**Ujian Akhir Semester**

**Simulasi Statistika (STK472)**

**Selasa, 10 Januari 2017**



**Muhamad Iqbal Hidayad**

**G14130090**

**A.**

Multikolinearitas adalah kondisi terdapatnya hubungan linier atau korelasi yang tinggi antara masing-masing variabel independen dalam model regresi. Indikasi terjadinya Multikolinieritas salah satunya adalah nilai yang tinggi (signifikan, namun nilai standar *error* dan tingkat signifikansi masing-masing variabel sangat rendah. Multikolinearitas juga terjadi ketika adanya korelasi atau hubungan linear di antara peubah penjelas. Dengan kata lain, peubah penjelas yang satu merupakan kombinasi linear dari peubah-peubah penjelas yang lain. Masalah multikolinearitas sering terjadi pada penerapan Analisis Regresi Berganda. Salah satu asumsi yang harus dipenuhi pada Analisis Regresi Berganda adalah tidak ada multikolinearitas pada peubah penjelas. Namun, asumsi ini tergantung pada tujuan dari penerapan Analisis Regresi Berganda. Jika tujuannya ingin mengetahui pola hubungan setiap peubah penjelas dengan peubah respon, asumsi ini harus dipenuhi sedangkan jika tujuannya ingin memprediksi nilai peubah respon berdasarkan peubah-peubah penjelas yang ada, asumsi ini tidak harus dipenuhi asalkan nilai *R-Square* besar.

Terdapat berbagai cara untuk mendeteksi adanya multikolinearitas pada peubah penjelas, antara lain dengan *Condition Index* (CI), *Multicolinearity Index* (MCI), dan *Variance Inflation Factor* (VIF). Namun, cara yang paling sering digunakan untuk mendeteksi adanya multikolinearitas adalah dengan menggunakan VIF. VIF yang lebih besar dari 5 atau 10 mengindikasikan adanya multikolinearitas.

Dampak buruk dari terjadinya multikolinearitas pada peubah penjelas, yaitu:

1. Dugaan parameter yang dihasilkan tidak stabil atau tidak konsisten. Hal ini dikarenakan tanda dugaan parameter regresi, misalnya dugaan koefisien regresi dapat berubah atau tidak sesuai dengan teori. Artinya, yang seharusnya bertanda positif menjadi negatif atau sebaliknya, yang seharusnya bertanda negatif menjadi positif.
2. Matriks akan memiliki sifat kondisi buruk (*ill-conditioned*) atau hampir singular karena memiliki determinan yang mendekati nol yang pada akhirnya akan menyebabkan dugaan parameter memiliki dugaan ragam yang menduga lebih (*overestimate*) walaupun tetap tak bias.
3. Karena dugaan parameter memiliki dugaan ragam yang menduga lebih (*overestimated*), maka nilai statistik uji t hitung akan menduga kurang (*underestimate*). Hal ini berdampak pada hasil pengujian parameter regresi secara parsial (uji-t). Hasil pengujian menjadi tidak valid karena keputusan pengujian lebih cenderung tak tolak H0, artinya peubah penjelas tidak berpengaruh terhadap respon padahal pada pengujian parameter regresi secara simultan (uji F) mengindikasikan tolak H0 (model nyata).
4. Nilai *R-Square Adjusted* bernilai tinggi padahal banyak peubah penjelas yang tidak berpengaruh terhadap respon.

Karena multikolinearitas membuat matriks memiliki determinan yang mendekati nol sehingga matriks tersebut dikatakan berkondisi buruk, maka parameter dari model regresi berganda yang terpengaruh langsung oleh pelanggaran asumsi tidak ada multikolinearitas pada peubah penjelas adalah parameter regresinya. Dengan metode kuadrat terkecil, dugaan parameter regresi diperoleh dengan persamaan

Dugaan parameter regresi tersebut tidak mungkin diperoleh jika terjadi multikolinearitas sempurna karena matriks tidak memiliki *invers*.

