Asumsi Multikolinearitas

Makalah Ini Disusun untuk Memenuhi Salah Satu Tugas Mata Kuliah

Model Linear

Dosen Pengampu : Ary Santoso, M.Si

Universitas Islam Negeri
SYARIF HIDAYATULLAH JAKARTA

Di susun oleh :

Rizki Aulia	11190940000002
Annisa Wulan M	11190940000003
Khoirunnisa Fi Nurdin	11190940000004
Herman	11190940000016
Wisnu Anggoro Putro	11190940000028
Salsabila Farah H.	11190940000029
Rosa Amalia Nursinta	11190940000041
Ikrom Al Furqon	11190940000046
Meissy Astariva Putri	11190940000063

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

2021

KATA PENGANTAR

Assalammu'alaikum Warahmatullahi Wabarakatuh

Puji syukur kita panjatkan kepada Allah Yang Maha Esa karena atas berkat rahmat dan karunia-Nya kami dapat menyelesaikan penyusunan tugas dengan judul **"Asumsi Multikolinearitas"** dengan baik.

Tugas ini merupakan salah satu tugas besar bagi kami, karena tugas ini merupakan tugas akhir di dalam mata kuliah Model Linear dengan dosen pengampunya adalah Bapak Ary Santoso.

Kami sadar bahwa laporan ini dapat diselesaikan karena adanya dukungan dan bantuan dari berbagai pihak. Karenanya, pada kesempatan kali ini kami sangat berterimakasih kepada semua yang berperan penting dalam penyelesaian Ujian Akhir Semester (UAS) ini, yang bukan lain adalah :

1. Bapak Ary Santoso selaku Dosen Pengampu mata kuliah Model Linear di Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta.
2. Keluarga kami yang selalu memberikan dukungan berupa materi maupun moril berbentuk doa serta semangat.
3. Teman sekelompok sekaligus teman seperjuangan Ujian Akhir Semester (UAS).
4. Seluruh pihak yang telah membantu kami dalam menyusun serta mengerjakan Tugas pada mata kuliah Model Linear ini yang tanpa mengurangi rasa hormat kami tidak dapat sebutkan satu-persatu.

Kami menyadari penuh bahwa dalam penyusunan project ini masih terdapat banyak kekurangan. Kami mengharapkan kritik dan saran yang membangun dari pembaca agar kedepannya kami bisa melakukan perbaikan untuk mendapatkan hasil yang lebih baik. Terakhir, kami berharap semoga laporan ini dapat bermanfaat.

Wassalammu'alaikum Warahmatullahi Wabarakatuh

Jakarta, 10 November 2021

Penulis

DAFTAR ISI

BAB I
PENDAHULUAN

A. Latar Belakang

Analisis regresi adalah teknik analisis statistik untuk mengetahui pola hubungan antara variabel tak bebas dengan variabel bebas. Hubungan antara variabel-variabel dalam regresi linier dinyatakan dalam model tersebut juga disebut model regresi klasik. Analisis regresi linear berganda digunakan untuk menganalisis hubungan linear antara dua atau lebih variabel bebas secara bersama-sama dengan satu variabel terikat. Adanya korelasi antar variabel yang cukup tinggi menimbulkan multikolinearitas yang menyebabkan model persamaan regresi yang diperoleh kurang layak.

Multikolinieritas adalah kondisi terdapatnya hubungan linier atau korelasi yang tinggi antara masing-masing variabel independent dalam model regresi. Multikolinieritas biasanya terjadi ketika sebagian besar variabel yang digunakan salang terkait dalam satu model regresi. Oleh karena itu, masalah multikolinieritas tidak terjadi pada regresi linier sederhana yang hanya melibatkan satu variabel independent. Salah satu ukuran untuk mendeteksi adanya multikolinearitas adalah dengan menguji koefisien korelasi (r) antar variabel prediktor. Jika koefisien korelasi diatas 0,85 maka diduga terdapat kasus multikolinearitas. Sebaliknya jika koefisien korelasi relatif rendah maka diduga tidak mengandung multikolinearitas. Deteksi ini diperlukan kehati-hatian.

Uji asumsi klasik merupakan uji asumsi yang tujuannya untuk memberikan kepastian bahwa persamaan regresi yang didapatkan memiliki ketepatan dalam estimasi, konsisten, dan tidak bias. Agar persamaan yang diestimasi dapat menghasilkan estimator yang *BLUE* (*Best Linear Unbiased Estimator*), perlu dilakukan uji asumsi klasik untuk memastikan bahwa model yang digunakan bersifat *robust*. Penyimpangan yang terjadi terhadap berbagai asumsi klasik menjadikan estimasi dari variabel yang diharapkan kurang tepat (Zaenuddin, 2015). Salah satu jenis uji asumsi klasik yang digunakan yaitu uji asumsi klasik multikoleniaritas.

Menurut Hamdi (2014), uji asumsi klasik multikolinearitas adalah korelasi linear yang sempurna atau eksak diantara variabel penjelas yang dimasukkan ke dalam model. Uji multikolinearitas digunakan untuk mengetahui apakah di dalam model regresi terjadi hubungan linear yang sempurna atau mendekati sempurna di antara beberapa atau semua variabel bebas.

B. Rumusan Masalah

- Apa pengertian Uji Asumsi Multikolinearitas
- Apa dampak (akibat) bagi model regresi yang dihasilkan jika uji asumsi Multikolinearitas dilanggar?
- Bagaimana cara untuk memeriksa atau melakukan pengujian/pemeriksaan pada Uji Asumsi Multikolinearitas?
- Bagaimana script R yang digunakan untuk melakukan pengujian/pemeriksaan pada Uji Asumsi Multikolinearitas?
- Apa yang perlu dilakukan jika terjadi pelanggaran pada Uji Asumsi Multikolinearitas?
- Buatlah sebuah contoh studi kasus dan terapkan mulai dari langkah awal pemeriksaan hingga mengatasi pelanggaran Uji Asumsi Multikolinearitas!

C. Tujuan dan Manfaat

Setelah membaca makalah ini diharapkan bagi penulis ataupun pembaca lainnya dapat lebih memahami Uji Asumsi Multikolinearitas secara lebih baik sehingga kedepannya baik dapat membuat Model Regresi yang baik dan dapat dipertanggung jawabkan.

