

**PRINCIPAL COMPONENT ANALYSIS (PCA)  
SEBAGAI SALAH SATU METODE UNTUK MENGATASI  
MASALAH MULTIKOLINEARITAS**

**OLEH :  
SOEMARTINI**



**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PADJADJARAN  
JATINANGOR  
2008**

## DAFTAR ISI

<b>Daftar Isi</b> .....	i
<b>BAB 1. Pendahuluan</b>	
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	2
1.3 Maksud dan Tujuan .....	2
1.4 Manfaat.....	2
<b>BAB 2. TINJAUAN PUSTAKA</b>	
2.1 Landasan Teori .....	3
2.2 Analisis Contoh Kasus .....	6
<b>BAB 3. KESIMPULAN DAN SARAN</b>	
3.1 Kesimpulan .....	17
3.2 Saran .....	17
<b>Daftar Pustaka</b> .....	18

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi merupakan analisis yang mempelajari bagaimana membangun sebuah model fungsional dari data untuk dapat menjelaskan ataupun meramalkan suatu fenomena alami atas dasar fenomena yang lain. Ada juga yang menyatakan bahwa analisis regresi merupakan suatu analisis mengenai hubungan antara dua variable atau lebih yang umumnya dinyatakan dalam persamaan matematik.

Dalam statistika sebuah model regresi dikatakan baik atau cocok, jika dipenuhi asumsi-asumsi ideal (klasik), yakni tidak adanya otokorelasi, heteroskedastisitas dan multikolinieritas. Sehingga proses control terhadap model perlu dilakukan untuk menelaah dipenuhi tidaknya asumsi tersebut.

Salah satu dari ketiga asumsi model regresi linier klasik adalah yakni tidak terdapat multikolinieritas di antara variabel yang menjelaskan yang termasuk dalam model. Ketika menentukan model regresi populasi ada kemungkinan bahwa dalam sampel tertentu, beberapa atau semua variable X sangat kolinear (mempunyai hubungan linear sempurna atau hampir sempurna).

Ada beberapa prosedur yang dapat digunakan untuk mengatasi masalah multikolinieritas, seperti : penggunaan informasi apriori dari hubungan beberapa variable yang berkolinear, menghubungkan data cross-sectional dan data time series, mengeluarkan suatu variabel atau beberapa variabel bebas yang terlibat hubungan kolinear, melakukan transformasi variabel dengan prosedur first difference, melalui  $\ln$  (logaritma) dan penambahan data baru dan juga melalui ridge regression. Akan tetapi pada prakteknya prosedur penanggulangan yang telah disebutkan di atas sangat tergantung sekali pada kondisi penelitian, misalnya : penggunaan informasi apriori sangat tergantung dari ada atau tidaknya dasar teori (literatur) yang sangat kuat untuk mendukung hubungan matematis antara variabel bebas yang saling berkolinear,

prosedur mengeluarkan variabel bebas yang berkolinear seringkali membuat banyak peneliti keberatan karena prosedur ini akan mengurangi obyek penelitian yang diangkat, sedangkan prosedur lainnya seperti menghubungkan data cross sectional dan time series, prosedur first difference dan penambahan data baru seringkali hanya memberikan efek penanggulangan yang kecil pada masalah multikolinearitas.

## **1.2 Identifikasi Masalah**

Berdasarkan paparan yang telah dijelaskan pada bagian latar belakang, maka dapat diidentifikasi masalah sebagai berikut :

- a. Bagaimana mendeteksi ada tidaknya salah satu Pelanggaran Asumsi Regresi Linear Klasik (multikolinearitas) ?
- b. Bagaimana prosedur penanggulangan masalah multikolinearitas dengan metode Principal Component Analysis (PCA) sebagai metode paling jitu ?
- c. Bagaimana menaksir Model Regresi Linear dalam kondisi ideal ?

## **1.3 Maksud dan Tujuan**

Maksud dan tujuan penyusunan makalah ini adalah sebagai berikut :

- a. Untuk mendeteksi ada tidaknya Pelanggaran Asumsi Regresi Linear Klasik (multikolinearitas).
- b. Untuk mendeskripsikan prosedur penanggulangan masalah multikolinearitas dengan metode Principal Component Analysis (PCA)
- c. Untuk menaksir Model Regresi Linear Multiple dalam kondisi ideal.

