PENGGUNAAN METODE *RIDGE TRACE* DAN

VARIANCE INFLATION FACTORS* (*VIF*) PADA REGRESI *RIDGE

SKRIPSI

Diajukan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta untuk memenuhi sebagian persyaratan guna memperoleh gelar Sarjana Sains

Oleh:

Agriska Prenadita Putri

05305144038

PROGRAM STUDI MATEMATIKA

JURUSAN PENDIDIKAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI YOGYAKARTA

2011

PERSETUJUAN

PENGGUNAAN METODE *RIDGE TRACE* DAN
VARIANCE INFLATION FACTORS (*VIF*) PADA REGRESI *RIDGE*

SKRIPSI

Telah disetujui
Pada tanggal 10 Maret 2011
Untuk diujikan di depan Panitia Penguji Skripsi
Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta

Menyetujui

Pembimbing

Endang Listyani, M.S.
NIP.195911151986012001

PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Agriska Prenadita Putri

NIM : 05305144038

Jurusan/ Prodi : Pendidikan Matematika/ Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Judul TAS : Penggunaan *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge*

Dengan ini menyatakan bahwa skripsi ini benar-benar hasil pekerjaan saya sendiri. Sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang ditulis atau diterbitkan orang lain atau telah digunakan sebagai persyaratan penyelesaian studi di perguruan tinggi lain kecuali pada bagian tertentu yang saya ambil sebagai acuan atau kutipan dengan mengikuti tata penulisan karya ilmiah yang telah lazim. Apabila terbukti pernyataan saya ini tidak benar, maka sepenuhnya menjadi tanggung jawab saya dan saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Yogyakarta, 10 Maret 2011
yang menyatakan,

Agriska Prenadita Putri

PENGESAHAN

Skripsi yang berjudul "Penggunaan Metode *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge* " ini telah diujikan di depan Dewan Penguji Skripsi pada tanggal 23 Maret 2011 dan dinyatakan lulus.

DEWAN PENGUJI

Nama	Jabatan	Tanda Tangan	Tanggal
Endang Listyani, M.S NIP. 195911151986012001	Ketua Penguji	……………….	25/04/2011
Retno Subekti, M.Sc NIP. 198111162005012002	Sekretaris Penguji	……………….	25/04/2011
Elly Arliani, M.Si NIP. 196708161992032001	Penguji Utama	……………….	18/04/2011
Kismiantini, M.Si NIP. 197908162001122001	Penguji Pendamping	……………….	19/04/2011

Yogyakarta, April 2011
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta

Dekan,
Dr. Ariswan
NIP.195909141988031003

iv

MOTTO

"Sesungguhnya sesudah kesulitan itu pasti ada kemudahan. Maka apabila kamu telah selesai (urusan dunia), bersungguh-sungguhlah (dalam beribadah)".

(Qs. Al Insyiroh : 6-7)

" ....Ya Tuhanku, tunjukilah aku untuk mensyukuri nikmat Engkau yang telah Engkau berikan kepadaku dan kepada ibu bapakku dan supaya aku dapat berbuat amal yang soleh yang Engkau ridhoi.."

(Q.s. Al-Ahqaaf (46) : 15)

Jangan pernah percaya pada apapun. Percayalah pada logikamu sendiri walaupun logikamu tidak selalu benar. Maka dalam hidup ini asahlah logikamu agar ia sealu benar".

do the right thing and do it right

yang menetapkan hati adalah keinginan

dan ...... yang menguatkan langkah adalah cinta

PERSEMBAHAN

Karya ini kupersembahkan untuk

Papa dan Mama tercinta yang mencurahkan kasih sayangnya,, menjadikan hidupku lebih bermakna dengan segenap pengorbanan dan kasih sayang yang akan selalu terjaga.

Kedua adikku Intan Paramita Sari dan Gita Astri Kusumadewi yang memberikan semangat dalam hidupku agar aku menjadi kakak yang baik..

Sahabat-sahabatku Diana, Dwi, Titis, Gita dan Dian yang selalu memberikan motivasi untukku.

Ms Rizky yang telah membantu dan selalu memberi semangat untukku.

Seluruh keluarga dan orang-orang yang telah memberikan kasih sayang dan selalu memberikan doa untukku.

KATA PENGANTAR

Alhamdulillah puji syukur tak henti penulis panjatkan kehadirat Allah SWT atas limpahan nikmat iman, islam, serta ikhsan yang tak pernah henti selalu dan selalu tercurah untuk seluruh hamba-NYA. Sungguh, tanpa kekuatan dan pertolongan dari-NYA, penulis takkan mampu menyelesaikan skripsi ini hingga paripurna. Shalawat dan salam semoga senantiasa tetap tercurahkan kepada insan Rasulullah Muhammad SAW atas kasih sayang dan perjuangannya dalam menghantarkan umat manusia dari gelapnya kejahiliyahan menuju benderang cahaya Islam.

Semua ada saatnya, begitupun dalam penyusunan skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Matematika, Jurusan Pendidikan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta dengan judul "Penggunaan Metode *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge*" ini, pada akhirnya paripurna sudah. Semua tak lepas dari dukungan serta bantuan dari berbagai pihak.

Pada kesempatan ini, ungkapan terima kasih yang tulus penulis sampaikan kepada :

1. Dr. Ariswan selaku Dekan FMIPA UNY yang telah mengesahkan skripsi ini.
2. Dr. Hartono selaku Kajurdik Matematika yang telah memberikan izin penulisan skripsi ini.
3. Atmini Dhoruri, M.Si selaku Kaprodi Matematika yang telah memberikan izin penulisan skripsi ini.

4. Kusprihantoso, S.Si selaku Penasehat Akademik yang telah memberikan bimbingan, motivasi, dan arahan kepada penulis.
5. Endang Listyani, M.S selaku Dosen Pembimbing yang telah memberikan masukan, bimbingan, dan arahan kepada penulis.
6. Elly Arliani, M.Si selaku Dosen Penguji Utama yang telah menguji skripsi penulis.
7. Kismiantini, M.Si selaku Dosen Penguji Penamping yang telah menguji skripsi penulis.
8. Retno Subekti, M.Sc selaku Dosen Sekretaris Penguji yang telah menguji skripsi penulis.
9. Seluruh Dosen Jurusan Pendidikan Matematika FMIPA UNY yang telah memberikan ilmu dan pengalaman kepada peneliti.
10. Teman-teman Matematika 2005 yang sama-sama berjuang menyelesaikan skripsi atas bantuan dan motivasinya kepada peneliti.

Tak ada gading yang tak retak, mungkin kata-kata ini peribahasa yang paling tepat untuk menggambarkan keterbatasan yang ada pada diri penulis. Penulis menyadari, skripsi ini masih jauh dari sempurna. Oleh karena itu saran, masukan, dan kritik yang membangun dari berbagai pihak sangat penulis harapkan guna dijadikan pertimbangan dan perbaikan pada penulisan selanjutnya. Akhirnya, penulis berharap semoga skripsi ini dapat dimanfaatkan sebagaimana mestinya.

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Secara umum model regresi linear ganda dengan k variabel bebas dinyatakan :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

dengan:

Y_i	=	variabel terikat
X_{ki}	=	variabel bebas
$\beta_1, \beta_2, \ldots, \beta_k$	=	parameter
ε_i	=	variabel gangguan / error

Analisis regresi merupakan sebuah teknik analisis statistik untuk membuat model dan mengetahui hubungan antara dua variabel atau lebih. Salah satu dari model statistika yang sering digunakan dalam pemecahan suatu permasalahan adalah model Regresi Linear *(Linear Regression)*. Hubungan dalam model tersebut dapat dinyatakan dalam bentuk persamaan yang menghubungkan variabel terikat (Y) dengan satu atau lebih variabel bebas $X_1, X_2, X_3, \ldots, X_k$. Jika variabel terikat (Y) hanya dihubungkan dengan satu variabel bebas (X), maka akan menghasilkan model regresi linear sederhana *(Simple Linear Regression).* Sedangkan jika variabel bebas (X) yang digunakan lebih dari satu, maka akan menghasilkan model regresi linear berganda *(Multiple Linear Regression)*.

$\beta_1, \beta_2, ..., \beta_k$ adalah parameter-parameter yang akan diduga, dan metode kuadrat terkecil biasa *(Ordinary Least Square/OLS)* adalah metode yang sering digunakan untuk menduga parameter tersebut. Asumsi-asumsi yang harus dipenuhi dalam pengujian parameter pada model regresi linear ganda adalah:

1. ε_i variabel random riil dan berdistribusi normal
2. Harga harapan atau rata-rata dari variabel gangguan adalah nol atau E(ε_i) = 0, untuk $i = 1,2, ..., n$
3. Variansi (ε_i) = σ^2 , sama untuk semua variabel pengganggu
4. Tidak ada hubungan linear antara variabel bebas X (multikolinearitas).

Estimasi dari β dengan menggunakan metode kuadrat terkecil akan menghasilkan estimator kuadrat terkecil $\hat{\beta}$ dengan sifat-sifat yang baik, maksudnya $\hat{\beta}$ merupakan estimator tak bias dan mempunyai variansi minimum.

Akan tetapi apabila terjadi penyimpangan salah satu asumsi, yaitu jika diantara variabel bebas X terjadi hubungan linear, mengakibatkan terjadinya multikolinearitas. Adanya multikolinearitas dapat menyebabkan det $(\boldsymbol{X^t X}) = \boldsymbol{0}$ (matriks singular), atau $\boldsymbol{X^t X}$ mendekati nol (hampir singular). Dapat dideteksi juga dengan menggunakan akar karakteristik. Montmogery dan Peck (1992: 198) menyatakan bahwa adanya multikolinearitas pada variabel bebas X menyebabkan terdapat akar karakteristik yang kecil dalam matriks $\boldsymbol{X^t X}$. Hal ini mengakibatkan variansi $\hat{\beta}$ besar. Draper dan Smith (1982: 258) menyatakan bahwa dengan adanya multikolinearitas yang tinggi maka $\hat{\beta}$ yang dihasilkan dengan metode kuadrat terkecil menjadi tidak stabil (peka tehadap perubahan kecil pada data yang kelihatannya tidak penting).

Sembiring (1995: 239) berpendapat bahwa salah satu akibat dari adanya multikolinearitas adalah variansi estimator β sangat besar mendekati tak hingga, bahkan untuk keadaan multikolinearitas sempurna tidak lagi sangat besar melainkan tak hingga. Menurut Somodiningrat (1994, 292-293) ada beberapa cara untuk mengatasi masalah ini, diantaranya adalah dengan memperbesar ukuran sampel sehingga kovariansi diantara parameter-parameternya dapat dikurangi. Jika variabel-variabel ini berkolinear dalam populasi maka prosedur memperbesar ukuran sampel tidak akan mengurangi multikolinearitas. Kedua adalah dengan pemilihan model, yaitu dengan mendefinisikan kembali variabel-variabel yang mengalami masalah multikolinearitas atau dengan menghapus satu atau lebih variabel bebas. Penghapusan variabel bebas ini tidak akan memberikan solusi yang memuaskan jika variabel yang dikeluarkan dari model mempunyai pengaruh yang relatif signifikan terhadap variabel terikat, karena dapat mengurangi ketepatan prediksi dari model. Ketiga adalah dengan metode Regresi Ridge (*Ridge Regression*), metode ini pertama kali dikemukakan oleh A. E. Hoerl pada tahun 1962. Regresi *Ridge* digunakan untuk mengatasi masalah multikolinearitas dengan memberikan suatu estimator $\hat{\beta}_R$ yang bersifat bias tetapi memiliki nilai $VIF < 10$, yang artinya masalah multikolinearitas sudah hilang dalam data, standar error setiap penduga parameter Regresi *Ridge* juga kecil, selain itu juga lebih stabil terhadap perubahan dalam data dibanding estimator tak bias $\hat{\beta}$. Metode estimator ini melalui pendekatan metode kuadrat terkecil, yakni melalui pendekatan metode Langrange dengan menambahkan suku $c\boldsymbol{I}$ pada matriks $\boldsymbol{X}^t\boldsymbol{X}$ sebelum diinverskan sehingga melemahkan multikolinearitas.

Pada Regresi *Ridge* terdapat dua metode untuk mengatasi masalah multikolinearitas yang digunakan dalam tulisan ini yaitu metode *Ridge Trace* dan *Variance Inflation Factors* (*VIF*). *Ridge Trace* adalah plot dari estimator Regresi *Ridge* dengan berbagai kemungkinan nilai tetapan bias *c*. Nilai yang dihasilkan dengan menggunakan *Ridge Trace* dan Nilai *VIF* yang mendekati satu dapat mengatasi masalah multikolinearitas.

B. Pembatasan Masalah

Dalam penulisan skripsi ini, pembatasan masalah sangat diperlukan agar penyelesaian masalah tidak menyimpang dari pembahasan yaitu dengan menganggap bahwa asumsi klasik yang lain terpenuhi dan difokuskan pada penggunaan metode *Ridge Trace* dan *VIF* pada Regresi *Ridge* untuk mengatasi masalah multikolinearitas.

C. Rumusan Masalah

1. Bagaimanakah penggunaan *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge* untuk mengatasi masalah multikolinearitas yang terjadi pada regresi linear berganda?
2. Bagaimanakah contoh penerapan Regresi *Ridge* untuk mengatasi multikolinearitas?

D. Tujuan

a. Menjelaskan penggunaan *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge* untuk mengatasi masalah multikolinearitas yang terjadi pada regresi linear berganda

b. Menunjukkan contoh penerapan Regresi *Ridge* untuk mengatasi multikolinearitas.

E. Manfaat

Skripsi ini dibuat agar mahasiswa mengetahui tentang Regresi *Ridge* sebagai salah satu metode estimasi alternatif untuk mengatasi masalah multikolinearitas yang terjadi pada regresi linear berganda. Dan bagi penulis agar lebih memahami tentang penggunaan metode *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) untuk mengatasi masalah multikolinearitas.

BAB II

LANDASAN TEORI

Pada pembahasan skripsi ini diperlukan beberapa definisi dan teorema. Bab ini membahas tentang definisi-definisi dan teorema-teorema yang diperlukan dalam pembahasan bab-bab selanjutnya

A. Matriks

Persamaan regresi linear berganda dapat dinyatakan dalam bentuk matriks. Hal tersebut sangat membantu dalam proses pembuktian sifat-sifat dan perhitungan matematis dari analisis regresi ganda tersebut. Untuk itu akan dibahas beberapa hal mengenai matriks.

Definisi 2.1.1 (Ruminta: 2009, 1): Matriks adalah himpunan skalar yang disusun/ dijajarkan secara khusus dalam bentuk baris dan kolom sehingga membentuk empat persegi panjang atau persegi.

Suatu matriks $\boldsymbol{A}$ berukuran $m \times n$ didefinisikan sebagai susunan angka-angka dengan *m* baris dan *n* kolom, yang anggotanya a_{ij} dimana a pada baris ke-*i* dan kolom ke-*j* dari matriks $\boldsymbol{A}$ tersebut, yang dapat dituliskan sebagai berikut :

$$\boldsymbol{A}_{\boldsymbol{m}\times\boldsymbol{n}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \tag{2.1}$$

atau dapat juga ditulis :

$$\boldsymbol{A_{m\times n}} = \left[a_{ij}\right]_{m\times n} \quad , \quad i = 1,2,\dots,m \quad j = 1,2,\dots,n$$

Sebuah matriks ***A*** dapat dikalikan dengan suatu bilangan konstan ataupun terhadap suatu matriks andaikan matriks ***A*** bukan matriks baris dan matriks ***B*** juga bukan matriks kolom, maka hasil kali dari ***AB*** tetap bisa dicari dengan mengingat bahwa matriks ***A*** bisa dipandang sebagai kumpulan matriks baris dan matriks ***B*** bisa dipandang sebagai kumpulan matriks kolom. Tetapi harus diingat bahwa supaya perkalian bisa dipenuhi syarat jumlah elemen matriks baris dari matriks ***A*** harus sama dengan banyaknya elemen matriks kolom dari matriks ***B***. Hal tersebut didefinisikan sebagai berikut:

Definisi 2.1.2 (Anton: 1987, 24): jika ***A*** adalah sebarang matriks dan *c* adalah sebarang skalar, maka hasil kali *c****A*** adalah matriks yang diperoleh dengan mengalikan setiap elemen dalam ***A*** dengan *c*.

Definisi 2.1.3. (Anton: 1987, 25): Jika ***A*** adalah sebuah matriks $m\times r$ dan ***B*** adalah sebuah matriks $r\times n$, maka hasil kali anggota ***AB*** adalah matriks $m\times n$ 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.

$$\boldsymbol{AB} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj}$$

$$= \sum_{k=1}^{r} a_{ir} b_{rj}$$

$$untuk \ (1 \le i \le m, 1 \le j \le n)$$

Simbol untuk suatu matriks yang ditranspose, kadang berbeda-beda. Untuk itu diperlukan pendefinisian simbol matriks transpose sebagai berikut:

Definisi 2.1.4 (Ruminta: 2009, 11): Jika $\boldsymbol{A}_{\mathbf{m\times n}} = [a_{ij}]$ adalah sebuah matriks $m \times n$, maka transpose $\boldsymbol{A}$ dinyatakan dengan $\boldsymbol{A}^t$ didefinisikan sebagai matriks $n \times m$ yang merupakan hasil dari pertukaran baris dan kolom dari matriks $\boldsymbol{A}$.