BAB II
METODOLOGI

A. Lembar Kerja Tim

No	Nama Anggota	NIM Anggota	Deskripsi Tugas
1	Rizki Aulia	11190940000002	Editor PPT, Cara Pemeriksaan Uji Asumsi, mencari referensi Jurnal
2	Annisa Wulan M	11190940000003	Pengertian dari Uji Asumsi, mencari referensi Jurnal
3	Khoirunnisa Fi Nurdin	11190940000004	Cara Pemeriksaan Uji Asumsi, mencari referensi Jurnal
4	Herman	11190940000016	Editor PPT, Dampak pelanggaran Uji Asumsi, mencari referensi Jurnal
5	Wisnu Anggoro Putro	11190940000028	Pengertian dari Uji Asumsi, Mencari Script R
6	Salsabila Farah H.	11190940000029	Dampak pelanggaran Uji Asumsi, mencari referensi Jurnal
7	Rosa Amalia Nursinta	11190940000041	Cara Pemeriksaan Uji Asumsi, Mencari Script R
8	Ikrom Al Furqon	11190940000046	Editor makalah, Dampak pelanggaran Uji Asumsi, , mencari referensi Jurnal
9	Meissy Astariva Putri	11190940000063	Editor PPT, Pengertian dari Uji Asumsi, Mencari Script R

Note : Jobdesk akan selalu diupdate pada telemeeting selanjutnya

B. Laporan Telemeeting

No	Agenda Telemeeting	Daftar Hadir	Link Telemeeting	Link Record	Foto Telemeeting
1	Diskusi Materi dan Pembagian Jobdesk.	Lengkap	https://zoom.us/j/7649417692?pwd=Y0FRTWIwRWs1T2dyZzNadmhxVkZBUT09	https://drive.google.com/drive/folders/177SUtrYsHdU58txXoIHcVVJvjZ_mAndb?usp=sharing	
2	Diskusi Materi (Tahapan pengujian & Script R)	Lengkap	https://zoom.us/j/7649417692?pwd=Y0FRTWIwRWs1T2dyZzNadmhxVkZBUT09	https://drive.google.com/drive/folders/177SUtrYsHdU58txXoIHcVVJvjZ_mAndb?usp=sharing	
3	Diskusi Materi (Script R Pengujian)	Lengkap	https://zoom.us/j/7649417692?pwd=Y0FRTWIwRWs1T2dyZzNadmhxVkZBUT09	https://drive.google.com/drive/folders/177SUtrYsHdU58txXoIHcVVJvjZ_mAndb?usp=sharing	

4	Diskusi Materi (Script R Perbaikan)	Lengkap	https://zoom.us/j/7649417692?pwd=Y0FRTWIwRWs1T2dyZzNadmhxVkZBUT09	https://drive.google.com/drive/folders/177SUtrYsHdU58txXoIHcVVJvjZ_mAndb?usp=sharing	

C. Tahapan Kegiatan dan Rancangan Waktu

No.	**Kegiatan**	**November 2021**													
		5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	Telemeeting														
2	Laporan Progress 1														
3	Laporan Progress 2														
4	Laporan Progress 3														

BAB III
PEMBAHASAN

A. Pengertian Asusmsi Multikolinearitas

Istilah multikolinearitas mula-mula ditemukan oleh Ragnar Frisch. Pada mulanya multikolinearitas berarti adanya hubungan linear yang sempurna atau pasti, diantara beberapa atau semua variabel bebas dari model regresi ganda Gujarati, 1995:157.

Multikonearitas adalah adanya hubungan eksak linier antar variabel penjelas. Multikonearitas diduga terjadi bila nilai R^2tinggi, nilai t semua variabel penjelas tidak signifikan, dan nilai F tinggi.

Multikolinearitas adalah suatu kondisi dimana terjadi korelasi antara variabel bebas atau antar variabel bebas tidak bersifat saling bebas.

Multikolinearitas adalah sebuah situasi yang menunjukkan adanya korelasi atau hubungan kuat antara dua variabel bebas atau lebih dalam sebuah model regresi berganda. Model regresi yang dimaksud dalam hal ini antara lain: regresi linear, regresi logistik, regresi data panel dan cox regression.

Menurut Hamdi (2014), uji asumsi klasik multikolinearitas adalah korelasi linear yang sempurna atau eksak diantara variabel penjelas yang dimasukkan ke dalam model. Uji multikolinearitas digunakan untuk mengetahui apakah di dalam model regresi terjadi hubungan linear yang sempurna atau mendekati sempurna di antara beberapa atau semua variabel bebas.

B. Dampak Pelanggaran Asumsi Multikolinearitas

Jika terdapat korelasi yang cukup tinggi maka dapat dikatakan terjadi multikolinearitas. Efek dari multikolinearitas ini dapat mengakibatkan estimasi parameter regresi yang dihasilkan menjadi tidak efisien karena mempunyai bias dan variansi yang besar.

Jika terjadi multikolineritas, estimasi koefisien regresi pada kondisi eror berdistribusi simetris yang diperoleh menjadi tidak valid. Pada kondisi error berdistribusi non simetris (exponensial, weibull dan gamma) standar error estimasi dan MSE yang diperoleh jauh lebih kecil daripada kondisi eror berdistribusi simetris. Dengan demikian dampak multikolineritas lebih berbahaya pada kondisi asumsi normalitas terpenuhi, Lebih umum dampak multikolinearitas lebih berbahaya pada kondisi eror berdistribusi simetri.

Dalam situasi terjadi multikolinearitas dalam sebuah model regresi berganda, maka nilai koefisien beta dari sebuah variabel bebas atau variabel predictor dapat berubah secara dramatis apabila ada penambahan atau pengurangan variabel bebas di dalam model. Oleh karena itu, multikolinearitas tidak mengurangi kekuatan prediksi secara simultan, namun mempengaruhi nilai prediksi dari sebuah variabel bebas. Nilai prediksi sebuah variabel bebas disini adalah koefisien beta. Oleh karena itu, sering kali kita bisa mendeteksi adanya multikolinearitas dengan adanya nilai standar error yang besar dari sebuah variabel bebas dalam model regresi.

Maka, jika terjadi multikolinearitas, sebuah variabel yang berkorelasi kuat dengan variabel lainnya di dalam model, kekuatan prediksinya jadi tidak handal dan tidak stabil.

Penyebab multikolinearitas adalah adanya korelasi atau hubungan yang kuat antara dua variabel bebas atau lebih, seperti yang sudah dijelaskan di atas. Namun penyebab lainnya yang dapat menyebabkan hal tersebut secara tidak langsung adalah, antara lain:

1. Penggunaan variabel dummy yang tidak akurat di dalam model regresi. Akan lebih beresiko terjadi multikolinearitas jika ada lebih dari 1 variabel dummy di dalam model.
2. Adanya perhitungan sebuah variabel bebas yang didasarkan pada variabel bebas lainnya di dalam model. Hal ini bisa dicontohkan sebagai berikut: dalam model regresi anda, ada variabel X_1, X_2 dan Perkalian antara X_1dan X_2 (X_1*X_2). Dalam situasi tersebut bisa dipastikan, terdapat kolinearitas antara X_1 dan X_1*X_2 serta kolinearitas antara X_2 dengan X_1*X_2.
3. Adanya pengulangan variabel bebas di dalam model, misalkan:

$$Y = \alpha + \beta_1 X_1 \ + \ \beta_2 X_1 + \beta_3 X_3 + \ \varepsilon.$$

Dampak dari multikolinearitas antara lain:

1. Koefisien Partial Regresi tidak terukur secara presisi. Oleh karena itu nilai standar errornya besar.
2. Perubahan pada satu variabel dapat menyebabkan perubahan besar pada nilai koefisien regresi parsial variabel lainnya.
3. Nilai Confidence Interval sangat lebar, sehingga akan menjadi sangat sulit untuk menolak hipotesis nol pada sebuah penelitian jika dalam penelitian tersebut terdapat multikolinearitas.