## **1.4 Manfaat**

Manfaat yang dapat diperoleh dari penyusunan makalah ini adalah kita dapat memahami tindakan yang perlu dilakukan dalam mengidentifikasi serta menanggulangi keberadaan mulikolineritas dan juga sebagai bahan referensi untuk penelitian-penelitian yang akan datang dalam data yang akan dianalisis.

## BAB II TINJAUAN PUSTAKA

### 2.1 Landasan Teori

Istilah Multikolinearitas pertama kali ditemukan oleh Ragnar Frisch yang berarti adanya hubungan linier yang “sempurna” atau pasti diantara beberapa atau semua variabel bebas dari model regresi berganda.

Multikolinearitas dapat terjadi karena:

1. Terdapat kecenderungan variabel ekonomi bergerak secara bersama-sama sepanjang waktu —————> Pertumbuhan kecenderungan factor-faktor dalam deret waktu dapat sebagai penyebab terjadinya multikolinearitas.

2. Penggunaan Lag —————> Sehingga terdapat model distribusi lag

Misal :  $C_t = f(Y_t, Y_{t-1}, \dots, Y_1)$  —————> Mungkin terdapat korelasi yang kuat antara  $Y_t$  dan  $Y_{t-1}$

- Multikolinearitas diperkirakan akan muncul dalam kebanyakan hubungan –hubungan ekonomi
- Lebih sering muncul dalam data deret waktu —————> bisa pula muncul dalam data cross sectional.

Menurut *Gujarati (1978)* gejala Multikolinearitas ini dapat didiagnosis dengan beberapa cara antara lain :

1. Menghitung koefisien korelasi sederhana (simple correlation) antara sesama variabel bebas, jika terdapat koefisien korelasi sederhana yang mencapai atau melebihi 0,8 maka hal tersebut menunjukkan terjadinya masalah multikolinearitas dalam regresi.
2. Menghitung nilai Toleransi atau VIF (Variance Inflation Factor), jika nilai Toleransi kurang dari 0,1 atau nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel bebas.

3. TOL yakni Ukuran toleransi untuk mendeteksi Multikoleniaritas

$$TOL_i = \frac{1}{VIF_i} = 1 - R^2_i$$

1 Jika  $X_i$  tidak berkorelasi dengan regresor lainnya

$TOL_i$

0 Jika  $X_i$  berkorelasi dengan regresor lainnya  $\longrightarrow$  korelasi sempurna

4. Lakukan regresi antar variabel bebas, kemudian melakukan uji-F dan bandingkan dengan F tabel. Jika nilai F hitung melebihi nilai F tabel berarti dapat dinyatakan bahwa  $X_i$  kolinier dengan  $X$  yang lain.

5. Dengan Nilai Eigen dan Indeks Kondisi (IK)

Output SAS dari F. Produksi Cobb-douglas menggunakan nilai Eigen dan Indeks Kondisi untuk mengdiagnosis Multikolinearitas

*Bilangan Kondisi :*

$$K = \frac{\lambda_{Max}}{\lambda_{Min}}$$

$\lambda$  : Nilai eigen

*Indeks Kondisi :*

$$ID = \sqrt{K}$$

Jika :  $100 \leq K \leq 1000$  : Dari sedang menuju Kuat

$K > 100$  : sangat Kuat

$ID < 10$  : Lemah

$10 \leq ID \leq 30$  : Sedang menuju Kuat

ID > 30 : Sangat Kuat

## 6. Melalui metoda Ridge Regression

Montgomery dan Hines (1990) menjelaskan bahwa dampak multikolinearitas dapat mengakibatkan koefisien regresi yang dihasilkan oleh analisis regresi berganda menjadi sangat lemah atau tidak dapat memberikan hasil analisis yang mewakili sifat atau pengaruh dari variabel bebas yang bersangkutan. Dalam banyak hal masalah Multikolinearitas dapat menyebabkan uji T menjadi tidak signifikan padahal jika masing-masing variabel bebas diregresikan secara terpisah dengan variabel tak bebas (simple regression) uji T menunjukkan hasil yang signifikan. Hal tersebutlah yang sering kali membuat pusing para peneliti karena hasil analisis yang dilakukan pada regresi berganda dan regresi sederhana tidaklah sejalan atau bahkan sangat bertentangan.

