REGRESI RIDGE UNTUK MENGATASI MULTIKOLINIERITAS

Reny Mulyasari[1], Sigit Nugroho[2] dan Jose Rizal[2]

[1]Alumni Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

[2]Dosen Jurusan Matematika, Fakultas MIPA, Universitas Bengkulu

ABSTRACT

This research aim to study about multicollinearity problem, Ridge Regression and its application. Method used in this research is literature study using simulation data. Variance Inflation Factor (VIF) is used to test if there is multicollinearity. The result shows that Ridge Regression can be used to handle the problem and give the biased estimator though having better variance than those one derived using ordinary least square.

Keyword : *Multicollinearity, Ridge Regression, Biased, Variance Inflation Factor (VIF)*

1. PENDAHULUAN

Regresi merupakan suatu teknik statistika yang dapat digunakan untuk menggambarkan hubungan diantara dua peubah atau lebih, yaitu hubungan antara satu atau lebih peubah bebas (*independent*) dengan satu peubah tak bebas (*dependent*). Menurut Draper dan Smith (1992) analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan peubah tehadap peubah lainnya.

Dalam perkembangannya terdapat dua jenis regresi yang sangat terkenal, yaitu regresi linier sederhana dan regresi linier berganda. Regresi linear sederhana digunakan untuk menggambarkan hubungan antara satu peubah bebas (X) dengan satu peubah tak bebas (Y) dalam bentuk persamaan linier sederhana.

$$\beta_0 + \beta_1 X_{i1} \quad ; \quad i = 1,2,...,n \tag{1.1}$$

Sedangkan regresi linier berganda menggambarkan hubungan antara dua atau lebih peubah bebas $(X_1, X_2, \ldots, X_k)$ dengan satu peubah tak bebas (Y).

$$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \quad ; \quad i = 1,2,...,n \tag{1.2}$$

Keterangan :

Y_i	=	peubah tak bebas (respon)
$X_{i1},...,X_{i2}$	=	peubah bebas (*predictor*)
$\beta_0,...,\beta_k$	=	parameter regresi
ε_i	=	galat (error), $N(0,\sigma^2)$

Parameter regresi linier sederhana dan regresi linier berganda pada persamaan di atas dicari penduganya dengan menggunakan metode kuadrat terkecil. Penduga kuadrat terkecil diperoleh dengan meminimumkan jumlah kuadrat galat untuk model regresinya. Penduga yang dihasilkan oleh metode kuadrat terkecil ini bersifat *BLUE (Best Linear Unbiased Estimator)* dengan beberapa persyaratan.

Persyaratan yang harus dipenuhi ini lebih dikenal dengan nama asumsi klasik regresi linier. Asumsi klasik regresi linier tersebut adalah sebagai berikut : (1) model regresi benar-benar linier, (2) nilai peubah bebas (X) tetap pada sampel berulang, (3) nilai rata-rata dari error adalah nol, (4) homoskedastisitas sama untuk setiap observasi, (5) tidak ada otokorelasi antara unsur pengganggu, (6) kovarian antar unsur penggangu regresi adalah nol, (7) jumlah observasi harus lebih besar dari jumlah parameter yang yang diobservasi, (8) nilai peubah bebas bervariasi, (9) Spesifikasi model harus benar dan (10) tidak ada multikolinieritas (Manurung, *et al* (2005) *dalam* Naftali (2007)).

Akan tetapi, yang menjadi permasalahan pada regresi linier berganda adalah terjadinya multikolinieritas. Multikolinieritas menyebabkan penduga yang dihasilkan dari metode kuadrat terkecil tidak bersifat *BLUE*. Penduga yang dihasilkan masih tetap tak bias dan konsisten, tetapi tidak efisien sehingga varian dari parameter regresi menjadi tidak minimum.

Berdasarkan hal tersebut maka penulis tertarik untuk mempelajari masalah multikolinieritas dan metode dalam menanggulangi masalah multikolinieritas ini, yaitu dengan menggunakan metode Regresi Ridge serta memberikan aplikasinya.

Metode Regresi Ridge ini merupakan salah satu dari metode regresi bias yang mempunyai tujuan sebagai solusi untuk masalah multikolinieritas. Regresi Ridge mngatasi masalah multikolinieritas ini dengan menentukan penduga yang bias tetapi mempunyai varian yang lebih kecil dari varian penduga regresi linier berganda.

2. LANDASAN TEORI

2.1 Regresi Linier Berganda

Regresi linier berganda adalah perluasan langsung dari regresi linier sederhana. Perluasannya terlihat dari banyaknya peubah bebas pada model regresi tersebut. Pada model regresi linier sederhana hanya ada satu peubah bebas (X), sedangkan pada regresi linier berganda ada lebih dari satu peubah bebas (Paulson, 2007). Model regesi linier berganda tersebut dapat dilihat pada persamaan (1.2).

Model regresi linier berganda pada (1.2) dapat juga ditulis dalam bentuk matriks, yaitu

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

(Montgomery, 1976).