C. Tahapan Pengujian Asumsi Multikolinearitas

Cara mendeteksi adanya Multikolinearitas di dalam model regresi adalah dengan cara:

1. Melihat kekuatan korelasi antar variabel bebas.

   Jika ada korelasi antar variabel bebas > 0,8 dapat diindikasikan adanya multikolinearitas.

2. Melihat nilai standar error koefisien regresi parsial.

   Jika ada nilai standar error > 1, maka dapat diindikasikan adanya multikolinearitas.

3. Melihat nilai Tolerance dan Variance Inflating Factor (VIF).

   Jika nilai Tolerance < 0,1 dan VIF > 10 dapat diindikasikan adanya multikolinearitas. Sebagian pakar menggunakan batasan Tolerance < 0,2 dan VIF > 5 dalam menentukan adanya multikolinearitas. Para pakar juga lebih banyak menggunakan nilai Tolerance dan VIF dalam menentukan adanya Multikolinearitas di dalam model regresi linear berganda dibandingkan menggunakan parameter-parameter yang lainnya.

D. Script R Pengujian Asumsi Multikolinearitas

```
> library(psych)
> library(factoextra)
> library(dplyr)
> library(lmtest)
> library(car)
> library(stats)
> library(olsrr)
>
> # Periksa Multikolinearitas
>
> # Regresi
> Data_Modlin <- read.csv("D:/Data5.csv")
>
> # View(Data_Modlin)
> model=lm(formula=y~x1+x2+x3+x4+x5, data=Data_Modlin)
> summary(model)
> anova(model)
> attach(Data_Modlin)
>
> #Melihat grafik linearitas antar variable
> coef(summary(model))
>
> #Mengetahui nilai VIF dan tolerance tiap variabel bebas
> ols_vif_tol(model)
>
> #Mengetahui koefisien korelasi antar variabel bebas
> korelasi <- cor(Data_Modlin)
> korelasi
>
```

E. Perbaikan Jika Terjadi Pelanggaran Asumsi Multikolinearitas

Terdapat beberapa metode yaitu

1. Metode Eliminasi Variabel (Script)

Mengatasi Multikolinearitas Menggunakan Metode Seleksi Variabel Stepwise Backward Elimination

Langkah Mengatasi:

1. Seleksi variabel menggunakan Uji-F Parsial : menseleksi variabel dengan metode menghilangkan satu persatu variabel bebas untuk mendapatkan model terbaik dengan memperhatikan nilai R-squared dan AIC yang diperoleh.
2. Membentuk Model Baru: membentuk model terbaru yang sudah terseleksi dan juga melihat R-squared nya untuk mengetahui seberapa baik model ini.
3. Uji Multikolinearitas model baru

```
> # Seleksi Variabel
>
> #Uji-F Parsial (Backward elimination) menghilangkan satu per
 satu variabel penjelas
> full.model <- lm(y~., data=Data_Modlin)
> reduced.model <- step(full.model, direction="backward")
> reduced.model$anova #interpretasikan
```

2. Penambahan Jumlah Data

Masalah multikolinieritas pada dasarnya merupakan persoalan sampel. Oleh karena itu, masalah multikolinieritas seringkali bisa diatasi jika kita menambah jumlah data. jika multikolinieritas menyebabkan variabel independen tidak signifikan mempengaruhi variabel dependen melalui uji t maka dengan penambahan jumlah data maka sekarang variabel independen menjadi signifikan mempengaruhi variabel dependen

3. Metode PCA

Principal Component Analysis (PCA) atau disebut juga transformasi KarhunenLoeve adalah tekhnik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransormasi linear sehingga terbentuk system koordinat baru dengan variansi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan (Cahyadi, 2007: 93).

Principal Component Analysis adalah suatu teknik stalistik yang banyak digunakan dalam psikologi, untuk pengembangan test objektif, pengukuran kepribadian dan intelegensi. Principal Component Analysis (PCA) adalah suatu teknik statistik yang secara linear mengubah bentuk sekumpulan variabel asli nienjadi kumpulan variabel yang lebih kecil yangtidak berkorelasi yang dapat mevvakili informasi dari kumpulan variabel asli (Dunteman. 1989:7). Sedangkan menurutTabachnick(2001 : 582)

PCA (PrincipalComponent Analysis) dan FA(FactorAnalysis) adalah teknik statistik yang diaplikasikan untuk satu kumpulan variabel ketika peneliti tertarik untuk menemukan variabel mana dalam kumpulan tersebut yang berhubungan dengan lainnya. Variabel berkorelasi satu dengan yang lainnya tetapi independen dengan subset lain yang merupakan kombinasi variabel-variabei di dalam faktor. Faktor adalah yang mencerminkan proses yang mendasari yang mempunyai korelasi antar variabel.

Tujuan PCA adalah untuk menjelaskan bagian dari variasi dalam kumpulan variabel yang diamati atas dasar beberapa dimensi. Dari variabel yang banyak dirubah menjadi sedikit variabel. Tujuan khusus PCA yaitu:

1. untuk meringkas pola korelasi antar variabel yang diobservasi.
2. mereduksi sejumlah besar variabel menjadi sejumlah kecil faktor,
4. memberikan sebuah definisi operasional (sebuah persamaan regresi) dimensi pokok penggunaan variabel yang diobservasi
5. menguji teori yang mendasarinya (Tabachnick ,2001)

Langkah Mengatasi :

1. Memeriksa nilai KMO : KMO digunakan untuk mengukur kecukupan sampling (sampling adequacy). Nilai ini membandingkan besarnya koefisien korelasi terobservasi dengan koefisien korelasi parsial. Nilai KMO yang kecil menunjukkan bahwa korelasi antar pasangan variabel tidak bisa diterangkan oleh variabel lainnya dan analisis faktor mungkin tidak tepat. Jika nilai KMO >0.5 maka model layak untuk dibentuk menjadi beberapa komponen utama / principal component (PC).
2. Membentuk komponen utama yaitu persamaan yang memuat variabel y, x1, x2, x3, x4, x5 dengan bantuan Software R
3. Memilih komponen utama :memilih PC mana saja yang akan digunakan untuk dibentuk ke dalam model baru dengan mempertimbangkan eigenvalue dan persentase variansi kumulatif.
4. Membentuk model baru : persamaan regresi yang memuat komponen utama sebagai variabel independent.

5. Memeriksa multikolinearitas model baru : melihat koefisien standar error, signifikansi, nilai VIF tiap komponen utama, dan melihat koefisien korelasi antar komponen utama.

