PENYELESAIAN MULTIKOLINEARITAS MELALUI METODE RIDGE REGRESSION

Oleh :

SOEMARTINI

JURUSAN STATISTIKA

FAKULTAS MATEMATIKA dan ILMU PENGETAHUAN ALAM

UNIVERSITAS PADJADJARAN

JATINANGOR

2008

DAFTAR ISI

BAB I
PENDAHULUAN

Analisis regresi adalah salah satu metode statistika yang sering digunakan untuk mengetahui sejauh mana ketergantungan atau hubungan sebuah variabel tak bebas (*regressand*) dengan sebuah atau lebih variabel bebas (*regressor*). Bila dalam analisisnya hanya melibatkan sebuah variabel bebas ,maka analisis yang digunakan adalah Analisis Regresi Linier Sederhana. Sedangkan bila dalam analisisnya melibatkan dua atau lebih variabel bebas , maka analisis yang digunakan adalah Analisis Linier Berganda.

Dalam kehidupan sehari-hari banyak permasalahan yang dapat dipecahkan dengan Analisis Regresi Linier Berganda, salah satu contohnya adalah mengenai tingkat konsumsi yang diduga dipengaruhi oleh pendapatan dan kekayaan. Dalam hal ini, tingkat konsumsi bertindak sebagai *regressand* serta pendapatan dan kekayaan bertindak sebagai *regressor*.

Di dalam analisis linier ganda yang mempunyai banyak variable *regressor*, sering timbul masalah karena terjadinya hubungan antara dua atau lebih variable *regressor*-nya. Variabel *regressor* yang saling berkorelasi disebut kolinieritas ganda (*multicollinearity).* Gejala ini menimbulkan masalah dalam pemodelan regresi. Korelasi yang sangat tinggi akan menghasilkan penaksir yang berbias, tidak stabil dan mungkin jauh dari nilai sasaran (*Gonst and Mason,* 1977). Metode kuadrat terkecil akan memberikan efek dari kolinieritas yaitu tingginya nilai koefisien determinasi tetapi tidak diikuti dengan hasil uji hipotesis yang signifikan.

Satu dari asumsi model regresi linier adalah bahwa tidak terdapat multikolinearitas diantara variabel *regressor* yang termasuk dalam model. Multikolinearitas terjadi apabila terdapat hubungan atau korelasi diantara beberapa atau seluruh variabel *regressor*. Masalah yang akan dibahas dalam makalah ini adalah penyelesaian masalah multikolinieritas antara variable-variabel *regressor*.

Salah satu cara untuk mendapatkan koefisien regresi pada persamaan regresi linier berganda adalah melalui metode kuadrat terkecil. Metode ini menghasilkan penaksir terbaik (tak bias dan bervarians minimum) jika saja tidak ada korelasi antar variable *regressor*. Namun jika hal itu terjadi, maka salah satu cara untuk mengatasi masalah tersebut adalah melalui metode *Ridge regression.* Pada dasarnya metode ini juga merupakan metode kuadrat terkecil. Perbedaannya adalah bahwa pada metode *ridge regression*, nilai variabel *regressor*nya ditransformasikan dahulu melalui prosedur *centering and rescaling*. Kemudian pada diagonal utama matriks korelasi variable *regressor* ditambahkan *Ridge Parameter* θ dimana nilainya antara 0 dan 1 (*Neter et al.*, 1990).

Metode yang dibahas dimaksudkan untuk mengatasi masalah dalam regresi linier ganda, yaitu terjadinya multi kolinieritas.

Metode *ridge regression* dapat digunakan dengan asumsi matriks korelasi dari variable *regressor* dapat diinverskan. Akibatnya nilai dugaan koefisien regresi dan variable *regressand* mudah didapat. Nilai dugaan variable *regressand* sangat ditentukan oleh besar kecilnya nilai *Ridge Parameter* θ.

BAB II
TINJAUAN PUSTAKA

2 .1 Metode Penaksiran Koefisien Regresi

Dalam menentukan koefisien untuk suatu persamaan regresi digunakan berbagai metode, diantaranya metode kuadrat terkecil dan metode kemungkinan maksimum. Pada makalah ini hanya akan dibahas mengenai metode kuadrat terkecil.

Metode kuadrat terkecil adalah suatu metode yang dipergunakan untuk menaksir parameter suatu persamaan regresi atau koefisien regresi dengan jalan meminimumkan jumlah kuadrat residunya.

Bentuk persamaan regresinya secara umum adalah :

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

dengan,

i	= 1,2,3,…,k
β_0	= *intercept*
$\beta_1, \beta_2, ..., \beta_3$	= *slope*
ε_i	= faktor residual

Jika persamaan di atas ditulis dalam bentuk matriks, maka akan menjadi :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

untuk

$\underset{\rightarrow}{Y}$ = vektor kolom n x 1

$\underset{\sim}{X}$ = matriks n x (k+1)

$\underset{\rightarrow}{\beta}$ = vektor kolom (k+1) x 1

$\underset{\rightarrow}{\varepsilon}$ = vektor n x 1

Penyajian matriks model regresi linear dengan k variabel, yaitu :

$$\underset{\rightarrow}{Y} = \underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\rightarrow}{\varepsilon}$$