Akan tetapi, pada prakteknya prosedur penanggulangan yang telah disebutkan di atas

sangat tergantung sekali pada kondisi penelitian, misalnya prosedur penggunaan informasi apriori sangat tergantung dari ada atau tidaknya dasar teori (literatur) yang sangat kuat untuk mendukung hubungan matematis antara variabel bebas yang saling berkolinear, prosedur mengeluarkan variabel bebas yang berkolinear seringkali membuat banyak peneliti keberatan karena prosedur ini akan mengurangi obyek penelitian yang diangkat, sedangkan prosedur lainnya seperti menghubungkan data cross sectional dan time series, prosedur first difference dan penambahan data baru seringkali hanya memberikan efek penanggulangan yang kecil pada masalah multikolinearitas.

Oleh karena itu, kita dapat menggunakan teknik lain yang dapat digunakan untuk meminimumkan masalah multikolinearitas tanpa harus mengeluarkan variabel bebas yang terlibat hubungan kolinear, yaitu dengan metode *Principal Component Analysis (PCA)* yang ada dalam analisis faktor.

Prosedur PCA pada dasarnya adalah bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi

variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan principal component. Setelah beberapa komponen hasil PCA yang bebas multikolinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi.

***Keuntungan penggunaan Principal Component Analysis (PCA) dibandingkan metode lain :***

1. Dapat menghilangkan korelasi secara bersih (korelasi = 0) sehingga masalah multikolinearitas dapat benar-benar teratasi secara bersih.
2. Dapat digunakan untuk segala kondisi data / penelitian
3. Dapat dipergunakan tanpa mengurangi jumlah variabel asal
4. Walaupun metode Regresi dengan PCA ini memiliki tingkat kesulitan yang tinggi akan tetapi kesimpulan yang diberikan lebih akurat dibandingkan dengan penggunaan metode lain.

## **2.2 Analisis Contoh Kasus**

Klein dan Goldberger mencoba mencocokkan model regresi berikut ini terhadap ekonomi Amerika Serikat :

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U_i$$

Dimana Y = Konsumsi ( dalam milyar dolar)

$X_1$  = Pendapatan upah ( dalam milyar dolar)

$X_2$  = Pendapatan non upah dan non pertanian ( dalam milyar dolar)

$X_3$  = Pendapatan pertanian ( dalam milyar dolar)

Berikut ini adalah data pengamatan untuk kasus di atas :

**Tabel 1**  
**Data Mengenai Konsumsi (Y), Pendapatan Upah (X<sub>1</sub>), Pendapatan Non Upah dan Non Pertanian (X<sub>2</sub>), dan Pendapatan Pertanian (X<sub>3</sub>) di Amerika Serikat**  
**Dalam Milyar dollar**

<b>Tahun</b>	<b>Y</b>	<b>X<sub>1</sub></b>	<b>X<sub>2</sub></b>	<b>X<sub>3</sub></b>
1936	62,8	43,41	17,1	3,96
1937	65	46,44	18,65	5,48
1938	63,9	44,35	17,09	4,37
1939	67,5	47,82	19,28	4,51
1940	71,3	51,02	23,24	4,88
1941	76,6	58,71	28,11	6,37
1945*	86,3	87,69	30,29	8,96
1946	95,7	76,73	28,26	9,76
1947	98,3	75,91	27,91	9,31
1948	100,3	77,62	32,3	9,85
1949	103,2	78,01	31,39	7,21
1950	108,9	83,57	35,61	7,39
1951	108,5	90,59	37,58	7,98
1952	111,4	95,47	35,17	7,42

Sumber : L.R. Klein dan A.S. Goldberger, *An Economic Model of The United States, 1929 - 1952*, North Holland Publishing Company, Amsterdam, 1964, Hal. 131

\* Data-data untuk tahun-tahun perang 1942 - 1944 hilang. Data untuk tahun-tahun lain dalam Milyar Dolar tahun 1939

Berdasarkan data di atas, kita ingin mengetahui ada atau tidaknya pelanggaran salah satu asumsi klasik multikolinearitas dalam data tersebut, sekaligus cara penanggulangannya jika pada data tersebut memang terdapat pelanggaran multikolinearitas, sehingga kita dapat memperoleh model regresi yang cocok untuk menggambarkan keadaan ekonomi di Amerika Serikat.

Untuk menyelesaikannya, kita harus melakukan langkah-langkah sebagai berikut :

**a. Uji Multikolinearitas**

Untuk menguji ada tidaknya multikolinearitas, kita dapat menggunakan nilai Toleransi atau VIF (Variance Inflation Factor), dengan rumus sebagai berikut :

$$VIF = \frac{1}{1-r_{12}^2} \qquad \text{Tolerance} = \frac{1}{VIF} = (1-r_{12}^2)$$

Jika nilai Toleransi kurang dari 0,1 atau nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel bebas.