Script R

1. Nilai KMO

```
> library(psych)
> library(factoextra)
> library(dplyr)
> # periksa KMO, jika KMO >0.5 maka model sudah layak untuk
dibentuk menjadi beberapa PC
> KMO(Data_Modlin)
>
```

2. Membentuk Komponen Utama (PC)

```
> # Komponen utama / Principal Component (PC)
> pca <- prcomp(Data_Modlin, scale=TRUE)
> pca
>
```

3. Memilih Komponen Utama

```
> # melihat eigen value dari tiap PC
> get_eigenvalue(pca)
>
```

4. Membentuk Model Baru

```
 > # dalam model baru ini akan digunakan PC1, PC2, PC3 yang
bisa mejelaskan 99.09362% keragaman data
> # PC1 = 0.4071798y + 0.4107108x1 + 0.4092664x2 +
0.4090370x3 + 0.4061373x4 + 0.4071406x5 dan seterusnya
> pcs <- as.data.frame(pca$x)
> pcs1 <- pcs %>% select(1:3)
> y1 = Data_Modlin %>% select(y)
> datapca <- cbind(y1, pcs1)
>
> # Model regresi linear berganda dengan Principal Component
> modelpca <- lm(y~., data = datapca)
```

Model baru yang akan dibuat adalah model dengan y sebagai variabel dependen dan komponen utama PC1, PC2, PC3 sebagai variabel independent nya.

F. Studi Kasus

Studi Kasus Uji Multikolinearitas Pada Data Dan Cara Mengatasinya

Data yang akan diuji :

y	x1	x2	x3	x4	x5
1.95	1.31	1.07	0.44	0.75	0.35
2.9	1.55	1.49	0.53	0.9	0.47
0.72	0.99	0.84	0.34	0.57	0.32
0.81	0.99	0.83	0.34	0.54	0.27
1.09	1.05	0.9	0.36	0.64	0.3
1.22	1.09	0.93	0.42	0.61	0.31
1.02	1.08	0.9	0.4	0.51	0.31
1.93	1.27	1.08	0.44	0.77	0.34
0.64	0.99	0.85	0.36	0.56	0.29
2.08	1.34	1.13	0.45	0.77	0.37
1.98	1.3	1.1	0.45	0.76	0.38
1.9	1.33	1.1	0.48	0.77	0.38
8.56	1.86	1.47	0.6	1.01	0.65
4.49	1.58	1.34	0.52	0.95	0.5
8.49	1.97	1.59	0.67	1.2	0.59
6.17	1.8	1.56	0.66	1.02	0.59
7.54	1.75	1.58	0.63	1.09	0.59
6.36	1.72	1.43	0.64	1.02	0.63
7.63	1.68	1.57	0.72	0.96	0.68
7.78	1.75	1.59	0.68	1.08	0.62
10.15	2.19	1.86	0.75	1.24	0.72
6.88	1.73	1.67	0.64	1.14	0.55

Langkah Uji Multikolinearitas :

1. Melihat nilai R-squared model utama dan signifikansi variabel bebas : Jika diperoleh R-squared yang besar namun hanya sedikit variabel bebas yang signifikan maka dapat mengindikasikan adanya multikolinearitas pada model.
2. Melihat koefisien standar error : jika terdapat koefisien standar error >1 maka dapat mengindikasikan adanya multikolienearitas.
3. Melihat grafik plot antar variabel : untuk melihat secara visual hubungan linear antara variabel x_i dan x_j
4. Uji asumsi multikolineritas menggunakan nilai VIF : model mengalami multikolineritas jika nilai VIF yang diperoleh >10 dan nilai tolerance <0.1. Model yang baik adalah model yang tidak terjadi multikolinearitas antara variabel-variabel bebas.
5. Melihat koefisien korelasi antar variabel-variabel bebas : untuk melihat seberapa besar hubungan korelasi antara variabel bebas X_i dan X_j. model mengalami multikolinearitas jika koefisien korelasi antar variabel bebas >0.80.

1. Nilai R-Squared Dan Signifikansi Tiap Variabel Bebas

```
> library(lmtest)
> library(car)
> library(stats)
> library(olsrr)
> #Regresi
> Data_Modlin <- read.csv("D:/Data5.csv")
> View(Data_Modlin)
> model=lm(formula=y~x1+x2+x3+x4+x5, data=Data_Modlin)
> summary(model)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5, data = Data_Modlin)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2610 -0.5373  0.1355  0.5120  0.8611 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -6.5122     0.9336  -6.976 3.13e-06 ***
x1            1.9994     2.5733   0.777  0.44851    
x2           -3.6751     2.7737  -1.325  0.20378    
x3            2.5245     6.3475   0.398  0.69610    
x4            5.1581     3.6603   1.409  0.17791    
x5           14.4012     4.8560   2.966  0.00911 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7035 on 16 degrees of freedom
Multiple R-squared:  0.9633,	Adjusted R-squared:  0.9519 
F-statistic: 84.07 on 5 and 16 DF,  p-value: 6.575e-11
```

Persamaan regresi awal yang diperoleh adalah $\hat{Y}$ = -6.512215 + 1.999413X_1 - 3.675096X_2 + 2.524486X_3 + 5.158082X_4 + 14.401162X_5 dan didapatkan R-squared adjusted sebesar 95.18733

- Hal ini berarti :
  - b_1 = 1.999413 menunjukkan bahwa ketika terjadi perubahan satu satuan x1 maka akan meningkatkan rata-rata Y sebesar 1.999413, dengan asumsi x2, x3, x4, x5 tetap.
  - b_2 = - 3.675096 menunjukkan bahwa ketika terjadi perubahan satu satuan x2 maka akan menurunkan rata-rata Y sebesar 3.675096 ,dengan asumsi x1, x3, x4, x5 tetap.
  - b_3 = 2.524486 menunjukkan bahwa ketika terjadi perubahan satu satuan x3 maka akan meningkatkan rata-rata Y sebesar 2.524486, dengan asumsi x1, x2, x4, x5 tetap.

- b_4 = 5.158082 menunjukkan bahwa ketika terjadi perubahan satu satuan x4 maka akan meningkatkan rata-rata Y sebesar 5.158082, dengan asumsi x1, x2, x3, x5 tetap.
- b_5 = 14.401162 menunjukkan bahwa ketika terjadi perubahan satu satuan x5 maka akan meningkatkan rata-rata Y sebesar 14.401162, dengan asumsi x1, x2, x3, x4 tetap.
- Sebesar 95.2% model mampu menjelaskan keragaman variabel y dan sekitar 4.8% keragaman yang tidak mampu dijelaskan oleh variabel penjelas dan terletak pada komponen error.

Pada model awal ini diperoleh nilai R-squared adjusted yaitu 95,19%, tentu merupakan angka yang besar untuk menggambarkan pengaruh dari variabel-variabel bebas terhadap variabel respon, Namun dapat dilihat juga bahwa berdasarkan p-value, variabel bebas yang signifikan hanyalah x5. Akibatnya, hal ini menunjukkan adanya gejala multikolinearitas pada model.