Dengan asumsi $E(\underset{\to}{\varepsilon}) = 0$ dan $E(\underset{\to}{\varepsilon}\,\underset{\to}{\varepsilon}') = \underset{\sim}{I}\,\sigma^2$, bisa dihitung jumlah kuadrat $\underset{\to}{\varepsilon}$ yaitu :

$$\begin{aligned}\underset{\to}{\varepsilon}'\underset{\to}{\varepsilon} &= (\underset{\to}{Y} - \underset{\sim}{X}\,\underset{\to}{\beta})'(\underset{\to}{Y} - \underset{\sim}{X}\,\underset{\to}{\beta}) \\ &= \underset{\to}{Y}'\underset{\to}{Y} - \underset{\to}{Y}'\underset{\sim}{X}\,\underset{\to}{\beta} - \underset{\to}{\beta}'\underset{\sim}{X}'\underset{\to}{Y} + \underset{\to}{\beta}'\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta}\end{aligned}$$

karena

$$\underset{\to}{\beta}'\underset{\sim}{X}'\underset{\to}{Y} = \text{skalar}$$

maka

$$\underset{\to}{\beta}'\underset{\sim}{X}'\underset{\to}{Y} = (\underset{\to}{\beta}'\underset{\sim}{X}'\underset{\to}{Y})'$$

sehingga

$$\underset{\to}{\varepsilon}'\underset{\to}{\varepsilon} = \underset{\to}{Y}'\underset{\to}{Y} - 2\underset{\to}{\beta}'\underset{\sim}{X}'\underset{\to}{Y} + \underset{\to}{\beta}'\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta}$$

Turunan pertama $\underset{\to}{\varepsilon}'\underset{\to}{\varepsilon}$ terhadap $\underset{\to}{\beta}$ adalah :

$$\frac{d\underset{\to}{\varepsilon}'\underset{\to}{\varepsilon}}{d\underset{\to}{\beta}} = -2\underset{\sim}{X}'\underset{\to}{Y} + 2\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta}$$

Jika turunan pertama disamakan dengan nol, maka diperoleh :

$$\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta} = \underset{\sim}{X}'\underset{\to}{Y}$$

Jika $\underset{\to}{b}$ merupakan jawab persamaan normal, akan didapat :

$$\underset{\to}{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{Y}$$

Untuk membuktikan bahwa $\underset{\to}{b}$ merupakan penaksir tak bias, dapat diperlihatkan sebagai berikut :

$$\begin{aligned}E(\underset{\to}{b}) &= E\,[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{Y}] \\ &= E\,[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'(\underset{\sim}{X}\,\underset{\to}{\beta} + \underset{\to}{\varepsilon}) \\ &= E\,[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta}\} + \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}\}] \\ &= E\,[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta}] + E\,[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}] \\ &= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,E(\underset{\to}{\beta}) + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'E(\underset{\to}{\varepsilon}) \\ &= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\to}{\beta} \\ &= \underset{\sim}{I}\,\underset{\to}{\beta} \\ &= \underset{\to}{\beta}\end{aligned}$$

Jadi, $\underset{\to}{b}$ merupakan penaksir tak bias bagi $\underset{\to}{\beta}$.

Varians $\underset{\to}{b}$ dapat dicari sebagai berikut :

$$\begin{aligned}
\text{Var}(\underset{\to}{b}) &= \text{E}\,[\{\underset{\to}{b} - \text{E}(\underset{\to}{b})\}\,\{\underset{\to}{b} - \text{E}(\underset{\to}{b})\}']\\
&= \text{E}\,[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{Y} - \underset{\to}{\beta}\}\,\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{Y} - \underset{\to}{\beta}\}]'\\
&= \text{E}\,[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'(\underset{\sim}{X}\underset{\to}{\beta}+\underset{\to}{\varepsilon}) - \underset{\to}{\beta}\}\,\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'(\underset{\sim}{X}\underset{\to}{\beta}+\underset{\to}{\varepsilon}) - \underset{\to}{\beta}\}'\\
&=\text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\to}{\beta}+(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}-\underset{\to}{\beta}\}\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\to}{\beta}+(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}-\\
&\quad \underset{\to}{\beta}\}]'\\
&= \text{E}\,[\{\underset{\to}{\beta}+(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon} - \underset{\to}{\beta}\}\,\{\underset{\to}{\beta}+(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon} - \underset{\to}{\beta}\}']\\
&= \text{E}\,[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}\}\,\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}\}']
\end{aligned}$$

Karena $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ adalah matriks simetri, maka $(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}$,sehingga

$$\begin{aligned}
\text{var}(\underset{\to}{b}) &= \text{E}\,[(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\to}{\varepsilon}\,\underset{\to}{\varepsilon}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}]\\
&= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\,\text{E}(\underset{\to}{\varepsilon}\,\underset{\to}{\varepsilon}')\,\underset{\sim}{X}\,(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\\
&= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2\\
&= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2
\end{aligned}$$

Selanjutnya akan diperlihatkan bahwa $\underset{\to}{b}$ mempunyai varians minimum.Untuk itu misalkan diambil suatu penduga lain yaitu

$$\underset{\to}{b}^{*} = \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\to}{Y}$$

untuk,

$\underset{\sim}{B}$ adalah matriks (k+1) x n

maka,

$$\begin{aligned}
\text{E}(\underset{\to}{b}^{*}) &= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\to}{Y}]\\
&= \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\text{E}(\underset{\to}{Y})\\
&= \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\text{E}(\underset{\sim}{X}\underset{\to}{\beta} + \underset{\to}{\varepsilon})
\end{aligned}$$

Karena,

$$\text{E}(\underset{\to}{\varepsilon}) = 0$$

maka,

$$\begin{aligned}
\text{E}(\underset{\to}{b}^{*}) &= \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\sim}{X}\,\text{E}(\underset{\to}{\beta})\\
&= \{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\sim}{X}\underset{\to}{\beta}\\
&= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\underset{\to}{\beta} + \underset{\sim}{B}\underset{\sim}{X}\underset{\to}{\beta}
\end{aligned}$$

$$= \underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}$$

$\underset{\rightarrow}{b}^{*}$ merupakan penduga tak bias bagi $\underset{\rightarrow}{\beta}$, bila $\underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} = 0$