2.1.1 Asumsi Regresi Linier Berganda

Berdasarkan model regresi linier berganda pada (2.1) maka asumsi untuk ε_i pada dasarnya sama dengan asumsi untuk regresi linier sederhana, yaitu

1. $E(\varepsilon_i) = 0$ untuk $i = 1,2,...,n$ atau ini sama dengan
   $E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$
2. $\text{var}(\varepsilon_i) = \sigma^2$ untuk $i = 1,2,...,n$
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ untuk setiap $i \neq j$

2.1.2 Metode Kuadrat Terkecil

Metode kuadrat terkecil merupakan salah satu metode yang paling banyak digunakan untuk menduga parameter-parameter regresi. Pada model regresi linier berganda juga digunakan metode kuadrat terkecil untuk mencari penduga $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ dari parameter $\beta_0, \beta_1, ..., \beta_k$. Biasanya penduga kuadrat terkecil ini diperoleh dengan meminimumkan jumlah kuadrat galat.

$$JKG = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \tag{2.2}$$

Sehingga diperoleh penduga kuadarat terkecil dari $\boldsymbol{\beta}$ sebagai berikut

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'Y} \tag{2.3}$$

(Montgomery, 1976).

2.1.3 Sifat Penduga Kuadrat Terkecil

Menurut Sembiring (2003) kuadrat terkecil memiliki beberapa sifat yang baik. Untuk menyelidiki sifatnya, pandang kembali model umum regresi linier pada persamaan (2.1). Disini dianggap bahwa ε bebas satu sama lain dan $E(\varepsilon)=0$, $\mathrm{var}(\varepsilon)=\sigma^2$. Dengan demikian maka $\boldsymbol{E}(\boldsymbol{Y})=\boldsymbol{X\beta}$ dan $\mathrm{cov}(\boldsymbol{Y})=\sigma^2\boldsymbol{I}$. Jadi sifat penduga kuadrat terkecil adalah

1. Takbias

   Jika $\boldsymbol{E}(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$ maka $\hat{\boldsymbol{\beta}}$ adalah penduga takbias untuk $\boldsymbol{\beta}$.
2. Varian Minimum

   Jika $\mathrm{cov}(\boldsymbol{Y})=\sigma^2\boldsymbol{I}$ maka matriks kovarian untuk $\hat{\boldsymbol{\beta}}$ diberikan oleh $\sigma^2(\boldsymbol{X'X})^{-1}$. Jika $\boldsymbol{E}(\boldsymbol{Y})=\boldsymbol{X\beta}$ dan $\mathrm{cov}(\boldsymbol{Y})=\sigma^2\boldsymbol{I}$ maka penduga kuadrat terkecil $\hat{\beta}$ mempunyai varian minimum diantara semua penduga tak bias linier (Rencher dan Schaalje, 2008).

2.2 Multikolinieritas

Istilah multikolinieritas pertama kali diperkenalkan oleh Ragnar Frisch pada tahun 1934, yang menyatakan bahwa multikolinieritas terjadi jika adanya hubungan linier yang sempurna *(perfect)* atau pasti *(exact)* diantara beberapa atau semua peubah bebas dari model regresi berganda (Rahardiantoro, 2008).

Multikolinieritas menyebabkan $\boldsymbol{X'X}$ -singular sehingga $(\boldsymbol{X'X})^{-1}$-nya tidak ada. Hal ini berarti bahwa satu atau lebih kolom merupakan kombinasi linier dari kolom-kolom lainnya sehingga $|\boldsymbol{X'X}|=0$. Akibatnya nilai dugaan yang diperoleh tidak unik atau tunggal lagi (Draper dan Smith, 1992).

Menurut Montgomery dan Hines (1990) *dalam* Rahardiantoro (2008) multikolinieritas dapat mengakibatkan parameter regresi yang dihasilkan oleh analisis regresi berganda menjadi sangat lemah atau tidak dapat memberikan hasil analisis yang mewakili sifat atau pengaruh dari peubah bebas yang bersangkutan. Hal ini disebabkan oleh penduga yang dihasilkan dari metode kuadrat terkecil tidak lagi bersifat *BLUE*.

Pendugaan dengan metode kuadrat terkecil akan menghasilkan penduga yang masih tak bias dan konsisten tetapi tidak efisien lagi. Karena varian yang dihasilkan menjadi lebih besar atau tidak minimum.

Menurut Montgomery dan Peck *dalam* Naftali (2007) adanya multikolinieritas dalam regresi linier berganda disebabkan oleh berbagai hal antara lain metode pengumpulan data yang digunakan, kendala model pada populasi yang diamati, spesifikasi model, dan penentuan jumlah peubah bebas yang lebih banyak dari jumlah observasi. Oleh karena itu, dalam suatu penelitian harus benar-benar diperhatikan metode, model, spesifikasi model dan jumlah peubah bebas yang digunakan. Akan tetapi, jika hal itu terjadi, gejala multikolinieritas ini masih dapat dideteksi.

Terdapat beberapa cara untuk mendeteksi adanya multikolinieritas ini baik secara informal maupun formal. Menurut Nachrowi dan Usman (2006) *dalam* Naftali (2007), multikolinieritas dapat dideteksi dengan adanya koefisien determinasi (R^2) yang tinggi dan uji-F yang signifikan. Ini merupakan pendeteksian adanya multikolinieritas secara informal.