Dengan bantuan *software SPSS 13*, kita dapat memperoleh nilai Toleransi atau VIF untuk data di atas pada tabel berikut ini.

**Tabel 2**  
**Nilai Toleransi dan VIF**

**Coefficients<sup>a</sup>**

Model		Collinearity Statistics	
		Tolerance	VIF
1	Pendapatan Upah	,081	12,297
	Pendapatan Non Upah dan Non Pertanian	,108	9,230
	Pendapatan Pertanian	,336	2,977

a. Dependent Variable: Konsumsi

Berdasarkan tabel 2 di atas, dapat dilihat bahwa terdapat nilai VIF (12,297) yang melebihi 10 dan nilai Tolerance (0,081) yang kurang dari 0,1. Oleh karena itu, kita menduga adanya pelanggaran asumsi klasik multikolinearitas pada data tersebut.

Selain itu, kita dapat mendeteksi multikolinearitas melalui koefisien korelasi sederhana di antara variabel bebas. Jika koefisien korelasi sederhana mencapai atau melebihi 0,8 maka hal tersebut menunjukkan terjadinya masalah multikolinearitas dalam regresi.

Melalui SPSS 13 berdasarkan table 1 , diperoleh nilai-nilai korelasi ( korelasi parsial ) antara Pendapatan upah, Pendapatan non upah dan Pendapatan pertanian

**Tabel 3. Koefesien korelasi**

**Correlations**

		Pendapatan Upah	Pendapatan Non Upah dan Non Pertanian	Pendapatan Pertanian
Pendapatan Upah	Pearson Correlation	1	,943**	,811**
	Sig. (2-tailed)		,000	,000
	N	14	14	14
Pendapatan Non Upah dan Non Pertanian	Pearson Correlation	,943**	1	,737**
	Sig. (2-tailed)	,000		,003
	N	14	14	14
Pendapatan Pertanian	Pearson Correlation	,811**	,737**	1
	Sig. (2-tailed)	,000	,003	
	N	14	14	14

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Berdasarkan tabel 3 di atas, dapat dilihat bahwa terdapat korelasi yang tinggi antara variable bebas pendapatan upah ( $X_1$ ) dengan variabel pendapatan non upah dan non pertanian ( $X_2$ ) yaitu sebesar 0,943, dan antara variable bebas pendapatan upah ( $X_1$ ) dengan variabel pendapatan pertanian ( $X_3$ ) yaitu sebesar 0,811. Dengan adanya nilai korelasi yang tinggi tersebut (melebihi 0,8), maka kita dapat menduga adanya pelanggaran asumsi klasik multikolinearitas pada data tersebut.

#### **b. Penanggulangan Multikolinearitas**

Karena pada pengujian sebelumnya telah menunjukkan bahwa terdapat permasalahan multikolinearitas dalam data tersebut, maka dilakukan penanggulangan untuk mengatasi masalah multikolinearitas tersebut. Dalam hal ini kita menggunakan prosedur *Principal Component Analysis (PCA)* untuk mengatasi multikolinearitas. Prosedur PCA pada dasarnya bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali.

Dengan bantuan software SPSS 13 , kita dapat menggunakan analisis faktor (prosedur PCA) untuk mereduksi variabel – variabel bebas yang berkorelasi tinggi sehingga kita dapat menilai variabel mana saja yang benar – benar dianggap layak untuk dimasukkan dalam analisis selanjutnya.

### KMO and Barlett Test

**Tabel 4**

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.666
Bartlett's Test of Sphericity	Approx. Chi-Square	36,774
	df	3
	Sig.	,000

Kesimpulan tentang layak tidaknya analisis factor dilakukan, baru sah secara statistik dengan menggunakan uji *Kaiser-Meyer-Olkin (KMO) measure of adequacy* dan *Barlett Test of Sphericity*. Apabila nilai KMO berkisar antara 0,5 sampai 1, maka analisis factor layak dilakukan. Sebaliknya, jika nilai KMO di bawah 0,5 maka analisis factor tidak layak dilakukan. Dari tabel 4 di atas terlihat bahwa nilai KMO adalah 0,666, maka analisis factor layak dilakukan.

