapat dilihat bahwa ada korelasi antara variabel bebas $Z_3\ dan\ Z_4$ dengan nilai korelasi hampir menuju nilai satu. Maka dalam kasus ini akan dicari salah satu variabel yang akan dihilangkan dengan melihat masing-masing konsekuensi yang dihasilkan apabila masing-masing variable bebas yang berkorelasi dihilangkan. Konsekuensinya dilihat dari berkurangnya nilai VIF pada variabel bebas.

✓ Konsekuensi apabila variabel bebas Z_3 yang dihilangkan

Tabel 4.3.10 Tabel koefisien regresi

variabel	Koefisien regresi
Z_1	-0.119
Z_2	0.161
Z_4	0.768

Maka didapatkan persamaan

$$Y^* = -0.119Z_1 + 0.161Z_2 + 0.768Z_3$$

Dikembalikan ke variabel semula didapatkan

$$Y = 20.2287 - 0.3007X_1 + 0.1444X_2 + 0.2643X_3$$

Tabel 4.3.10 konsekuensi apabila variabel bebas X_3 dihilangkan

variabel	VIF	variansi
konstan		176.0568
X_1	1.477	0.1890
X_2	2.151	0.0290
X_4	2.064	0.0002

✓ Konsekuensi apabila variabel bebas Z_4 yang dihilangkan

Tabel 4.3.10 Tabel konsekuensi apabila variabel bebas Z_4 yang dihilangkan

variabel	Koefisien regresi	VIF
Z_1	-0.067	1.513
Z_2	0.099	2.185
Z_3	0.881	2.198

Sehingga didapatkan persamaan

$$Y^* = -0.0668Z_1 + 0.0989Z_2 + 0.8805Z_3$$

Dikembalikan ke variabel semula, didapatkan

$$Y = 20.2287 - 0.3007X_1 + 0.1444X_2 + 0.2643X_3$$

Tabel anava

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*
regresi	2.1486e+003	3	716.1981	279.2121
Residual	33.3459	13	2.5651	
Total (Corr.)	2.1819e+003	16		

Hasil, dengan taraf signifikan $\alpha = 0.05$ maka $F_{tabel\ (3,13,0.05)} = 3.41$, karena $F_{hit} > F_{tab}$ dinyatakan bahwa regresi signifikan.

Tabel 4.3.11 Tabel nilai VIF dan variansi

variabel	VIF	variansi
konstan		26.6037
X_1	1.513	0.0360
X_2	2.185	0.0055
X_3	2.198	0.0002

Disini, dengan menghilangkan variabel bebas X_3 ataupun X_4 nilai VIF nya menandakan multikolinearitas sudah teratasi sehingga akan dilihat dari nilai variansinya dimana dengan menghilangkan variabel bebas X_4 variansinya lebih kecil dibandingkan dengan menghilangkan variabel bebas X_3. Sehingga untuk kasus ini akan dihilangkan variabel bebas X_4. Dan terdapat alasan yang cukup jelas bahwa dengan melihat variabel bebas X_4 yang sebenarnya merupakan variabel yang tidak relevan bila diregresikan kedalam kasus tersebut memang harus dihilangkan.

- Perbandingan Setiap Metode

  Persamaan yang dihasilkan dari setiap metode yaitu:

  Persamaan RLB :

  $$Y = 32.6893 - 0.2716X_1 + 0.1535X_2 + 0.3314X_3 - 0.0270X_4$$

Persamaan metode RKU :

$$Y = -11.7685 - 0.2231X_1 + 0.4011X_2 + 0.1093X_3 + 0.0432X_4$$

Persamaan metode RR :

$$Y = 8.3832 - 0.4983X_1 + 0.2401X_2 + 0.1165X_3 + 0.0336X_4$$

Persamaan metode PV :

$$Y = 20.2287 - 0.3007X_1 + 0.1444X_2 + 0.2643X_3$$

Perbandingan metode Regresi Komponen Utama, Regresi Ridge, dan penghilangan variabel dilihat dari nilai VIF dan nilai variansi:

Tabel 4.3.12 Tabel perbandingan nilai VIF kasus ketiga

Variabel bebas	Nilai VIF dari metode		
	RKU	RR	PV
X_1	0.6780	0,578975	1.513
X_2	0.4559	0,580647	2.185
X_3	0.1589	0,301758	2.198
X_4	0.1882	0,320529	Tidak ada