Sedangkan secara formal, pertama dapat dilihat pada matriks korelasi (korelasi antar peubah bebas), yaitu jika korelasi antar peubah melebihi 0,50 diduga terdapat gejala multikolinieritas (Gujarati (2003) *dalam* Naftali, 2007). Kedua, dapat dilihat dari nilai *Variance Inflation Factor* (*VIF*), yaitu jika nilai *VIF* kurang dari 10 maka tidak terdapat multikolinieritas. Ketiga, dengan menggunakan bilangan kondisi yang memerlukan nilai eigen matriks $\boldsymbol{X'X}$, yaitu makin kecil nilai eigen maka makin tinggi multikolinieritas antar peubah bebas (Sembiring, 2003).

2.3 Regresi Ridge

Regresi Ridge adalah suatu teknik yang dikembangkan untuk menstabilkan parameter regresi karena adanya multikolinieritas. Metode Regresi Ridge pertama kali dikemukakan oleh A.E. Hoerl pada tahun 1962. Metode ini ditujukan untuk mengatasi kondisi buruk (*ill-conditioned*) yang diakibatkan oleh korelasi yang tinggi antara beberapa peubah bebas di dalam model regresi, sehingga menyebabkan matriks $\boldsymbol{X'X}$ -nya hampir singular, yang pada gilirannya menghasilkan nilai dugaan parameter model regresi yang tidak stabil (Draper dan Smith, 1992).

Regresi Ridge merupakan modifikasi dari metode kuadrat terkecil yang menghasilkan penduga bias dari parameter regresi (Kutner, *et al*., 2005).

2.3.1 Pemusatan dan Penskalaan

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

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

Dalam hal ini yang akan dibakukan (distandarisasi) adalah model regresi linier berganda yang ditunjukkan pada persamaan (1.2). Berikut ini merupakan pembakuan peubah respon Y dan peubah prediktor $X_1, \ldots, X_k$

$$\frac{Y_i - \bar{Y}}{S_Y} \quad \text{, dimana} \quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}{n-1}} \tag{2.4}$$

$$\frac{X_{ij} - \bar{X}_j}{S_{X_j}} \quad \text{, dimana} \quad S_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2}{n-1}} \qquad j = 1,2,\ldots,k \tag{2.5}$$

Keterangan :

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

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

S_Y = standar deviasi dari Y

S_{X_j} = standar deviasi dari X_j

Seperti penjelasan sebelumnya bahwa transformasi korelasi merupakan fungsi sederhana dari pembakuan peubah pada persamaan (2.4) dan (2.5). Sehingga melalui transformasi korelasi diperoleh

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{S_Y}\right) \tag{2.6}$$

$$X_{ij}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ij} - \bar{X}_j}{S_{X_j}}\right) \qquad , j = 1,2,\ldots,k \tag{2.7}$$

Berdasarkan transformasi peubah Y_i^* dan X_{ij}^* yang didefenisikan dengan transformasi korelasi pada persamaan (2.6) dan (2.7) diatas diperoleh model regresi sebagai berikut

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \cdots + \beta_k^* X_{ik}^* + \varepsilon_i \tag{2.8}$$

Persamaan (2.8) diatas disebut sebagai model regresi yang baku (*standardized regression model*).

Diantara parameter $\beta_1^*, \beta_2^*, \ldots, \beta_k^*$ pada model regresi yang baku dengan parameter asli $\beta_1, \beta_2, \ldots, \beta_k$ pada model regresi linier berganda yang biasa terdapat suatu hubungan. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan seperti dibawah ini.

$$\beta_j = \left(\frac{S_Y}{S_{X_j}}\right)\beta_j^* \qquad , j = 1,2,\ldots,k \tag{2.9}$$

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

$$= \quad \bar{Y} - \sum_{j=1}^{k} \beta_j \bar{X}_j \qquad (2.10)$$

(Kutner, *et al*., 2005).

2.3.2 Matriks Korelasi

Model regresi untuk model yang dibakukan pada persamaan (2.8) dapat dibuat dalam bentuk matriks seperti berikut ini

$$\begin{bmatrix} Y_1^* \\ Y_2^* \\ \vdots \\ Y_n^* \end{bmatrix} = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_k^* \end{bmatrix} \begin{bmatrix} X_{11}^* & X_{12}^* & \cdots & X_{1k}^* \\ X_{21}^* & X_{22}^* & \cdots & X_{2k}^* \\ \vdots & \vdots & & \vdots \\ X_{n1}^* & X_{n2}^* & \cdots & X_{nk}^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{bmatrix} \qquad (2.11)$$

Dari persamaan (2.11) diperoleh matriks $\boldsymbol{X}^{*'}\boldsymbol{X}^*$ dan $\boldsymbol{X}^{*'}\boldsymbol{Y}^*$, yaitu

$$\boldsymbol{X}^{*'}\boldsymbol{X}^* = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^{*2} & \sum_{i=1}^{n} X_{i1}^* X_{i2}^* & \cdots & \sum_{i=1}^{n} X_{i1}^* X_{ik}^* \\ \sum_{i=1}^{n} X_{i2}^* X_{i1}^* & \sum_{i=1}^{n} X_{i2}^{*2} & \cdots & \sum_{i=1}^{n} X_{i2}^* X_{ik}^* \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ik}^* X_{i1}^* & \sum_{i=1}^{n} X_{ik}^* X_{i2}^* & \cdots & \sum_{i=1}^{n} X_{ik}^{*2} \end{bmatrix} \qquad (2.12)$$

dan

$$\boldsymbol{X}^{*'}\boldsymbol{Y}^* = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^* Y_i^* \\ \sum_{i=1}^{n} X_{i2}^* Y_i^* \\ \vdots \\ \sum_{i=1}^{n} X_{ik}^* Y_i^* \end{bmatrix} \qquad (2.13)$$