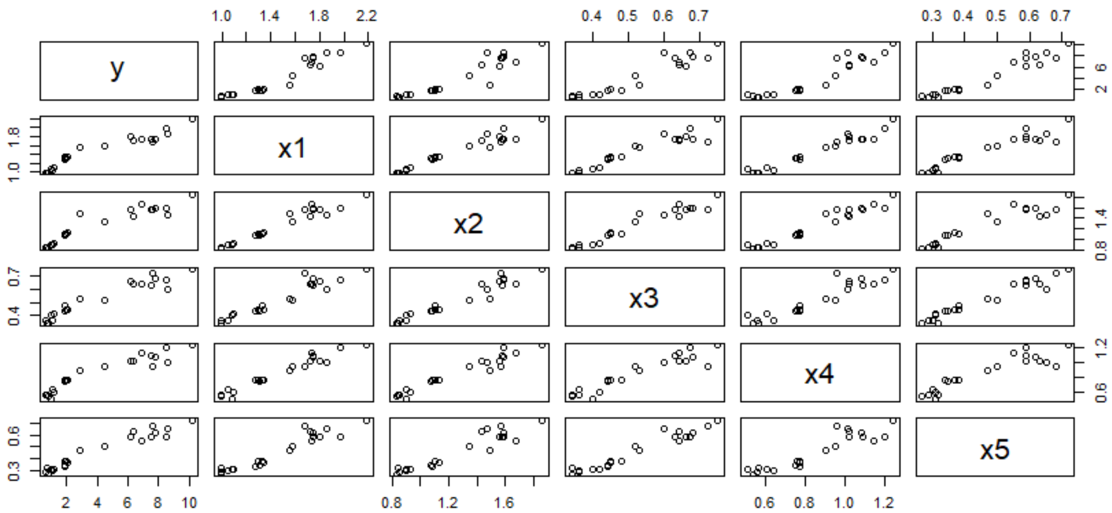
Maka diperlukan pemeriksaan asumsi multikolinearitas lebih lanjut menggunakan koefisien standar error, nilai VIF dan nilai koefisien korelasi antar variabel bebas.

2. Koefisien Standar Error Tiap Variabel Bebas

```
> #Melihat koefisien standar error tiap variabel bebas
> coef(summary(model))
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept) -6.512215  0.9335607 -6.9756734 3.126569e-06
x1           1.999413  2.5733384  0.7769726 4.485105e-01
x2          -3.675096  2.7736596 -1.3249989 2.037838e-01
x3           2.524486  6.3474950  0.3977137 6.960971e-01
x4           5.158082  3.6602832  1.4092029 1.779143e-01
x5          14.401162  4.8559938  2.9656468 9.108927e-03
>
```

- Nilai koefisien standar error untuk tiap variabel bebas yaitu :
  - x1 : 2.5733384
  - x2 : 2.7736596
  - x3 : 6.3474950
  - x4 : 3.6602832
  - x5 : 4.8559938
- Kita telah mengetahui bahwa adanya multikolenaritas pada model mengakibatkan koefisien standar error yang besar yaitu >1. Model yang baik adalah model yang memiliki koefisisen standar error variabel bebas nya <1, namun berdasarkan output yang diperoleh, semua variabel bebas memiliki koefisien standar error >1, maka hal ini dapat mengindikasikan adanya multikolienaritas pada model.

3. Plot Linearitas Antar Variabel



Secara visual, dapat dilihat bahwa terdapat hubungan linear yang relatif kuat antar variabel-variabel bebas, sehingga ini semakin mendukung indikasi adanya multikolinearitas pada model.

4. Nilai VIF dan Tolarance Tiap Variabel Bebas

```
> #Mengetahui nilai VIF dan tolerance tiap variabel bebas
> ols_vif_tol(model)
  Variables  Tolerance      VIF
1        x1 0.02794583 35.78351
2        x2 0.02924234 34.19699
3        x3 0.03298838 30.31371
4        x4 0.03390922 29.49050
5        x5 0.04471293 22.36490
>
```

- Nilai VIF tiap variabel bebas
  - Nilai VIF x1 : 35.78351
  - Nilai VIF x2 : 34.19699
  - Nilai VIF x3 : 30.31371
  - Nilai VIF x4 : 29.49050
  - Nilai VIF x5 ; 22.36490
- Nilai Tolerance tiap variabel bebas
  - Nilai tolerance x1 : 0.02794583
  - Nilai tolerance x2 : 0.02924234
  - Nilai tolerance x3 : 0.03298838
  - Nilai tolerance x4 : 0.03390922
  - Nilai tolerance x5 : 0.04471293
- Berdasarkan nilai VIF dan Tolerance yang diperoleh, semua variabel bebas memiliki nilai VIF >10 dan tolerance <0.1. Maka ini menunjukkan adanya multikolinearitas antar variabel bebas.

5. Koefisien Korelasi Antar Variabel Bebas

```
> korelasi <- cor(Data_Modlin)
> korelasi
           y        x1        x2        x3        x4        x5
y  1.0000000 0.9603595 0.9387393 0.9560244 0.9411316 0.9723701
x1 0.9603595 1.0000000 0.9715103 0.9568051 0.9758073 0.9538686
x2 0.9387393 0.9715103 1.0000000 0.9683549 0.9723228 0.9469462
x3 0.9560244 0.9568051 0.9683549 1.0000000 0.9415902 0.9719096
x4 0.9411316 0.9758073 0.9723228 0.9415902 1.0000000 0.9228714
x5 0.9723701 0.9538686 0.9469462 0.9719096 0.9228714 1.0000000
>
```

- Nilai koefisien korelasi antar variabel bebas
    - Koef. korelasi antara x1 dan x2 : 0.9715103
    - Koef. korelasi antara x1 dan x3 : 0.9568051
    - Koef. korelasi antara x1 dan x4 : 0.9758073
    - Koef. korelasi antara x1 dan x5 : 0.9538686
    - Koef. korelasi antara x2 dan x3 : 0.9683549
    - Koef. korelasi antara x2 dan x4 : 0.9723228
    - Koef. korelasi antara x2 dan x5 : 0.9469462
    - Koef. korelasi antara x3 dan x4 : 0.9415902
    - Koef. korelasi antara x3 dan x5 : 0.9719096
    - Koef. korelasi antara x4 dan x5 : 0.9228714
- Berdasarkan koefisien korelasi, semua korelasi antar variabel bebas diperoleh koefisien >0.80, maka dapat dipastikan telah terjadi multikolinearitas pada model.

Mengatasi Multikolinearitas Menggunakan Metode Seleksi Variabel Stepwise Backward Elimination

Langkah Mengatasi:

1. Seleksi variabel menggunakan Uji-F Parsial : menseleksi variabel dengan metode menghilangkan satu persatu variabel bebas untuk mendapatkan model terbaik dengan memperhatikan nilai R-squared dan AIC yang diperoleh.