Tabel 4.3.13 Tabel perbandingan nilai variansi kasus ketiga

Variabel bebas	Nilai variansi dari metode		
	RKU	RR	PV
konstan	277.9720	822.8131	26.6037
X_1	0.1232	0.3648	0.0360
X_2	0.0187	0.0552	0.0055
X_3	0.0062	0.0185	0.0002
X_4	0.0009	0.0026	$-$

Ket:
RLB: regresi linear berganda
RKU: regresi komponen utama
RR: regresi ridge

Dilihat dari nilai VIF setiap variabel bebas antara metode Regresi Komponen Utama metode Regresi Ridge dan metode penghilangan variabel, terlihat ketiga metode ini dapat mengatasi multikolinearitas meskipun nilai VIFnya beragam dan tidak stabil menuju nilai VIF satu, sehingga apabila dilihat dari nilai variansinya, yang paling kecil adalah variansi pada metode penghilangan variabel , maka untuk kasus ini akan lebih efektif menggunakan metode penghilangan variabel.

BAB V
KESIMPULAN DAN SARAN

4.1 Kesimpulan

Dari pembahasan yang telah dibahas, dapat disimpulkan bahwa:

1. Cara pembentukan model regresi komponen utama dari regresi linear berganda terdapat dua cara yaitu dengan matriks kovarian apabila skala pengukuran variabelnya sama dengan bentuk komponen utamanya $W_i = e_i' X = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p \quad i = 1,2,\ldots,p$. Kedua dengan matriks korelasi apabila sekala pengukuran variabelnya tidak sama dengan bentuk komponen utamanya $W_p = e_{1p}Z_1 + e_{2p}Z_2 + \cdots + e_{pp}Z_p$
2. Cara pembentukan model Regresi Ridge dari model regresi linear berganda adalah dengan cara menambahkan tetapan bias c pada diagonal utama matriks *X'X.* Sehingga penduga koefisiennya menjadi $\hat{\beta}^*(c) = (X'X + cI)^{-1}X'Y$, dengan nilai c ini dipilih berdasarkan Ridge Trace dan nilai VIF yang makin kecil yang didapat dari nilai c yang semakin besar.
3. Cara pembentukan model regresi linear berganda setelah dilakukan penghilangan variabel bebas adalah dengan menghilagkan salah satu variabel bebas yang diduga menjadi penyebab utama adanya multikolinearitas.
4. Kelebihan dan kekurangan masing-masing metode dapat dilihat pada bentuk tabel berikut:

Kekurangan/kelebihan metode yang dilihat dari:	Metode Penghilangan Multikolinearitas		
	RKU	RR	Penghilangan Variabel
Tingkat kesulitan pembuatan model	sulit	sedang	sederhana
Sifat pembuatan model	objektif	subjektif	subjektif
Sifat penaksir koefisien regresi	bias dan variansi minimum	Bias dan variansi minimum	Bias dan variansi minimum
Dampak multikolinearitas	Mengurangi multikolinearitas	Mengurangi multikolinearitas	Mengurangi multikolinearitas
Jenis kasus yang seuai	Kasus dengan data besar dan variabel banyak	Kasus dengan data kecil dan variabel sedikit	Kasus dengan data terdapat variabel multikolinear yang tidak relevan

5. Dari ketiga metode tersebut, metode yang paling efektif yaitu tergantung pada suatu kondisi tertentu, dimana metode Regresi komponen Utama dapat efektif apabila terdapat kasus dengan sampel besar dan variabel banyak, untuk metode Regresi Ridge dapat efektif apabila terdapat kasus kecil

dengan variabel sedikit, dan untuk metode penghilangan variabel dapat menjadi metode paling efektif apabila terdapat kasus dengan variabel bebas yang berkorelasi kuat dengan variabel bebas lain tersebut mempengruhi variabel bebas yag tidak penting.

4.2 Saran

Terdapatnya multikolinearitas pada regresi linear berganda menyebabkan model regresi yang tidak baik, oleh karena itu sebelum dilakukan pembuatan modelnya perlu dilakukan uji multikolinearitas. Sehingga bila diindikasikan adanya multikolinearitas, maka dapat diatasi dengan metode Regresi Komponen Utama, metode Regresi Ridge, atau metode penghilangan variabel dengan melihat keefektifitasan penggunaan metode tersebut dalam jenis kasus.