Matriks $\boldsymbol{X}^{*'}\boldsymbol{X}^*$ dan $\boldsymbol{X}^{*'}\boldsymbol{Y}^*$ diatas dapat juga ditulis dalam bentuk matriks korelasi. Hal ini berkaitan dengan penjelasan sebelumnya bahwa peubah X^* dan Y^* diperoleh dari transformasi korelasi peubah X dan Y sehingga kedua bentuk matriks diatas dapat ditulis dalam bentuk matriks korelasi. Matriks pertama yaitu $\boldsymbol{X}^{*'}\boldsymbol{X}^*$ dapat ditulis dalam bentuk matriks korelasi dari peubah X dan dinotasikan dengan $\boldsymbol{r}_{XX}$. Matriks ini didefenisikan sebagai berikut :

$$\boldsymbol{r}_{XX} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & & \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix} \qquad (2.14)$$

Berdasarkan persamaan (2.12) dan (2.14) diperoleh persamaan seperti berikut ini

$$\boldsymbol{X}^{*'}\boldsymbol{X}^* = \boldsymbol{r}_{XX} \qquad (2.15)$$

Sedangkan matriks kedua, yaitu $\boldsymbol{X}^{*'}\boldsymbol{Y}^*$ adalah vektor yang berisi koefisien korelasi sederhana diantara peubah respon Y dan setiap peubah X, yang dinotasikan dengan $\boldsymbol{r_{YX}}$. Matriks korelasinya didefenisikan sebagai berikut :

$$\boldsymbol{r_{YX}} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \\ \vdots \\ r_{Yk} \end{bmatrix} \tag{2.16}$$

Setelah itu sama seperti matriks $\boldsymbol{X}^{*'}\boldsymbol{X}^*$, disini berdasarkan persamaan (2.13) dan (2.16) dapat ditulis bahwa

$$\boldsymbol{X}^{*'}\boldsymbol{Y}^* = \boldsymbol{r_{YX}} \tag{2.17}$$

(Kutner, *et al*., 2005).

2.3.3 Penduga Regresi Ridge

Dalam menduga parameter Regresi Ridge yaitu $\beta_1^*, \beta_2^*, \cdots, \beta_k^*$ dapat diperoleh dengan menggunakan metode *penalized least square*. Kriteria dari metode *penalized least square* ini adalah mengkombinasikan jumlah kuadrat galat (JKG) dengan *penalty* atau kendala tunggal, yaitu

$$\begin{aligned} PLS &= \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \left(\boldsymbol{Y}_{\boldsymbol{i}}^* - \hat{\boldsymbol{Y}}_{\boldsymbol{i}}^*\right)^2 + c\sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}^*\right)^{\gamma} \\ &= \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \left(\boldsymbol{Y}_{\boldsymbol{i}}^* - \boldsymbol{X}_{\boldsymbol{i}}^*\hat{\boldsymbol{\beta}}^*\right)^2 + c\sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}^*\right)^{\gamma} \quad ,\gamma > 0 \end{aligned} \tag{2.18}$$

Menurut ketentuannya untuk Regresi Ridge digunakan $\gamma = 2$, sehingga metode *penalized least square* bentuknya ditulis seperti berikut ini

$$PLS = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \left(\boldsymbol{Y}_{\boldsymbol{i}}^* - \boldsymbol{X}_{\boldsymbol{i}}^*\hat{\boldsymbol{\beta}}^*\right)^2 + c\sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}^*\right)^2 \tag{2.19}$$

(Tutz dan Ulbricht, 2006).

Untuk menghasilkan penduga Regresi Ridge dari metode *penalized least square*, caranya sama seperti pada metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat galatnya. Akan tetapi, karena metode *penalized least square* terdiri dari kombinasi jumlah kuadrat galat dan *penalty* maka yang diminimumkan adalah jumlah kuadrat galat dan *penalty*-nya.

Jumlah kuadrat galat yang diminimumkan adalah jumlah kuadrat galat untuk model pada (2.8). Sehingga *penalized least square* dapat ditulis seperti berikut ini

$$PLS = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \left(\boldsymbol{Y}_{\boldsymbol{i}}^* - \hat{\boldsymbol{\beta}}_1^*\boldsymbol{X}_{\boldsymbol{i}1}^* - \hat{\boldsymbol{\beta}}_2^*\boldsymbol{X}_{\boldsymbol{i}2}^* - \cdots - \hat{\boldsymbol{\beta}}_{\boldsymbol{k}}^*\boldsymbol{X}_{\boldsymbol{ik}}^*\right)^2 + c\left[\sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{j}}^*\right)^2\right] \tag{2.20}$$

(Kutner, *et al*., 2005;Walpole dan Myers,1995).