Barlett Test merupakan tes statistik untuk menguji apakah betul variabel-variabel bebas yang dilibatkan berkorelasi.

**H<sub>0</sub>** : Tidak ada korelasi antarvariabel bebas

**H<sub>1</sub>** : Ada korelasi antarvariabel bebas

Kriteria uji dengan melihat p-value (signifikansi) :

Terima H<sub>0</sub> jika Sig. > 0,05 atau tolak H<sub>0</sub> jika Sig. < 0,05

Pada tabel 4 di atas dapat dilihat bahwa nilai Chi-Square adalah 36,774, dengan derajat bebas sebesar 3, dan p-value (sig) sebesar 0,000. Karena p-value (0,000) < 0,05 maka H<sub>0</sub> di tolak. Artinya, benar-benar terdapat korelasi antarvariabel bebas.

## Anti-image Matriks

**Tabel 5**

**Anti-image Matrices**

		Pendapatan Upah	Pendapatan Non Upah dan Non Pertanian	Pendapatan Pertanian
Anti-image Covariance	Pendapatan Upah	,081	-,082	-,085
	Pendapatan Non Upah dan Non Pertanian	-,082	,108	,027
	Pendapatan Pertanian	-,085	,027	,336
Anti-image Correlation	Pendapatan Upah	,601 <sup>a</sup>	-,873	-,514
	Pendapatan Non Upah dan Non Pertanian	-,873	,647 <sup>a</sup>	,141
	Pendapatan Pertanian	-,514	,141	,809 <sup>a</sup>

a. Measures of Sampling Adequacy(MSA)

Perhatikan bagian Anti-image Correlation, khususnya pada angka korelasi yang bertanda a (arah diagonal dari kiri atas ke kanan bawah). Angka MSA (*Measure of Sampling Adequacy*) berkisar dari 0 sampai 1, dengan kriteria :

- MSA = 1, variabel tersebut dapat diprediksi tanpa kesalahan oleh variabel lain.
- MSA > 0.5, variabel masih bisa diprediksi dan bisa dianalisa lebih lanjut.
- MSA < 0.5, variabel tidak bisa diprediksi dan tidak bisa dianalisis lebih lanjut, atau dikeluarkan dari variabel lainnya.

Dengan melihat kriteria angka MSA di atas, terlihat pada tabel 5 bahwa semua angka MSA memiliki nilai di atas 0,5. Artinya, analisis dapat dilanjutkan.

## Communalities

**Tabel 6**

**Communalities**

	Initial	Extraction
Pendapatan Upah	1,000	,954
Pendapatan Non Upah dan Non Pertanian	1,000	,905
Pendapatan Pertanian	1,000	,805

Extraction Method: Principal Component Analysis.

Communalities menunjukkan berapa varians yang dapat dijelaskan oleh factor yang diekstrak (factor yang terbentuk). Cara memperolehnya adalah dengan mengkuadratkan nilai korelasi yang terdapat pada tabel 8 (Component Matrix). Setiap variabel berkorelasi dengan faktor-faktor yang terbentuk.

Untuk variabel pendapatan upah, diperoleh nilai sebesar 0,954. Hal ini berarti sekitar 95,4% variabel pendapatan upah dapat dijelaskan oleh faktor yang terbentuk. Untuk variabel pendapatan non upah dan non pertanian, diperoleh nilai sebesar 0,905. Hal ini berarti sekitar 90,5% variabel pendapatan non upah dan non pertanian dapat dijelaskan oleh faktor yang terbentuk. Untuk variabel pendapatan pertanian, diperoleh nilai sebesar 0,805. Hal ini berarti sekitar 80,5% variabel pendapatan pertanian dapat dijelaskan oleh faktor yang terbentuk.

### Total Variance Explained

**Tabel 7**

**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,664	88,787	88,787	2,664	88,787	88,787
2	,288	9,599	98,386			
3	,048	1,614	100,000			

Extraction Method: Principal Component Analysis.

Jika ada 3 variabel yang dilibatkan, maka akan ada 3 faktor (disebut juga component) yang diusulkan dalam analisis factor. Setiap factor mewakili variabel-variabel yang dianalisis. Kemampuan setiap factor mewakili variabel-variabel yang dianalisis ditunjukkan oleh besarnya varians yang dijelaskan, yang disebut dengan *eigenvalue*. Varians yang dimaksud adalah varians variabel-variabel yang sudah distandardisasi. Dengan standardisasi, nilai rata-rata setiap variabel menjadi nol dan variansnya menjadi satu. Karena varians setiap variabel adalah satu, maka varians totalnya ada 3 karena dalam kasus ini ada 3 variabel bebas.