Jika terjadi multikolinearitas yang tidak sempurna tetapi tinggi, nilai dugaan parameter regresi masih mungkin diperoleh tetapi interpretasinya sulit. Nilai dugaan koefisien regresi dari suatu peubah penjelas diinterpretasi untuk mengukur perubahan nilai dugaan rataan peubah respon akibat perubahan nilai peubah penjelas tersebut dengan asumsi nilai peubah penjelas lainnya konstan. Apabila terjadi multikolinearitas yang tidak sempurna tetapi tinggi, sulit untuk mempertahankan nilai peubah penjelas lainnya yang memiliki korelasi dengan peubah penjelas tersebut. Hal ini dikarenakan ketika peubah penjelas tersebut berubah maka nilai peubah penjelas lainnya yang berkorelasi dengan peubah penjelas tersebut akan ikut berubah sesuai dengan arah korelasinya.

Selain itu, karena matriks merupakan komponen perhitungan dugaan ragam bagi dugaan parameter regresi, yaitu

maka parameter ragam bagi dugaan parameter regresi juga akan terpengaruh. Akibat yang ditimbulkan akan hal ini dijelaskan pada poin 2 di atas.

**B.**

Adanya pelanggaran terhadap asumsi tidak ada multikolinearitas pada peubah penjelas dapat diatasi dengan Regresi Komponen Utama (RKU) dan Regresi *Ridge* (Regresi Gulud).

**Regresi Komponen Utama (RKU)**

Analisis regresi komponen utama adalah teknik statistik yang dapat digunakan untuk menjelaskan struktur variansi-kovariansi dari sekumpulan variabel melalui beberapa variabel baru dimana variabel baru ini saling bebas, dan merupakan kombinasi linier dari variabel lain. Prinsip dari regresi komponen utama adalah meregresikan peubah respon dengan skor beberapa komponen utama yang terpilih. Komponen utama merupakan kombinasi linear dari peubah penjelas yang diamati. Informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah penjelas dengan bobot tertentu. Banyaknya komponen utama yang terbentuk akan sama dengan banyaknya peubah penjelas. Antarkomponen utama bersifat ortogonal atau saling bebas dan informasinya tidak tumpang tindih. Oleh karena itu, Regresi Komponen Utama dapat digunakan untuk mengatasi masalah ketika informasi yang ada pada peubah-peubah penjelas saling tumpang tindih.

Banyaknya komponen utama yang dipilih berjumlah sedikit tetapi komponen tersebut mampu mempertahankan sebagian besar informasi yang terkandung pada data peubah penjelas asal. Dengan demikian, komponen utama yang dipilih adalah komponen utama yang memiliki ragam besar. Ragam dari komponen utama itu sendiri merupakan akar ciri matriks ragam peragam atau matriks korelasi. Ragam dari komponen utama pertama adalah akar ciri terbesar pertama, ragam dari komponen utama kedua adalah akar ciri terbesar kedua, dan seterusnya. Sementara itu, bobot kombinasi linear semua peubah penjelas pada setiap komponen utama merupakan vektor ciri dari akar ciri yang bersesuaian.

Matriks yang digunakan dapat berupa matriks korelasi atau matriks ragam peragam. Secara umum, cukup sulit untuk menentukan kapan menggunakan matriks korelasi atau matriks ragam peragam. Namun, hal umum yang menjadi patokan adalah satuan dari peubah penjelasnya. Jika satuan peubah penjelas sudah sama, gunakan matriks ragam peragam tetapi jika satuan peubah penjelas berbeda, dapat menggunakan matriks korelasi. Penggunaan matriks korelasi memang cukup efektif kecuali pada dua hal, yaitu secara teori pengujian statistik terhadap akar ciri dan vektor ciri matriks korelasi jauh lebih rumit dan dengan menggunakan matriks korelasi kita memaksakan setiap peubah memiliki ragam yang sama sehingga tujuan mendapatkan peubah yang kontribusinya paling besar tidak tercapai.

Penentuan banyaknya komponen utama yang terpilih dapat menggunakan tiga metode, yaitu:

1. Proporsi kumulatif keragaman total yang dapat dijelaskan oleh komponen utama. Metode ini dapat diterapkan pada penggunaan matriks korelasi maupun matriks ragam peragam. Batas minimum persentase kumulatif keragaman yang mampu dijelaskan dapat ditentukan oleh peneliti, misalnya 70%, 80%, bahkan 90%.
2. Akar ciri yang lebih besar atau sama dengan satu. Metode ini dapat diterapkan pada penggunaan matriks korelasi. Metode ini didasarkan pada argumen bahwa jika peubah penjelas asal saling bebas, komponen utama tidak lain adalah peubah asal, dan setiap komponen utama akan memiliki ragam satu. Dengan cara ini, komponen utama yang berpadanan dengan akar ciri kurang dari satu tidak digunakan.
3. *Scree Plot*, yaitu plot antara akar ciri dengan k. Metode ini dapat diterapkan ketika menggunakan matriks korelasi maupun matriks ragam peragam. Banyaknya komponen utama yang dipilih adalah k jika pada titik k tersebut kurvanya curam di sebelah kiri tetapi tidak curam di sebelah kanan. Pada intinya, banyaknya komponen utama yang dipilih sedemikian rupa sehingga selisih antara dua akar ciri yang berurutan sudah tidak besar lagi.