Sehingga diperoleh penduga Regresi Ridge Ridge sebagai berikut

$$\hat{\boldsymbol{\beta}}^* = \left(\boldsymbol{X}^{*'}\boldsymbol{X}^* + c\boldsymbol{I}\right)^{-1}\boldsymbol{X}^{*'}\boldsymbol{Y}^* \tag{2.21}$$

Untuk $c \geq 0$, c menggambarkan jumlah bias dalam penduga, c adalah $0 \leq c \leq 1$. Jika $c = 0$ maka prosedur Ridge kembali ke persamaan normal kuadrat terkecil dan begitu juga penduganya menjadi sama dengan penduga kuadrat terkecil. Tetapi jika $c > 0$, koefisien regresi adalah bias dan cenderung lebih stabil daripada penduga kuadrat terkecil biasa (Kutner, *et al*., 2005).

Menurut Hoerl dan Kennard (1970) tedapat hubungan antara penduga Ridge dengan penduga kuadrat terkecil biasa, yang dituliskan sebagai berikut

$$\hat{\boldsymbol{\beta}}^* = \left(\boldsymbol{I} + c\left(\boldsymbol{X}^{*'}\boldsymbol{X}^*\right)^{-1}\right)^{-1}\hat{\boldsymbol{\beta}} \tag{2.22}$$

Berdasarkan hubungan antara penduga Ridge dan penduga kuadrat terkecil biasa pada persamaan (2.22), dimisalkan bahwa $\boldsymbol{Z} = \left(\boldsymbol{I} + c\left(\boldsymbol{X}^{*'}\boldsymbol{X}^*\right)^{-1}\right)^{-1}$. Sehingga persamaan (2.22) dapat ditulis sebagai $\hat{\boldsymbol{\beta}}^* = \boldsymbol{Z}\hat{\boldsymbol{\beta}}$. Sehingga varian dan nilai harapan penduga Regresi Ridge diperoleh sebagai berikut

$$Var\left(\hat{\boldsymbol{\beta}}^*\right) = \boldsymbol{\sigma}^2\boldsymbol{Z}\left(\boldsymbol{X}^{*'}\boldsymbol{X}^*\right)^{-1}\boldsymbol{Z}' \tag{2.23}$$

$$E\left(\hat{\boldsymbol{\beta}}^*\right) = \boldsymbol{Z\beta} \tag{2.24}$$

Berdasarkan persamaan (2.24) berarti bahwa penduga Regresi Ridge bukan merupakan penduga takbias dari $\boldsymbol{\beta}^*$. Penduga Regresi Ridge merupakan penduga bias.

Oleh karena, Regresi Ridge merupakan penduga bias maka kuadrat tengah galatnya didasarkan pada penduga yang bias sehingga kuadrat tengah galatnya terdiri dari varian bagi unsur-unsur penduga Ridge dan bias kuadrat Ridge, seperti dijelaskan berikut ini

$$KTG = \boldsymbol{\sigma}^2 trace\boldsymbol{Z}\left(\boldsymbol{X}^{*'}\boldsymbol{X}^*\right)^{-1}\boldsymbol{Z}' + \left[\boldsymbol{\beta}'(\boldsymbol{Z}-\boldsymbol{\beta})'(\boldsymbol{Z}-\boldsymbol{\beta})\boldsymbol{\beta}\right] \tag{2.25}$$

2.3.4 Analisis varian

Tabel 1. Tabel ANAVA untuk Regresi Ridge

Sumber Keragaman	Jumlah Kuadrat (JK)	Derajat bebas (db)	Kuadrat Tengah (KT)
Regresi	$\hat{\boldsymbol{\beta}}^{*'}\boldsymbol{X}^{*'}\boldsymbol{Y}^* - \left(\frac{1}{n}\right)\boldsymbol{Y}^{*'}\boldsymbol{JY}^*$	k	$\frac{JKR}{k}$
Galat	$\boldsymbol{Y}^{*'}\boldsymbol{Y}^* - \hat{\boldsymbol{\beta}}^{*'}\boldsymbol{X}^{*'}\boldsymbol{Y}^*$	$n-k-1$	$\frac{JKG}{n-k-1}$
TOTAL	$\boldsymbol{Y}^{*'}\boldsymbol{Y}^* - \left(\frac{1}{n}\right)\boldsymbol{Y}^{*'}\boldsymbol{JY}^*$	$n-1$	

Sumber : Paulson, D.S. 2007. Handbook of Regression and Modeling

Keterangan :

$\boldsymbol{J}$ adalah matriks kuadrat $n \times n$ yang didefenisikan sebagai berikut $\begin{bmatrix} 1 & \cdots & \cdots & 1 \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ 1 & \cdots & \cdots & 1 \end{bmatrix}$

2.3.5 Pengujian Hipotesis

Langkah-langkah pengujiannya sama seperti pada model regresi linier berganda, yaitu sebagai berikut :

1) Menentukan uji hipotesis, yang mana selalu menggunakan uji dua arah.

$$H_0 \quad : \quad \beta_1^* = \beta_2^* = \cdots = \beta_k^* = 0 \tag{2.26}$$

$H_1 \quad : \quad \beta_j^* \neq 0,\ j = 1,2,\cdots,k$ (tidak semua $\beta_j^* = 0$)

2) Menentukan taraf signifikan (α) dan ukuran sampel (*n*).
3) Memilih dan menuliskan uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-*F*.

$$F_{hitung} = \frac{KTR}{KTG} \tag{2.27}$$

4) Menentukan aturan pengambilan keputusan untuk F_{tabel} atau $F_{(1-\alpha;k,(n-k-1))}$.