sehingga,

$$\underset{\sim}{B}\,\underset{\sim}{X} = 0$$

atau

$$(\underset{\sim}{B}\,\underset{\sim}{X})' = 0$$

sehingga,

` $$\underset{\sim}{X}'\,\underset{\sim}{B}' = 0$$

dengan demikian

$$\text{var}(\underset{\rightarrow}{b}^{*}) = \text{E}[\{\underset{\rightarrow}{b}^{*} - \text{E}(\underset{\rightarrow}{b}^{*})\}\{\underset{\rightarrow}{b}^{*} - \text{E}(\underset{\rightarrow}{b}^{*})\}]'$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\rightarrow}{Y} - (\underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta})][\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}\underset{\rightarrow}{Y} - (\underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta})]'$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}(\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\rightarrow}{\varepsilon}) - \underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}]\,[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}' + \underset{\sim}{B}\}(\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\rightarrow}{\varepsilon}) - (\underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}]'$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + (\underset{\sim}{X}\,\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon} - \underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}\}\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon} - \underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}\}]'$$

$$= \text{E}[\{\underset{\rightarrow}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon} - \underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}\}\{\underset{\rightarrow}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta} + \underset{\rightarrow}{\beta}\,\underset{\rightarrow}{\varepsilon} - \underset{\rightarrow}{\beta} - \underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\rightarrow}{\beta}\}]'$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon}\}\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon}\}]'$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon}\}\{\underset{\rightarrow}{\varepsilon}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} + \underset{\rightarrow}{\varepsilon}'\underset{\sim}{B}'\}]$$

$$= \text{E}[\{(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}'\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}'\underset{\sim}{B}' + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}'\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} + \underset{\sim}{B}\,\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}'\underset{\sim}{B}'\}]$$

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}')\underset{\sim}{B}' + \underset{\sim}{B}\,\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}')\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} + \underset{\sim}{B}\,\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}')\underset{\sim}{B}'$$

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}') + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{B}'\,\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}') + \underset{\sim}{B}\,\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}') + \underset{\sim}{B}\,\underset{\sim}{B}'\text{E}(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}')$$

karena,

$$E(\underset{\rightarrow}{\varepsilon}\,\underset{\rightarrow}{\varepsilon}') = \underset{\sim}{I}\,\sigma^2$$

dan

$$\underset{\sim}{B}\,\underset{\sim}{X} = (\underset{\sim}{B}\,\underset{\sim}{X})' = \underset{\sim}{X}'\,\underset{\sim}{B}' = 0$$

maka

$$\text{Var}(\underset{\rightarrow}{b}^{*}) = (\underset{\sim}{X}'\,\underset{\sim}{X})^{-1}\sigma^2 + \underset{\sim}{B}\,\underset{\sim}{B}'\sigma^2$$

$$= \text{Var}(\underset{\rightarrow}{b}) + \underset{\sim}{B}\,\underset{\sim}{B}'\sigma^2$$

Karena $\underset{\sim}{B}\,\underset{\sim}{B}'$ adalah matriks kuadratik, maka

$$\text{Var}(\underset{\rightarrow}{b}*) \geq \text{Var}(\underset{\rightarrow}{b})$$

Jadi $\underset{\rightarrow}{b}$ yang diperoleh dengan Metode Kuadrat Terkecil merupakan penaksir linier tak bias bervarians minimum (terbaik) untuk $\underset{\rightarrow}{\beta}$.

2.2 Metode *Centering and Rescaling* dan Matriks Korelasi

2.2.1 Metode *Centering and Rescaling*

Dalam persamaan regresi yang memiliki model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Persamaan tersebut di atas dapat dibentuk menjadi :

$$Y_i = \beta_0 + \beta_1 (X_{1i} - \bar{X}) + \beta_1 \bar{X}_1 + \beta_2 (X_{2i} - \bar{X}_2) + \beta_2 \bar{X}_2 + \varepsilon_i$$

$$= (\beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2) + \beta_1 (X_{1i} - \bar{X}) + \beta_2 (X_{2i} - \bar{X}_2) + \varepsilon_i$$

menurut rumus untuk mendapatkan β_0 yaitu :

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

maka berlaku

$$= \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2$$

sehingga

$$Y_i - (\beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2) = \beta_1 (X_{1i} - \bar{X}) + \beta_2 (X_{2i} - \bar{X}_2) + \varepsilon_i$$

$$Y_i - \bar{Y} = \beta_1 (X_{1i} - \bar{X}) + \beta_2 (X_{2i} - \bar{X}_2) + \varepsilon_i$$

Jika $y_i = Y_i - \bar{Y}$

$x_{1i} = X_{1i} - \bar{X}_1$

$x_{21} = X_{2i} - \bar{X}_2$

maka kita dapat persamaan baru yaitu :

$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_I$

Prosedur untuk membentuk persamaan pertama menjadi persamaan terakhir disebut dengan prosedur *centering*. Prosedur ini mengakibatkan hilangnya β_0 (intercept) yang membuat perhitungan untuk mencari model regresi menjadi lebih sederhana.

Bila dari persamaan di atas kita bentuk persamaan :

$Y_i^* = \beta_1 Z_{1i} + \beta_2 Z_{2i} + \varepsilon_i'$

dengan

$$Y_i^* = \frac{y_i}{\sqrt{n-1}\,Sy} = \frac{Y_i - \overline{Y}_i}{\sqrt{n-1}\;\;Sy}$$

$$Z_{1i} = \frac{x_{1i}}{\sqrt{n-1}\,S_1} = \frac{X_{1i} - \overline{X}_1}{\sqrt{n-1}\,S_1}$$

$$Z_{2i} = \frac{x_{2i}}{\sqrt{n-1}\,S_2} = \frac{X_{2i} - \overline{X}_2}{\sqrt{n-1}\,S_2}$$

maka prosedur ini disebut dengan prosedur *Rescaling*. Keseluruhan dari prosedur di atas disebut prosedur *centering and rescaling*.