Matriks $m \times n$ dapat ditulis :

$$\boldsymbol{A}_{\boldsymbol{m\times n}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}] \quad , \quad i = 1,2,\ldots,m \quad j = 1,2,\ldots,n$$

$$\boldsymbol{A}^t_{\boldsymbol{m\times n}} = \boldsymbol{A}_{\boldsymbol{n\times m}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix} = [a_{ji}] \quad , \quad i = 1,2,\ldots,m \quad j = 1,2,\ldots,n$$

Jika $\boldsymbol{X}$ merupakan vektor kolom yang beranggotakan $x_1, x_2, \ldots, x_n$ dan $\boldsymbol{Y}$ merupakan vektor kolom yang beranggotakan $y_1, y_2, \ldots, y_n$ yang dapat dinyatakan dalam bentuk notasi:

$$\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \; dan \;\; \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\boldsymbol{X^t} = [x_1 \quad x_2 \quad \cdots \quad x_n] \;\; dan \;\; \boldsymbol{Y^t} = [y_1 \quad y_2 \quad \cdots \quad y_n] \tag{2.2}$$

Sehingga jumlah dari perkalian elemen-elemen dalam vektor ***X*** dan ***Y*** dapat dituliskan sebagai berikut:

$$\sum_{i=1}^{n} \boldsymbol{X_i Y_i} = \boldsymbol{X^t Y} \tag{2.3}$$

Jumlah kuadrat dari elemen-elemen dalam vektor ***X*** merupakan perkalian vektor ***X*** baris dengan vektor ***X***.

$$\sum_{i=1}^{n} \boldsymbol{X^2} = \boldsymbol{X^t X} \tag{2.4}$$

Sifat transpose matriks yang berguna dalam analisis regresi ganda :

$$(\boldsymbol{AB})^{\boldsymbol{t}} = \boldsymbol{B^t A^t} \tag{2.5}$$

Definisi 2.1.5 (Ruminta: 2009, 5): Matriks persegi adalah matriks yang memiliki baris dan kolom yang sama banyak atau matriks yang berukuran $n \times n$.

Definisi 2.1.6 (Ruminta: 2009, 9): Suatu matriks persegi A berukuran $n \times n$ disebut *orthogonal* jika $\boldsymbol{AA^{-1}} = \boldsymbol{I} \Leftrightarrow \boldsymbol{A^{-1}A} = \boldsymbol{I}.$ Jadi $\boldsymbol{A^{-1}} = \boldsymbol{A^t}$

Definisi 2.1.7 (Ruminta: 2009, 7): Suatu matriks persegi dikatakan matriks *non singular* atau *invertible* , jika nilai determinan matriks ≠ 0 dan dikatakan *singular* jika nilai detrminan matriks = 0 sehingga tidak mempunyai invers.

Definisi 2.1.7 (Ruminta: 2009, 5): Matriks Identitas adalah suatu matriks di mana semua elemen pada diagonal utamanya bernilai satu (1) dan elemen di luar diagonal utama bernilai nol. Matriks identitas biasa diberi symbol ***I***.

$\boldsymbol{A} = [a_{ij}] = I \Leftrightarrow m = n$ dan untuk

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

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

Suatu matriks Identitas umum dapat dituliskan sebagai berikut :

$$\boldsymbol{I_{n \times n}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Definisi 2.1.9 (Budi Murtiyasa: 2010, 61): Andaikan $\boldsymbol{A}$ adalah matriks persegi berdimensi *n.* Matriks $\boldsymbol{A}$ dikatakan simetri, jika berlaku $= \boldsymbol{A^t}$. Dengan definisi ini berarti bahwa matriks persegi $\boldsymbol{A}$ adalah simetri jika dan hanya jika $\boldsymbol{a_{ij}} = \boldsymbol{a_{ji}}$ untuk semua *i* dan *j.*

B. Metode Kuadrat Terkecil

Secara umum bentuk persamaan regresi linear berganda adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i \qquad (2.6)$$

dengan:

Y_i	= variabel terikat
X_{ki}	= variabel bebas
$\beta_1, \beta_2, \ldots, \beta_k$	= parameter
ε_i	= variabel gangguan / error

Dengan mean $E\{\varepsilon_i\} = 0$ dan variansinya $\sigma^2\{\varepsilon_i\} = \sigma^2$, ε_i dan ε_j tidak berkorelasi sehingga kovariansinya $E\{\varepsilon_i, \varepsilon_j\} = 0$, untuk semua nilai i dan j; $i \neq j$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$.

Model regresi linear berganda diatas dapat dinyatakan dalam bentuk matriks berikut:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.7)$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1i} & x_{2i} & \cdots & x_{ki} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

dengan $\boldsymbol{Y}$ adalah vektor $i \times 1$ dari observasi-observasi pada variabel terikat, $\boldsymbol{X}$ adalah matriks $i \times k$ dari observasi-observasi pada k-variabel bebas, β adalah vektor $k \times 1$ dari koefisien regresi dan ε adalah vektor $i \times 1$ dari error dengan ε_i $\varepsilon_i \sim NID(0, \sigma^2)$.

Jika diasumsikan $\boldsymbol{X}$ dan $\boldsymbol{Y}$ telah dipusatkan dan diskalakan sehingga $\mathbf{1}^t\boldsymbol{X} = \mathbf{0}$ dan $\boldsymbol{X}^t\boldsymbol{X} = \mathbf{1}$ maka $\boldsymbol{X}^t\boldsymbol{X}$ dan $\boldsymbol{X}^t\boldsymbol{Y}$ adalah matriks korelasi dari koefisien-koefisien.

Estimator kuadrat terkecil untuk β adalah:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y} \tag{2.8}$$

Bukti:

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

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta}$$

$$\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})^t(\boldsymbol{Y} - \boldsymbol{X\beta})$$

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

$$= \boldsymbol{Y}^t\boldsymbol{Y} - \mathbf{2}\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y} + \boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{X\beta}$$

Catatan: $\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y}$ adalah matriks 1×1 atau suatu skalar dan nilai transposenya $(\boldsymbol{\beta}^t\boldsymbol{X}^t\boldsymbol{Y})^t = \boldsymbol{Y}^t\boldsymbol{X\beta}$ adalah skalar yang sama.

Dalam metode kuadrat terkecil harus dipenuhi:

$$\left.\frac{\partial \boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}} = -\mathbf{2}\boldsymbol{X}^t\boldsymbol{Y} + \mathbf{2}\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}$$

sehingga diperoleh:

$$\boldsymbol{X^t X \hat{\beta} = X^t Y}$$

$$\boldsymbol{\hat{\beta}} = (\boldsymbol{X^t X})^{-1} \boldsymbol{X^t Y}$$

Matriks $\boldsymbol{X^t X}$ dari persamaan (2.8) diatas adalah matriks simetris $(\boldsymbol{k}+\boldsymbol{1}) \times (\boldsymbol{k}+\boldsymbol{1})$ yang mana elemen-elemen diagonalnya adalah jumlah kuadrat dari elemen-elemen kolom dalam matriks $\boldsymbol{X}$ dan elemen-eleman diluar diagonal adalah jumlahan hasil kali *product* dari elemen-elemen dalam kolom yang sama. Pada dasarnya $\boldsymbol{X^t X}$ mempunyai peranan penting dalam sifat-sifat estimator β dan sering menjadi faktor utama dalam kesuksesan atau kegagalan estimasi kuadrat terkecil.

C. Sifat-Sifat Estimator Kuadrat Terkecil

Metode kuadrat terkecil memiliki suatu estimator yang mempunyai sifat-sifat yang sangat baik, yang menjadikan metode estimasi ini sangat popular dikalangan para peneliti. Sifat-sifat tersebut antara lain adalah sebagai berikut:

Sifat estimator $\hat{\beta}$ = $\boldsymbol{(X^t X)^{-1} X^t Y}$ (Bain, L.J. 1992: 516).

1. Linear

   Dengan memperhatikan nilai $\hat{\beta}$ pada persamaan (2.8) dapat ditunjukkan bahwa $\hat{\beta}$ adalah fungsi linear dalam $\boldsymbol{Y}$ dan atau $\hat{\beta}$ adalah fungsi linear dalam observasi/ sampel random.

2. Takbias

$$\mathbf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \tag{2.9}$$

Bukti:

$\mathbf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{E}[(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}]$

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

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

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

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

$= \boldsymbol{\beta} + \mathbf{0}$

$= \boldsymbol{\beta}$

jadi $\hat{\beta}$ merupakan penaksir tak bias dari β.

3. Best

Untuk menunjukkan bahwa $\hat{\beta}$ adalah *best* estimator, maka harus dibuktikan bahwa Var ($\hat{\beta}$) $\leq$ Var ($\hat{\beta}_*$). Beberapa persamaan yang berguna untuk membuktikan hal tersebut adalah sebagai berikut:

a. Jarak $\hat{\beta}$ ke β.

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\varepsilon \tag{2.10}$$

Bukti:

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

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

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

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

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \quad = (\boldsymbol{X^tX})^{-1}\boldsymbol{X^t\varepsilon}$$

b. Matriks kovariansi

$$\boldsymbol{E}(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\boldsymbol{t}}) = \boldsymbol{\sigma}^{\boldsymbol{2}}(\boldsymbol{X^tX})^{-1} \tag{2.11}$$

Bukti:

$$\boldsymbol{E}(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\boldsymbol{t}}) = \boldsymbol{E}[\ ((\boldsymbol{X^tX})^{-1}\boldsymbol{X^tY})((\boldsymbol{X^tX})^{-1}\boldsymbol{X^tY})^{\boldsymbol{t}}]$$

$$= (\boldsymbol{X^tX})^{-1}\boldsymbol{X^t}\boldsymbol{E}(\boldsymbol{Y^tY})\ \boldsymbol{X}(\boldsymbol{X^tX})^{-1}$$

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

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

c. Variansi {MSE ($\hat{\beta}$)}

$$E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\boldsymbol{t}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] = \sigma^2 trace\ (\boldsymbol{X^tX})^{-1} \tag{2.12}$$

Bukti:

$$E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\boldsymbol{t}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] = E[\{(\boldsymbol{X^tX})^{-1}\boldsymbol{X^t\varepsilon}\}^t\{(\boldsymbol{X^tX})^{-1}\boldsymbol{X^t\varepsilon}\}]$$

$$= E[\{\boldsymbol{\varepsilon^tX}(\boldsymbol{X^tX})^{-1}\boldsymbol{X^t}\}^t\{(\boldsymbol{X^tX})^{-1}\boldsymbol{X^t\varepsilon}\}]$$

Diselesaikan dengan teorema 1.7 (Seber, 2003: 13). Untuk ε vektor $i \times 1$ dari variabel random dan $\boldsymbol{A}$ matriks simetris $k \times k$, jika $E(\varepsilon) = 0$ dan $Kov\ (\varepsilon) = \sigma^2\boldsymbol{I}$.

$$\boldsymbol{E}(\boldsymbol{\varepsilon^tA\varepsilon}) = Trace\ (\boldsymbol{A\sigma^2I}) + \boldsymbol{\theta^tA\theta}$$

Karena $E(\varepsilon) = 0$ maka

$$E\left[\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^{\boldsymbol{t}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]=\sigma^2\ trace\ [\boldsymbol{X(X^tX)^{-1}}(\boldsymbol{X^tX})^{-\mathbf{1}}\boldsymbol{X^t}+0$$

$$=\sigma^2\ trace\ [(\boldsymbol{X^tX})^{-\mathbf{1}}]$$

Dari persamaan (2.11) dan (2.12) terlihat bahwa jika $\boldsymbol{C}=(\boldsymbol{X^tX})^{-\mathbf{1}}$, maka variansi dari $\hat{\beta}_j$ adalah $\sigma^2\boldsymbol{C_{jj}}$ dan kovariansi antara $\hat{\beta}_i$ dan $\hat{\beta}_j$ adalah $\sigma^2\boldsymbol{C_{jj}}$.

Bukti bahwa Var ($\hat{\beta}$) ≤ Var ($\hat{\beta}_*$) dapat diperoleh dengan memisalkan $\hat{\beta}_*$ sebagai *LUE (Linear Unbiased Estimator)* untuk β. Caranya adalah dengan memisalkan bahwa $\hat{\beta}_*=\boldsymbol{a^tY}$ agar $\hat{\beta}_*$ tak bias untuk β, karena E($\hat{\beta}_*$) = $\boldsymbol{a^tXB}$, agar E($\hat{\beta}_*$) = β, haruslah $\boldsymbol{a^tX}=\boldsymbol{I}$. Jika dipilih $\boldsymbol{a^t}=(\boldsymbol{X^tX})^{-\mathbf{1}}+\boldsymbol{c^t}$ maka:

$$Var\left(\hat{\beta}_*\right)=Var\ [\boldsymbol{a^tY}]$$

$$=\boldsymbol{a^t}\ Var\ (\boldsymbol{Y})(\boldsymbol{a^t})^{\boldsymbol{t}}$$

$$=\boldsymbol{a^t}[\sigma^2]\boldsymbol{a}$$

$$=\sigma^2\boldsymbol{a^ta}$$

$$=\sigma^2[\boldsymbol{(X^tX)^{-1}X^t}+\boldsymbol{c^t}][\boldsymbol{X}(\boldsymbol{X^tX})^{-\mathbf{1}}+\boldsymbol{c}]$$

$$=\sigma^2\boldsymbol{(X^tX)^{-1}X^t}+\sigma^2(\boldsymbol{X^tX})^{-\mathbf{1}}\boldsymbol{c^t}+\sigma^2\boldsymbol{c^tX}(\boldsymbol{X^tX})^{-\mathbf{1}}+\sigma^2\boldsymbol{c^tc}]$$

$$=Var\left(\hat{\beta}\right)+\sigma^2\boldsymbol{c^tc}$$

Karena $\boldsymbol{c^tc}\geq 0$ maka $Var\left(\hat{\beta}\right)\leq Var\ (\hat{\beta}_*)$.

Ketiga sifat diatas lebih dikenal dengan istilah *BLUE (Best Linear Unbiased Estimator).*

D. Nilai Eigen dan Vektor Eigen

Berikut ini akan dibahas beberapa hal yang berhubungan dengan nilai eigen.

Definisi 2.4.1 (Anton: 1987, 277): Jika $\boldsymbol{A}$ adalah matriks $n \times n$, maka vektor tak nol $\boldsymbol{X}$ di dalam R^n dinamakan vektor eigen dari $\boldsymbol{A}$ jika $\boldsymbol{AX}$ adalah kelipatan skalar dari $\boldsymbol{X}$; yakni,

$$\boldsymbol{AX} = \lambda \boldsymbol{X} \tag{2.13}$$

Untuk suatu skalar λ. Skalar λ dinamakan *nilai eigen* dari $\boldsymbol{A}$ dan $\boldsymbol{X}$ dinamakan *vektor eigen* yang bersesuaian dengan λ .

Eigen dalam bahasa Jerman berarti asli, jadi nilai eigen adalah nilai asli atau nilai karakteristik.

Untuk mencari nilai eigen matriks $\boldsymbol{A}$ yang berukuran $n \times n$:

$$\boldsymbol{A_{n\times n}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad \boldsymbol{I_{n\times n}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$\boldsymbol{AX} = \lambda \boldsymbol{X}, \qquad \boldsymbol{X} \neq \boldsymbol{0}$$

$$\boldsymbol{AX} = \lambda \boldsymbol{IX}$$

$$\boldsymbol{IX} - \boldsymbol{AX} = 0$$

$$(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$$

$$\boldsymbol{X} \neq \boldsymbol{0}, \rightarrow |\lambda \boldsymbol{I} - \boldsymbol{A}| = 0 \tag{2.14}$$

Untuk memperoleh nilai λ.

$$|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$$

$$\begin{bmatrix} \lambda - a_{11} & . & . & . & -a_{1n} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ -a_{n1} & . & . & . & \lambda - a_{nn} \end{bmatrix} = 0$$

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

n buah akar $\lambda_1, \lambda_2, \cdots, \lambda_n$

Jika eigen value λ_n disubstitusi pada persamaan $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = \boldsymbol{0}$, maka solusi dari eigen vektor adalah $(\lambda_n \boldsymbol{I} - \boldsymbol{A})\boldsymbol{X_n} = \boldsymbol{0}$

Misalkan $\boldsymbol{A} = [\boldsymbol{a_{ij}}]$ matriks $n \times n$. Determinan

$$f(\lambda) = \boldsymbol{det}\ (\lambda_n \boldsymbol{I} - \boldsymbol{A}) = \begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix}$$

Dikatakan karakteristik polinom dari $\boldsymbol{A}$. Persamaan $f(\lambda) = \boldsymbol{det}\ (\lambda_n \boldsymbol{I} - \boldsymbol{A}) = \boldsymbol{0}$ dikatakan persamaan karakteristik dari $\boldsymbol{A}$.

Teorema 2.4.1 (Ruminta: 2009, 199): Jika A adalah matriks $n \times n$ dengan akar-akar karakteristik $\lambda_1, \lambda_2, \cdots, \lambda_n$ maka

$$\boldsymbol{Trace\ A} = \sum_{j=1}^{n} \boldsymbol{\lambda_j}$$

Bukti:

$\boldsymbol{A}$ adalah matriks $n \times n$. Nilai eigen dari matriks $\boldsymbol{A}$ yaitu $\lambda_1, \lambda_2, \cdots, \lambda_n$ adalah solusi dari persamaan $det|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$. Bersesuaian dengan j nilai eigen, λ_j , didefinisikan vektor eigen yang merupakan solusi dari

$$(\lambda_j \boldsymbol{I} - \boldsymbol{A})\boldsymbol{V_j} = \boldsymbol{0}$$

Matriks $\boldsymbol{V} = [\boldsymbol{v_1} \quad \boldsymbol{v_2} \quad \cdots \quad \boldsymbol{v_n}]$ dapat digunakan untuk mendiagonalkan $\boldsymbol{A}$, sehingga

$$\boldsymbol{V^t A V} = \boldsymbol{diag}\,[\boldsymbol{\lambda_1} \quad \boldsymbol{\lambda_2} \quad \dots \quad \boldsymbol{\lambda_n}] \tag{2.15}$$

Sebagai tambahan, *V* adalah matriks *orthogonal* yakni $\boldsymbol{V^t V} = \boldsymbol{V V^t} = \boldsymbol{I_n}$. Pada kasus dimana $\boldsymbol{A} = (\boldsymbol{X^{*t} X^*})$ adalah matriks korelasi setelah proses pemusatan dan penyekalaan. Jika matriks korelasi tersebut adalah matriks diagonal yakni tidak ada hubungan linear diantara variabel-variabel penjelas maka semua nilai eigen akan mendekati 1, ini adalah bentuk ideal. Jika matriks korelasi mendekati singular yakni terdapat paling sedikit satu ketergantungan linear erat diantara kolom-kolom pada $\boldsymbol{X^*}$ maka determinan dari $(\boldsymbol{X^{*t} X^*})$ akan mendekati nol. Jadi nilai eigen mempunyai hubungan yang penting dengan determinan. Perhatikan persamaan (2.15) dimana $\boldsymbol{A} = (\boldsymbol{X^{*t} X^*})$ dan *V orthogonal*, $|\boldsymbol{V}| = \mathbf{1.0}, |\boldsymbol{V^t}| = \mathbf{1.0}$ maka diperoleh:

$$\begin{aligned} |\boldsymbol{V^t A V}| &= |\boldsymbol{V^t}(\boldsymbol{X^{*t} X^* V})| \\ &= |\boldsymbol{V^t}||\boldsymbol{X^{*t} X^*}||\boldsymbol{V}| \\ &= |\boldsymbol{X^{*t} X^*}| \end{aligned} \tag{2.16}$$

$$|\boldsymbol{X^{*t} X^*}| = \begin{vmatrix} \boldsymbol{\lambda_1} & \dots & \dots & \mathbf{0} \\ \vdots & \boldsymbol{\lambda_2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \boldsymbol{\lambda_n} \end{vmatrix}$$

$$= \boldsymbol{tr}\,(\boldsymbol{A})$$

$$= \prod_{j=1}^{n} \lambda_j$$

E. Ukuran Pemusatan dan Penskalaan (*centering and scaling*)

Statistika adalah pengetahuan yang berhubungan dengan cara-cara pengumpulan data, pengolahan atau penganalisisan dan penarikan kesimpulan berdasarkan kumpulan data. Kumpulan data yang lengkap dan jelas yang ingin dipelajari sifat-sifatnya dinamakan populasi sedangkan sebagian data yang diambil dari populasi dinamakan sampel, dimana sampel diharapkan dapat mewakili populasi.