   Jika $F_{hitung} > F_{tabel}$ untuk derajat bebas k dan $n-k-1$ maka hipotesis H_0 ditolak pada taraf α dan berarti juga bahwa H_1 diterima.

   Jika $F_{hitung} \leq F_{tabel}$ maka H_0 diterima pada taraf α.
5) Kemudian hitung nilai F_{hitung}.
6) Tuliskan kesimpulan ketika F_{hitung} dibandingkan dengan F_{tabel}.

2.3.6 Koefisien Determinasi

Koefisien determinasi merupakan salah satu patokan yang biasa digunakan untuk melihat apakah suatu model regresi yang dicocokan belum atau sudah memadai, yang dinotasikan dengan R^2. Koefisien determinasi ini hanya menunjukkan ukuran proporsi variasi total dalam respon Y yang diterangkan oleh model yang dicocokan (Walpole dan Myers, 1995).

Koefisien determinasi untuk regresi ridge dapat didefenisikan sama dengan koefisien determinasi untuk kuadrat terkecil, yaitu

$$R^2 = \frac{JKR}{JKT} \tag{2.28}$$

(Kutner, *et al.,* 2005).

3. TELADAN PENERAPAN

Pada teladan penerapan ini, data yang digunakan adalah data simulasi yang dibangkitkan dengan program Microsoft Excel. Data simulasi yang dibangkitkan dengan program Microsoft Excel ini kemudian akan diuji multikolinieritasnya dengan menggunakan V*ariance Inflation Factor (VIF)*. Setelah diketahui bahwa data simulasi yang dibangkitkan ini terdapat multikolinieritas, maka langkah selanjutnya adalah mengatasi multikolinieritas ini dengan Regresi Ridge.

Simulasi yang dilakukan pada penelitian ini sebanyak 1000 simulasi. Berdasarkan setiap simulasi inilah nanti akan terlihat bahwa Regresi Ridge dapat mengatasi multikolinieritas dengan menghasilkan nilai kuadrat tengah galat yang lebih kecil daripada yang dihasilkan oleh metode kuadrat terkecil.

3.1 Hasil

Berikut ini merupakan salah satu contoh hasil simulasi data dengan program Microsoft Excel. Simulasi data yang dilakukan adalah untuk model regresi linier berganda yang terdiri dari 5 peubah bebas (X_1, X_2, X_3, X_4, X_5) dan peubah tak bebas (Y) dengan banyaknya pengamatan 30. Secara umum, model regresi linier berganda untuk kasus diatas dapat ditulis sebagai berikut

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \varepsilon_i \quad ; i = 1,2,\ldots,30 \tag{3.1}$$

Tabel 2. Hasil Simulasi Data

No	Simulation Data						
	X_0	X_1	X_2	X_3	X_4	X_5	Y
1	1	0.889	11.614	11.711	12.459	13.746	248.898
2	1	1.830	22.067	23.797	25.827	27.632	480.477
3	1	4.892	59.402	63.814	69.091	74.050	1253.516
4	1	0.726	9.255	9.919	10.643	11.521	211.594
5	1	2.336	28.374	30.863	33.631	35.758	616.363
6	1	3.323	40.179	43.253	46.981	50.748	860.575
7	1	0.987	12.507	12.965	14.688	15.694	279.023
8	1	1.169	14.104	15.737	16.895	18.128	320.879
9	1	2.382	29.248	31.798	33.548	36.669	628.400
10	1	4.561	55.164	59.369	64.157	68.424	1164.252
11	1	2.698	32.410	35.632	38.325	41.384	705.337
12	1	2.547	30.568	33.913	36.458	38.937	668.414
13	1	2.390	28.830	31.264	33.991	36.098	623.380
14	1	1.317	16.316	17.708	19.267	20.480	361.633
15	1	1.610	19.949	21.041	22.752	24.596	429.559
16	1	2.184	26.406	29.251	30.906	33.323	575.223
17	1	2.605	31.288	34.219	37.240	39.222	677.550
18	1	1.128	14.404	15.351	16.157	17.525	312.973
19	1	2.231	26.998	29.541	31.971	34.253	589.019
20	1	4.275	51.809	56.010	60.766	64.595	1099.546
21	1	0.077	1.012	1.494	1.558	1.196	44.193
22	1	0.882	10.736	11.657	12.397	13.309	242.977
23	1	1.596	19.596	21.163	23.277	24.743	432.382
24	1	3.164	38.068	41.291	44.375	47.924	816.005
25	1	2.652	32.779	35.188	37.644	40.753	697.350
26	1	2.544	30.566	33.091	35.932	38.424	660.258
27	1	2.259	27.116	30.268	31.930	33.934	590.798
28	1	3.494	42.358	45.957	49.633	53.086	905.327
29	1	4.750	57.086	62.072	67.247	71.292	1213.270
30	1	3.371	41.405	44.241	47.256	51.552	874.306

Tabel 3. Beberapa Hasil Perhitungan Nilai R^2 dan *VIF*

No	R^2	*VIF*
1	0.97640	42.36465
2	0.97160	35.20917
3	0.97676	43.03044
4	0.97709	43.64217
5	0.97847	46.44601
6	0.97518	40.29362
7	0.97403	38.50679
8	0.97725	43.96312
9	0.97357	37.84049
10	0.97038	33.76264