2.2.2 Matriks Korelasi

Persamaan yang didapat melalui prosedur *Centering and Rescaling* di atas bila dituliskan dalam bentuk matriks adalah :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} z_{11} & z_{21} \\ z_{12} & z_{22} \\ z_{13} & z_{23} \\ \vdots & \vdots \\ z_{1k} & z_{2k} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} z_{11} & z_{12} & z_{13} & \cdots & z_{1k} \\ z_{21} & z_{22} & z_{23} & \cdots & z_{2k} \end{bmatrix} \begin{bmatrix} z_{11} & z_{21} \\ z_{12} & z_{22} \\ z_{13} & z_{23} \\ \vdots & \vdots \\ z_{1k} & z_{2k} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{k} z_{1i}^{2} & \sum_{i=1}^{k} z_{1i} z_{2i} \\ \sum_{i=1}^{k} z_{1i} z_{2i} & \sum_{i=1}^{k} z_{2i}^{2} \end{bmatrix}$$

untuk,

$$\sum_{i=1}^{k} z_{1i}^{2} = \sum \left\{ \frac{x_{1i} - \bar{x}_1}{\sqrt{n-1}\, S_1} \right\}^2$$

$$= \frac{\sum_{i=1}^{k} \left(x_{1i} - \bar{x}_1 \right)^2}{(n-1) S_1^{2}}$$

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

$$= 1$$

Hal ini berlaku juga untuk $\sum_{i=1}^{k} z_{2i}^{2} = 1$

Sedangkan untuk

$$\sum_{i=1}^{k} z_{1i} z_{2i} = \sum_{i=1}^{k} \left\{ \frac{x_{1i} - \bar{x}_1}{\sqrt{n-1} S_1} \right\} \left\{ \frac{x_{2i} - \bar{x}_2}{\sqrt{n-1} S_2} \right\}$$

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

$$= \frac{\sum_{i=1}^{k} \left(x_{1i} - \bar{x}_1 \right)\left(x_{2i} - \bar{x}_2 \right)}{(n-1) \sqrt{\frac{\sum_{i=1}^{k} \left(x_{1i} - \bar{x}_1 \right)^2}{(n-1)}} \sqrt{\frac{\sum_{i=1}^{k} \left(x_{2i} - \bar{x}_{1\backslash 2} \right)^2}{(n-1)}}}$$

$$= \frac{\sum_{i=1}^{k} \left(x_{1i} - \bar{x}_1 \right)\left(x_{2i} - \bar{x}_2 \right)}{\sqrt{\sum_{i=1}^{k} \left(x_{1i} - \bar{x}_1 \right)^2} \sqrt{\sum_{i=1}^{k} \left(x_{2i} - \bar{x}_2 \right)^2}}$$

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

Sehingga matriks korelasi untuk persamaan regresinya adalah :

$\underset{\sim}{Z}' \underset{\sim}{Z} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}$ Matriks $\underset{\sim}{Z}' \underset{\sim}{Z}$ yang diperoleh disebut matriks korelasi.

2.3 Koefisien Determinasi

Koefisien determinasi (*Coefficient of Determination*), R^2 didefinisikan sebagai berikut :

$$R^2 = \frac{JKR}{JKT(terkoreksi)} = \frac{\underline{b}'\underset{\sim}{X}'\underline{Y} - n\overline{Y}^2}{\underline{Y}'\underline{Y} - n\overline{Y}^2} \quad \text{(Neter et al., 1990)}$$

Dengan JKR = Jumlah kuadrat residu regresi

JKT = Jumlah kuadrat total antara JKR dan Jumlah kuadrat galat

$\underline{b}$ = vektor taksiran parameter β

$\underset{\sim}{X}$ = matriks variable *regressor* berukuran ($n \times k$)

$\underline{Y}$ = vektor variable *regressand berukuran (* $n \times 1$ *)*

$\overline{Y}$ =vektor rata-rata variable *regressand*

Koefisien Determinasi (R^2) adalah besaran yang mengukur proporsi variable *regressor* dalam model yang mampu menerangkan jumlah kuadrat total variable *regressand* Y(terkoreksi). R^2 bernilai antara 0 sampai 1. Apabila nilai R^2 semakin besar, ini menunjukkan bahwa ketepatan model semakin besar dalam menerangkan keragaman data.

2.4 Distribusi $t - student$

Untuk mengetahui signifikansi masing-masing individu koefisien regresi bisa dilakukan uji t-student dengan hipotesis sebagai berikut :

$H_0 : \beta_i = 0$ (koefisien regresi tidak berarti)

$H_1 : \beta_i \neq 0$ (koefisien regresi berarti)

Dengan statistik uji : $t_{hitung} = \frac{b_i}{S_i}$

untuk : b_i = koefisien regresi

S_i = galat baku b_k

i = 1,2,3

Dengan kriteria uji, tolak H_0 jika: $t > t_{(n-1;\alpha)}$

$t < -t_{(n-k-1;\,\alpha)}$

$t \geq t_{(n-k-1;\,\alpha/2)}$ atau $t \leq -\,t_{(n-k-1;\,\alpha/2)}$

Untuk menguji kecocokan model regresi linier ganda secara *simultan* atau bersama-sama melalui uji ANAVA dengan bentuk hipotesis:

H_0 : β = 0 (vektor koefisien regresi ganda bernilai nol)

H_1 : β ≠ 0 (vektor koefisien regresi ganda tidak bernilai nol)

Tabel 2.1 ANAVA Regresi

Sumber	dk	JK	RJK	F_{hitung}
Regresi	k	JK (R)	JK(R) / k	(JK(R) / k) / (JK(ε)/(n-k-1))
Galat	n-k-1	JK (ε)	JK(ε)/ (n-k-1)	
Total	n-1	JK (T)		

Dengan :

$$JK\,(R) = \underset{\to}{b}'(\underset{\sim}{X}'\underset{\to}{Y}) - \frac{(\sum Y)^2}{n}$$

$$JK\,(T) = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$$

$$JK\,(\varepsilon) = JK\,(T) - JK\,(R)$$

$$\underset{\to}{b} = \left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\to}{Y}$$

Kriteria uji : Dengan kriteria uji: Tolak H_0 jika $F_{hitung} \geq F_{(k;n-k-1;\alpha/2)}$ atau bisa juga dilihat dari nilai p, tolak Ho jika nilai $p \leq \alpha$.