2. Membentuk Model Baru: membentuk model terbaru yang sudah terseleksi dan juga melihat R-squared nya untuk mengetahui seberapa baik model ini. Uji Multikolinearitas model baru

```
> # Seleksi Variabel
> 
> #Uji-F Parsial (Backward elimination) menghilangkan satu per satu
 variabel penjelas
> full.model <- lm(y~., data=Data_Modlin)
> reduced.model <- step(full.model, direction="backward")
Start:  AIC=-10.48
y ~ x1 + x2 + x3 + x4 + x5

       Df Sum of Sq     RSS      AIC
- x3    1    0.0783  7.9958 -12.2668
- x1    1    0.2987  8.2163 -11.6684
<none>               7.9175 -10.4832
- x2    1    0.8688  8.7863 -10.1927
- x4    1    0.9827  8.9002  -9.9093
- x5    1    4.3522 12.2697  -2.8460

Step:  AIC=-12.27
y ~ x1 + x2 + x4 + x5

       Df Sum of Sq     RSS     AIC
- x1    1    0.2856  8.2814 -13.495
<none>               7.9958 -12.267
- x2    1    0.8193  8.8151 -12.121
- x4    1    0.9869  8.9827 -11.706
- x5    1    8.6436 16.6394   1.856

Step:  AIC=-13.49
y ~ x2 + x4 + x5

       Df Sum of Sq     RSS      AIC
- x2    1    0.6978  8.9793 -13.7148
<none>               8.2814 -13.4946
- x4    1    2.8116 11.0931  -9.0639
- x5    1   14.1791 22.4606   6.4558

Step:  AIC=-13.71
y ~ x4 + x5

       Df Sum of Sq     RSS      AIC
<none>               8.9793 -13.7148
- x4    1    2.7879 11.7671  -9.7661
- x5    1   15.6948 24.6740   6.5236
> reduced.model$anova #interpretasikan
  Step Df   Deviance Resid. Df Resid. Dev       AIC
1      NA         NA        16   7.917523 -10.48321
2 - x3  1 0.07827274        17   7.995795 -12.26679
3 - x1  1 0.28564666        18   8.281442 -13.49456
4 - x2  1 0.69780985        19   8.979252 -13.71477
> 
> 
```

- Indikator kebaikan model :
  - Nilai AIC (Akaike Information Criterion) : Semakin kecil nilainya maka semakin baik modelnya
- Berdasarkan output yang dihasilkan diperoleh nilai AIC tiap seleksi variable yaitu
  - AIC = -10.48 untuk y = x1 + x2 + x3 + x4 + x5

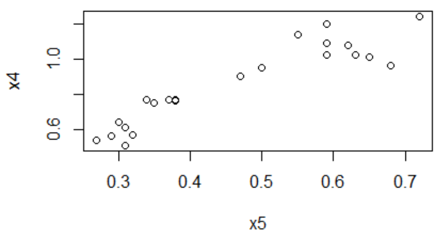
- AIC = -12.27 untuk y = x1 + x2 + x4 + x5
- AIC = -13.49 untuk y = x2 + x4 + x5
- AIC = -13.71 untuk y = x4 + x5

- Kesimpulan :

  Berdasarkan nilai AIC yang diperoleh dapat disimpulkan bahwa model terbaik untuk Data5 adalah y = x4 + x5. Dan variable yang tereliminasi yaitu x1, x2, dan x3.

Uji Multikolinearitas Model Baru

```
> # Model Baru Stepwise
> model_stepwise=lm(formula=y~x4+x5, data=Data5)
> summary(model_stepwise)

Call:
lm(formula = y ~ x4 + x5, data = Data5)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.56123 -0.49604  0.09069  0.45717  0.98057 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -6.3351     0.6009 -10.543 2.23e-09 ***
x4            4.1542     1.7104   2.429   0.0252 *  
x5           15.0160     2.6057   5.763 1.49e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6875 on 19 degrees of freedom
Multiple R-squared:  0.9584,	Adjusted R-squared:  0.954 
F-statistic: 218.9 on 2 and 19 DF,  p-value: 7.584e-14

> plot(x5,x4)
```




```
> ols_vif_tol(model_stepwise)
  Variables Tolerance      VIF
1        x4 0.1483085 6.742703
2        x5 0.1483085 6.742703
>

> #Mengetahui koefisien korelasi antara x4 dan x5
> korelasistepwise <- cor(x4,x5)
> korelasistepwise
[1] 0.9228714
>
```

- Setalah melakukan seleksi variable, didapat model regresi yang terbaik dari metode Stepwise yaitu $\hat{Y}$ = -6.3351 + 4.1542X4 + 15.0160X5.
- Dengan R-squared dan R-squared adjusted yang baru berturut-turut yaitu 0.9584 dan 0.954, artinya sebesar 95% model ini mampu menjelaskan keragaman variable y, dan sekitar 4% keragaman yang tidak mempu dijelaskan oleh variable x4 dan x5 dan terletak pada komponen error.

- Namun berdasarkan grafik plot yang dihasilkan, secara visual terlihat bahwa masih terdapat hubungan linear yang relatif kuat antara variabel x4 dan x5
- Berdasarkan nilai VIF dan tolerance, diperoleh nilai VIF $>$ 5 dan nilai tolerance $>$0.1 maka ini masih menunjukkan adanya multikolinearitas pada model
- Berdasarkan koeifisen korelasi diperoleh koefisien korelasi antara x4 dan x5 yang cukup kuat yaitu 0.9228714, sehingga pada model ini dapat dikatakan masih terjadi multikolinearitas.
- Selanjutnya untuk mendapatkan model yang terbaik, kita dapat menggunakan metode Analisis Komponen Utama / Principal Component (PCA)

Mengatasi Multikolinearitas Menggunakan Metode Analisis Komponen Utama / Principal Compenen (PCA)

Langkah Mengatasi :

1. Memeriksa nilai KMO : KMO digunakan untuk mengukur kecukupan sampling (sampling adequacy). Nilai ini membandingkan besarnya koefisien korelasi terobservasi dengan koefisien korelasi parsial. Nilai KMO yang kecil menunjukkan bahwa korelasi antar pasangan variabel tidak bisa diterangkan oleh variabel lainnya dan analisis faktor mungkin tidak tepat. Jika nilai KMO $>$0.5 maka model layak untuk dibentuk menjadi beberapa komponen utama / principal component (PC).
2. Membentuk komponen utama yaitu persamaan yang memuat variabel y, x1, x2, x3, x4, x5 dengan bantuan Software R
3. Memilih komponen utama :memilih PC mana saja yang akan digunakan untuk dibentuk ke dalam model baru dengan mempertimbangkan eigenvalue dan persentase variansi kumulatif.
4. Membentuk model baru : persamaan regresi yang memuat komponen utama sebagai variabel independent.
5. Memeriksa multikolinearitas model baru : melihat koefisien standar error, signifikansi, nilai VIF tiap komponen utama, dan melihat koefisien korelasi antar komponen utama.