Adapun langkah-langkah yang dilakukan dalam analisis regresi komponen utama adalah sebagai berikut:

1. Pembakuan peubah penjelas jika peubah penjelas memiliki satuan yang berbeda.
2. Menghitung nilai akar ciri dan vektor ciri dari matriks korelasi atau matriks ragam peragam sehingga akan terbentuk komponen utama yang merupakan kombinasi linear dari peubah asal.
3. Hitung skor komponen utama.
4. Pilih komponen utama yang akan digunakan.
5. Regresikan peubah respon terhadap skor komponen utama terpilih.
6. Perhatikan apakah komponen utama yang digunakan tersebut berpengaruh terhadap peubah respon atau tidak. Jika ada komponen utama yang tidak berpengaruh, regresikan kembali peubah respon terhadap komponen utama yang berpengaruh.
7. Dugaan persamaan regresi yang diperoleh merupakan fungsi dari komponen utama sehingga perlu dilakukan transformasi sedemikian rupa supaya menjadi fungsi dari peubah asal.

Dugaan persamaan regresi hasil transformasi tersebut merupakan dugaan persamaan regresi yang sudah bebas dari masalah multikolinearitas.

**Regresi *Ridge* (Regresi Gulud)**

Metode Regresi *Ridge* dapat digunakan untuk mengatasi korelasi yang tinggi antara beberapa variabel bebas.Metode Regresi *Ridge* diperoleh dengan cara yang sama seperti metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat sisaan. Pada Regresi *Ridge* terdapat penambahan kendala (tetapan bias) pada matriks sebelum di-*invers* di dalam perhitungan untuk memperoleh penduga parameter regresi pada metode kuadrat terkecil. Oleh karena itu, penduga parameter regresi dengan metode Regresi *Ridge* dapat dihitung dengan rumus:

dengan tetapan bias dan adalah matriks identitas. Tetapan bias yang diinginkan adalah tetapan bias yang relatif kecil dan menghasilkan dugaan parameter regresi, khususnya dugaan koefisien regresi yang relatif stabil. Pada umumnya nilai terletak pada interval .

Dengan metode Regresi *Ridge* dapat mengurangi masalah multikolinearitas karena penambahan tetapan bias pada matriks menjadikan det()tidak mendekati nol sehingga matriks tidak berkondisi buruk. Pada dasarnya Regresi *Ridge* memang ditujukan untuk mengatasi masalah kondisi buruk (*ill-conditioned*). Penggunaan Regresi *Ridge* dapat meminimumkan ragam dugaan bagi dugaan parameter regresi yang dihasilkan. Namun, penduga parameter regresi yang dihasilkan dengan metode ini menjadi bias karena adanya penambahan tetapan bias tersebut. Perbedaannya adalah bahwa pada metode Regresi *Ridge,* nilai variabel bebasnya ditransformasikan dahulu melalui prosedur *Centering* dan *Rescaling.*

**C.**

Berikut akan diberikan simulasi untuk mengkaji pengaruh multikolinearitas terhadap Analisis Regresi Berganda serta penanganan multikolinearitas dengan Regresi Komponen Utama dan Regresi *Ridge*:

Analisis Regresi Berganda yang diterapkan menggunakan tiga peubah penjelas, yaitu , , dan dengan satu peubah respon, yaitu . Model regresi linear bergandanya adalah:

Kajian simulasi menggunakan rancangan percobaan.

**Faktor** yang diperiksa adalah besarnya korelasi antara ketiga peubah penjelas tersebut (,, dan ) dengan tiga taraf (), yaitu:

1. Korelasi rendah ()
2. Korelasi sedang ()
3. Korelasi tinggi ()

Karena terdapat tiga taraf, maka akan ada tiga gugus data peubah penjelas yang terbentuk.