BAB III

METODE RIDGE REGRESSION

Dalam bab sebelumnya, telah dijelaskan salah satu metode yang dapat digunakan untuk menaksir parameter regresi dari model regresi linier berganda adalah Metode Kuadrat Terkecil. Dugaan parameter koefisien regresi dengan Metode Kuadrat Terkecil adalah

$$\underline{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'Y$$

Dengan membentuk $\underset{\sim}{X}'\underset{\sim}{X}$ menjadi bentuk matriks korelasi, maka kesalahan yang disebabkan pengaruh pembulatan menjadi lebih kecil (*Draper & Smith* ,1981).. Terutama jika variabel *regressor*nya lebih dari dua dan data yang ada besar. Jika $\underset{\sim}{X}'\underset{\sim}{X}$ yang merupakan matriks korelasi adalah matriks identitas maka nilai dugaan variabel *regressand* akan sama dengan nilai sebenarnya. Apabila $\underset{\sim}{X}'\underset{\sim}{X}$ tidak mendekati matriks identitas melainkan menjauhinya, maka dapat dikatakan $\underset{\sim}{X}'\underset{\sim}{X}$ hampir singular (buruk). Kondisi ini disebut sebagai *ill conditioned* (*Draper & Smith* ,1981). Kondisi ini terjadi apabila terdapat korelasi antar variabel *regressor* yang cukup tinggi sehingga menyebabkan determinan $\underset{\sim}{X}'\underset{\sim}{X}$ mendekati nol. Maka antara variabel *regressor* terjadi multikolinieritas ganda tidak sempurna.

Apabila terjadi situasi tersebut, penaksiran parameter koefisien regresi masih mungkin dilakukan dengan metode kuadrat terkecil, tetapi dengan konsekuensi simpangan bakunya menjadi sangat sensitif sekalipun terjadi perubahan yang sangat kecil dalam datanya. Simpangan baku ini cenderung membesar sejalan dengan meningkatnya multikolinieritas.

Apabila terjadi multikolinieritas tidak sempurna pada variabel *regressor* pada diagonal utama $\underset{\sim}{X}'\underset{\sim}{X}$ ditambah bilangan kecil positif θ yang bernilai antara 0 dan 1 (*Hoerl A.E*, 1962). Prosedur ini disebut *Ridge Trace.* Kemudian prosedur tersebut dikembangkan oleh A.E Hoerl dan Robert W Kennard (1970) dan Normon R. Draper dan Harry Smith (1981) dengan mentransformasikan matriks $\underset{\sim}{X}'\underset{\sim}{X}$ menjadi matriks korelasi $\underset{\sim}{Z}'\underset{\sim}{Z}$.

Sehingga dugaan koefisien regresi menjadi :

$$\underline{b}_z (\theta) = (\underset{\sim}{Z}'\underset{\sim}{Z} + \theta \underline{I}_k)^{-1} Z'\underline{Y}$$

dengan :

$\underline{b}_z(\theta)$ = estimator *ridge regression*

θ = *ridge parameter* (bilangan kecil positif terletak antara

0 dan 1)

$\underset{\sim}{Z}$ = matriks n x k yang merupakan hasil transformasi variabel *regressor* melalui metode *centering and rescaling*.

Sehingga nilai dugaan untuk variabel *regressand* menjadi

$$\hat{Y}(\theta) = Z\,\underline{b}_z(\theta)$$

Proses tersebut di atas disebut dengan *Ridge regression.* Analisis *ridge regression* dapat digunakan apabila $\underset{\sim}{Z}'\underset{\sim}{Z}$ tidak singular. Asumsi yang digunakan hanyalah $(\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}$ ada dan tidak sulit mendapatkannya (*Draper & Herzberg*, 1986).

Pemilihan nilai θ sebenarnya diserahkan kepada analis. Untuk memperoleh nilai θ , analis mencobakan nilai θ sampai keadaan stabil.. Ada beberapa metode yang bisa digunakan salah satunya yaitu dengan mencari nilai satistik *Cp Mallows* ($C\theta$) dengan rumus :

$$C\theta = \frac{SS_{res,k}}{\hat{\sigma}^2} - n + 2 + 2tr\left(H_\theta\right)$$

Dengan

$$tr[H\theta] = \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_i + \theta}$$

Keterangan :

$SS_{res,k}$ = Jumlah kuadrat residu dari persamaan *Ridge* Regression

n = banyaknya pengamatan

λ_i = Eigen value dari matriks ($\underset{\sim}{Z}'\underset{\sim}{Z} + \theta \underline{I}_k$)

$Tr\left(H_\theta\right)$ = Trace dari matriks H_θ

$\hat{\sigma}^2$ = penaksir varians metode kuadrat terkecil

Setelah memperoleh nilai Cθ , nilai θ terpilih adalah nilai θ yang dapat meminimumkan nilai Cθ.(*Mayers*, 1990).

Untuk memperoleh Koefisien regresi dalam variabel asal digunakan rumus sebagai berikut :

$$b_i = \left(\frac{S_Y}{S_i} \right) b_i^{'}$$

Dengan : i = 1,2,3

S_Y = Galat baku dari data awal Y

S_i = Galat baku dari data awal X ke-I

$b^{'}_i$ = koefisien regresi setelah melalui metode *ridge regression*

`

BAB IV

CONTOH PEMAKAIAN

4.1 Data dan Permasalahan

Data yang digunakan untuk contoh pemakaian ini adalah data dari Rumah Sakit Sardjito Yogyakarta. Data ini menyangkut tentang jam kerja pegawai rumah sakit (Y) yang diduga bergantung pada rata – rata peningkatan jumlah pasien (X_1), tempat tidur harian yang dipakai perbulan (X_2), dan populasi pasien yang memenuhi syarat pada area rumah sakit, dalam ribuan (X_3).

Tujuan kita disini adalah untuk memperoleh persamaan yang akan digunakan untuk menaksir dan atau mememprediksi tenaga kerja yang diperlukan untuk rumah sakit.