Pemusatan dan penskalaan data merupakan bagian dari membakukan (*standardized*) variabel. Modifikasi sederhana dari pembakuan atau standarisasi variabel ini adalah transformasi korelasi (*correlation transformation*). Pemusatan merupakan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Sedangkan penskalaan meliputi gambaran pengamatan pada kesatuan (unit) standar deviasi dari pengamatan untuk variabel (*Kutner, et al., 2005: 98*).

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

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \boldsymbol{\varepsilon}_i \tag{2.17}$$

Berikut ini merupakan pembakuan variabel terikat *Y* dan variabel bebas $X_1, X_2, \ldots, X_k$.

$$\frac{Y_i - \bar{Y}}{S_Y}\text{ , dengan } \quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} \tag{2.18}$$

$\frac{X_{ij}-\bar{X}_j}{S_{X_j}}$, dengan $S_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)^2}{n-1}}$ $j = 1, 2, \ldots, k$ (2.19)

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

Transformasi korelasi merupakan fungsi sederhana dari pembakuan variabel. Sehingga melalui transformasi diperoleh:

$$Y_i^* = \frac{Y_i-\bar{Y}_i}{\sqrt{n-1}S_Y} \tag{2.20}$$

$$X_{ij}^* = \frac{X_{ki}-\bar{X}_k}{\sqrt{n-1}S_X} \tag{2.21}$$

Berdasarkan transformasi peubah Y_i^* dan X_{ij}^* yang didefinisikan dengan transformasi korelasi pada model (2.20) dan (2.21) di atas diperoleh model regresi sebagai berikut:

$$Y_i^* = \beta_1^* X_{1i} + \beta_2^* X_{2i} + \cdots + \beta_k^* X_{ki} + \boldsymbol{\varepsilon}_i \tag{2.22}$$

Model (2.22) di atas disebut sebagai model regresi yang baku (*standardized regression model*). Diantara parameter $\beta_1^*, \beta_2^*, \ldots, \beta_k^*$ pada model regresi baku dengan parameter asli $\beta_1, \beta_2, \ldots, \beta_k$ pada model regresi linear berganda yang biasa terdapat suatu hubungan linear. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan seperti di bawah ini (*Kutner, et al., 2005: 99*):

$$\beta_j = \left(\frac{s_Y}{s_{X_j}}\right)\beta_j^{\,*}, \quad j = 1, 2, \ldots, k \tag{2.23}$$

$$\begin{aligned} \beta_0 &= \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \cdots - \beta_k\bar{X}_k \\ &= \bar{Y} - \textstyle\sum_{j=1}^{k}\beta_j\,\bar{X}_j \end{aligned} \tag{2.24}$$

Prosedur ini disebut dengan prosedur penskalaan.

Model (2.17) di atas dapat dibentuk menjadi:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_1\bar{X}_1 + \beta_2(X_{2i} - \bar{X}_2) + \beta_2\bar{X}_2 + \cdots + \beta_k(X_{ki} - \bar{X}_i) \\ &\quad + \beta_k\bar{X}_k + \boldsymbol{\varepsilon}_i \\ &= (\beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k) + \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \cdots \\ &\quad + \beta_k(X_{ki} - \bar{X}_k) + \boldsymbol{\varepsilon}_i \end{aligned}$$

Menurut rumus untuk mendapatkan β_0 yaitu:

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

Maka berlaku:

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

sehingga

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

$$= \beta_1 (X_{1i} - \bar{X}_1) + \beta_2 (X_{2i} - \bar{X}_2) + \cdots + \beta_k (X_{ki} - \bar{X}_k) + \boldsymbol{\varepsilon}_i$$

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

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

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

$$x_{ki} = X_{ki} - \bar{X}_k$$

maka didapat model baru yaitu:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \boldsymbol{\varepsilon}_i$$

Prosedur untuk membentuk model pertama menjadi model terakhir disebut dengan prosedur pemusatan. Prosedur ini mengakibatkan hilangnya β_0 (*intercept*) yang membuat perhitungan untuk mencari model regresi menjadi lebih sederhana. Keseluruhan dari prosedur di atas disebut prosedur pemusatan dan penskalaan.

F. Matriks Korelasi

Model yang didapat melalui prosedur pemusatan dan penskalaan di atas (2.22) bila dituliskan dalam matriks adalah:

$$\begin{bmatrix} {Y_1}^* \\ {Y_2}^* \\ \vdots \\ {Y_i}^* \end{bmatrix} = \begin{bmatrix} {\beta_1}^* \\ {\beta_2}^* \\ \vdots \\ {\beta_i}^* \end{bmatrix} \begin{bmatrix} {X_{11}}^* {X_{12}}^* \cdots {X_{1k}}^* \\ {X_{21}}^* {X_{22}}^* \cdots {X_{2k}}^* \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ {X_{ik}}^* {X_{ik}}^* \cdots {X_{ik}}^* \end{bmatrix} + \begin{bmatrix} {\varepsilon_1}^* \\ {\varepsilon_2}^* \\ \vdots \\ {\varepsilon_i}^* \end{bmatrix} \tag{2.25}$$

Dari persamaan (2.25) diperoleh matriks $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{X}^*$ dan $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{Y}^*$**,** yaitu:

$$\boldsymbol{X^tX} = \begin{bmatrix} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i1}}^{*\boldsymbol{2}} & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i1}}^{*\boldsymbol{2}} {X_{i2}}^* & \cdots & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i1}}^{*\boldsymbol{2}} {X_{i2}}^* \\ \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i2}}^{*\boldsymbol{2}} {X_{2i}}^* & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i2}}^{*2} & \cdots & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i2}}^{*\boldsymbol{2}} {X_{ik}}^* \\ \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{ik}}^{*\boldsymbol{2}} {X_{i1}}^* & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{ik}}^{*\boldsymbol{2}} {X_{i2}}^* & \cdots & \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{ik}}^{*\boldsymbol{2}} \end{bmatrix}$$

$$(2.26)$$

dan

$$\boldsymbol{X^tY} = \begin{bmatrix} \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i1}}^{*\boldsymbol{2}} Y_i^* \\ \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{i2}}^{*\boldsymbol{2}} Y_i^* \\ \vdots \\ \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} {X_{ik}}^{*\boldsymbol{2}} Y_i^* \end{bmatrix}$$

$$(2.27)$$

Matriks $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{X}^*$ dan $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{Y}^*$ di atas dapat juga ditulis dalam bentuk matriks korelasi. Hal ini berkaitan dengan penjelasan sebelumnya bahwa variabel $\boldsymbol{X}^*$dan

$\boldsymbol{Y}^*$ diperoleh dari transformasi korelasi peubah X dan Y sehingga kedua bentuk matriks di atas dapat ditulis dalam bentuk matriks korelasi. Matriks pertama yaitu $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{X}^*$ dapat ditulis dalam bentuk matriks korelasi dari peubah X dan dinotasikan dengan $\boldsymbol{r_{XX}}$. Matriks ini didefinisikan sebagai berikut:

$$\boldsymbol{r_{XX}} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix} \tag{2.28}$$

Berdasarkan persamaan (2.26) dan (2.28) diperoleh persamaan seperti berikut ini:

$$\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{X}^* = \boldsymbol{r_{XX}} \tag{2.29}$$

Sedangkan matriks kedua, yaitu $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{Y}^*$ adalah vektor yang berisi koefisien korelasi sederhana diantara variabel terikat Y dan setiap variabel bebas X, yang dinotasikan dengan $\boldsymbol{r_{YX}}$. Matriks korelasinya didefinisikan sebagai berikut:

$$\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{Y}^* = \boldsymbol{r_{YX}} \tag{2.30}$$

(*Kutner, et, al., 2005: 101*)

G. Uji Regresi Linear Ganda

Setelah model yang baik diperoleh, kemudian model akan dianalisis melalui hipotesis. Untuk mengujinya diperlukan dua macam jumlah kuadrat sisa (*JKS*) yang dapat dihitung dengan rumus (Sudjana: 2001, 91):

$$\left.\begin{aligned} JKR &= \hat{\beta}^t\left(\boldsymbol{X}^t\underline{\boldsymbol{Y}}\right) - n\bar{Y}^2 \\ JKS = JKT - JKR &= \underline{\boldsymbol{Y}}^t\underline{Y} - \hat{\beta}^t\left(\boldsymbol{X}^t\underline{Y}\right) \\ JKT &= \underline{\boldsymbol{Y}}^t\underline{Y} - n\bar{Y}^2 \end{aligned}\right\} \quad (2.31)$$

dengan JKR = Jumlah Kuadrat Regresi

JKS = Jumlah Kuadrat Sisa

JKT = Jumlah Kuadrat Total

$\hat{\beta}$ = Vektor estimasi parameter β

$\boldsymbol{X}$ = Matriks variabel bebas berukuran $(n \times k)$

$\underline{Y}$ = Vektor variabel terikat berukuran $(n \times 1)$

$\bar{Y}$ = Vektor rata-rata variabel terikat

Pengujian hipotesis untuk uji keberartian regresi sebagai berikut:

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

   $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ (tidak ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

   $H_1: \exists \beta_j \neq 0$, j = *1, 2, 3, ..., k* (ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

2. Menentukan tingkat signifikansi (α)

3. Memilih dan menuliskan uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-F.

$$R^2 = \frac{JKR}{JKT} \qquad F_{hitung} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

dengan k = banyaknya variabel bebas dalam model

n = banyaknya data

4. Menentukan aturan pengambilan keputusan untuk F_{hitung} atau $F_{(1-\alpha;k,(n-k-1))}$. Jika $F_{hitung} > F_{(1-\alpha;k,(n-k-1))}$ untuk derajat bebas k dan *n-k-1* maka hipótesis H_0 ditolak pada taraf α dan berarti juga bhwa H_1 diterima. Jika $F_{hitung} \leq F_{(1-\alpha;k,(n-k-1))}$ maka H_0 diterima pada taraf (α)

5. Kemudian hitung nilai F_{hitung}.

6. Tuliskan kesimpulan ketika F_{hitung} dibandingkan dengan $F_{(1-\alpha;k,(n-k-1))}$.

Sehingga *F* statistiknya dapat dicari dengan rumus:

$$F = \frac{JKR/k}{JKS/(n-k-1)} \tag{2.32}$$

F statistik inilah yang dipakai untuk menguji kelinearan suatu regresi ganda. Jika $F_{hit} > F_{(1-\alpha;k,(n-k-1))}$ dengan taraf signifikansi yang dipilih, maka dapat disimpulkan bahwa ada hubungan linear antara variabel-variabel bebas dengan variabel terikat.

H. Koefisien Determinasi

Koefisien determinasi (*Coefficient of Determination*), R^2 didefinisikan sebagai berikut (Sudjana: 2001, 107):

$$R^2 = \frac{JKR}{JKT} \tag{2.33}$$

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

I. Pengali Langrange

Menurut Siegel (1985: 145) metode *Pengali Langrange* adalah metode untuk mendapatkan nilai maksimum relatif atau nilai minimum relatif dari sebuah fungsi *F(x, y, z)* yang memenuhi syarat pembatas (*constrain condition*) $\phi(x, y, z)$.

$$G(x, y, z) = F(x, y, z) + k\phi(x, y, z)$$

dengan syarat :

$$\frac{\partial G}{\partial x} = 0 \qquad \frac{\partial G}{\partial y} = 0 \qquad \frac{\partial G}{\partial z} = 0$$

Yang merupakan syarat-syarat perlu untuk memaksimumkan relatif atau untuk meminimumkan relatif. k yang tergantung *x, y, z* dinamakan *Pengali Langrange*.

J. Multikolinearitas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch pada tahun 1934 yang berarti adanya hubungan linier diantara beberapa atau semua variabel bebas dalam model regresi.

1. Pengertian Multikolinearitas

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi yang kuat diantara variabel-variabel bebas (*X*) yang diikutsertakan dalam pembentukan model regresi linier. Jelas bahwa multikolinieritas adalah suatu kondisi yang menyalahi asumsi regresi linier. Tentu saja, multikolinieritas tidak mungkin terjadi apabila variabel bebas (*X*) yang diikutsertakan hanya satu.

Dalam bentuk matriks, multikolinearitas adalah suatu kondisi buruk atau *ill condition* dari matriks $\boldsymbol{X^tX}$ yaitu suatu kondisi yang tidak memenuhi asumsi klasik. Jika multikolinearitas terjadi antara dua variabel atau lebih dalam suatu persamaan regresi, maka nilai perkiraan koefisien dari variabel yang bersangkutan menjadi tak berhingga, sehingga tidak mungkin lagi menduganya. Hal ini disebabkan $\boldsymbol{X^tX}$ menjadi singular atau $\boldsymbol{X^tX}$ mendekati nol. Dalam multikolinearitas terdapat dua jenis hubungan linear yang sempurna (multikolinearitas sempurna) dan hubungan linear kurang sempurna (multikolinearitas kurang sempurna).

a. Multikolinearitas Sempurna

Untuk hubungan yang terdiri dari *k* variabel, mencakup variabel bebas $X_1, X_2, \ldots, X_k$. Hubungan linier yang sempurna atau pasti terjadi jika berlaku hubungan berikut:

$$C_1X_1 + C_2X_2 + \ldots + C_kX_k = 0 \quad (2.34)$$

dengan $C_1, C_2, \ldots, C_k$ merupakan bilangan konstan dan tidak seluruhnya nol atau paling tidak ada satu yang tidak sama dengan nol, yaitu $\exists\, C_j \neq 0$ *(j = 1,2, . . .,k).*

b. Multikolinearitas Kurang Sempurna

Istilah multikolinearitas digunakan dalam arti lebih luas, yaitu mencakup hubungan linier sempurna dan juga dimana variabel-variabel bebas *X interkorelasi*, akan tetapi tidak sempurna seperti hubungan berikut:

$$C_1X_1 + C_2X_2 + \ldots + C_kX_k + \varepsilon_i = 0 \quad (2.35)$$

dengan ε_i : kesalahan pengganggu (*standar error*).

Untuk mengetahui perbedaan antara multikolinearitas sempurna dan multikolinearitas kurang sempurna, diasumsikan $C_2 \neq 0$. Dapat ditunjukkan uintuk setiap observasi ke-*i* persamaan (2.35) menjadi:

$$X_{2i} = -\frac{C_1}{C_2} X_{1i} - \frac{C_3}{C_2} X_{3i} - \ldots - \frac{C_k}{C_2} X_{ki} \quad (2.36)$$

yang menunjukkan bagaimana variabel bebas X_{2i} berhubungan linier secara sempurna dengan variabel lainnya secara keseluruhan atau bagaimana hubungan tersebut dapat diturunkan dari suatu hubungan linier antara variabel bebas-variabel bebas lainnya.

Diasumsikan $C_2 \neq 0$ maka persamaan (2.36) menjadi

$$X_{2i} = -\frac{C_1}{C_2} X_{1i} - \frac{C_3}{C_2} X_{3i} - \ldots - \frac{C_k}{C_2} X_{ki} - \frac{1}{C_2} \varepsilon_i \qquad (2.37)$$

Persamaan diatas menunjukkan bahwa X_{2i} tidak berhubungan linier sempurna dengan variabel lainnya, sebab masih tergantung pada kesalahan pengganggu (ε_i) (*Sumodiningrat, 2002: 281-282*).

2. Konsekuensi Multikolinearitas

a. Multikolinearitas Sempurna

Untuk multikolinearitas yang sempurna, perkiraan koefisien regresi tidak dapat ditentukan dan variansi serta standar errornya tidak terhingga.

Diasumsikan X_{1i} dan X_{2i} berhubungan sedemikian rupa sehingga $X_{2i} = \gamma X_{1i}$, dimana γ = bilangan konstan.