1. Nilai KMO

```
> library(psych)
> library(factoextra)
> library(dplyr)
> # periksa KMO, jika KMO >0.5 maka model sudah layak untuk
dibentuk menjadi beberapa PC
> KMO(Data_Modlin)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = Data_Modlin)
Overall MSA =  0.88
MSA for each item = 
   y   x1   x2   x3   x4   x5 
0.88 0.92 0.88 0.91 0.85 0.85 
>
```

- Diperoleh nilai KMO
  - y : 0.88
  - x1 : 0.92
  - x2 : 0.88
  - x3 : 0.91
  - x4 : 0.85
  - x5 : 0.85
- Karena semua nilai KMO >0.5 maka model layak untuk dibentuk ke dalam beberapa komponen utama / principal component (PC)

2. Membentuk Komponen Utama (PC)

```
> # Komponen utama / Principal Component (PC)
> pca <- prcomp(Data_Modlin, scale=TRUE)
> pca
Standard deviations (1, .., p=6):
[1] 2.4049160 0.3254282 0.2368399 0.1523877 0.1321517 0.1170319

Rotation (n x k) = (6 x 6):
         PC1        PC2         PC3         PC4         PC5         PC6
y  0.4071798 -0.3805299  0.58695024 -0.41955246  0.10548433  0.39716468
x1 0.4107108  0.2382112  0.26349075  0.64217010 -0.49132675  0.22662815
x2 0.4092664  0.3334670 -0.44607536  0.05171378  0.54811791  0.46819867
x3 0.4090370 -0.2263647 -0.56861034 -0.36707852 -0.56771544 -0.03290267
x4 0.4061373  0.5806191  0.24142943 -0.32374061  0.06197151 -0.57532368
x5 0.4071406 -0.5467108 -0.07397927  0.41153901  0.34770181 -0.48949972
>
```

Diperoleh persamaan komonen utama (PC) :

- PC1 = 0.4071798y + 0.4107108x1 + 0.4092664x2 + 0.4090370x3 + 0.4061373x4 + 0.4071406x5

- PC2 = -0.3805299y + 0.2382112x1 + 0.3334670x2 - 0.2263647x3 + 0.5806191x4 -0.5467108x5
- PC3 = 0.58695024y + 0.26349075x1 - 0.44607536x2 - 0.56861034x3 + 0.24142943x4 -0.07397927x5
- PC4 = -0.41955246y + 0.64217010x1 + 0.05171378x2 - 0.36707852x3 - 0.32374061x4 + 0.41153901x5
- PC5 = 0.10548433y - 0.49132675x1 + 0.54811791x2 - 0.56771544x3 + 0.06197151x4 + 0.34770181x5
- PC6 = 0.39716468y + 0.22662815x1 + 0.46819867x2 - 0.03290267x3 - 0.57532368x4 - 0.48949972x5

3. Memilih Komponen Utama

```
> # melihat eigen value dari tiap PC
> get_eigenvalue(pca)
      eigenvalue variance.percent cumulative.variance.percent
Dim.1 5.78362086       96.3936809                    96.39368
Dim.2 0.10590349        1.7650581                    98.15874
Dim.3 0.05609312        0.9348853                    99.09362
Dim.4 0.02322200        0.3870334                    99.48066
Dim.5 0.01746407        0.2910678                    99.77173
Dim.6 0.01369647        0.2282745                   100.00000
>
```

- Diperoleh eigen value
  - dimensi 1 : 5.78362086
  - dimensi 2 : 0.10590349
  - dimensi 3 : 0.05609312
  - dimensi 4 : 0.02322200
  - dimensi 5 : 0.01746407
  - dimensi 6 : 0.01369647
- Persentase variansi kumulatif :
  - dimensi 1 : 96.39368
  - dimensi 2 : 98.15874
  - dimensi 3 : 99.09362
  - dimensi 4 : 99.48066
  - dimensi 5 : 99.77173
  - dimensi 6 : 100.00000

- Berdasarkan nilai eigenvalue yang diperoleh, PC yang akan digunakan dalam model baru ini adalah PC1, PC2, dan PC3 yang akan menjelaskan 99.09362% keragaman pada model

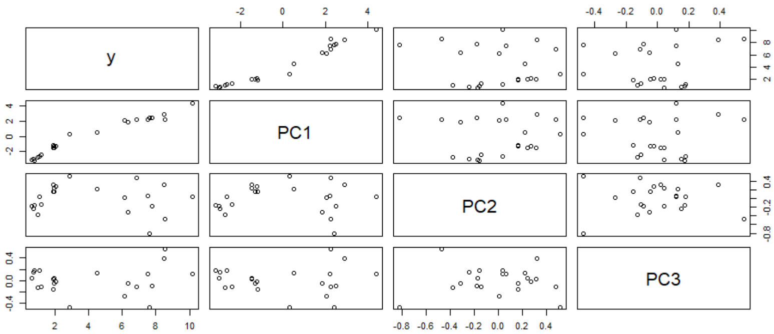
4. Membentuk Model Baru

```
 > # dalam model baru ini akan digunakan PC1, PC2, PC3 yang bisa
mejelaskan 99.09362% keragaman data
> # PC1 = 0.4071798y + 0.4107108x1 + 0.4092664x2 + 0.4090370x3 +
0.4061373x4 + 0.4071406x5 dan seterusnya
> pcs <- as.data.frame(pca$x)
> pcs1 <- pcs %>% select(1:3)
> y1 = Data_Modlin %>% select(y)
> datapca <- cbind(y1, pcs1)
>
> # Model regresi linear berganda dengan Principal Component
> modelpca <- lm(y~., data = datapca)
```

Model baru yang akan dibuat adalah model dengan y sebagai variabel dependen dan komponen utama PC1, PC2, PC3 sebagai variabel independent nya.

5. Uji Multikolinearitas Model Baru

```
> summary(modelpca)

Call:
lm(formula = y ~ ., data = datapca)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.56261 -0.20094  0.03328  0.16820  0.54542 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.19500    0.05927  70.779  < 2e-16 ***
PC1          1.30565    0.02522  51.760  < 2e-16 ***
PC2         -1.22020    0.18641  -6.546 3.76e-06 ***
PC3          1.88210    0.25614   7.348 8.05e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.278 on 18 degrees of freedom
Multiple R-squared:  0.9936,	Adjusted R-squared:  0.9925 
F-statistic: 925.3 on 3 and 18 DF,  p-value: < 2.2e-16

> ols_vif_tol(modelpca)
  Variables Tolerance VIF
1       PC1         1   1
2       PC2         1   1
3       PC3         1   1
> 
> #Melihat grafik linearitas antar PC
> plot(datapca)
```




```
> `
> #Mengetahui koefisien korelasi antar PC
> korelasipca <- cor(datapca)
> korelasipca
              y          PC1           PC2           PC3
y     1.0000000 9.792332e-01 -1.238352e-01  1.390132e-01
PC1   0.9792332 1.000000e+00  1.401037e-15  1.001355e-15
PC2  -0.1238352 1.401037e-15  1.000000e+00 -1.384717e-16
PC3   0.1390132 1.001355e-15 -1.384717e-16  1.000000e+00
> 
```