Dari setiap taraf, **respon** yang diamati pada simulasi ini adalah nilai dugaan parameter regresi, yaitu , , , dan . Selain itu, berdasarkan jawaban pada bagian A akan diamati pula nilai dugaan simpangan baku bagi dugaan parameter regresi, yaitu , , , dan . Respon tersebut lebih dari satu sehingga merupakan kasus multivariate pada rancangan percobaan.

Peubah respon akan dibangkitkan berdasarkan model regresi linear berganda dengan dan nilai parameter regresi tertentu. Akan ada tiga peubah respon sesuai dengan gugus data peubah penjelas. Kemudian, regresikan masing-masing peubah respon terhadap semua peubah penjelas padanannya dan amati respon yang dihasilkan. Lalu, lakukan Regresi Komponen Utama dan Regresi *Ridge* pada masing-masing peubah respon terhadap semua peubah penjelas padanannya dan amati respon yang dihasilkan.

Berdasarkan hal tersebut, rancangan percobaan yang digunakan adalah Faktorial Rancangan Acak Lengkap (**Faktorial RAL**) dengan dua faktor dan empat respon. Faktor pertama yang merupakan faktor tetap adalah besarnya korelasi antara ketiga peubah penjelas dengan tiga taraf () sedangkan faktor kedua adalah metode analisis regresi dengan tiga taraf (). Dari kedua faktor tersebut akan terbentuk sembilan kombinasi perlakuan. Setiap perlakuan diulang tiga kali ().

**Algoritma** simulasi yang dilakukan adalah sebagai berikut:

1. Tentukan banyaknya observasi
2. Bangkitkan nilai untuk
3. Membangkitkan gugus data peubah penjelas yang terdiri atas tiga peubah penjelas yang saling berkorelasi rendah. Langkah-langkah untuk mendapatkan ketiga peubah penjelas yang saling berkorelasi tersebut adalah:
4. Bangkitkan dengan berukuran untuk
5. Tentukan vektor nilai tengah peubah penjelas
6. Tentukan nilai , , dan dengan ragam dari setiap peubah penjelas adalah 1 sehingga matriks ragam peragamnya adalah
7. Cari matriks supaya
8. Hitung dengan untuk sehingga akan diperoleh matriks berukuran
9. Transpose matriks sehingga terbentuk matriks baru yang berukuran dengan ketiga kolomnya merupakan ketiga peubah penjelas (, , dan ) yang saling berkorelasi rendah.
10. Tentukan nilai parameter regresi , , , dan
11. Hitung nilai peubah respon untuk
12. Ulangi langkah 3c – 5 untuk membangkitkan ketiga peubah penjelas yang saling berkorelasi sedang dan tinggi serta peubah respon yang bersesuaian. Nilai korelasi antarpeubah penjelasnya ditentukan sebagai berikut:
13. Korelasi sedang

, , dan dengan matriks ragam peragamnya adalah

1. Korelasi tinggi

, , dan dengan matriks ragam peragamnya adalah

1. Regresikan masing-masing peubah respon terhadap tiga peubah penjelas padanannya.
2. Amati nilai respon yang dihasilkan dari tiap taraf korelasi.
3. Regresikan masing-masing peubah respon Y terhadap tiga peubah penjelas padanannya dengan metode Regresi Komponen Utama.
4. Amati nilai respon yang dihasilkan dari tiap taraf korelasi.
5. Regresikan masing-masing peubah respon Y terhadap tiga peubah penjelas padanannya dengan metode Regresi *Ridge*.
6. Amati nilai respon yang dihasilkan dari tiap taraf korelasi.
7. Ulangi langkah 1 – 12 di atas sebanyak kali sehingga diperoleh 3 kali ulangan untuk kombinasi taraf nilai korelasi dan metode regresi.

Berikut **syntax SAS** untuk proses penyiapan atau **pembangkitan tiga gugus data** yang setiap gugusnya terdiri atas peubah respon dan tiga peubah penjelas sesuai dengan taraf besarnya korelasi antarpeubah penjelas.