Dari $\hat{\beta} = (\boldsymbol{X^t X})^{-1} \boldsymbol{X^t Y}$, maka

$$\boldsymbol{X^t X} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ X_{11} & X_{12} & X_{13} & \ldots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \ldots & X_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{k1} & X_{k2} & X_{k3} & \ldots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} & \ldots & X_{k1} \\ 1 & X_{12} & X_{22} & \ldots & X_{k2} \\ 1 & X_{13} & X_{23} & \ldots & X_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \ldots & X_{kn} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \ldots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \ldots & \sum X_{1i}X_{ki} \\ \sum X_{2i} & \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \ldots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \ldots & \sum X_{ki}^2 \end{bmatrix}$$

Karena $X_{2i} = \gamma X_{1i}$, maka

$$= \begin{bmatrix} n & \sum X_{1i} & \gamma \sum X_{1i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \gamma \sum {X_{1i}}^2 & \dots & \sum X_{1i} X_{ki} \\ \gamma \sum X_{1i} & \gamma \sum {X_{1i}}^2 & \gamma^2 \sum {X_{1i}}^2 & \dots & \gamma \sum X_{1i} X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i} X_{ki} & \gamma \sum X_{2i} X_{ki} & \dots & \sum {X_{ki}}^2 \end{bmatrix}$$

Berdasarkan teori matriks, nilai determinan suatu matriks tidak berubah apabila suatu baris/ kolom dikalikan dengan suatu bilangan konstan, kemudian baris/ kolom lain dikurangi dengan baris/ kolom tersebut. Dalam hal ini kalikan baris kedua dengan γ kemudian baris ketiga dikurangi dengan baris kedua, maka diperoleh:

$$\boldsymbol{X^t X} = \begin{bmatrix} n & \sum X_{1i} & \gamma \sum X_{1i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \gamma \sum {X_{1i}}^2 & \dots & \sum X_{1i} X_{ki} \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i} X_{ki} & \gamma \sum X_{2i} X_{ki} & \dots & \sum {X_{ki}}^2 \end{bmatrix}$$

Menurut teori matriks, apabila baris/ kolom suatu matriks semua elemennya 0, maka determinan matriks yang bersangkutan nol. Oleh karena determinan $(\boldsymbol{X^t X}) = 0$, maka $\boldsymbol{X^t X}$ adalah matriks singular dan karenanya koefisien regresi tidak dapat ditentukan.

b. Multikolinearitas Kurang Sempurna

Untuk multikolinearitas yang kurang sempurna, masih mungkin untuk menghitung perkiraan koefisien regresi. Tetapi nilai variansi dan standar errornya besar.

Misal untuk regresi linear berganda

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \boldsymbol{\varepsilon}_i$$

Dari $\widehat{\beta} = (\boldsymbol{X^t X})^{-1} \boldsymbol{X^t Y}$, maka

$$\boldsymbol{X^t X} = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \sum X_{1i}X_{2i} & \dots & \sum X_{1i}X_{ki} \\ \sum X_{2i} & \sum X_{1i}X_{2i} & \sum {X_{2i}}^2 & \dots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \dots & \sum {X_{ki}}^2 \end{bmatrix}$$

Karena hubungan linear yang kurang sempurna, diambil $X_{2i} = \gamma X_{1i} + \boldsymbol{\varepsilon}_i$, maka $\boldsymbol{X^t X}$:

$$= \begin{bmatrix} n & \sum X_{1i} & \gamma \sum X_{1i} + \sum \boldsymbol{\varepsilon}_i & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \gamma \sum {X_{1i}}^2 & \dots & \sum X_{1i}X_{ki} \\ \gamma \sum X_{1i} + \sum \boldsymbol{\varepsilon}_i & \gamma \sum {X_{1i}}^2 & \gamma^2 \sum {X_{1i}}^2 + \sum \boldsymbol{\varepsilon}_i & \dots & \gamma \sum X_{1i}X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i}X_{ki} & \gamma \sum X_{2i}X_{ki} & \dots & \sum {X_{ki}}^2 \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{1i} & \gamma \sum X_{1i} + \sum \boldsymbol{\varepsilon}_i & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \gamma \sum {X_{1i}}^2 & \dots & \sum X_{1i}X_{ki} \\ \sum \boldsymbol{\varepsilon}_i & 0 & \sum \boldsymbol{\varepsilon}_i & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{1i}X_{ki} & \gamma \sum X_{2i}X_{ki} & \dots & \sum {X_{ki}}^2 \end{bmatrix}$$

Terlihat bahwa nilai dari $(\boldsymbol{X^t X})^{-1}$ tergantung dari kesalahan pengganggu. Apabila kesalahan pengganggu sangat kecil atau mendekati nol, maka berakibat tidak dapat ditentukan nilainya. Kemudian untuk variansi, karena nilai determinan dari $(\boldsymbol{X^t X})$ kecil, maka nilai variansinya akan cenderung besar. Pengaruh lain dari multikolinearitas adalah ketidakstabilan koefisien regresi. Yakni, koefisien regresi sangat tergantung pada fakta-fakta yang terdapat dalam kumpulan data. Ketidakstabilan tersebut dapat dideteksi dengan cara merubah atau "mengganggu" observasi dalam *Y* dan mengecek kestabilan relatif dalam koefisien-koefisien regresi. Sebagai tambahan dapat juga dengan mengeluarkan salah satu dari kumpulan variabel bebas dan jika multikolinearitas adalah masalah yang serius maka koefisien-koefisien yang lain bisa berubah dalam jumlah yang besar dan mungkin berubah tanda. Ketidakstabilan ini merupakan kondisi yang sangat tidak menguntungkan dalam analisis karena akan mempengaruhi estimasi yang dilakukan, hasil prediksi akan menjadi tidak valid dan juga mempengaruhi kualitas dari model yang dipilih akan sangat baik jika gangguan kecil dalam data tidak merubah keaslian data dan tidak merubah nilai koefisien regresi (*Sumodiningrat, 2002: 283-287*).

3. Mendeteksi Adanya Multikolinearitas

Beberapa teknik telah diperkenalkan untuk mendeteksi adanya multikolinearitas. Dalam hal ini sangat diperlukan sifat-sifat dari prosedur pengecekan yang bisa menunjukkan secara langsung derajat dari masalah multikolinearitas dan memberikan informasi yang dapat membantu dalam menentukan variabel-variabel bebas yang mana yang terlibat. Beberapa teknik tersebut adalah sebagai berikut:

a. Plot Variabel Bebas

Cara yang paling sederhana untuk mendeteksi adanya multikolinearitas adalah dengan memplot hubungan antara variabel-variabel bebas. Diagram yang sering digunakan untuk memplot hubungan antara variabel-variabel penjelas adalah *scater plot* atau diagram pencar, dengan menambahkan garis regresi. Karena *scater plot* hanya menampilkan hubungan dua variabel saja, maka didalam regresi ganda, pengujian dilakukan dengan berpasangan tiap dua variabel. Sehingga jika ada *p*-variabel bebas maka terdapat $\boldsymbol{C_1^p}$ cara. Jika hubungan antara kedua variabel mengikuti garis regresi atau membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut.

b. Pemeriksaan Matriks Korelasi

Langkah yang paling sederhana untuk mengukur adanya multikolinearitas adalah dengan melakukan pemeriksaan terhadap elemen-elemen di luar diagonal r_{ij} $(i \neq j)$ dalam matriks korelasi $\boldsymbol{X}^{*\boldsymbol{t}}\boldsymbol{X}^*$, yang mana kolom dari $\boldsymbol{X}^*$ adalah hasil dari pemusatan dan penyekalaan dari matriks $\boldsymbol{X}$. Jika variabel bebas X_i dan X_j mempunyai hubungan linear yang erat, maka $|r_{ij}|$ yang mengindikasikan korelasi berpasangan dari variabel-variabel bebas akan mendekati satu. Jika digunakan program SPSS dalam melakukan pemeriksaan matriks korelasi maka yang menjadi pedoman suatu model regresi ganda yang bebas multikolinearitas adalah koefisien korelasi antar variabel bebas adalah $-1 \leq r_{ij} \leq 1$. Jika dua variabel mempunyai nilai $r_{ij} = 0$ berarti antara dua variabel tidak terdapat hubungan, tetapi jika dua variabel mempunyai $r_{ij} = +1$ atau $r_{ij} = -1$ maka kedua variabel tersebut mempunyai hubungan sempurna. Menurut Budiono dan Koster (2002: 184), arti koefisien korelasi r_{ij} adalah sebagai berikut:

1. Jika $0{,}7 < r_{ij} < 0{,}9$ atau $-0{,}9 < r_{ij} < -0{,}7$ maka terdapat kolinearitas sangat kuat
2. Jika $0{,}5 < r_{ij} < 0{,}7$ atau $-0{,}7 < r_{ij} < -0{,}5$ maka terdapat kolinearitas kuat
3. Jika $0{,}3 < r_{ij} < 0{,}5$ atau $-0{,}5 < r_{ij} < -0{,}3$ maka terdapat kolinearitas lemah
4. Jika $0 < r_{ij} < 0{,}3$ atau $-0{,}3 < r_{ij} < 0$ maka terdapat kolinearitas sangat lemah

Pemeriksaan korelasi sederhana r_{ij} antar variabel-variabel bebas hanya membantu dalam mendeteksi adanya ketergantungan linear yang erat antar pasangan-pasangan variabel bebas. Sedangkan ketika terdapat lebih dari dua variabel bebas yang terlibat dalam ketergantungan linear yang erat, maka tidak ada jaminan akan terdapat korelasi berpasangan r_{ij} yang besar. Secara umum, pemeriksaan nilai r_{ij} tidak cukup untuk mendeteksi adanya multikolinearitas yang lebih kompleks.

c. VIF *(Variance Inflation Factors)* dan *Tolerance*

Menurut Montgomery, salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinearitas pada regresi linear berganda adalah *Variance Inflation Factors (VIF)*. Adanya multikolinearitas dinilai dari nilai VIF yang dihasilkan. 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.

$$\text{VIF} = \left(1 - {R_j}^2\right)^{-1} \qquad (2.38)$$

disebut sebagai *Variance Inflation Factors*. VIF pada setiap bagian (untuk setiap *j*) dalam model, mengukur kombinasi pengaruh ketergantungan antara variabel-variabel bebas pada variansi dalam bagian tersebut. VIF menunjukkan inflasi yang dialami oleh setiap koefisien regresi di atas nilai idealnya, yaitu di atas nilai yang dialami jika matriks korelasi adalah matriks identitas. Terdapat satu atau dua lebih nilai VIF yang besar menandakan adanya multikolinearitas. Dari praktek-praktek yang banyak dilakukan mengindikasikan bahwa jika ada nilai VIF yang melebihi 10, maka ini menandakan bahwa koefisien-koefisien regresi adalah estimasi yang kurang baik karena pengaruh multikolinearitas. Selama itu VIF juga dapat membantu mengidentifikasi variabel-variabel bebas yang mana yang terlihat dalam masalah multikolinearitas.

Jika digunakan program SPSS untuk menghitung niai VIF maka akan muncul tambahan kolom *Collinearity Statistics* pada tabel *Coefficients*. Di sana akan terdapat subkolom Tolerance yang merupakan indikator dari persentasi variansi dalam penduga yang tidak dapat dihitung oleh variabel bebas. Pedoman suatu model regresi ganda yang bebas multikolinearitas adalah mempunyai nilai VIF disekitar angka 1 dan mempunyai angka *Tollerance* mendekati satu.

$$Tollerance = \frac{1}{VIF} \qquad atau \qquad VIF = \frac{1}{Tollerance} \tag{2.39}$$

jika nilai Tollerance kurang dari 0.1 sebaiknya diselidiki lebih lanjut karena hal ini menandakan adanya multikolinearitas.

d. Sistem Nilai Eigen dari $\boldsymbol{X^tX}$

Nilai eigen dari vektor eigen dalam matriks korelasi $\boldsymbol{X^tX}$ mempunyai peranan penting dalam kasus adanya multikolinearitas dalam kumpulan data dari analisis regresi yang dilakukan. Akar-akar karakteristik atau eigenvalue dari $\boldsymbol{X^tX}$ adalah $\lambda_1, \lambda_2, \ldots, \lambda_k$ yang dapat digunakan untuk mengukur adanya multikolinearitas. Jika ada satu atau lebih hubungan linier di dalam data, maka satu atau lebih dari *eigenvalue* kecil. Sedangkan satu atau lebih *eigenvalue* yang kecil menandakan adanya hubungan linier di dalam kolom-kolom dari variabel bebas *X*. Jadi multikolinearitas akan terjadi jika ada satu atau lebih *eigenvalue* yang kecil.

Multikolinearitas dapat diukur dalam bentuk rasio dari nilai terbesar dan terkecil dari nilai eigen, yaitu $\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$ yang disebut nilai kondisi dari matriks korelasi. Nilai φ yang besar mengindikasikan multikolinearitas yang serius. Nilai kondisi yang terlalu besar menunjukkan ketidakstabilan koefisien regresi terhadap perubahan kecil dalam data variabel bebas.

Montgomery & Peck (1992) memberikan kategori multikolinearitas berdasarkan harga $\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$ jika:

$\varphi < 100$ maka disebut multikolinearitas rendah

$100 \leq \varphi < 1000$ maka disebut multikolinearitas agak kuat

$\varphi \geq 1000$ maka disebut multikolinearitas kuat

Nilai $\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$ mempunyai hubungan dengan nilai r_{ij}. Hal tersebut dapat ditunjukkan pada perumusan berikut ini. Untuk persamaan dengan dua variabel bebas dapat dicari nilai eigen dari matriks $\boldsymbol{X^tX}$ dengan menyelesaikan persamaan berikut:

$$\boldsymbol{det\,(\lambda I - X^tX) = 0}$$

$$\left|\begin{bmatrix} \boldsymbol{\lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\lambda} \end{bmatrix} - \begin{bmatrix} \boldsymbol{1} & \boldsymbol{r_{12}} \\ \boldsymbol{r_{12}} & \boldsymbol{1} \end{bmatrix}\right| = \boldsymbol{0}$$

$$\left|\begin{bmatrix} \boldsymbol{\lambda - 1} & \boldsymbol{-r_{12}} \\ \boldsymbol{-r_{12}} & \boldsymbol{\lambda - 1} \end{bmatrix}\right| = \boldsymbol{0}$$

$$(\lambda - 1)^2 - {r_{12}}^2 = 0$$

$$(\lambda - 1)^2 = {r_{12}}^2$$

$$(\lambda - 1) = r_{12}$$

$$\lambda = r_{12} + 1$$

$r_{12} \rightarrow -1$, maka λ semakin kecil $\rightarrow \lambda_{min}$

$r_{12} \rightarrow 1$, maka λ semakin besar $\rightarrow \lambda_{maks}$

BAB III

PEMBAHASAN

A. Penggunaan *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge* untuk mengatasi masalah multikolinearitas.

Berbagai masalah penyimpangan asumsi-asumsi dalam analisis regresi sering terjadi diantaranya multikolinearitas, autokorelasi, heteroskedastisitas. Dalam bab ini akan dibahas lebih lanjut tentang penggunaan *Ridge Trace* dan *Variance Inflation Factors* (*VIF*) pada Regresi *Ridge* untuk mengatasi masalah multikolinearitas yang terjadi pada regresi linear berganda. Sebelumnya akan dijelaskan terlebih dahulu tentang estimator Regresi *Ridge*.

1. Estimator Regresi Ridge

Estimasi Ridge untuk koefisien regresi dapat diperoleh dengan menyelesaikan suatu bentuk dari persamaan normal regresi. Asumsikan bahwa bentuk standar dari model regresi linear ganda adalah sebagai berikut:

$$Y_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \boldsymbol{\varepsilon}$$

Parameter penting yang membedakan regresi ridge dari metode kuadrat terkecil adalah *c*. Tetapan bias *c* yang relatif kecil ditambahkan pada diagonal utama matriks $\boldsymbol{X^t X}$, sehingga koefisien estimator *regresi ridge* dipenuhi dengan besarnya tetapan bias *c*. (Hoerl dan Kennard: 1970, 235).

Dalam prakteknya, perhitungan estimator regresi Ridge dengan menyelesaikan persamaan diatas sangatlah rumit, oleh sebab itu dilakukan penyederhanaan dengan membawanya kedalam bentuk notasi matriks. Estimator Ridge diperoleh dengan meminimumkan jumlah kuadrat galat untuk model:

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

atau

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta}$$

dengan menggunakan metode pengali Langrange yang meminimumkan fungsi:

$$\boldsymbol{\varepsilon}^{\boldsymbol{t}}\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta_R})^{\boldsymbol{t}}(\boldsymbol{Y} - \boldsymbol{X\beta_R})$$

dengan syarat pembatas

$$\boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{\beta_R} - \boldsymbol{k}^2 = \boldsymbol{0}$$

$$\boldsymbol{G} = (\boldsymbol{Y} - \boldsymbol{X\beta_R})^{\boldsymbol{t}}(\boldsymbol{Y} - \boldsymbol{X\beta_R}) + \boldsymbol{c}(\boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{\beta_R} - \boldsymbol{k}^2)$$

yang memenuhi syarat

$$\left.\frac{\partial G}{\partial \beta_R}\right|_{\widehat{\beta}_R} = 0$$

Dimana *c* pengali Langrange tidak bergantung pada β_R dan *c* konstanta positif berhingga.

Dicari $\hat{\beta}_R$ dengan memecah $\left.\frac{\partial G}{\partial \beta_R}\right|_{\widehat{\beta}_R} = 0$

$$\boldsymbol{G} = (\boldsymbol{Y} - \boldsymbol{X\beta_R})^{\boldsymbol{t}}(\boldsymbol{Y} - \boldsymbol{X\beta_R}) + \boldsymbol{c}\left(\boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{\beta_R} - \boldsymbol{k^2}\right)$$

$$= \boldsymbol{Y^tY} - \boldsymbol{Y^tX\beta_R} - \boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{X^tY} + \boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{X^tX\beta_R} + \boldsymbol{c}(\boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{\beta_R} - \boldsymbol{k^2})$$

$$= \boldsymbol{Y^tY} + \boldsymbol{2\beta_R}^{\boldsymbol{t}}\boldsymbol{X^tY} + \boldsymbol{2\beta_R}^{\boldsymbol{t}}\boldsymbol{X^tX\beta_R} + \ \boldsymbol{c}(\boldsymbol{\beta_R}^{\boldsymbol{t}}\boldsymbol{\beta_R} - \boldsymbol{k^2})$$

$$\left.\frac{\partial G}{\partial \beta_R}\right|_{\hat{\beta}_R} = 0$$

$$-\boldsymbol{2X^tY} + \boldsymbol{2X^tX\widehat{\beta}_R} + \boldsymbol{2cI\widehat{\beta}_R} = \boldsymbol{0}$$

$$-\boldsymbol{X^tY} + \boldsymbol{X^tX\widehat{\beta}_R} + \boldsymbol{cI\widehat{\beta}_R} \ = \boldsymbol{0}$$

$$\boldsymbol{X^tX\widehat{\beta}_R} + \boldsymbol{cI\widehat{\beta}_R} \quad = \boldsymbol{X^tY}$$

$$(\boldsymbol{X^tX} + \boldsymbol{cI})\boldsymbol{\widehat{\beta}_R} \quad = \boldsymbol{X^tY}$$

$$\boldsymbol{\widehat{\beta}_R} = (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X^tY} \tag{3.1}$$

dimana $\boldsymbol{\widehat{\beta}_R} = (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X^tY}$ dengan $0 \leq c \leq \infty$, itulah yang disebut sebagai Estimator Regresi Ridge. $c \geq 0$ adalah nilai konstan yang dipilih sebagai indeks dari kelas estimator.

$$\boldsymbol{L} = -\boldsymbol{2X^tY} + \boldsymbol{2X^tX\widehat{\beta}_R} + \boldsymbol{2cI\widehat{\beta}_R}$$

adalah turunan pertama dari *G* terhadap β_R, agar *G* minimum maka $\frac{\partial L}{\partial \widehat{\beta}_R}$ harus lebih besar dari nol atau bernilai positif.

$$\frac{\partial L}{\partial \hat{\beta}_R} = \boldsymbol{2X^tX} + \boldsymbol{2cI} > 0$$

$$\boldsymbol{X^tX} + \boldsymbol{cI} > 0$$

agar diperoleh fungsi G minimum maka dipilih $c > 0$ sedemikian sehingga $\boldsymbol{X^tX} + \boldsymbol{cI} > 0$.