4.2 Metode Regresi Linier Ganda

Langkah-langkah untuk mendapatkan koefisien regresi dengan data awal adalah sebagai berikut :

1. Hitung nilai penaksir parameter β, kemudian hitung galat baku dan hitung t, buat suatu model.
2. Hitung ŷ dan menganalisa table ANAVA.

4.3 Metode Regresi Ridge

Tahapan penaksiran koefisien *ridge regression*.

1. Lakukan transformasi tehadap matriks X dan vektor Y, melalui *centering and rescaling*.
2. Hitung matriks $\underset{\sim}{Z}'\underset{\sim}{Z}$ = r_{xx} = matriks korelasi dari variable bebas, serta hitung $\underset{\sim}{Z}'\underset{\rightarrow}{Y}^*$ =korelasi dari variable bebas terhadap variable tak bebas y.
3. Hitung nilai penaksir parameter $\beta*$ dengan berbagai kemungkinan tetapan bias θ, θ≥0.
4. hitung nilai C_θ dengan berbagai nilai θ.

5. Tentukan nilai θ dengan mempertimbangkan nilai C_{θ}.

   Tentukan koefisien penaksir ridge regression dari nilai θ yang bersesuaian.

6. Hitung nilai $\hat{y}$ dan menganalisa ANAVA.

BAB V

HASIL DAN PEMBAHASAN

5.1 Penaksiran Model Regresi Linier Ganda

Hasil Analisis Regresi dengan menggunakan Metode Kuadrat Terkecil terhadap data pada lampiran 1 tercantum pada tabel nilai penaksir parameter (tabel 5.1).

Pengujian keberartian model regresi ganda yang dilakukan secara *parsial* atau individu , dengan hipotesis

H_0 : $\beta_i = 0$, untuk i=1,2,3 (variabel *regressor* X secara individu tidak berpengaruh secara signifikan terhadap nilai taksiran Y)

H_1 : $\beta_i \neq 0$, untuk i=1,2,3 (variabel *regressor* X secara individu tidak perpengaruh secara signifikan terhadap nilai taksiran Y)

$\alpha = 5\%$

Dengan statistik uji t-student, maka kita peroleh nilai T_{hitung} dari masing-masing variabel X secara individu adalah sebagai berikut

Tabel 5.1. Penaksir Parameter Metode Kuadrat Terkecil

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	p	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-12,414	326,884		-,038	,970		
	X1	-163,950	119,013	-4,748	-1,378	,192	,000	7182,317
	X2	6,230	3,834	5,498	1,625	,128	,000	6920,164
	X3	13,023	6,845	,253	1,903	,079	,094	10,678

a. Dependent Variable: Y

Tabel 5.2. Deskripsi Data

Descriptive Statistics

	Mean	Std. Deviation	N
Y	4978,4800	5560,53359	17
X1	148,2759	161,03858	17
X2	4480,6182	4906,64206	17
X3	106,3176	107,95415	17

Dengan kriteria uji :

Tolak Ho jika $t_{hitung} \leq -t_{(n-2;\alpha/2)}$ atau $t_{hitung} \geq t_{(n-2;\alpha/2)}$, terima dalam hal lainnya.

Kriteria uji ini bisa juga dilihat dari nilai p. Tolak H_0 jika nilai $p \leq \alpha$, terima dalam hal lainnya.

Dari tabel 4.1 diatas diperoleh model regresi sebagai berikut :

$\hat{Y} = -12.414 - 163.950\ X_1 + 6.230\ X_2 + 13.023\ X_3$

Dilihat dari Tabel 5.1 dan 5.2 diatas maka dapat disimpulkan koefisien penaksir tidak bisa ditaksir secara tepat, hal ini ditunjukkan oleh nilai galat baku yang cukup besar dan nilai p yang lebih besar dari α menunjukkan bahwa tidak ada satu pun variabel *regressor* X secara individu yang berpengaruh secara signifikan terhadap nilai taksiran Y. Begitu juga apabila dilihat dari t_{hitung} yang lebih kecil dari t_{tabel} berarti semua variabel regressor X secara individu tidak berpengaruh secara signifikan terhadap nilai taksiran Y.

Sedangkan apabila kita uji keberartian model secara simultan atau bersama-sama untuk semua β , maka hipotesisnya adalah sebagai berikut

Ho : $\underline{\beta} = 0$ (Variabel X secara simultan tidak bergantung terhadap nilai taksiran Y)

H1 : $\underline{\beta} \neq 0$ (Variabel X secara simultan bergantung terhadap nilai taksiran Y)

$\alpha = 5\%$

Dengan menggunakan statistik uji ANAVA atau uji F, maka berdasarkan taksiran parameter melalui Metode Kuadrat Terkecil untuk regresi linier ganda pada data dalam lampiran 1 diperoleh tabel ANAVA sebagai berikut:

Tabel 5.3 ANAVA Metode Kuadrat Terkecil

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	p
1	Regression	4,84E+08	3	161358255,6	197,190	,000[a]
	Residual	10637774	13	818290,285		
	Total	4,95E+08	16			

a. Predictors: (Constant), X3, X2, X1

b. Dependent Variable: Y

$R^2 = 0.978$

Dengan kriteria uji: Tolak H_0 jika $F_{hitung} \geq F_{(p;n-p-1;\alpha/2)}$ atau bisa juga dilihat dari nilai p, tolak Ho jika nilai $p \leq \alpha$.

Dari tabel diatas terlihat bahwa nilai p kurang dari α. Ini berarti semua variabel X secara *simultan* berpengaruh terhadap nilai taksiran Y. hal ini berbeda jika pengujian dilakukan secara *parsial* atau individu. Dari tabel 5.3 dilihat bahwa R^2 mendekati satu, tidak diikuti dengan hasil uji hipotesis yang signifikan dari koefisien β. Hal ini menunjukkan adanya kolinieritas.

Di bawah ini disajikan tabel hasil perhitungan nilai korelasi antar variabel *regressor*.