**data** data1 (keep=z1 z2 z3);

n = **1000**;

do i=**1** to n;

z1 = rannor(**0**);

z2 = rannor(**0**);

z3 = rannor(**0**);

output;

end;

**run**;

**proc** **iml**;

use data1;

read all var{z1 z2 z3} into z;

mu1 = **2**;

mu2 = **3**;

mu3 = **4**;

mu = mu1//mu2//mu3;

sigmar = {**1** **0.1** **0.15**,

**0.1** **1** **0.2**,

**0.15** **0.2** **1**};

sigmas = {**1** **0.4** **0.5**,

**0.4** **1** **0.6**,

**0.5** **0.6** **1**};

sigmat = {**1** **0.8** **0.85**,

**0.8** **1** **0.9**,

**0.85** **0.9** **1**};

Ar = half(sigmar);

As = half(sigmas);

At = half(sigmat);

n = **1000**;

xr = J(**3**,n,**.**);

xs = J(**3**,n,**.**);

xt = J(**3**,n,**.**);

do j=**1** to n;

xr[,j] = mu+(Ar)`\*z[j,]`;

xs[,j] = mu+(As)`\*z[j,]`;

xt[,j] = mu+(At)`\*z[j,]`;

end;

xr = (xr)`;

xs = (xs)`;

xt = (xt)`;

b0=**4**; b1=**0.5**; b2=**0.8**; b3=**0.7**;

e = J(n,**1**,**.**);

yr = J(n,**1**,**.**);

ys = J(n,**1**,**.**);

yt = J(n,**1**,**.**);

do i=**1** to n;

e[i,] = rannor(**0**);

yr[i,] = b0+b1\*xr[i,**1**]+b2\*xr[i,**2**]+b3\*xr[i,**3**]+e[i,];

ys[i,] = b0+b1\*xs[i,**1**]+b2\*xs[i,**2**]+b3\*xs[i,**3**]+e[i,];

yt[i,] = b0+b1\*xt[i,**1**]+b2\*xt[i,**2**]+b3\*xt[i,**3**]+e[i,];

end;

r = yr||xr;

s = ys||xs;

t = yt||xt;

create datar from r;

append from r;

create datas from s;

append from s;

create datat from t;

append from t;

**quit**;

**data** data\_r;

set datar;

rename COL1=Y COL2=X1 COL3=X2 COL4=X3;

**run**;

**data** data\_s;

set datas;

rename COL1=Y COL2=X1 COL3=X2 COL4=X3;

**run**;

**data** data\_t;

set datat;

rename COL1=Y COL2=X1 COL3=X2 COL4=X3;

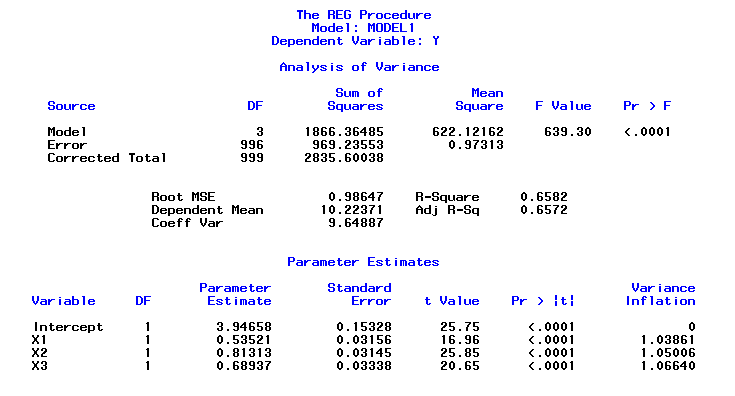
**run**;

* Regresikan terhadap , , dan untuk yang korelasi rendah:

**proc** **reg** data=data\_r;

model Y = X1 X2 X3/vif;