*Eigenvalues* menunjukkan kepentingan relatif masing-masing factor dalam menghitung varians ketiga variabel yang dianalisis. Susunan eigenvalues selalu

diurutkan dari yang terbesar sampai ke yang terkecil, dengan kriteria bahwa angka eigenvalues di bawah 1 tidak digunakan dalam menghitung jumlah faktor yang terbentuk.

Dari tabel 7 di atas terlihat bahwa hanya satu factor yang terbentuk, karena dengan satu faktor, angka *eigenvalues* memiliki nilai di atas 1. Sedangkan untuk 2 faktor, angka *eigenvalues* sudah di bawah 1. Sehingga proses factoring seharusnya berhenti pada satu faktor saja. Faktor 1 memiliki *eigenvalue* sebesar 2,664, artinya faktor 1 ini dapat menjelaskan 2,664 atau 88,787% dari total *communalities*.

### Component Matrix

**Tabel 8**

**Component Matrix<sup>a</sup>**

	Component
	1
Pendapatan Upah	,977
Pendapatan Non Upah dan Non Pertanian	,951
Pendapatan Pertanian	,897

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Tabel ini berisikan *factor loading* (nilai korelasi) antara variabel-variabel analisis dengan faktor yang terbentuk. Berdasarkan tabel 8 di atas, terlihat bahwa hanya satu faktor yang terbentuk dari ketiga variabel. Hal ini menunjukkan bahwa satu faktor adalah jumlah yang paling optimal untuk mereduksi ketiga variabel bebas tersebut.

### Component Score Coefficient Matrix

**Tabel 9**

**Component Score Coefficient Matrix**

	Component
	1
Pendapatan Upah	,367
Pendapatan Non Upah dan Non Pertanian	,357
Pendapatan Pertanian	,337

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Component Scores.

Setelah kita mendapatkan faktor yang terbentuk melalui proses reduksi, maka kita perlu menari persamannya. Dengan persamannya tersebut, kita bisa mencari skor setiap faktor secara manual. Persamaan yang dibuat mirip dengan regresi linier berganda, hanya dalam persamaan faktornya tidak terdapat konstanta.

Dengan menggunakan hasil dari tabel 9 di atas, maka persamaan untuk faktor baru yang terbentuk adalah sebagai berikut :

$$F_1 = 0,367 X_1 + 0,357 X_2 + 0,337 X_3$$

Skor-skor faktor yang dihasilkan dapat digunakan untuk menggantikan skor-skor pada variabel bebas yang asli. Setelah komponen hasil PCA yang bebas multikolinearitas diperoleh maka komponen-komponen tersebut diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi linier.

### c. Model Regresi yang Cocok

Setelah kita mendapatkan variabel bebas baru ( $F_1$ ) yang bebas multikolinearitas melalui teknik *Principal Component Analysis* (PCA), maka kita akan meregresikan variabel bebas yang baru ( $F_1$ ) tersebut terhadap variabel tak bebas (Y). Kita misalkan saja variabel bebas baru ( $F_1$ ) tersebut dengan nama variabel pendapatan. Karena variabel bebas baru ( $F_1$ ) yang terbentuk hanya satu, maka pada model tersebut digunakan analisis regresi linier sederhana dengan persamaan sebagai berikut :

$$Y_i = \beta_0 + \beta_1 F_1 + U_i$$

Dimana : Y = Konsumsi ( dalam milyar dolar)

$F_1$  = Pendapatan ( dalam milyar dolar)

$$= 0,367 X_1 + 0,357 X_2 + 0,337 X_3$$

**Tabel 10****Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	87.121	1.736		50.191	.000
	F1	17.568	1.801	.942	9.753	.000

a. Dependent Variable: Konsumsi

Dari tabel 10 di atas, diperoleh model regresi sebagai berikut :

$$Y = 87,121 + 17,568 F_1$$

dengan  $F_1 = 0,367 X_1 + 0,357 X_2 + 0,337 X_3$

Artinya,

- Jika Konsumsi (Y) sama sekali tidak dipengaruhi oleh variabel pendapatan ( $F_1$ ), maka konsumsi di Amerika Serikat akan bernilai 87,121 milyar dolar.
- Untuk setiap kenaikan variabel pendapatan ( $F_1$ ) sebesar satu milyar dolar, akan mengakibatkan meningkatnya konsumsi (Y) di Amerika Serikat sebesar 17,568 milyar dolar.