Tabel 4. Hasil Perhitungan Penduga Regresi Ridge

No	c	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_3^*$	$\hat{\beta}_4^*$	$\hat{\beta}_5^*$
1	$c = 0.01$	0.19919	0.19642	0.19878	0.20116	0.20253
2	$c = 0.02$	0.19901	0.19759	0.19879	0.20000	0.20069
3	$c = 0.03$	0.19869	0.19773	0.19854	0.19935	0.19981
4	$c = 0.04$	0.19833	0.19760	0.19821	0.19882	0.19917
5	$c = 0.05$	0.19796	0.19576	0.19601	0.19625	0.19639
6	$c = 0.1$	0.19606	0.19576	0.19601	0.19625	0.19639
7	$c = 0.2$	0.19231	0.19215	0.19228	0.19240	0.19247
8	$c = 0.3$	0.18869	0.18857	0.18866	0.18874	0.18879
9	$c = 0.4$	0.18520	0.18511	0.18517	0.18523	0.18527
10	$c = 0.5$	0.18183	0.18176	0.18181	0.18186	0.18189
11	$c = 0.6$	0.17859	0.17852	0.17857	0.17861	0.17863
12	$c = 0.7$	0.17545	0.17539	0.17544	0.17547	0.17549
13	$c = 0.8$	0.17243	0.17238	0.17241	0.17244	0.17246
14	$c = 0.9$	0.16951	0.16946	0.16949	0.16952	0.16953
15	$c = 1$	0.16668	0.16664	0.16667	0.16669	0.16670

3.2 Pembahasan

Salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinieritas pada regresi linier berganda adalah *Variance Inflation Factor (VIF)*. Adanya multikolinieritas dinilai dari nilai *VIF* yang dihasilkan. Nilai *VIF* yang melebihi 10 akan menunjukkan adanya multikolinieritas dan sebaliknya.

Besarnya nilai *VIF* ini bergantung pada nilai koefisien determinasi (R^2) yang dihasilkan. Jika nilai *VIF* melebihi 10 maka koefisien determinasi bernilai lebih besar dari 0.9. Hal ini menunjukkan adanya pengaruh nilai R^2 terhadap nilai *VIF* yang dihasilkan, yaitu semakin besar nilai R^2 maka semakin besar pula nilai *VIF* yang dihasilkan.

Berdasarkan hasil perhitungan uji multikolinieritas dengan menggunakan *VIF* dan perhitungan R^2 terlihat bahwa semua data hasil simulasi, yaitu sebanyak 1000 simulasi data mempunyai nilai *VIF* yang lebih besar dari 30 dan nilai R^2 yang lebih besar dari 0.9. Ini berarti bahwa terdapat multikolinieritas pada seluruh data simulasi tersebut.

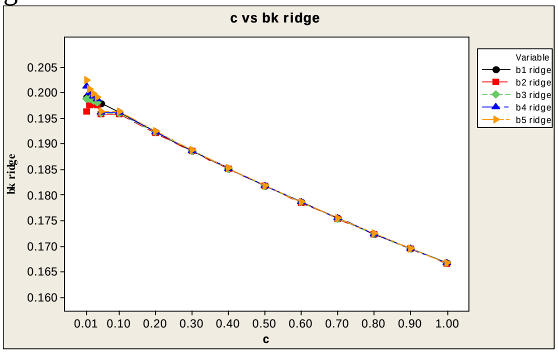
Masalah multikolinieritas pada kasus ini dapat diatasi dengan menggunakan Regresi Ridge. Regresi Ridge dapat menghasilkan penduga yang lebih baik daripada penduga yang dihasilkan kuadrat terkecil. Penduga yang dihasilkan oleh Regresi Ridge merupakan penduga. Hal ini disebabkan karena terdapat penambahan nilai c pada perhitungan Regresi Ridge (persamaan 2.21).

Untuk setiap nilai c yang digunakan, penduga Regresi Ridge yang dihasilkan bergerak menuju ke nol. Jika dibuat grafik antara nilai c dengan penduga Ridge yang dihasilkan persamaan (2.21), maka akan terlihat lebih jelas penduga Ridge ($\hat{\beta}_k^*$) yang dihasilkan bergerak menuju ke nol. Gambar 1. dibawah ini menunjukkan kondisi tersebut.

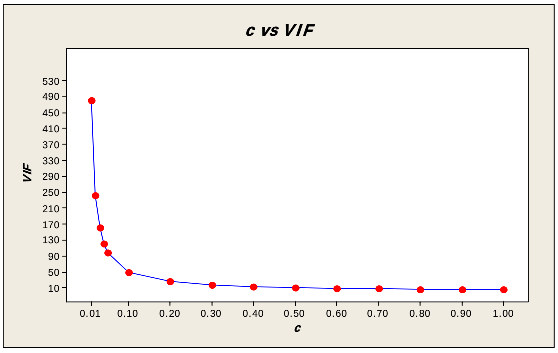
Selain Regresi Ridge dapat menghasilkan penduga yang bias dengan adanya penambahan nilai c. Regresi Ridge juga menghasilkan varian yang lebih minimum daripada varian yang dihasilkan kuadrat terkecil. Varian yang minimum ini dapat dilihat dari nilai kuadrat tengah galat (KTG) yang dihasilkan.