Estimator regresi ridge juga dapat dinyatakan dalam bentuk:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \boldsymbol{WX^tY}$$

dimana $\boldsymbol{W} = (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}$ (3.2)

2. Hubungan Estimator Regresi Ridge dengan Penduga Kuadrat Terkecil

Secara matematis terdapat hubungan antara $\hat{\beta}_R$ dengan $\hat{\beta}$. Estimator regresi Ridge merupakan suatu bentuk transformasi linear dari penduga kuadrat terkecil, karena:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} &= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X^tY} \\
&= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}(\boldsymbol{X^tX})(\boldsymbol{X^tX})^{-\boldsymbol{1}}\boldsymbol{X^tY} \\
&= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}(\boldsymbol{X^tX})\widehat{\boldsymbol{\beta}} \\
&= \boldsymbol{W}(\boldsymbol{X^tX})\widehat{\boldsymbol{\beta}}
\end{aligned}$$

Atau

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = (\boldsymbol{X^tX} + \boldsymbol{cI})^{-\boldsymbol{1}}[(\boldsymbol{X^tX} + \boldsymbol{cI}) - \boldsymbol{cI}]\widehat{\boldsymbol{\beta}}$$

$$= [(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}(\boldsymbol{X^tX} + \boldsymbol{cI}) - \boldsymbol{cI}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}]\hat{\boldsymbol{\beta}}$$

$$= [(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}(\boldsymbol{X^tX} + \boldsymbol{cI}) - \boldsymbol{c}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}]\hat{\boldsymbol{\beta}}$$

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

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \boldsymbol{Z_k}\hat{\boldsymbol{\beta}}$$

dengan $\boldsymbol{Z_k} = [\boldsymbol{I} + \boldsymbol{c}(\boldsymbol{X^tX})^{-1}]^{-1}$

Atau

$$\boldsymbol{Z_k} = \boldsymbol{W}(\boldsymbol{X^tX})$$

$$= [\boldsymbol{I} - \boldsymbol{c}(\boldsymbol{X^tX} + \boldsymbol{kI})^{-1}]$$

$$= \boldsymbol{I} - \boldsymbol{kW} \qquad (3.3)$$

dengan W dinyatakan dalam persamaan (3.2).

3. Sifat-Sifat Estimator Regresi *Ridge*

Ada beberapa sifat yang dimiliki estimator regresi ridge, diantaranya adalah sebagai berikut:

a. Nilai Ekspektasi dari Estimator Regresi Ridge

$$\boldsymbol{E}(\hat{\boldsymbol{\beta}}_{\boldsymbol{R}}) = \boldsymbol{E}[(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tY}]$$

$$= \boldsymbol{E}[(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}(\boldsymbol{X^tX})(\boldsymbol{X^tX})^{-1}\boldsymbol{X^tY}]$$

$$= \boldsymbol{E}\left[(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}(\boldsymbol{X^tX})\widehat{\boldsymbol{\beta}}\right]$$

$$= \boldsymbol{E}[(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}(\boldsymbol{X^tX})]\boldsymbol{E}(\widehat{\boldsymbol{\beta}})$$

$$= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tX\beta}$$

Bias sebesar $(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tX}$ dari β

Jika memanfaatkan hubungan antara $\hat{\beta}_R$ dengan β, maka diperoleh nilai ekspektasi dari estimator Regresi Ridge adalah sebagai berikut:

$$\boldsymbol{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) = [\boldsymbol{I} - \boldsymbol{c}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}]\boldsymbol{\beta}$$

$$= \boldsymbol{\beta} - \boldsymbol{c}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{\beta}$$

Bias sebesar $\boldsymbol{c}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{\beta}$, $\quad 0 \leq c \leq \infty$

b. Matriks Variansi – Kovariansi

$$\boldsymbol{kov}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) = \boldsymbol{E}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{t}}\right)$$

$$= \boldsymbol{E}[((\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tY})((\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tY})^{\boldsymbol{t}}]$$

$$= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^t}\boldsymbol{E}(\boldsymbol{Y^tY})\boldsymbol{X}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}$$

$$= (\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tX}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{E}(\boldsymbol{Y^tY})$$

Diketahui $\boldsymbol{E}(\boldsymbol{Y^tY}) = \boldsymbol{\sigma^2}$ sehingga diperoleh:

$$\boldsymbol{kov}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}) = (\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{X^tX}(\boldsymbol{X^tX} + \boldsymbol{cI})^{-1}\boldsymbol{\sigma^2}$$

c. Jumlah Kuadrat Kesalahan

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

$$= \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right)^{\boldsymbol{t}}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right) + \left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \widehat{\boldsymbol{\beta}}\right)^{\boldsymbol{t}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \widehat{\boldsymbol{\beta}}\right)$$

Bukti:

$$\boldsymbol{SSE}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right) = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})^{\boldsymbol{t}}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}})$$

$$= \boldsymbol{Y}^{\boldsymbol{t}}\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{t}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{Y} - \boldsymbol{Y}^{\boldsymbol{t}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{t}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}$$

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

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}{}^{\boldsymbol{t}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}$$

d. Variansi

Menurut hasil yang diperoleh pada bagian (2.11) dan (2.12) dapat disimpulkan bahwa jika kovariansi antara $\hat{\beta}_i$ dan $\hat{\beta}_j$ adalah σ^2 maka variansi dari $\hat{\beta}_j$ adalah $\sigma^2 trace$, sehingga:

$$\boldsymbol{var}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right) = \boldsymbol{\sigma}^{\boldsymbol{2}}\boldsymbol{trace}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}$$

Bukti:

$$\boldsymbol{var}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}\right) = (\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}[\boldsymbol{\sigma}^{\boldsymbol{2}}\boldsymbol{trace}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}]$$

$$= (\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}[\boldsymbol{\sigma}^{\boldsymbol{2}}\boldsymbol{traceIX}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}]$$

$$= \boldsymbol{\sigma}^{\boldsymbol{2}}\boldsymbol{trace}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{X} + \boldsymbol{cI})^{-\boldsymbol{1}}$$

4. Mendeteksi Multikolinearitas dengan Metode Ridge

Ada dua metode untuk mendeteksi multikolinearitas yang berhubungan dengan Regresi Ridge. Metode pertama dihubungkan dengan efek multikolinearitas terhadap error antara penduga kuadrat terkecil dan nilai sebenarnya dari koefisien regresi. Metode kedua berhubungan dengan ketidakstabilan penduga kuadrat terkecil dalam menghadapi perubahan kecil dalam data.

a. *Variance Inflation Factors (VIF)*

Metode pertama diasosiasikan dengan *Variance Inflation Factors (VIF)*. Ketelitian estimasi kuadrat terkecil dari koefisien regresi diukur oleh variansi ini, yang mana proporsional dengan σ^2, variansi dari bentuk residual dalam model regresi. Konstanta kesebandingan tersebut mengacu pada suatu bentuk yang disebut *Variance Inflation Factors (VIF)*. Menurut Montgomery, salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinearitas pada regresi linear berganda adalah *Variance Inflation Factors (VIF)*. Adanya multikolinearitas dinilai dari nilai VIF yang dihasilkan. 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. VIF berkorespondensi dengan setiap estimasi kuadrat terkecil dari koefisien regresi. Nilai VIF dapat dilihat dari persamaan di bawah ini:

$$VIF = \frac{1}{1-{R_j}^2}$$

Dengan ${R_j}^2$ adalah koefisien determinasi ganda, jika ${R_j}^2$ besar maka VIF akan besar pula dan jika VIF > 10 maka terdapat multikolinearitas. VIF menunjukkan inflasi yang dialami oleh setiap koefisien regresi di atas nilai idealnya. Yaitu, di atas nilai yang dialami jika matriks korelasi adalah matriks identitas. Terdapat satu atau dua lebih nilai VIF yang besar menandakan adanya multikolinearitas. Dari praktek-praktek yang banyak dilakukan mengindikasikan bahwa jika ada nilai VIF yang melebihi 10, maka ini menandakan bahwa koefisien-koefisien regresi adalah estimasi yang kurang baik karena pengaruh multikolinearitas. Selama itu VIF juga dapat membantu mengidentifikasi variabel-variabel bebas yang mana yang terlihat dalam masalah multikolinearitas.

b. Ridge Trace

Metode yang kedua untuk mendeteksi multikolinearitas dengan menggunakan *Ridge Trace* atau jejak *Ridge*. Salah satu kesulitan utama dalam menggunakan Regresi *Ridge* adalah dalam menentukan nilai *c* yang tepat. (*Hoerl and Kennard, 1970, 237*), pencipta Regresi *Ridge* menganjurkan untuk menggunakan suatu grafik yang mereka sebut sebagai *Ridge Trace*. *Ridge Trace* adalah plot dari estimator Regresi *Ridge* dengan berbagai kemungkinan nilai tetapan bias *c,* konstanta *c* mencerminkan jumlah bias dalam estimator $\hat{\beta}(c)$. Bila $c = 0$ maka estimator $\hat{\beta}(c)$ akan bernilai sama dengan kuadrat terkecil β, tetapi cenderung lebih stabil dari pada estimator kuadrat terkecil.

Plot ini menggambarkan koefisien *Regresi Ridge* sebagai fungsi dari *c*. Nilai dari *c* berada pada interval (0.1). Pemilihan tetapan bias *c* merupakan masalah yang perlu diperhatikan. Tetapan bias yang diinginkan adalah tetapan bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien yang relatif stabil.

5. Pengujian Hipotesis

Untuk uji keberartian regresi langkah-langkah pengujiannya sama seperti pada model regresi linear berganda tanpa intersep, yaitu sebagai berikut:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \boldsymbol{\varepsilon}$$

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

   $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ (tidak ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

   $H_1: \exists \beta_j \neq 0$, $j = 1, 2, 3, \ldots, k$ (ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

2. Menentukan tingkat signifikansi (α)

3. Memilih dan menuliskan uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-F.

$$R^2 = \frac{JKR}{JKT} \qquad\qquad F_{hitung} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

dengan k = banyaknya variabel bebas dalam model

n = banyaknya data

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

   Jika $F_{hitung} > F_{(1-\alpha;k,(n-k-1))}$ untuk derajat bebas k dan *n-k-1* maka hipótesis H_0 ditolak pada taraf a dan berarti juga bhwa H_1 diterima.

   Jika $F_{hitung} \leq F_{(1-\alpha;k,(n-k-1))}$ maka H_0 diterima pada taraf (α)

5. Kemudian hitung nilai F_{hitung}.

6. Tuliskan kesimpulan ketika F_{hitung} dibandingkan dengan $F_{(1-\alpha;k,(n-k-1))}$.

Untuk uji keberartian koefisien korelasi digunakan :

1. Hipotesis

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

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

2. Tingkat signifikansi (α)

3. Uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-t.

$$t = \frac{b_i}{s_{b_i}}$$

b_i= Koefisien regresi variabel-i

S_{b_i}= Standar error variabel-i

4. Kriteria keputusan

H_0 diterima jika $|t_{hitung}| \leq t_{(\alpha/2,n-k-1)}$

H_0 ditolak jika $|t_{hitung}| > t_{(\alpha/2,n-k-1)}$

dengan *k* = banyaknya variabel bebas dalam model

n = banyaknya data

B. Penerapan Regresi Ridge

1. Data

Contoh kasus untuk memperjelas penggunaan regresi *Ridge* dalam memperbaiki model regresi yang memuat multikolinearitas pada tulisan ini adalah menggunakan data yang diambil dari "*Gujarati (2006: 58)*" . Sumber data yang digunakan dalam penelitian ini adalah sumber data sekunder dengan perubahan tahun.

"Seorang peneliti melakukan suatu survey terhadap permintaan ayam di AS tahun 1982-2004. Data yang diambil adalah konsumsi ayam per kapita (pon) sebagai Y, pendapatan siap konsumsi riil per kapita (Dollar AS) sebagai X_1, harga eceran ayam riil per pon (sen) sebagai X_2, harga babi eceran riil per pon (sen) sebagai X_3, harga sapi eceran riil per pon (sen) sebagai X_4". Babi dan sapi dipilih sebagai faktor yang mungkin mempengaruhi permintaan ayam karena di AS babi dan sapi merupakan barang pengganti ayam. Akan diselidiki apakah X_1, X_2, X_3, X_4 berpengaruh secara individual terhadap Y. Data survey jumlah permintaan ayam di AS tahun 1982-2004 ditunjukkan dalam tabel berikut:

Tabel 1. Data survey jumlah permintaan ayam di AS tahun 1982-2004

Tahun	Y	X_1	X_2	X_3	X_4
1982	27,8	397,5	42,2	50,7	78,3
1983	29,9	413,3	38,1	52	79,2
1984	29,8	439,2	40,3	54	79,2
1985	30,8	459,7	39,5	55,3	79,2
1986	31,2	492,9	37,3	54,7	77,4
1987	33,3	528,6	38,1	63,7	80,2
1988	35,6	560,3	39,3	69,8	80,4
1989	36,4	624,6	37,8	65,9	83,9
1990	36,7	666,4	38,4	64,5	85,5
1991	38,4	717,8	40,1	70	93,7
1992	40,4	768,2	38,6	73,2	106,1
1993	40,3	843,3	39,8	67,8	104,8
1994	41,8	911,6	39,7	79,1	114

1995	40,4	931,1	52,1	95,4	124,1
1996	40,7	1021,5	48,9	94,2	127,6
1997	40,1	1165,9	58,3	12,5	142,9
1998	42,7	1349,6	57,9	129,9	143,6
1999	44,1	1449,4	56,5	117,6	139,2
2000	46,7	1575,5	63,7	130,9	165,5
2001	50,6	1759,1	61,6	129,8	203,3
2002	50,1	1994,2	58,9	128	219,6
2003	51,7	2258,1	66,4	141	221,6
2004	52,9	2478,7	70,4	168,2	232,6

Sumber: *Gujarati (2006: 58)*

2. Pendeteksian Multikolinearitas

Akan dibuat suatu model yang sesuai dan diperiksa apakah terdapat mulltikolinearitas diantara variabel bebas X_i. Analisa regresi dengan metode kuadrat terkecil terhadap data menghasilkan nilai estimator parameter (tabel 2) dengan daftar anava (tabel 3).

Tabel 2. Estimator Parameter Regresi Kuadrat Terkecil

Peubah	Penduga Parameter	Simpangan Baku
Konstan	37,232	3,718
X_1	0,005	0,005
X_2	-0,611	0,163

X_3	0,198	0,064
X_4	0,070	0,051

Tabel 3. ANAVA Untuk Data Awal

Sumber Variansi	Jumlah Kuadrat	DK	RJK	Fhitung
Regresi	1127,259	4	281,815	73,871
Sisa	68,670	18	3,815	
Total	1195,929	22		

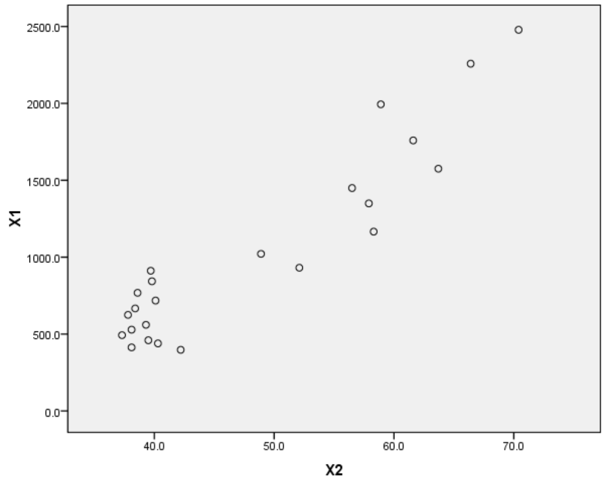
Dari data di atas diperoleh persamaan regresi linear berganda seperti pada persamaan berikut:

$$\hat{Y} = 37{,}232 + 0{,}005X_1 - 0{,}611X_2 + 0{,}198X_3 + 0{,}070X_4$$

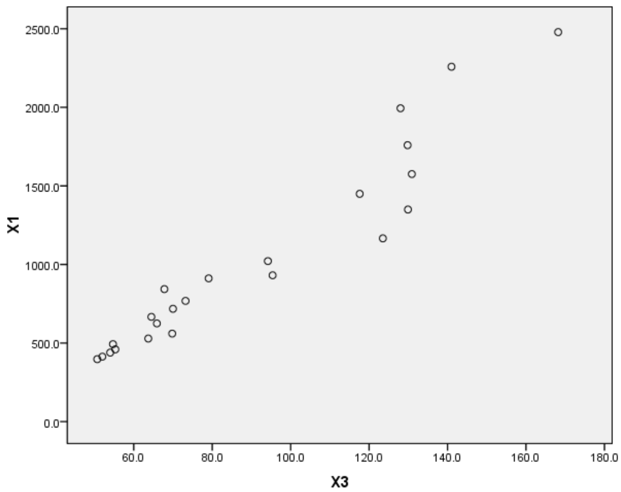
Untuk pendeteksian multikolinearitas ada beberapa cara yang dapat digunakan , antara lain:

1. **Plot Variabel Bebas**.