Dari tabel 10 pun dapat dilihat bahwa sig. bernilai  $0,000 < 0,05$ , maka dapat disimpulkan bahwa pendapatan memiliki pengaruh yang signifikan terhadap konsumsi di Amerika Serikat.

Jika kita ingin mengetahui seberapa kuat hubungan yang terjadi antara variabel pendapatan dengan konsumsi di Amerika Serikat, maka kita dapat melihatnya melalui koefisien korelasi Pearson.

**Tabel 11****Koefisien Korelasi Pearson****Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.942 <sup>a</sup>	.888	.879	6.49478

a. Predictors: (Constant), F1

Dari tabel 11 didapatkan nilai koefisien korelasi (r) sebesar 0,942. Artinya, terdapat hubungan yang sangat kuat antara variabel pendapatan dengan konsumsi. Selain itu

pun, kita dapat mengetahui seberapa besar pengaruh yang dapat diberikan variabel pendapatan terhadap konsumsi di Amerika Serikat melalui koefisien determinasi, dengan rumus sebagai berikut :

$$KD = r^2 \times 100\%$$

$$KD = (0,942)^2 \times 100\%$$

$$KD = 88,8\%$$

Artinya, sebesar 88,8% variabel pendapatan dapat mempengaruhi konsumsi di Amerika Serikat. Sedangkan sisanya sebesar 11,2% menyatakan bahwa variabel konsumsi dapat dipengaruhi oleh variabel-variabel bebas lainnya yang tidak diteliti.

## **BAB III**

### **KESIMPULAN DAN SARAN**

#### **3.1 Kesimpulan**

Berdasarkan hasil analisis dari contoh kasus sebelumnya, dapat disimpulkan bahwa metode *Principal Component Analysis* (PCA) terbukti dapat mengatasi masalah pelanggaran asumsi klasik multikolinearitas tanpa perlu membuang variabel bebas yang berkolinear tinggi. Sehingga setelah diperoleh variabel bebas baru dari hasil reduksi, kita dapat meramalkan pengaruh dari variabel bebas (pendapatan) terhadap variabel tak bebas (konsumsi) melalui analisis regresi linier. Dengan metode PCA, kita akan mendapatkan variabel bebas baru yang tidak berkorelasi, bebas satu sama lainnya, lebih sedikit jumlahnya daripada variabel asli, akan tetapi bisa menyerap sebagian besar informasi yang terkandung dalam variabel asli atau yang bisa memberikan kontribusi terhadap varian seluruh variabel.

#### **3.2 Saran**

Principal Component Analysis (PCA) merupakan salah satu metode dasar dari analisis faktor. Oleh karena itu, metode ini akan lebih baik digunakan jika kita telah mempelajari terlebih dahulu mengenai analisis faktor.

Walaupun tampak canggih, analisis faktor bukan segalanya. Teknik ini tidak terlepas dari berbagai kelemahan. Keterbatasan utama adalah tingginya subjektivitas dalam penentuan jumlah faktor dan interpretasi setiap faktor. Untuk itu, diperlukan suatu teori-teori yang mendukung untuk menentukan banyaknya faktor yang terbentuk secara tepat.

## DAFTAR PUSTAKA

Dien Sukardinah, Soemartini, I.Gde.Mindra .2005. Bahan Kuliah Regresi Lanjutan. Bandung.

Gujarati, Damodar. *Ekonometrika Dasar*. Jakarta : Penerbit Erlangga.

[http://dickyrahardi.blogspot.com/2006/12/principal component analisis-pca.html](http://dickyrahardi.blogspot.com/2006/12/principal-component-analysis-pca.html).

Santoso, Singgih. 2002. *SPSS Statistik Multivariat*. Jakarta : PT. Elex Media Komputundo.

Simamora, Bilson. 2005. *Analisis Multivariat Pemasaran*. Jakarta : PT. Gramedia Pustaka Utama.

Supranto, M.A. 2004. *Analisis Multivariat (Arti & Interpretasi)*. Jakarta : Rineka Cipta.

[www.Google.com](http://www.Google.com) /Multikolinearitas/Principal Component Analysis (PCA) Metode Jitu Untuk Mengatasi Masalah Multikolinearitas.