Pada kasus untuk 1000 simulasi data yang dilakukan, Regresi Ridge menghasilkan KTG yang lebih minimum daripada kuadrat terkecil. Hal ini disebabkan karena terdapat nilai c yang dipilih untuk meminimumkan kuadrat tengah galat.

Peningkatan nilai *c* yang digunakan juga mengakibatkan penurunan nilai *VIF* dan R^2 yang dihasilkan Regresi Ridge. Penurunan nilai *VIF* terhadap kenaikan nilai c tersebut digambarkan pada grafik berikut ini :



Gambar 1. Grafik Antara *c vs* b_k Ridge



Gambar 2. Grafik Antara *c vs VIF*

Pada Gambar 2. terlihat bahwa nilai *VIF* yang dihasilkan semakin kecil dengan semakin meningkatnya nilai *c* yang digunakan.

Seperti yang dijelaskan sebelumnya bahwa nilai *VIF* ini digunakan untuk mengukur multikolinieritas. Oleh karena itu, pada kasus ini Regresi Ridge dapat mengatasi multikolinieritas dimulai pada saat $c = 0.6$. Karena dimulai pada saat $c = 0.6$, *VIF*nya bernilai kurang dari 10.

Berdasarkan uraian diatas Regresi Ridge ini dapat mengatasi multikolinieritas dengan menghasilkan penduga yang bias tetapi mempunyai varian yang minimum daripada kuadrat terkecil yang dapat dilihat dari nilai kuadrat tengah galat yang dihasilkan.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil simulasi data dan pembahasan sebelumnya, maka dapat diambil kesimpulan sebagai berikut :

1. Multikolinieritas terjadi ketika dua atau lebih peubah bebas saling berkorelasi secara signifikan satu sama lain.
2. Adanya multikolinieritas dapat dilihat dari nilai R^2 dan *VIF* yang dihasilkan dari data tersebut , yaitu semakin besar nilai R^2 yang dihasilkan maka semakin besar pula nilai *VIF*nya.
3. Regresi Ridge merupakan metode untuk mengatasi multikolinieritas dengan menghasilkan penduga yang bias tetapi mempunyai varian yang minimum dari varian kuadrat terkecil.
4. Regresi Ridge menghasilkan penduga yang bias karena adanya penambahan nilai *c* pada perhitungan Regresi Ridge.
5. Nilai kuadrat tengah galat (KTG) menjadi ukuran bahwa varian yang dihasilkan minimum. KTG yang minimum menunjukkan juga bahwa variannya minimum.

4.2 Saran

1. Untuk mendapatkan nilai penduga Regresi Ridge dan nilai Kuadrat Tengah Galat (KTG) yang terbaik, maka perlu dilakukan kajian lebih lanjut mengenai pemilihan nilai c sehingga Regresi Ridge dapat benar-benar menjadi solusi yang baik dalam mengatasi masalah multikolinieritas.
2. Agar masalah multikolinieritas ini dapat teratasi dengan lebih tepat, maka perlu dilakukan kajian terhadap metode-metode lain yang dapat juga mengatasi masalah multikolinieritas, seperti *Principal Component Regression (PCR)*, *Partial Least Square (PLS)* dan lain-lain.

DAFTAR PUSTAKA

Draper, N dan H. Smith. 1992. *Analisis Regresi Terapan, Terjemahan Edisi Kedua*. PT. Gramedia Pustaka Utama. Jakarta.

Hoerl, A.E. dan R.W. Kennard. 1970. *Ridge Regression : Biased Estimation for Nonorthogonal Problems*. http://statgen.ucr.edu/file/STAT288/hoerl70a.pdf
Kutner, M.H., *et al.* 2005. *Applied Linear Statistical Models, Fifth Edition.* McGraw-Hill. New York.
Montgomery, D.C. 1976. *Design and Analysis of Experiments*. Jhon Wiley & Sons. New York.
Naftali, Y. 2007. *Regresi dan Multikolinieritas dalam Regresi*. http://yohanli.wordpress.com/category/science/
Paulson, D.S. 2007. *Handbook of Regression and Modeling*. Chapman & Hall/CRC. United State of America.
Rahardiantoro, D. 2008. *Principal Component Analysis (PCA) sebagai Metode Jitu Untuk Mengatasi Masalah Multikolinearitas*. http://dickyrahardi.wordpress.com/2006/2/9/principal-component-analysis-pca-sebagai-metode-jitu-untuk-mengatasi multikolinieritas
Rencher, A.C. dan G.B. Schaalje. 2008. *Linier Models in Statistics, Second Edition*. Jhon Wiley & Sons. New York.
Sembiring, R.K. 2003. *Analisis Regresi*. ITB Bandung. Bandung.
Tutz, G. dan J. Ulbricht. 2006. *Penalized Regression with Correlation Based Penalty*. http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper486.pdf
Walpole, R.E dan R.H. Myers 1995. *Ilmu Peluang dan Statistik untuk Insyinyur dan Ilmuwan, Terjemahan Edisi Keempat*. ITB. Bandung.
Weisberg, S. 1980. *Applied Linear Regression*. John Wiley & Sons. New York.