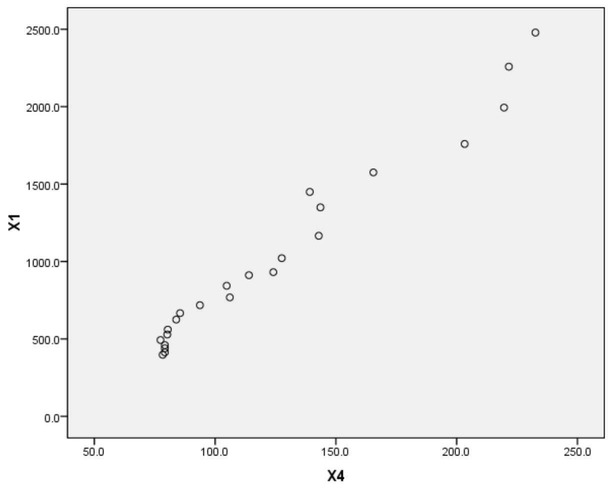
Gambar 1. Plot Variabel Bebas



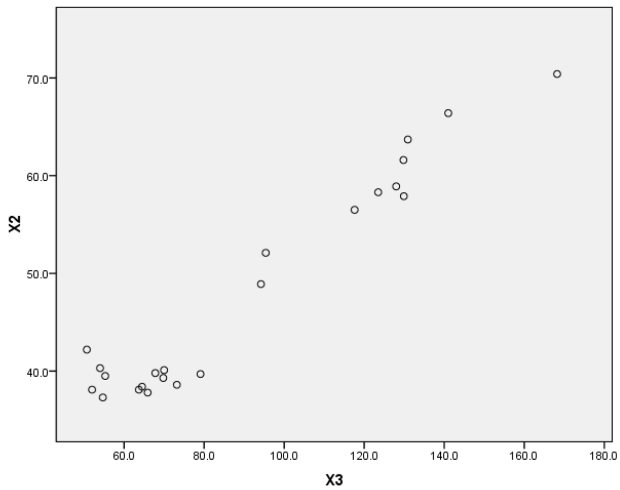
Hubungan antara pendapatan siap konsumsi riil per kapita (X_1) dengan harga eceran ayam riil per pon (X_2) terlihat membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut.



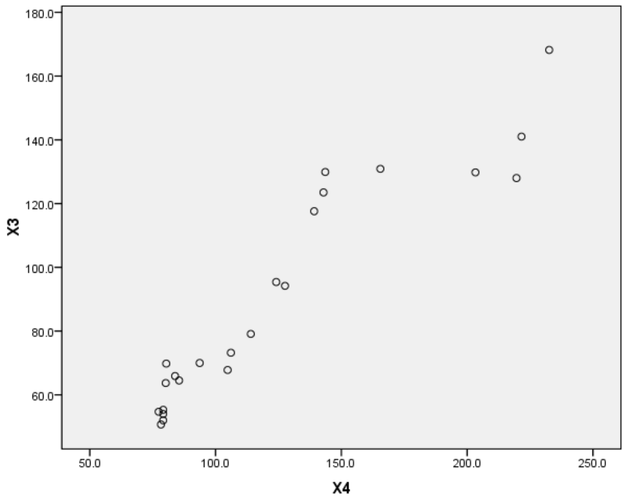
Hubungan antara pendapatan siap konsumsi riil per kapita (X_1) dengan harga babi eceran riil per pon (X_3) terlihat membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut



Hubungan antara pendapatan siap konsumsi riil per kapita (X_1) dengan harga sapi eceran riil per pon (X_4) terlihat membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut



Hubungan antara harga eceran ayam riil per pon (X_2) dengan harga babi eceran riil per pon (X_3) terlihat membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut



Hubungan antara harga babi eceran riil per pon (X_3) dengan harga sapi eceran riil per pon (X_4) terlihat membentuk pola garis lurus maka terdapat hubungan linear diantara kedua variabel tersebut

Model Summary[b]

Model	R	R Square	Adjusted R Square
1	.971[a]	.943	.930

a. Predictors: (Constant), X4, X2, X3, X1

b. Dependent Variable: Y

Pada *scater plot–scater plo*t pasangan variabel bebas diatas terlihat bahwa hubungan semua pasangan variabel cenderung mengikuti garis regresi dan memiliki *R*-square yang tinggi mendekati 1, maka dapat disimpulkan bahwa terdapat hubungan linear diantara pasangan-pasangan variabel bebas tersebut.

2. Koefisien Korelasi Parsial

Cara kedua adalah dengan melihat keeratan hubungan antara dua variabel bebas atau yang lebih dikenal dengan istilah korelasi. Untuk memperoleh koefisien korelasi parsial antar variabel bebasnya dihitung dengan menggunakan persamaan (2.22), maka diperoleh:

$r_{X_1X_1} = 1$

$r_{X_1X_2} = 0{,}782$ $\quad r_{X_2X_2} = 1$

$r_{X_1X_3} = 0{,}708$ $\quad r_{X_2X_3} = 0{,}917$ $\quad r_{X_3X_3} = 1$

$r_{X_1X_4} = 0{,}881$ $\quad r_{X_2X_4} = 0{,}744$ $\quad r_{X_3X_4} = 0{,}602$ $\quad r_{X_4X_4} = 1$

Sehingga korelasi parsial antar variabel bebasnya dapat dibuat dalam bentuk matriks korelasi C sebagai berikut:

$$C = \begin{bmatrix} 1 & 0{,}782 & 0{,}708 & 0{,}881 \\ 0{,}782 & 1 & 0{,}917 & 0{,}744 \\ 0{,}708 & 0{,}917 & 1 & 0{,}602 \\ 0{,}881 & 0{,}744 & 0{,}602 & 1 \end{bmatrix}$$

Dari matriks C terlihat bahwa korelasi antara variabel bebasnya sangat tinggi mendekati nilai 1. Ini menunjukkan bahwa adanya multikolinearitas antara variabel bebasnya.

3. Determinan Matriks Korelasi

Dari matriks korelasi C dapat dihitung determinannya, yaitu:

$$C = \begin{bmatrix} 1 & 0,782 & 0,708 & 0,881 \\ 0,782 & 1 & 0,917 & 0,744 \\ 0,708 & 0,917 & 1 & 0,602 \\ 0,881 & 0,744 & 0,602 & 1 \end{bmatrix}$$

$$|C| = 0,109022$$

Nilai determinan dari matriks korelasi C mendekati 0, ini menunjukkan bahwa tingkat multikolinearitas tinggi.

4. Sistem nilai eigen dari X^tX

Pendeteksian multikolinearitas dilakukan dengan metode *Eigenvalue*. Jika terdapat satu atau lebih nilai eigen yang kecil, menandakan adanya ketergantungan linear. Multikolinearitas terjadi jika ada satu atau lebih nilai eigen yang kecil mendekati nol. Berikut adalah output dengan SPSS:

Eigenvalues of Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	3.856994	96.42	96.42	1.00
2	0.101874	2.55	98.97	37.86
3	0.030928	0.77	99.74	124.71
4	0.010203	0.26	100.00	378.01

Some Condition Numbers greater than 100. Multicollinearity is a MILD problem.

Eigenvector of Correlations

No.	Eigenvalue	X1	X2	X3	X4
1	3.856994	-0.502336	-0.496482	-0.501402	-0.499760
2	0.101874	0.443851	-0.616016	-0.369007	0.536055
3	0.030928	-0.223004	0.572766	-0.702094	0.359545
4	0.010203	-0.707760	-0.214405	0.345680	0.577592

Dari output SPSS di atas, dapat dilihat nilai eigen $\lambda_1 = 3{,}856994$, $\lambda_2 = 0{,}101874$, $\lambda_3 = 0{,}030928$, $\lambda_4 = 0{,}010203$. Keempat nilai eigennya mendekati nol ini berarti terdapat multikolinearitas. Selain itu multikolinearitas dapat diukur dalam bentuk rasio dan nilai terbesar dan terkecil dari nilai eigen, yaitu $= \frac{\lambda_{maks}}{\lambda_{min}}$, yang disebut nilai kondisi dari matriks korelasi. Nilai φ yang besar mengindikasikan multikolinearitas yang serius.

$$\varphi = \frac{\lambda_{maks}}{\lambda_{min}}$$

$$= \frac{3{,}856994}{0{,}010203}$$

$$= 378{,}0255$$

Dari semua proses pemeriksaan di atas, ternyata setiap proses pemeriksaan menunjukkan masalah multikolinearitas dalam data yang dianalisis. Untuk mengatasi masalah multikolinearitas tersebut, maka dilakukan analisis Regresi *Ridge*.

3. Penaksiran Model Regresi Ridge

1. Uji VIF

Untuk menentukan apakah suatu model memiliki gejala multikolinearitas, akan digunakan cara VIF dan uji korelasi. Dengan cara ini akan dilihat apakah nilai VIF untuk masing-masing variabel lebih besar dari 10 atau tidak. Bila nilai VIF lebih besar dari 10, maka diindikasikan model tersebut memiliki gejala multikolinearitas.

Untuk data di atas dilakukan uji VIF dengan hasil sebagai berikut:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	37.232	3.718		10.015	.000		
	X1	.005	.005	.420	1.024	.319	.019	52.701
	X2	-.611	.163	-.922	-3.753	.001	.053	18.901
	X3	.198	.064	.948	3.114	.006	.034	29.051
	X4	.070	.051	.485	1.363	.190	.025	39.761

a. Dependent Variable: Y

Dapat dilihat bahwa seluruh variabel bebas memiliki nilai VIF lebih besar dari 10, maka dapat disimpulkan bahwa model regresi ini memiliki masalah multikolinearitas.

2. Proses Pemusatan dan Penskalaan

Untuk menghilangkan kondisi buruk yang tidak menguntungkan yang diakibatkan oleh adanya multikolinearitas dalam data yang dianalisis dan juga untuk memudahkan kerja dalam proses pendeteksian maupun dalam penanganan multikolinearitas maka dilakukan proses pemusatan dan penskalaan terhadap data atau variabel dengan program NCSS.

Tabel 4. Hasil Proses Pemusatan dan Penskalaan

NO	Y^*	${X_1}^*$	${X_2}^*$	${X_3}^*$	${X_4}^*$
1	-0,33568	-0,21517	-0,1087	-0,23501	-0,18677
2	-0,27629	-0,20984	-0,1856	-0,22732	-0,18313
3	-0,27912	-0,2011	-0,14434	-0,21548	-0,18313
4	-0,25084	-0,19418	-0,15934	-0,20778	-0,18313
5	-0,23953	-0,18297	-0,20061	-0,21133	-0,19042
6	-0,18014	-0,17092	-0,1856	-0,15806	-0,17908
7	-0,11509	-0,16023	-0,1631	-0,12195	-0,17827
8	-0,09247	-0,13853	-0,19123	-0,14503	-0,1641
9	-0,08398	-0,12442	-0,17998	-0,15332	-0,15762
10	-0,0359	-0,10707	-0,14809	-0,12076	-0,12442
11	0,020657	-0,09006	-0,17622	-0,10182	-0,07422
12	0,017829	-0,06472	-0,15372	-0,13379	-0,07948
13	0,060251	-0,04167	-0,15559	-0,06689	-0,04223
14	0,020657	-0,03509	0,076981	0,029599	-0,00134
15	0,029142	-0,00458	0,016962	0,022495	0,012833

16	0,012173	0,044155	0,193268	0,195943	0,07478
17	0,085704	0,106151	0,185766	0,233829	0,077615
18	0,125297	0,139832	0,159508	0,161016	0,0598
19	0,198828	0,182389	0,294551	0,239749	0,166284
20	0,309124	0,244351	0,255163	0,233237	0,319331
21	0,294983	0,323694	0,204522	0,222582	0,385327
22	0,340233	0,412757	0,345192	0,299538	0,393424
23	0,37417	0,487206	0,420216	0,460554	0,437962

Dalam proses pengestimasian regresi ridge, pemilihan tetapan bias *c* merupakan hal yang paling penting dalam penelitian ini, penentuan tetapan bias *c* ditempuh melalui pendekatan nilai VIF dan gambar Ridge Trace. Nilai dari koefisien $\hat{\beta}(c)$ dengan berbagai kemungkinan tetapan bias *c* dapat dilihat pada tabel.

Tabel 5. Nilai VIF $\hat{\beta}\ (c)$ dengan berbagai nilai *c*

Nilai *c*	VIF $\hat{\beta}_1(c)$	VIF $\hat{\beta}_2(c)$	VIF $\hat{\beta}_3(c)$	VIF $\hat{\beta}_4(c)$
0,000	52,7010	18,9013	29,0510	39,7614
0,001	44,1915	17,4068	26,0450	33,8727
0,002	37,6646	16,1540	23,5985	29,3226
0,003	32,5450	15,0835	21,5643	25,7256
0,004	28,4521	14,1544	19,8429	22,8268

0,005	25,1262	13,3378	18,3649	20,4516
0,006	22,3850	12,6126	17,0805	18,4771
0,007	20,0973	11,9630	15,9528	16,8151
0,008	18,1671	11,3768	14,9538	15,4003
0,009	16,5224	10,8445	14,0623	14,1842
0,010	15,1086	10,3585	13,2614	13,1296
0,020	7,6117	7,0921	8,2129	7,3079
0,030	4,7950	5,3093	5,7230	4,9301
0,040	3,3948	4,1862	4,2672	3,6631
0,050	2,5789	3,4169	3,3291	2,8819
0,060	2,0524	2,8597	2,6839	2,3542
0,070	1,6881	2,4393	2,2187	1,9751
0,080	1,4227	2,1123	1,8707	1,6904
0,090	1,2218	1,8516	1,6030	1,4694
0,100	1,0652	1,6398	1,3921	1,2934
0,200	0,4239	0,6829	0,5246	0,5318
0,300	0,2478	0,3919	0,2939	0,3089
0,400	0,1720	0,2634	0,1979	0,2111
0,500	0,1318	0,1946	0,1481	0,1588
0,600	0,1073	0,1531	0,1185	0,1271
0,700	0,0911	0,1258	0,0992	0,1061
0,800	0,0796	0,1068	0,0857	0,0914

0,900	0,0709	0,0928	0,0757	0,0805
1,000	0,0642	0,0821	0,0681	0,0720

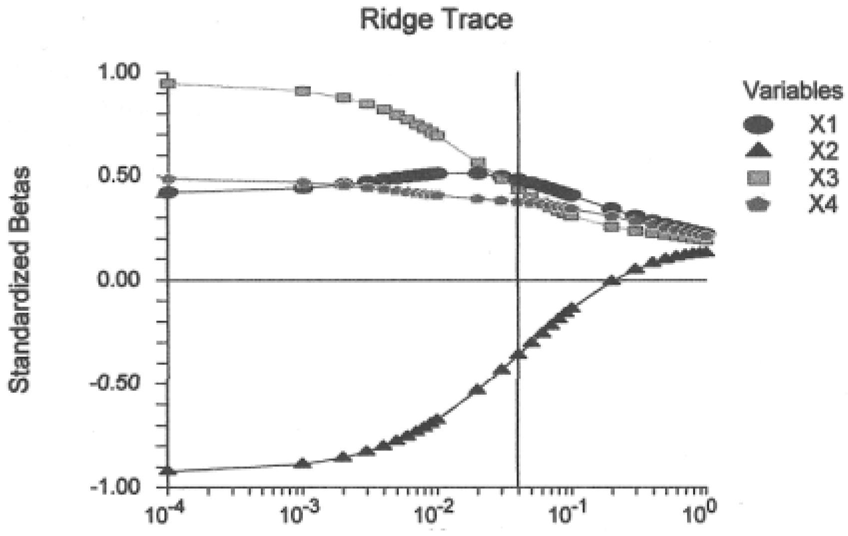
Dari tabel di atas tampak bahwa mulai tetapan bias ***c*** = 0,000 sampai pada *c* = 1,000, VIF koefisien estimator $\hat{\beta}\ (c)$ semakin lama semakin kecil. Nilai VIF yang diambil adalah VIF yang relatif dekat dengan satu, sedangkan nilai koefisien estimator parameter $\hat{\beta}\ (c)$ dengan berbagai kemungkinan tetapan bias *c* dapat dilihat pada tabel berikut:

Tabel 6. Nilai $\hat{\beta}\ (c)$ dengan berbagai harga c

Nilai *c*	$\hat{\beta}_1(c)$	$\hat{\beta}_2(c)$	$\hat{\beta}_3(c)$	$\hat{\beta}_4(c)$
0,000	0,4199	-0,9216	0,9479	0,4855
0,001	0,4413	-0,8879	0,9101	0,4683
0,002	0,4580	-0,8572	0,8762	0,4547
0,003	0,4711	-0,8289	0,8457	0,4437
0,004	0,4814	-0,8027	0,8180	0,4349
0,005	0,4896	-0,7782	0,7926	0,4276
0,006	0,4961	-0,7553	0,7692	0,4215
0,007	0,5011	-0,7337	0,7476	0,4165
0,008	0,5051	-0,7134	0,7275	0,4122
0,009	0,5081	-0,6942	0,7088	0,4085
0,010	0,5104	-0,6759	0,6914	0,4054

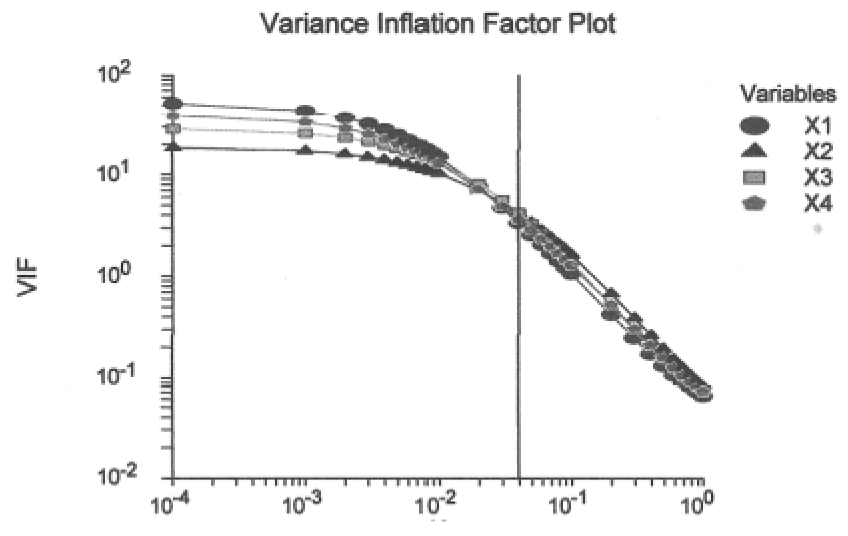
0,020	0,5112	-0,5332	0,5638	0,3884
0,030	0,4974	-0,4352	0,4861	0,3804
0,040	0,4811	-0,3624	0,4341	0,3742
0,050	0,4654	-0,3056	0,3971	0,3684
0,060	0,4508	-0,2598	0,3696	0,3627
0,070	0,4376	-0,2219	0,3484	0,3573
0,080	0,4256	-0,1900	0,3318	0,3520
0,090	0,4147	-0,1628	0,3184	0,3469
0,100	0,4049	-0,1391	0,3074	0,3421
0,200	0,3404	-0,0063	0,2550	0,3048
0,300	0,3058	0,0502	0,2359	0,2811
0,400	0,2832	0,0805	0,2248	0,2643
0,500	0,2668	0,0985	0,2167	0,2514
0,600	0,2538	0,1100	0,2101	0,2408
0,700	0,2432	0,1175	0,2044	0,2319
0,800	0,2341	0,1225	0,1993	0,2241
0,900	0,2262	0,1257	0,1946	0,2172
1,000	0,2191	0,1277	0,1902	0,2109