- Diperoleh nilai R-squared adjusted yaitu 99.25%, dan berdasarkan p-value, semua variabel bebas PC1, PC2, PC3 sudah signifikan terhadap model . Maka hal ini menunjukkan tidak adanya gejala multikolinearitas pada model.
- Nilai koefisien standar error untuk tiap variabel bebas PC yaitu :
  - PC1 : 0.02522
  - PC2 : 0.18641
  - PC3 : 0.25614
- Berdasarkan output yang diperoleh, semua koefisien standar error <1, maka dapat diindikasikan tidak adanya multikolienaritas.
- Secara visual, dapat dilihat bahwa tidak terdapat hubungan yang relatif linear antar variabel-variabel PC, sehingga ini semakin mendukung indikasi tidak adanya multikolinearitas pada model.
- Nilai koefisien korelasi antar variabel bebas
  - Koef. korelasi antara PC1 dan PC2 : 1.401037e-15
  - Koef. korelasi antara PC1 dan PC3 : 1.001355e-15
  - Koef. korelasi antara PC2 dan PC3 : -1.384717e-16
- Berdasarkan koefisien korelasi, semua korelasi antar variabel bebas diperoleh koefisien <0.80, maka dapat disimpulkan sudah tidak terjadi multikolinearitas pada model.
- Nilai VIF tiap variabel bebas
  - Nilai VIF PC1 : 1
  - Nilai VIF PC2 : 1
  - Nilai VIF PC3 : 1
- Nilai Tolerance tiap variabel bebas
  - Nilai tolerance PC1 : 1
  - Nilai tolerance PC2 : 1
  - Nilai tolerance PC3 : 1
- Berdasarkan nilai VIF dan Tolerance yang diperoleh, semua variabel bebas memiliki nilai VIF <10 dan tolerance >0.1. Maka ini menunjukkan sudah tidak adanya multikolinearitas pada model
- Berdasarkan 5 pengujian ini, maka dapat dipastikan sudah tidak terjadi multikolinearitas pada model $y = 1.30565PC1 - 1.22020PC2 + 1.88210PC3$
- Kemudian variable PC kita transformasikan kedalam ke x1, x2, x3, x4, x5 sehingga persamaan regresi yang baru menjadi $y = -0.673684814x1 + 0.646974381x2 - 1.708482028x3 - 0.250937701x4 - 0.962556x5$

- Pada model ini diperoleh R-squared adjusted sebesar 0.9925 artinya sebesar 99.25% model ini mampu menjelaskan keragaman variable y, dan sekitar 0.75% keragaman yang tidak mampu dijelaskan oleh variable x1, x2, x3, x4, x5 dan terletak pada komponen error.

BAB IV
KESIMPULAN

Berdasarkan hasil analisis data diperoleh kesimpulan sebagai berikut :

1. Persamaan regresi awal yang diperoleh adalah $\hat{Y} = -6.512215 + 1.999413X_1 - 3.675096X_2 + 2.524486X_3 + 5.158082X_4 + 14.401162X_5$ dan didapatkan R-squared adjusted sebesar 95.18733 Hal ini berarti sebesar 95.2% model mampu menjelaskan keragaman variabel y dan sekitar 4.8% keragaman yang tidak mampu dijelaskan oleh variabel penjelas dan terletak pada komponen error.
2. Berdasarkan uji asumsi multikolinearitas dengan melihat koefisien standar error, nilai VIF, Tolerance dan nilai koefisien korelasi antar variabel bebas, disimpulkan bahwa telah terjadi pelanggaran asumsi multikolinearitas dengan adanya korelasi linear antar variabel bebas yang cukup kuat yaitu >0.90 atau di atas 90%
3. Untuk mengatasi pelanggaran multikolinearitas ini, dengan metode Stepwise Backward Elimination, berdasarkan nilai AIC yang diperoleh didapatkan variabel bebas yang terseleksi yaitu x4 dan x4 sehingga diperoleh persamaan baru yaitu $\hat{Y} = -6.3351 + 4.1542X4 + 15.0160X5$.
4. Namun pada model kedua ini, nyatanya belum bisa mengatasi pelanggaran yang terjadi. Koefisien korelasi antara x4 dan x5 masih cukup kuat yaitu 0.9228714 atau di atas 92.3%
5. Dengan metode kedua yaitu analisis komponen utama (PCA) dengan mentransformasi variabel x1, x2, x3, x4, x5 ke dalam beberapa komponen utama diperoleh persamaan baru yaitu $\hat{Y} = 1.30565PC1 - 1.22020PC2 + 1.88210PC3$ dengan
   - $PC1 = 0.4071798y + 0.4107108x1 + 0.4092664x2 + 0.4090370x3 + 0.4061373x4 + 0.4071406x5$
   - $PC2 = -0.3805299y + 0.2382112x1 + 0.3334670x2 - 0.2263647x3 + 0.5806191x4 - 0.5467108x5$
   - $PC3 = 0.58695024y + 0.26349075x1 - 0.44607536x2 - 0.56861034x3 + 0.24142943x4 -0.07397927x5$
6. Setelah dilakukan uji asumsi multikolinearitas pada model ini, dengan melihat koefisien standar error, nilai VIF, Tolerance dan nilai koefisien korelasi antar variabel bebas, pelanggaran multikolinearitas telah berhasil diatasi dengan koefisien korelasi antar komonen utama yang relatif lemah yaitu <0.01 atau di bawah 1%

7. Sehingga didapatkan model terbaik untuk data ini yaitu Ŷ = -0.673684814x1 + 0.646974381x2 -1.708482028x3 -0.250937701x4 -0.962556x5. Dengan adjusted R squared 0.9925.
   - Hal ini berarti:
     - b_1 = -0.673684814 menunjukkan bahwa ketika terjadi perubahan satu satuan x1 maka akan meningkatkan rata-rata Y sebesar -0.673684814, dengan asumsi x2, x3, x4, x5 tetap.
     - b_2 = 0.646974381 menunjukkan bahwa ketika terjadi perubahan satu satuan x2 maka akan menurunkan rata-rata Y sebesar 0.646974381, dengan asumsi x1, x3, x4, x5 tetap.
     - b_3 = -1.708482028 menunjukkan bahwa ketika terjadi perubahan satu satuan x3 maka akan meningkatkan rata-rata Y sebesar -1.708482028, dengan asumsi x1, x2, x4, x5 tetap.
     - b_4 = -0.250937701 menunjukkan bahwa ketika terjadi perubahan satu satuan x4 maka akan meningkatkan rata-rata Y sebesar -0.250937701, dengan asumsi x1, x2, x3, x5 tetap.
     - b_5 = -0.962556 menunjukkan bahwa ketika terjadi perubahan satu satuan x5 maka akan meningkatkan rata-rata Y sebesar -0.962556, dengan asumsi x1, x2, x3, x4 tetap.
     - Sebesar 99.25% model mampu menjelaskan keragaman variabel y dan sekitar 0,75% keragaman yang tidak mampu dijelaskan oleh variabel penjelas dan terletak pada komponen error.

DAFTAR PUSTAKA

Gujarati, D. N, dkk. 1995. Basics Ekonometrics, Mc Graw Hill, Inc. New York.

Hamdi, Asep Saepul dan E. Baharuddin. 2014. Metode Penelitian Kuantitaf Aplikasi Dalam Pendidikan. Yogyakarta: Deepublish.

Hidayat, Anwar. 2016. Pengertian Multikolinearitas dan dampaknya. Jakarta: Statistikian

Dunteman, H. George , (1989). Principal Component Analysis. Sage Publications., Newbury Park London New Delhi. (Reseach Triangle Institute).

Cahyadi, Daniel. 2007. Ekstraksi dan Kemiripan. Universitas Indonesia.

Tabachnick B 7 Fidell L.S (2001).Using Multivariate Statistics, 4rd ed. Boston : Allyn & Bacon.