Tabel 5.4. Matriks Korelasi dari Variabel X

	X1	X2	X3
X1	1	0,999	0,936
X2	0,999	1	0,933
X3	0,936	0,933	1

Dari tabel 5.4 terlihat korelasi yang sangat tinggi antar variabel *regressor*-nya. Hal ini menunjukkan adanya multikolinieritas.

Adanya multikolinieritas juga bisa dilihat melalui perhitungan determinan matriks $\underset{\sim}{Z}'\underset{\sim}{Z}$. Dari perhitungan, diperoleh :

$$\underset{\sim}{Z}'\underset{\sim}{Z} = \begin{pmatrix} 1 & 0.99 & 0.936 \\ 0.99 & 1 & 0.933 \\ 0.936 & 0.933 & 1 \end{pmatrix}$$

Matriks $\underset{\sim}{Z}'\underset{\sim}{Z}$ ini merupakan matriks korelasi antar peubah prediktor. Terlihat bahwa korelasi antar variabel *regressor* sangat tinggi ini juga bisa dilihat dari determinan matriks $\underset{\sim}{Z}'\underset{\sim}{Z}$ = 0.00242524 yang mendekati 0. hal ini berarti matriks $\underset{\sim}{Z}'\underset{\sim}{Z}$ hampir singular dan memperlihatkan adanya multikolinieritas dengan *ill conditioned.*

5.2 Penaksiran Model *Ridge Regression*

Dalam analisis *ridge regression* digunakan data yang sudah ditransformasi melalui metode *centering and rescaling* (lampiran 2).

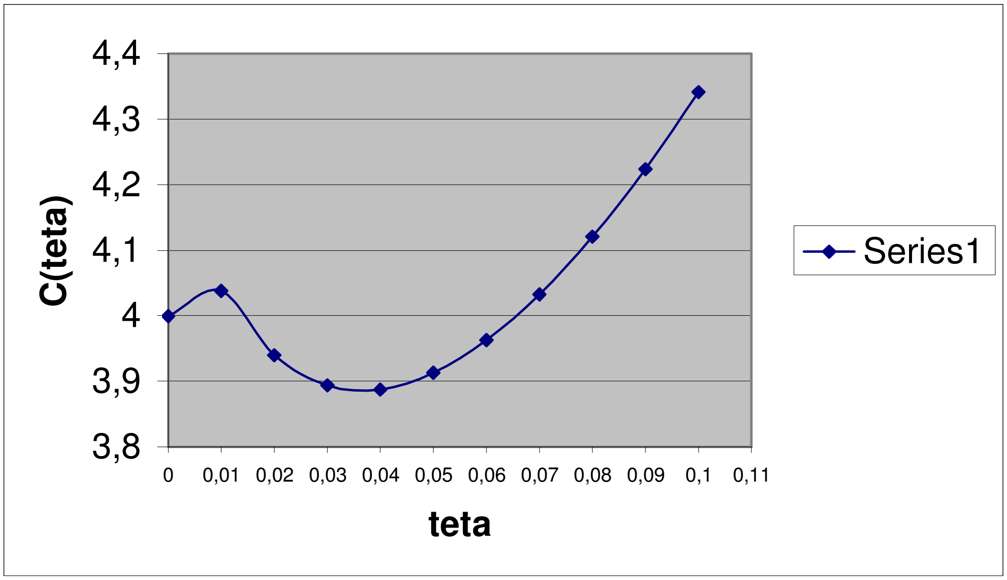
Dalam memilih tetapan θ untuk dapat menaksir *ridge regression* digunakan statistik $C_p Mallows(C_\theta)$. Nilai C_θ dengan berbagai nilai kemungkinan tetapan θ disajikan dalam tabel 5.5 berikut :

Tabel 5.5 Nilai C_θ dengan berbagai nilai θ

θ	C_θ
0	3,99972
0,01	4,038074
0,02	3,939568
0,03	3,893453
0,04	**3,887591**
0,05	3,912661
0,06	3,962384
0,07	4,03261
0,08	4,120363
0,09	4,223615
0,1	4,340929

Dari tabel 5.5 dibuat grafik dengan sumbu datar θ dan sumbu tegak C_θ, hasilnya disajikan dalam Grafik 5.1

Grafik 5.1



Nilai θ yang terpilih adalah pada saat C_θ minimum. Dari tabel 5.5 dan grafik 5.1 terlihat bahwa nilai θ yang meminimumkan C_θ adalah $\theta = 0.04$.

Sehingga persamaan regresinya menjadi :

$\hat{Y}^* = 0.3797\ Z_1 + 0.4013\ Z_2 + 0.2025\ Z_3$

Setelah dikembalikan ke variabel-variabel asal diperoleh persamaan regresinya :

$\hat{Y} = -112.1599 + 13.1107\ X_1 + 0.4548\ X_2 + 10.4304\ X_3$

Jika kita uji secara *simultan* untuk semua β, maka hipotesisnya adalah sebagai berikut :

$H_0 : \underline{\beta} = 0$ (Variabel X secara simultan tidak bergantung terhadap nilai taksiran Y)

H1 : $\underline{\beta} \neq 0$ (Variabel X secara simultan bergantung terhadap nilai taksiran Y)

$\alpha = 5\%$

Dengan menggunakan statistik uji ANAVA atau uji F, maka kita dapatkan tabel untuk metode *ridge regression* yang disajikan dalam tabel 5.6 berikut :

Tabel 5.6 ANAVA *Ridge Regression*

Sumber Regresi	dk	JK	RJK	F_{hitung}
Regresi b	3	0,9604	0,320133	105,0943
Galat	13	0,0396	0,003046	
Total	16	1		

Dengan kriteria uji :

Tolak H_0 jika $F_{hitung} > F_{(p;n-p-1;\alpha/2)}$ atau bisa juga dilihat melalui nilai p, tolak H_0 jika nilai $p \le \alpha$. Dari table distribusi F diperoleh $F_{Tabel} = F_{(3;13;0,025)} = 3,41$. Ternyata $F_{hitung} > F_{(p;n-p-1;\alpha/2)}$ sehingga tolak H_0. Ini menunjukkan bahwa regresi linier ganda Y atas X bersifat signifikan.