Atas dasar koefisien estimator pada tabel 6 dapat dibuat suatu gambar *Ridge Trace.*



c

Gambar 2. Ridge Trace



c

Gambar 3. VIF Plot

Dari berbagai harga *c* yang ada, nilai VIF mulai tampak ada penurunan pada *c* sebesar 0,02. harga *c* yang memberikan nilai VIF relatif dekat dengan

1, yaitu pada c = 0,02 ini menunjukkan bahwa pada c = 0,02 koefisien $\hat{\beta}$ lebih stabil. Dengan demikian persamaan Regresi Ridge yang diperoleh jika c yang diambil sebesar 0,02 yaitu:

$$\hat{Y}^* = 0{,}5112X_1{}^* - 0{,}5332X_2{}^* + 0{,}5638X_3{}^* + 0{,}3884X_4{}^*$$

4. Uji Keberartian Regresi

Persamaan regresi yang diperoleh:

$$\hat{Y}^* = 0{,}5112X_1{}^* - 0{,}5332X_2{}^* + 0{,}5638X_3{}^* + 0{,}3884X_4{}^*$$

Kemudian akan diuji keberartian dari model tersebut, untuk melakukan pengujian regresi linear dilakukan sebagai berikut:

1. Hipotesis

   $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (tidak ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

   $H_1: \exists\beta_j \neq 0$, j = *1, 2, 3, 4* (ada hubungan linear antara variabel-variabel bebas dengan variabel terikat)

2. Tingkat signifikansi (α)

   $\alpha = 0.05$

3. Uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-F.

$$F_{hitung} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

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

   Jika $F_{hitung} > F_{(1-\alpha;k,(n-k-1))}$ untuk derajat bebas k dan n-k-1 maka hipótesis H_0 ditolak pada taraf α dan berarti juga bahwa H_1 diterima.

5. Nilai F_{hitung}. Hasil pengujian tersebut dapat dibentuk dalam tabel ANAVA sebagai berikut:

Tabel 7. ANAVA Ridge

S. Varians	JK	DK	KT	F_{hitung}	F_{tabel}
Regresi	0,9426	4	0,23565	73,41	2,928
Sisa	0,0574	18	0,0032		
Total	1	22			

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

$$= 0{,}9426$$

$$F_{hitung} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

$$= \frac{0{,}9426/4}{(1-0{,}9426)/(23-4-1)}$$

$$= \frac{0{,}23565}{0{,}0032}$$

$$= 73{,}641$$

6. Hasil: dengan taraf nyata α = 0,05, maka $F_{tabel} = 2{,}928$ karena $F_{hitung} > F_{tabel}$, maka dapat dinyatakan ada hubungan linear antara variabel-variabel bebas X_1, X_2, X_3, X_4 dengan variabel terikat *Y*.

Untuk mengetahui apakah koefisien yang diperoleh berarti atau tidak dilakukan pengujian sebagai berikut:

1. Hipotesis

   $H_0 \colon \beta_i = 0$ (koefisien regresi tidak signifikan)

   $H_1 \colon \beta_i \neq 0$ (koefisien regresi signifikan)

2. Tingkat signifikansi (α)

   $\alpha = 0{,}05$

3. Uji statistik yang digunakan, dalam hal ini uji statistik yang digunakan adalah uji-t.

$$t = \frac{b_i}{S_{b_i}}$$

   b_i= Koefisien regresi variabel-*i*

   S_{b_i}= Standar error variabel-*i*

4. Kriteria keputusan

H_0 diterima jika $|t_{hitung}| \leq t_{(\alpha/2,n-k-1)}$

H_0 ditolak jika $|t_{hitung}| > t_{(\alpha/2,n-k-1)}$

5. Hasil: dengan taraf nyata $\alpha = 0{,}05$ maka $t_{(0{,}025,18)} = 2{,}101$, maka disimpulkan

	b_i	$S\{b_i\}$	$\|t_{hitung}\|$	t_{tabel}	Kesimpulan
1	0,5112	0,4992	1,024	2,101	Tidak Signifikan
2	-0,5332	0,1421	3,753		Signifikan
3	0,5638	0,1811	3,114		Signifikan
4	0,3884	0,2850	1,363		Tidak Signifikan

Proses pengembalian X^* ke X bentuk semula dengan $\bar{Y} = 39{,}66957$, $\bar{X}_1 = 1035{,}065$, $\bar{X}_2 = 47{,}99565$, $\bar{X}_3 = 90{,}4$, $\bar{X}_4 = 124{,}4304$, $S_Y = 7{,}37295$, $S_{X_1} = 617{,}847$, $S_{X_2} = 11{,}11721$, $S_{X_3} = 35{,}22369$, $S_{X_4} = 51{,}49974$.

Sehingga persamaan regresinya menjadi:

$$\hat{Y} = 30{,}38082 + 0{,}00057X_1 - 0{,}24031X_2 + 0{,}00909X_3 + 0{,}00536X_4$$

Dari berbagai analisis data di atas dapat disimpulkan :

1. Estimasi yang di peroleh dengan menggunakan metode kuadrat terkecil yaitu:

$$\hat{Y} = 37{,}232 + 0{,}005X_1 - 0{,}611X_2 + 0{,}198X_3 + 0{,}070X_4$$

2. Adanya multikolinearitas dalam persamaan regresi tersebut, ini terlihat dari besarnya nilai korelasi antar variabel bebas yang mendekati 1, nilai determinan dari matriks korelasi mendekati nol dan seluruh nilai VIF dari $(\boldsymbol{X^tX})^{-1}$ **b**esar (lebih dari 10) yaitu VIF untuk $X_1 = 52{,}701$, $X_2 = 18{,}901$, $X_3 = 29{,}051$, $X_4 = 39{,}761$
3. Dengan menggunakan metode regresi Ridge, yaitu dengan menambah tetapan bias c pada diagonal matriks $(\boldsymbol{X^tX})$ diperoleh pada nilai *c* = 0,02 nilai VIF relatif dekat dengan 1 (kurang dari 10), yaitu VIF untuk $X_1 = 7{,}6117$, $X_2 = 7{,}0921$, $X_3 = 8{,}2129$, $X_4 = 7{,}3079$
4. sehingga pada c = 0,02 koefisien $\hat{\beta}$ lebih stabil. Dengan demikian persamaan Regresi Ridge yang diperoleh jika *c* yang diambil sebesar 0,02 yaitu:

$$\hat{Y}^* = 0{,}5112{X_1}^* - 0{,}5332{X_2}^* + 0{,}5638{X_3}^* + 0{,}3884{X_4}^*$$

5. Estimasi yang diproleh dengan menggunakan Regresi *Ridge* yaitu:

$$\hat{Y} = 30{,}38082 + 0{,}00057X_1 - 0{,}24031X_2 + 0{,}00909X_3 + 0{,}00536X_4$$

6. Nilai korelasi determinasi estimator mendekati 1 yaitu R^2 = 94,26 %, hal ini menunjukkan bahwa estimator yang diperoleh sudah dapat digunakan dan variansi jumlah konsumsi ayam di AS dapat dijelaskan oleh jumlah pendapatan siap konsumsi riil perkapita, harga ayam eceran per pon, harga babi eceran per pon dan harga sapi eceran per pon.

BAB IV

PENUTUP

Dari serangkaian pembahasan yang dilakukan pada bab-bab sebelumnya, maka dapat diambil suatu kesimpulan dan saran.

A. Kesimpulan

1. Masalah multikolinearitas yang terjadi pada regresi linear berganda pada skripsi ini diselesaikan dengan metode Regresi Ridge yang dikhususkan dengan menggunakan *Ridge Trace* dan VIF (*Variance Inflation Factors).* Dengan menggunakan Ridge Trace, yaitu dengan menambah tetapan bias *c* pada diagonal matriks $(\boldsymbol{X^t X})$ akan diperoleh pada nilai *c* tertentu nilai VIF relatif dekat dengan 1 (kurang dari 10), sehingga koefisien $\hat{\beta}$ lebih stabil.
2. Contoh penerapan Regresi *Ridge* pada skripsi ini diambil kasus permintaan ayam di AS tahun 1982-2004 yang mengalami masalah multikolinearitas, yaitu hubungan antara konsumsi ayam per kapita (pon) sebagai *Y* dengan pendapatan siap konsumsi riil per kapita (Dollar AS) sebagai X_1, harga eceran ayam riil per pon (sen) sebagai X_2, harga babi eceran riil per pon (sen) sebagai X_3, harga sapi eceran riil per pon (sen) sebagai X_4. Dengan penggunaan *Ridge Trace* dan *VIF* diperoleh nilai *VIF* > 10 yang relatif mendekati 1 pada *c* = 0,02. Ini menunjukkan koefisien $\hat{\beta}$ lebih stabil, sehingga diperoleh

persamaan Regresi *Ridge* : $\hat{Y} = 30{,}38082 + 0{,}00057X_1 - 0{,}24031X_2 + 0{,}00909X_3 + 0{,}00536X_4$. Dengan demikian penggunaan *Ridge Trace* dan *VIF* pada Regresi *Ridge* sudah dapat mengatasi masalah multikolinearitas pada kasus di atas.

B. Saran

Banyak metode untuk mengatasi masalah multikolinearitas yaitu memperbesar ukuran sampel, spesifikasi model, penggunaan informasi tambahan, pengaplikasian estimator bias dan metode lainnya. Analis dapat memilih salah satu diantara semua metode yang lebih baik dari Metode Kuadrat Terkecil. *Ridge Trace* dan *VIF* pada Regresi *Ridge* dapat digunakan untuk menyelesaikan model yang mengandung multikolinearitas dengan menambah tetapan bias *c* pada diagonal matriks $(\boldsymbol{X^tX})$**.** Ini dikarenakan melalui model ini diusahakan memperoleh variansi yang mengecil dengan menentukan nilai β sehingga diperoleh keadaan yang lebih stabil.

DAFTAR PUSTAKA

Algifari. 2000. *Analisis Regresi (teori, kasus dan solusi).* Yogyakarta: BPFE

Andi. 2009. *SPSS 17 untuk Pengolahan Data Statistik*. Yogyakarta: ANDI

Bain, L.J. 1992. *Introduction to Probability and Mathematical Statistics*. Jakarta: Gramedia Pustaka.

Budi Murtiyasa 2010. *Matriks dan Sistem Persamaan Linear*. Surakarta: UMS

Draper N & Smith H. 1992. *Analisis Regresi Terapan*. Edisi 2. (Terjemahan: Bambang- Sumantri). Jakarta. PT. Gramedia Pustaka Utama.

Dwi Priyatno. 2010. *Paham Analisa Statistik Data dengan SPSS.* Yogyakarta: MEDIAKOM.

Damodar, Gujarati. 1999. *Ekonometrika Dasar*. Terjemahan: Sumarno Zain. Jakarta. Erlangga.

Damodar, Gujarati. 2006. *Dasar-Dasar Ekonometrika* Edisi ketiga Terjemahan: Mc. Graw Hill. Jakarta. Erlangga.

Hoerl, A.E. & R.W. Kennard. 1970. *Ridge Regression: Biased Estimation for Nonorthogonal Problems*. http://statgen.ucr.edu/file/STAT288/hoerl70a.pdf (Diakses 15 Nov. 2010).

Anton. 1987. *Aljabar Linier Elementer.* Terjemahan: Pantur-Nyoman Susila. Jakarta. Erlangga.

Joko Sulistiyo. 2010. *6 Hari Jago SPSS 17.* Yogyakarta: Cakrawala.

Montgomery, D.C & Peck. 1992. *Introduction to linear Regression Analysis*. New York. John Wiley & sons.

Sembiring R.K. 2003. *Analisis Regresi edisi kedua.* Bandung: ITB Bandung.

Siegel. 1985. *Statistik Nonparametrik*. Jakarta. Gramedia Pustaka.

Sudjana. 2003. *Teknik Analisis Regresi dan Korelasi*. Bandung: Tarsito Bandung.

Ruminta. 2009. *Matriks Persamaan Linier dan Pemrograman Linier*. Bandung: Rekayasa Sains.

Seber, G.A.F & Lee, A.J. 2003. Linear Regression Analysis. Ney Jersey. John Willey & Sons.

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

Sumodiningrat. 1994. *Ekonometrika Pengantar*. Yogyakarta: BPFE.

Tutz, G. & Ulbricht, J. 2006. *Penalized with Correlation Based Penalty*. http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper486.pdf (Diakses 16 Oktober 2010).

Weisberg, S. 1980. *Applied Linear Regression*. New York: John Wiley & Sons

Walpole, R.E & Raymond, H. 1995. *Ilmu Peluang dan Statistika untuk Insiyur dan Ilmuwan.* Terjemahan: Sembiring, R.K. Edisi 4. Jakarta: Gramedia Pustaka.

http://www.stat.purdue.edu/~jennings/stat512/notes/topic5a.pdf (Diakses 16 Oktober 2010)

Lampiran 1

Output Pendeteksian Multikolinearitas

1. Uji Multikolinearitas dengan VIF

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1127.259	4	281.815	73.871	.000[a]
	Residual	68.670	18	3.815		
	Total	1195.929	22			

a. Predictors: (Constant), X4, X2, X3, X1

b. Dependent Variable: Y

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	37.232	3.718		10.015	.000		
	X1	.005	.005	.420	1.024	.319	.019	52.701
	X2	-.611	.163	-.922	-3.753	.001	.053	18.901
	X3	.198	.064	.948	3.114	.006	.034	29.051
	X4	.070	.051	.485	1.363	.190	.025	39.761

a. Dependent Variable: Y

Collinearity Diagnostics[a]

Model	Dimension	Condition Index	Variance Proportions				
			(Constant)	X1	X2	X3	X4
1	1	1.000	.00	.00	.00	.00	.00
	2	5.567	.04	.01	.00	.00	.00
	3	22.239	.12	.03	.03	.23	.08
	4	43.117	.75	.45	.23	.06	.45
	5	57.961	.09	.50	.74	.72	.47

a. Dependent Variable: Y

2. Uji Korelasi Parsial

Correlations

Control Variables			X1	X2	X3	X4
Y	X1	Correlation	1.000	.782	.708	.881
		Significance (2-tailed)	.	.000	.000	.000
		df	0	20	20	20
	X2	Correlation	.782	1.000	.917	.744
		Significance (2-tailed)	.000	.	.000	.000
		df	20	0	20	20

X3	Correlation	.708	.917	1.000	.602
	Significance (2-tailed)	.000	.000	.	.003
	df	20	20	0	20
X4	Correlation	.881	.744	.602	1.000
	Significance (2-tailed)	.000	.000	.003	.
	df	20	20	20	0

3. Sistem Nilai Eigen

Eigenvalues of Correlations

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1	3.856994	96.42	96.42	1.00
2	0.101874	2.55	98.97	37.86
3	0.030928	0.77	99.74	124.71
4	0.010203	0.26	100.00	378.01

Some Condition Numbers greater than 100. Multicollinearity is a MILD problem.

Eigenvector of Correlations

No.	Eigenvalue	X1	X2	X3	X4
1	3.856994	-0.502336	-0.496482	-0.501402	-0.499760
2	0.101874	0.443851	-0.616016	-0.369007	0.536055
3	0.030928	-0.223004	0.572766	-0.702094	0.359545
4	0.010203	-0.707760	-0.214405	0.345680	0.577592

Lampiran 2

Output Ridge Regression dengan Program NCSS

1. Output Regresi Ridge

Ridge Regression Report

Descriptive Statistics Section

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
X1	23	1035.065	617.847	397.5	2478.7
X2	23	47.99565	11.11721	37.3	70.4
X3	23	90.4	35.22369	50.7	168.2
X4	23	124.4304	51.49974	77.4	232.6
Y	23	39.66957	7.37295	27.8	52.9

Correlation Matrix Section

	X1	X2	X3	X4	Y
X1	1.000000	0.931681	0.957131	0.985878	0.947171
X2	0.931681	1.000000	0.970112	0.928469	0.839958
X3	0.957131	0.970112	1.000000	0.940567	0.912392
X4	0.985878	0.928469	0.940567	1.000000	0.935355
Y	0.947171	0.839958	0.912392	0.935355	1.000000

Least Squares Multicollinearity Section

Independent Variable	Variance Inflation	R-Squared Vs Other X's	Tolerance
X1	52.7010	0.9810	0.0190
X2	18.9013	0.9471	0.0529
X3	29.0510	0.9656	0.0344
X4	39.7614	0.9748	0.0252

Since some VIF's are greater than 10, multicollinearity is a problem.

Eigenvalues of Correlations

No.	*Eigenvalue*	Incremental *Percent*	Cumulative *Percent*	Condition *Number*
1	3.856994	96.42	96.42	1.00
2	0.101874	2.55	98.97	37.86
3	0.030928	0.77	99.74	124.71
4	0.010203	0.26	100.00	378.01

Some Condition Numbers greater than 100. Multicollinearity is a MILD problem.