**run**;



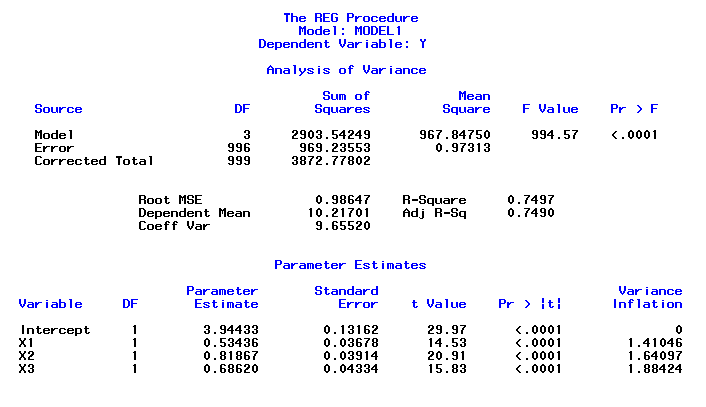
Nilai VIF ketiga peubah penjelas mendekati satu sehingga tidak ada indikasi terjadinya multikolinearitas. Nilai dugaan parameter regresi yang dihasilkan memiliki tanda yang sesuai dengan penentuan awal ketika simulasi dan nilainya mendekati nilai yang ditentukan tersebut. Berdasarkan uji-t pada semua peubah penjelas, *p-value* yang dihasilkan bernilai kurang dari taraf nyata yang digunakan, yaitu 5% sehingga keputusan pengujiannya adalah Tolak Ho dan kesimpulannya adalah ketiga peubah penjelas tersebut berpengaruh terhadap peubah respon secara parsial. Hal ini didukung oleh uji F yang menghasilkan *p-value* kurang dari taraf nyata sehingga keputusan pengujian yang diambil adalah Tolak H0 dan kesimpulannya adalah minimal ada satu peubah penjelas yang berpengaruh terhadap peubah respon. Nilai *R-Square* dan *R-Square Adjusted* yang dihasilkan pun cukup besar, yaitu 65.82% dan 65.72% dengan nilai RMSE sebesar 0.98647.

* Regresikan terhadap , , dan untuk yang korelasi sedang:

**proc** **reg** data=data\_s;

model Y = X1 X2 X3/vif;

**run**;



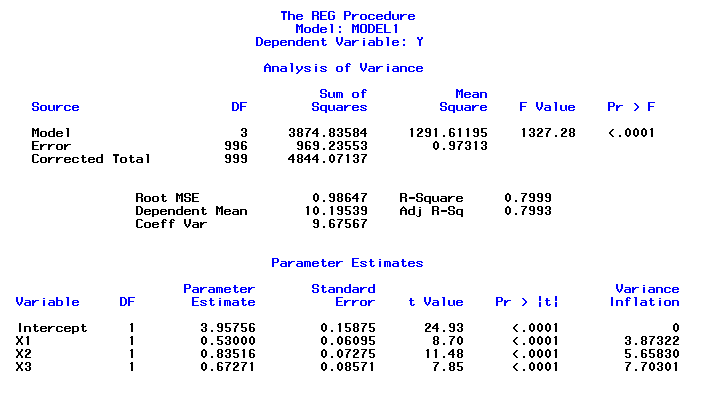
Data peubah penjelas yang dibangkitkan saling berkorelasi sedang. Nilai VIF ketiga peubah penjelas masih mendekati satu namun nilai VIF ini lebih besar dibandingkan untuk kasus peubah penjelas yang korelasinya rendah. Berdasarkan output di atas, tanda dari dugaan parameter regresi yang dihasilkan tidak berubah namun nilai dugaan simpangan bakunya lebih besar dibandingkan untuk kasus yang korelasi peubah penjelasnya rendah. Hasil uji F menunjukan bahwa minimal ada satu peubah penjelas yang berpengaruh terhadap peubah respon pada taraf nyata 5% dan menurut hasil uji-t untuk semua peubah penjelas memberikan kesimpulan bahwa semua peubah penjelas berpengaruh terhadap peubah respon secara parsial. Nilai RMSE yang dihasilkan sama dengan nilai RMSE untuk kasus peubah penjelas saling berkorelasi rendah, yaitu 0.98647 namun pada kasus ini memberikan nilai *R-Square* dan *R-Square Adjusted* yang lebih besar, yaitu 74.97% dan 74.90%.

* Regresikan terhadap , , dan untuk yang korelasi tinggi:

**proc** **reg** data=data\_t;

model Y = X1 X2 X3/vif;

**run**;



Nilai VIF peubah penjelas dan lebih besar dari lima. Hal ini dapat dijadikan indikasi terjadinya multikolinearitas. Namun, adanya multikolinearitas ini tidak memberikan dampak terhadap tanda dari dugaan parameter regresi yang dihasilkan. Kenaikan dugaan simpangan bagi dugaan parameter pun tidak terlalu besar sehingga hasil uji-t masih memberikan keputusan Tolak H0 yang berarti semua peubah penjelas berpengaruh terhadap peubah respon secara parsial pada taraf nyata 5%. Nilai RMSE yang dihasilkan masih sama dengan nilai RMSE untuk kasus peubah penjelas berkorelasi rendah dan berkorelasi sedang, yaitu 0.98647 tetapi baik nilai *R-Square* maupun *R-Square Adjusted* menjadi lebih tinggi, yaitu 79.99% dan 79.93%. Hal ini merupakan salah satu dampak dari terjadinya masalah multikolinearitas.