Pegujian keberartian model *ridge regression* yang dilakukan secara parsial atau individu dapat dilakukan melalui pengujian hipotesis sebagai berikut :

H_0 : $\beta_i^* = 0$, untuk i=1,2,3 (variabel *regressor* X secara individu tidak berpengaruh secara signifikan terhadap nilai taksiran Y)

H_1 : $\beta_i^* \neq 0$, untuk i=1,2,3 (variabel *regressor* X secara individu tidak perpengaruh secara signifikan terhadap nilai taksiran Y)

$\alpha = 5\%$

Dengan statistik uji t-student, maka kita peroleh nilai T_{hitung} dari masing-masing variabel X secara individu adalah sebagai berikut :

Tabel 5.7 T_{Hitung}

Penaksir	t_{hitung}
b_1^*	20,1657
b_2^*	20,1489
b_3^*	19,6797

Dengan kriteria uji :

Tolak Ho jika $t_{hitung} \leq -t_{(n-k-1;\alpha/2)}$ atau $t_{hitung} \geq t_{(n-k-1;\alpha/2)}$, terima dalam hal lainnya.

Kriteria uji ini bisa juga dilihat dari nilai p. Tolak H_0 jika nilai $p \leq \alpha$, terima dalam hal lainnya. Dari table distribusi t-student diperoleh $t_{Tabel} = t_{(13;\ 0,0025)} = 2,16$ dari table 5.7 terlihat bahwa semua nilai $t_{Hitung} > t_{tabel}$ sehingga tolak H_0. Hal ini menunjukkan bahwa setiap variable X secara individu berpengaruh secara signifikan terhadap nilai taksiran Y.

BAB VI
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penjelasan yang telah diuraikan sebelumnya, kita bisa menyimpulkan hal-hal sebagai berikut:

1. Nilai R^2 yang besar tidak diikuti oleh hasil uji hipotesis yang signifikan dari semua koefisien penaksir b_i serta eigen valuenya yang kecil. Hal ini menunjukan multikolinieritas dalam data.
2. Multikolinieritas tidak sempurna terjadi jika terdapat kondisi *ill conditioned*, yaitu kondisi dimana terjadi korelasi antar variable regressor cukup tinggi, sehingga menyebabkan determinan ($\underset{\sim}{X}'\underset{\sim}{X}$) mendekati tidak sempurna atau mendekati 0.
3. Jika antar variable regressor terjadi multikolinieritas, maka pemanfaatan aljabar matriks dapat digunakan melalui transformasi *centering and rescaling*.
4. Metode *Ridge Regression* digunakan untuk mengatasi multikolinieritas tidak sempurna atau *ill conditioned* yang terjadi antara variable regressor.

5.2 Saran

Banyak metode untuk mengatasi masalah multikolinieritas. Analis dapat memilih salah satu diantara semua metode yang lebih baik dari Metode Kuadrat Terkecil. Walaupun *ridge regression* belum tentu dapat digunakan untuk menyelesaikan semua model yang mengandung multikolinieritas, tetapi sudah cukup bukti bahwa *ridge regression* merupakan salah satu metode yang baik. Ini dikarenakan melalui model ini diusahakan memperoleh varians yang mengecil dengan menentukan nilai θ sehingga diperoleh keadaan yang lebih stabil.

Lampiran 1. Data Mengenai Tenaga Kerja di Rumah Sakit Sardjito Yogyakarta

Y	X1	X2	X3
566.52	15.57	472.92	18
696.82	44.02	1339.75	9.5
1033.15	20.42	620.25	12.8
1603.62	18.74	568.33	36.7
1611.37	49.2	1497.6	35.7
1613.27	44.92	1365.83	24
1854.17	55.48	1687	43.3
2160.55	59.28	1639.92	46.7
2305.58	94.39	2872.33	78.7
3503.93	128.02	3655.08	180.5
3571.89	96	2912	60.9
3741.4	131.42	3921	103.7
4026.52	127.21	3865.67	126.8
10343.81	252.9	7684.1	157.7
11732.17	409.2	12446.33	169.4
15414.94	463.7	14098.4	331.4
18854.45	510.22	15524	371.6

Lampiran 2. Data hasil Transformasi Melalui Metode *Centering and Rescaling*

Y'	Z_1	Z_2	Z_3
-0.19836	-0.20602	-0.2042	-0.20453
-0.1925	-0.16185	-0.16003	-0.22421
-0.17738	-0.19849	-0.19669	-0.21657
-0.15173	-0.20109	-0.19934	-0.16122
-0.15138	-0.15381	-0.15199	-0.16354
-0.1513	-0.16045	-0.1587	-0.19063
-0.14047	-0.14406	-0.14234	-0.14594
-0.12669	-0.13816	-0.14474	-0.13806
-0.12017	-0.08365	-0.08194	-0.06396
-0.0663	-0.03145	-0.04206	0.171791
-0.06324	-0.08115	-0.07992	-0.10518
-0.05562	-0.02617	-0.02851	-0.00606
-0.0428	-0.0327	-0.03133	0.047433
0.241224	0.162421	0.163222	0.118991
0.303644	0.405065	0.405864	0.146086
0.46922	0.489672	0.490039	0.521245
0.62386	0.56189	0.562675	0.61434

DAFTAR PUSTAKA

Gujarati,Damodar 1995 , Basic Econometrics,Mc Graw Hill Book Co.-Singapore.

Myers, R.H, 1990, *Classical and modern Regression With Application* . PWS-KENT Publishing Company Boston.

R.K. Sembiring ,1995 , *Analisis Regresi* , Bandung, ITB.

Supranto , 1984 ,Ekonomerika Edisi Kedua ,Jakarta: Lembaga Penerbit Fakultas Ekonomi , UI.