Eigenvector of Correlations

No.	Eigenvalue	X1	X2	X3	X4
1	3.856994	-0.502336	-0.496482	-0.501402	-0.499760
2	0.101874	0.443851	-0.616016	-0.369007	0.536055
3	0.030928	-0.223004	0.572766	-0.702094	0.359545
4	0.010203	-0.707760	-0.214405	0.345680	0.577592

2. Pendekatan Nilai VIF

Ridge Regression Report

Variance Inflation Factor Section

k	X1	X2	X3	X4
0.000000	52.7010	18.9013	29.0510	39.7614
0.001000	44.1915	17.4068	26.0450	33.8727
0.002000	37.6646	16.1540	23.5985	29.3226
0.003000	32.5450	15.0835	21.5643	25.7256
0.004000	28.4521	14.1544	19.8429	22.8268
0.005000	25.1262	13.3378	18.3649	20.4516
0.006000	22.3850	12.6126	17.0805	18.4771
0.007000	20.0973	11.9630	15.9528	16.8151
0.008000	18.1671	11.3768	14.9538	15.4003
0.009000	16.5224	10.8445	14.0623	14.1842
0.010000	15.1086	10.3585	13.2614	13.1296
0.020000	7.6117	7.0921	8.2129	7.3079
0.030000	4.7950	5.3093	5.7230	4.9301
0.040000	3.3948	4.1862	4.2672	3.6631
0.040000	3.3948	4.1862	4.2672	3.6631
0.050000	2.5789	3.4169	3.3291	2.8819
0.060000	2.0524	2.8597	2.6839	2.3542
0.070000	1.6881	2.4393	2.2187	1.9751
0.080000	1.4227	2.1123	1.8707	1.6904
0.090000	1.2218	1.8516	1.6030	1.4694
0.100000	1.0652	1.6398	1.3921	1.2934
0.200000	0.4239	0.6829	0.5246	0.5318
0.300000	0.2478	0.3919	0.2939	0.3089
0.400000	0.1720	0.2634	0.1979	0.2111
0.500000	0.1318	0.1946	0.1481	0.1588
0.600000	0.1073	0.1531	0.1185	0.1271
0.700000	0.0911	0.1258	0.0992	0.1061
0.800000	0.0796	0.1068	0.0857	0.0914
0.900000	0.0709	0.0928	0.0757	0.0805
1.000000	0.0642	0.0821	0.0681	0.0720

2. Pendekatan Koefisien Estimator Parameter $\hat{\beta}\ (c)$

Ridge Regression Report

Standardized Ridge Regression Coefficients Section

k	X1	X2	X3	X4
0.000000	0.4199	-0.9216	0.9479	0.4855
0.001000	0.4413	-0.8879	0.9101	0.4683
0.002000	0.4580	-0.8572	0.8762	0.4547
0.003000	0.4711	-0.8289	0.8457	0.4437
0.004000	0.4814	-0.8027	0.8180	0.4349
0.005000	0.4896	-0.7782	0.7926	0.4276
0.006000	0.4961	-0.7553	0.7692	0.4215
0.007000	0.5011	-0.7337	0.7476	0.4165
0.008000	0.5051	-0.7134	0.7275	0.4122
0.009000	0.5081	-0.6942	0.7088	0.4085
0.010000	0.5104	-0.6759	0.6914	0.4054
0.020000	0.5112	-0.5332	0.5638	0.3884
0.030000	0.4974	-0.4352	0.4861	0.3804
0.040000	0.4811	-0.3624	0.4341	0.3742
0.040000	0.4811	-0.3624	0.4341	0.3742
0.050000	0.4654	-0.3056	0.3971	0.3684
0.060000	0.4508	-0.2598	0.3696	0.3627
0.070000	0.4376	-0.2219	0.3484	0.3573
0.080000	0.4256	-0.1900	0.3318	0.3520
0.090000	0.4147	-0.1628	0.3184	0.3469
0.100000	0.4049	-0.1391	0.3074	0.3421
0.200000	0.3404	-0.0063	0.2550	0.3048
0.300000	0.3058	0.0502	0.2359	0.2811
0.400000	0.2832	0.0805	0.2248	0.2643
0.500000	0.2668	0.0985	0.2167	0.2514
0.600000	0.2538	0.1100	0.2101	0.2408
0.700000	0.2432	0.1175	0.2044	0.2319
0.800000	0.2341	0.1225	0.1993	0.2241
0.900000	0.2262	0.1257	0.1946	0.2172
1.000000	0.2191	0.1277	0.1902	0.2109

3. ANAVA Ridge

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971[a]	.9426	.930	.055239

a. Predictors: (Constant), X4, X2, X3, X1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.9426	4	.23565	73.640	.000[a]
	Residual	.0574	18	.0032		
	Total	1	22			

a. Predictors: (Constant), X4, X2, X3, X1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	VIF
1	(Constant)	30.38082	.00166		
	X1	.00057	. 00166	.4811	3.3948

X2	-.24031	. 10244	-.3624	4.1862
X3	.00909	. 00326	.4341	4.2672
X4	.00536	.00206	.3742	3.6631

a. Dependent Variable: Y

Lampiran 3

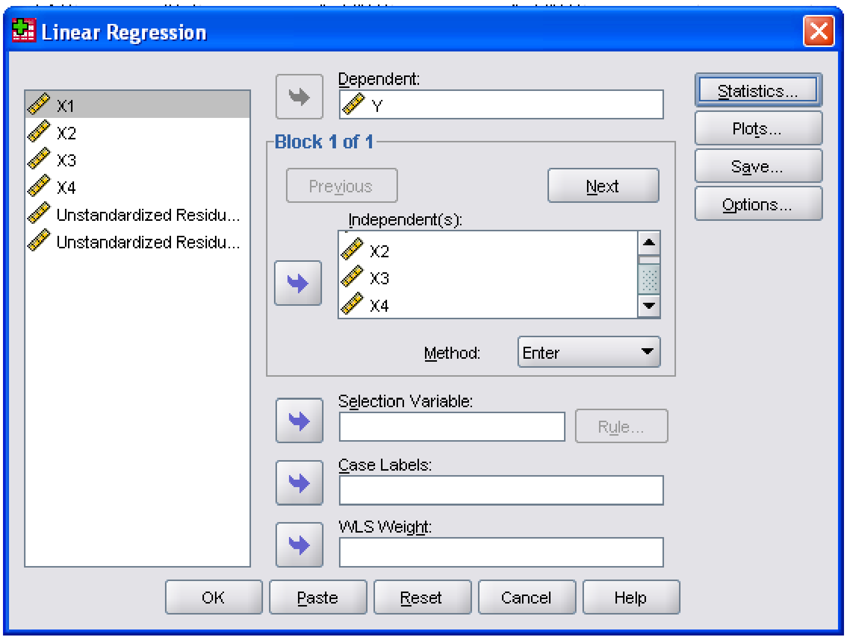
Langkah-Langkah Analisis pada Program SPSS dan NCSS

Uji Multikolinearitas dengan VIF

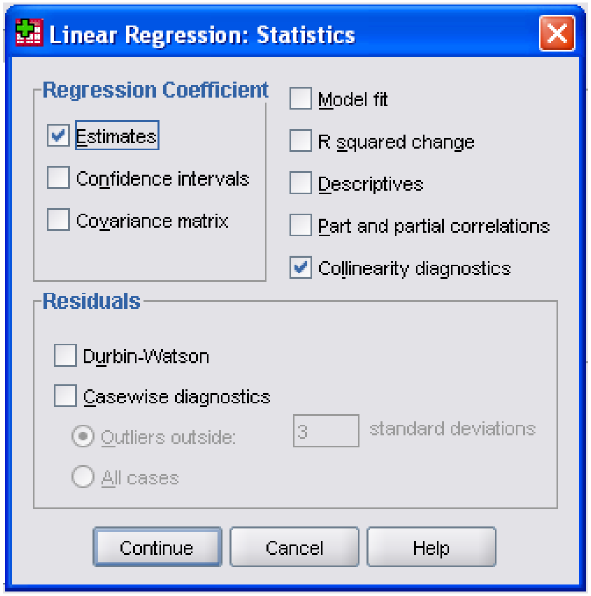
1. Definisikan variabel dan masukkan data ke SPSS.

2 Pilih menu ***Analyze → Regression → Linear***.

3. Masukkan variabel Y ke ***Dependent*** dan X1, X2, X3, X4 ke ***Independent***.



4. Klik ***Statistics*** lalu pilih ***Estimates*** dan ***Collinearity Diagnostics***.



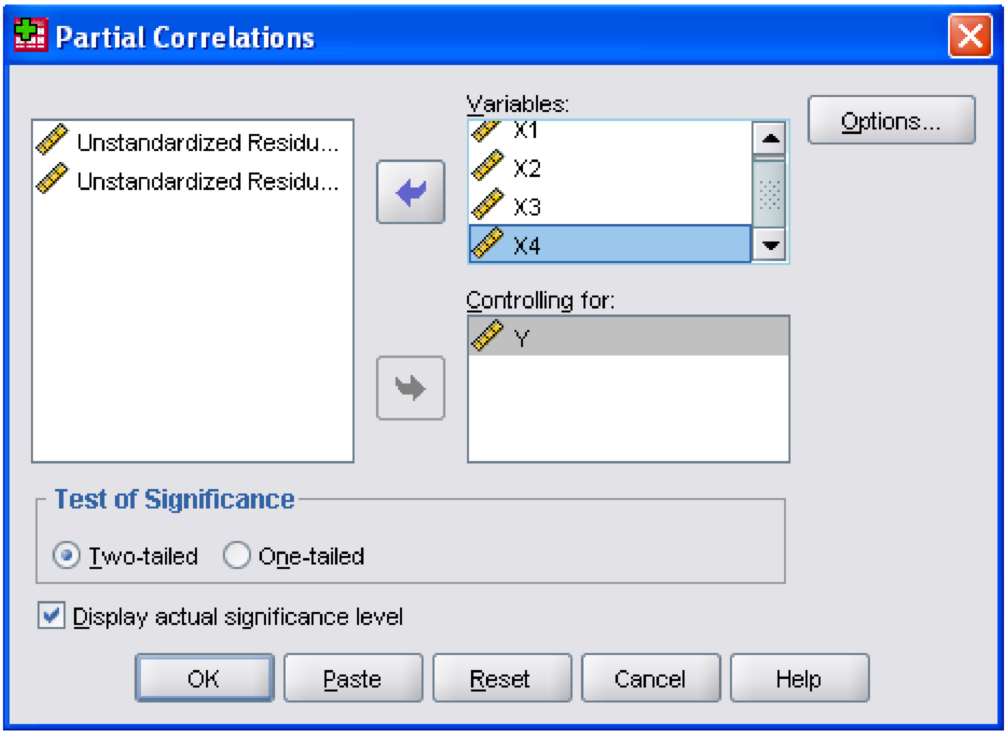
5. Klik ***Continue*** lalu ***OK***.

Uji Korelasi Parsial

1. Definisikan variabel dan masukkan data ke SPSS

2. Pilih menu ***Analyze → Correlate → Partial***

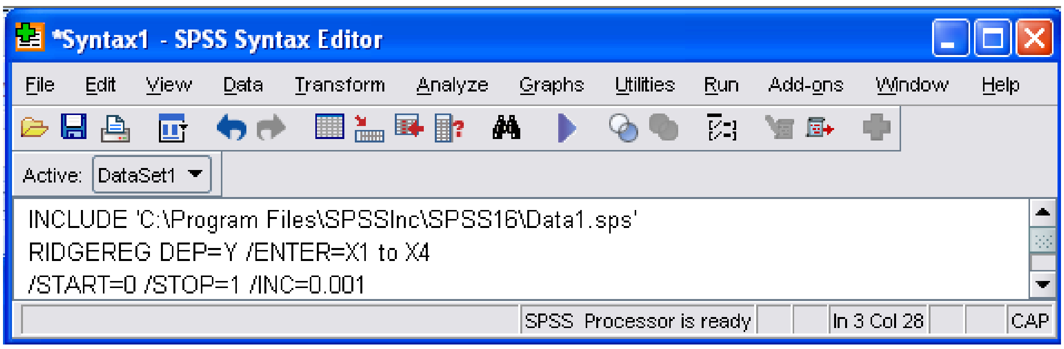
3. Masukkan variabel Y ke ***Controlling for*** dan X1, X2, X3, X4 ke ***Variables***



4. Klik ***OK.***

Mencari Nilai dari *c*

untuk mencari nilai dari *c* maka kita menggunakan Software SPSS 16 khususnya dengan menggunakan Syntax yang ada pada SPSS 16 tersebut adapun syntax yang digunakan untuk mendapatkan nilai *c* sebagai berikut:



Mencari Ridge Regression dengan program NCSS

1. Definisikan variabel dan masukkan data ke NCSS
2. Pilih menu ***Analyze → Regression → Ridge Regression***
3. klik **OK**

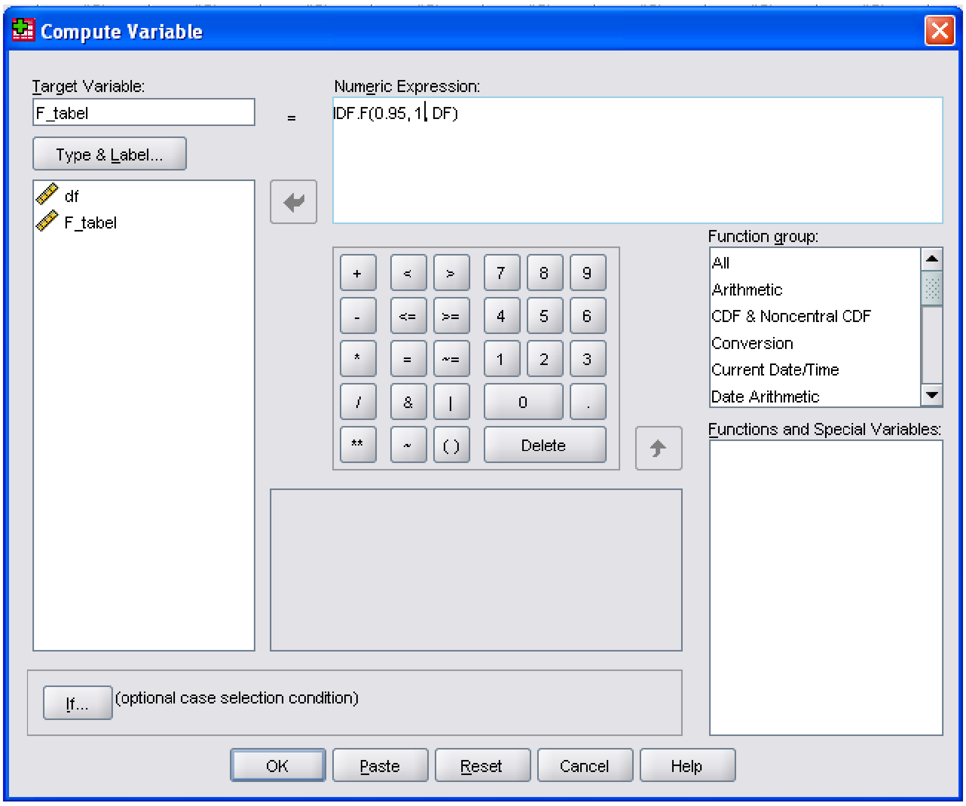
Membuat tabel F

Tabel F memiliki 2 degree of freedom (df) yaitu DF1 (numerator) dan DF2 (denominator).

Akan dibuat table F dengan tingkat Signifikansi (α) 5%, DF1 dari 1 sampai 7 dan DF2 dari 1 sampai 20.

1. Definisikan variabel dengan nama df.
2. Isi variabel dengan angka 1 sampai 20.
3. Klik ***Transform → Compute***.
4. Isi kolom ***Target Variable*** dengan F_tabel.
5. Isi kolom ***Numeric Expression*** dengan IDF.F(0.95, 1, df) sampai 20 atau bisa dipilih melalui kolom ***Function and Special Variables***.

Disini dipakai 95% (100% dikurangi tingkat signifikansi 5%) dan df adalah nama variabel yang sudah dibuat.



6. Klik ***OK***.

Tabel F

(Taraf Signifikansi 0.05)

Df 2	Df1						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	161.448	199.499	215.707	224.583	230.162	233.986	236.768
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353
3	10.128	9.552	9.277	9.117	9.0135	8.941	8.887
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876
6	5.987	5.143	4.757	4.533	4.387	4.284	4.207
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293
10	4.965	4.102	3.708	3.478	3.326	3.217	3.135
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913

13	4.667	3.806	3.410	3.179	3.025	2.915	2.832
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657
17	4.451	3.592	3.198	2.965	2.810	2.699	2.6142
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577
19	4.381	3.522	3.127	2.895	2.740	2.628	2.545
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514

Tabel t

(Pada taraf signifikansi 0,05) 1 sisi (0,05) dan 2 sisi (0,025)

Df	Signifikansi		Df	Signifikansi	
	0.025	0.05		0.025	0.05
1	12.706	6.314	46	2.013	1.679
2	4.303	2.920	47	2.012	1.678
3	3.182	2.353	48	2.011	1.677
4	2.776	2.132	49	2.010	1.677
5	2.571	2.015	50	2.009	1.676
6	2.447	1.943	51	2.008	1.675
7	2.365	1.895	52	2.007	1.675
8	2.306	1.860	53	2.006	1.674
9	2.262	1.833	54	2.005	1.674
10	2.228	1.812	55	2.004	1.673
11	2.201	1.796	56	2.003	1.673
12	2.179	1.782	57	2.002	1.672
13	2.160	1.771	58	2.002	1.672
14	2.145	1.761	59	2.001	1.671
15	2.131	1.753	60	2.000	1.671
16	2.120	1.746	61	2.000	1.670
17	2.110	1.740	62	1.999	1.670
18	2.101	1.734	63	1.998	1.669
19	2.093	1.729	64	1.998	1.669
20	2.086	1.725	65	1.997	1.669
21	2.080	1.721	66	1.997	1.668
22	2.074	1.717	67	1.996	1.668
23	2.069	1.714	68	1.995	1.668
24	2.064	1.711	69	1.995	1.667
25	2.060	1.708	70	1.994	1.667
26	2.056	1.706	71	1.994	1.667
27	2.052	1.703	72	1.993	1.666
28	2.048	1.701	73	1.993	1.666
29	2.045	1.699	74	1.993	1.666
30	2.042	1.697	75	1.992	1.665