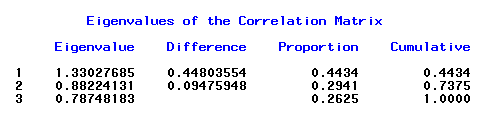
Dari ketiga gugus data yang sama, dilakukan regresi dengan metode **Regresi Komponen Utama**. Matriks yang digunakan adalah matriks korelasi dan penentuan banyaknya komponen utama yang digunakan menggunakan akar ciri, yaitu akar ciri yang lebih besar atau sama dengan satu.

* Regresikan terhadap , , dan dengan Regresi Komponen Utama untuk yang korelasi rendah:

**proc** **princomp** data=data\_r;

var X1 X2 X3;

**run**;



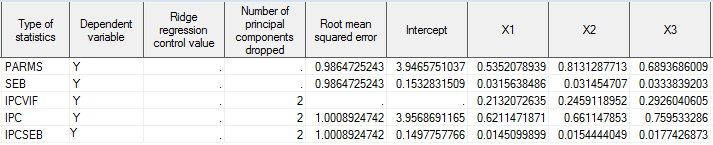
Output di atas menunjukan akar ciri dari setiap komponen utama. Hanya ada satu komponen yang memiliki akar ciri lebih dari atau sama dengan satu sehingga hanya komponen utama pertama yang akan digunakan. Hal ini berimplikasi bahwa banyaknya komponen utama yang dibuang adalah dua.

**proc** **reg** data=data\_r outest=rku\_r pcomit=**2** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;



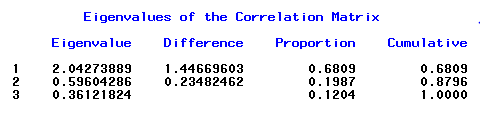
Regresi Komponen Utama menghasilkan nilai VIF yang kecil mendekati nol untuk semua peubah penjelas. Tanda dari dugaan parameter tidak mengalami perubahan dan dugaan parameter tersebut memiliki dugaan simpangan baku yang lebih kecil dibandingkan dengan regresi berganda biasa tetapi nilai RMSEnya lebih besar, yaitu 1.00089.

* Regresikan terhadap , , dan dengan Regresi Komponen Utama untuk yang korelasi sedang:

**proc** **princomp** data=data\_s;

var X1 X2 X3;

**run**;



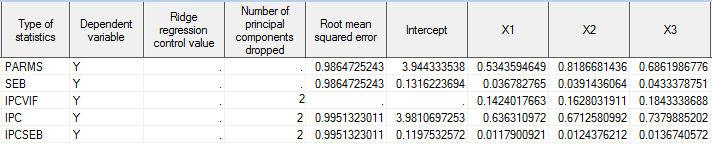
Berdasarkan output di atas, hanya komponen utama pertama yang memiliki akar ciri lebih dari atau sama dengan satu sehingga hanya akan digunakan komponen utama pertama.

**proc** **reg** data=data\_s outest=rku\_s pcomit=**2** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;

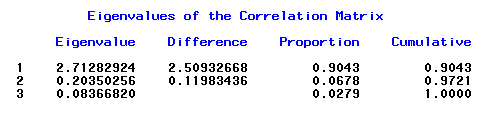


Nilai VIF untuk semua peubah penjelas mendekati nol. Dugaan parameter regresi yang dihasilkan berbeda dengan penentuan awal tetapi tanda dugaan parameter tidak berubah dan nilai dugaan ragam bagi dugaan parameter regresi yang dihasilkan lebih kecil. Namun, nilai RMSE yang dihasilkan sedikit lebih besar daripada nilai RMSE yang dihasilkan ketika menggunakan Analisis Regresi Berganda biasa, yaitu 0.99513 tetapi nilai RMSE ini lebih kecil daripada nilai RMSE yang dihasilkan dengan metode yang sama pada kasus peubah penjelas berkorelasi rendah.

* Regresikan terhadap , , dan dengan Regresi Komponen Utama untuk yang korelasi tinggi:

**proc** **princomp** data=data\_t;

var X1 X2 X3;

**run**;

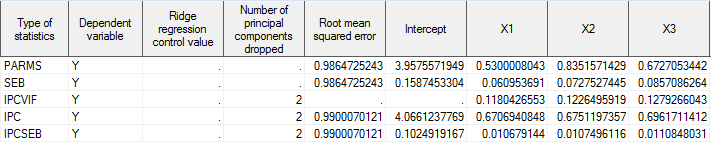
Berdasarkan akar ciri setiap komponen, banyaknya komponen utama yang terpilih adalah satu, yaitu komponen utama pertama. Besarnya keragaman yang dapat dijelaskan oleh komponen utama pertama tersebut adalah 90.43%. Dengan demikian, banyaknya komponen utama yang dibuang adalah dua.

**proc** **reg** data=data\_t outest=rku\_t pcomit=**2** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;



Regresi Komponen Utama pada gugus data dengan peubah penjelasnya saling berkorelasi menghasilkan nilai dugaan ragam bagi dugaan parameter yang lebih kecil dibandingkan dengan nilai dugaan ragam bagi dugaan parameter ketika menggunakan regresi berganda biasa. Tanda nilai dugaan koefisien regresinya stabil. Namun, nilai RMSEnya sedikit lebih besar dibandingkan dengan nilai RMSE hasil regresi berganda biasa, yaitu 0.99001 tetapi nilai RMSE ini lebih kecil daripada nilai RMSE yang dihasilkan dengan metode yang sama pada kasus peubah penjelas berkorelasi rendah dan sedang.

Kemudian, dari ketiga gugus data yang sama, dilakukan **Regresi *Ridge***.

Berikut syntax SAS untuk meregresikan terhadap , , dan dengan Regresi *Ridge* untuk yang korelasi rendah, sedang, dan tinggi:

**proc** **reg** data=data\_r outest=ridge\_r ridge=**0.1** to **0.7** by **0.05** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;

**proc** **reg** data=data\_s outest=ridge\_s ridge=**0.1** to **0.7** by **0.05** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;

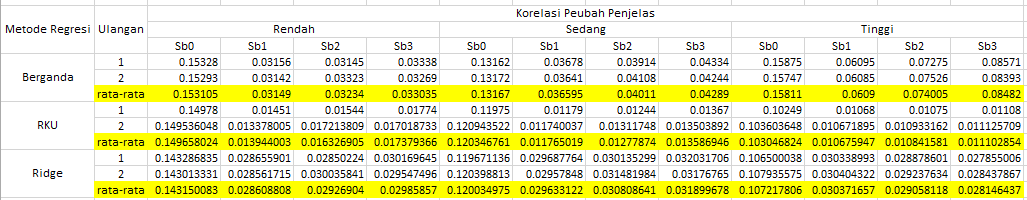
**proc** **reg** data=data\_t outest=ridge\_t ridge=**0.1** to **0.7** by **0.05** outseb outvif;

model Y = X1 X2 X3;

**run**;

**quit**;

Regresi *Ridge* tersebut menggunakan nilai *ridge* atau tetapan bias 0.1 sampai 0.7 dengan jarak 0.05 sehingga akan ada 12 nilai tetapan bias. Output akan menampilkan nilai VIF, dugaan koefisien parameter regresi, dan dugaan ragam bagi dugaan parameter regresi di masing-masing nilai tetapan bias. Nilai-nilai tersebut dipilih berdasarkan tetapan bias yang menghasilkan nilai VIF mendekati satu dari nol. Selain itu, setiap kombinasi taraf perlakuan diulang dua kali. Tabel faktorial RAL:



Pada output diatas dapat kita lihat bahwa nilai dugaan simpangan baku dari parameter. Dugaan simpangan baku terbaik adalah dugaan simpangan nilai baku yang memiliki nilai terkecil. Untuk model regresi *Ridge* nilai dugaan simpangan baku terkecilnya terdapat pada nilai Sb0. Untuk dugaan simpangan baku terkecil Sb1 terdapat pada metode RKU. Untuk dugaan simpangan baku terkecil Sb2 dan Sb3 nilai terkecil yaitu metode RKU. Untuk Sehingga untuk data yang berkorelasi rendah metode RKU merupakan metode terbaik.