

METODE PRINCIPAL COMPONENT ANALYSIS
UNTUK MENGATASI MULTIKOLINEARITAS PADA REGRESI LINIER BERGANDA
(STUDI KASUS FAKTOR YANG MEMPENGARUHI INDEK PEMBANGUNAN MANUSIA DI JAWA TIMUR)

Agung Sudrajat

Departemen Biostatistika
Fakultas Kesehatan Masyarakat Universitas Airlangga
agungsudrajat54@yahoo.com

ABSTRACT

This research is conducted in order to implement PCA (Principal component Analysis) to overcome multicollinearity in multiple linear regression. The case studies used are factors that affect the human development index (HDI) in East Java by 2013. These factors consist of the old school, literacy rates, spending per capita and life expectancy, so there are four (4) independent variables in this research. This research is using data publication because it is non-reactive research that is the kind of research is devoted to the publication of data. The result is obtained equation regression model $Y = 72,440 + 4,470 F_1$, with new independent variable $F_1 = 0,292$ (Old School) + 0,286 (Literacy Rates) + 0,267 (Spending Per Capita) + 0,249 (Life Expectancy).

The conclusion of this research is regression results obtained from the new independent variable (F_1) with the human development index variable (Y) indicate a strong influence because of the significant value $p = 0.000 < 0.05$. In addition the results of the analysis of PCA (Principal Component Analysis) can also be used to determine the dominant factor. The dominant factor of human development index case in East Java by 2013 that is old school with coefficient 0.292.

Keywords: multicollinearity; human development index; PCA (Principal component Analysis)

PENDAHULUAN

Di bidang kesehatan hubungan antara satu variabel bebas dengan variabel tak bebas sering terjadi. Hubungan ini disebut dengan regresi linier. Analisis regresi berhubungan dengan peramalan nilai rata hitung atau nilai rata (populasi) variabel tak bebas atas dasar nilai variabel bebas yang tetap. Garis lurus antara variabel tak bebas dengan satu variabel bebas yang dikenal dengan model regresi linear sederhana. Namun dikatakan model regresi linear berganda apabila model tersebut dibentuk dengan beberapa variabel bebas.

Secara kuantitatif dari masing-masing variabel bebas terhadap variabel tak bebas dapat mengestimasi besarnya pengaruh pada model regresi. Selain itu juga jika nilai variabel bebas telah diketahui juga dapat diestimasi nilai variabel tak bebasnya. Ada beberapa metode penyusunan model regresi, tetapi sejauh yang berhubungan dengan analisis regresi, metode kuadrat terkecil biasa (method of ordinary least square, OLS) yang paling sering digunakan. Asumsi model regresi linear klasik merupakan estimasi parameter model regresi yang diperoleh dengan OLS merupakan estimator yang baik jika model regresi memenuhi asumsi tertentu.

Tidak adanya multikolinearitas atau tidak adanya korelasi yang tinggi antara variabel bebas dalam suatu model regresi linear berganda merupakan satu asumsi kritis dari model regresi linear klasik. Hubungan antara variabel bebas terhadap variabel terikatnya menjadi terganggu jika ada korelasi yang tinggi di antara variabel bebasnya. Selain itu juga menimbulkan penggunaan

OLS dalam mengestimasi parameter/koefisien regresi akan terganggu. Galat yang dihasilkan akan menjadi besar, variansi dan kovariansi parameter tidak terhingga jika multikolinearitas yang hampir sempurna terjadi, meskipun metode kuadrat terkecil dapat digunakan. Jika mengalami multikolinearitas akibatnya variabel penjelas akan lebih mudah menjadi tidak signifikan karena selang interval cenderung besar. Oleh karena itu perlu adanya perbaikan jika terjadi multikolinearitas pada suatu data. Mengatasi masalah multikolinearitas, ada beberapa prosedur yang dapat digunakan seperti: melalui ln (logaritma) dan penambahan data baru, melakukan transformasi variabel dengan prosedur first difference, mengeluarkan suatu variabel atau beberapa variabel bebas yang terlibat hubungan kolinear, menghubungkan data cross-sectional dan data time series, penggunaan informasi apriori dari hubungan beberapa variabel yang berkolinier dan juga melalui ridge regression.

Namun pada prakteknya prosedur penanggulangan yang telah disebutkan di atas sangat tergantung sekali pada kondisi penelitian, misalnya: prosedur mengeluarkan variabel bebas yang berkolinier seringkali membuat banyak penelitian keberatan sebab prosedur ini akan mengurangi obyek penelitian yang diangkat, penggunaan informasi apriori sangat tergantung dari ada atau tidaknya dasar teori (literatur) yang sangat kuat untuk mendukung hubungan matematis antara variabel bebas yang saling berkolinier, 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 sebab itu, penelitian ini menggunakan metode Principal Component Analysis (PCA) dalam mengatasi multikolinearitas karena memiliki banyak kelebihan diantaranya dapat menghilangkan korelasi secara bersih sehingga masalah multikolinearitas dapat benar teratasi secara bersih, dapat dipergunakan tanpa mengurangi jumlah variabel asal, dapat digunakan untuk segala kondisi data/penelitian, dan kesimpulan yang diberikan lebih akurat dibandingkan dengan penggunaan metode lain walaupun metode regresi dengan PCA ini memiliki tingkat kesulitan yang tinggi.

Sebuah transformasi linier yang biasa digunakan pada kompresi data atau transformasi linear untuk menentukan sistem koordinat yang baru dari data dan merupakan teknik yang umum digunakan untuk menarik fitur-fitur dari data pada sebuah skala berdimensi tinggi dikenal dengan metode Principal Component Analysis (PCA). Data dapat diproyeksikan ke dalam subspace menggunakan metode Principal Component Analysis (PCA). Mengurangi dimensi dari data tanpa menghilangkan informasi penting dari data tersebut dapat digunakan teknik PCA.

Matriks ragam peragam () dan matriks korelasi dari $x_1, x_2, x_3, \dots, x_p$ dapat menentukan komponen utama. Apabila semua variabel yang diamati mempunyai satuan pengukuran yang sama dapat digunakan matriks kovarian untuk membentuk komponen utama. Sedangkan, apabila variabel yang diamati tidak mempunyai satuan pengukuran yang sama korelasi untuk membentuk komponen utama. Komponen utama berdasarkan matriks korelasi ditentukan dari variabel baku sehingga variabel tersebut perlu dibakukan.

Penelitian ini menggunakan studi kasus tentang faktor yang mempengaruhi indeks pembangunan manusia pada semua kabupaten/kota di Jawa Timur tahun 2013. Faktor yang mempengaruhi indeks pembangunan manusia meliputi lama sekolah, angka melek huruf, pengeluaran per kapita dan angka harapan hidup. Penelitian ini mengambil data pada tahun 2013, karena pada tahun tersebut terjadi peningkatan indeks pembangunan manusia yang begitu tinggi dibanding tahun sebelumnya. Pada tahun 2012 IPM di Jawa Timur sebesar 66.74 dan tahun 2013 sebesar 67.55, sehingga mengalami peningkatan 0.81 selama 2 (dua) tahun.

Berdasarkan Susenas tahun 2013, penduduk usia 10 tahun ke atas di Jawa Timur yang buta huruf (belum melek huruf) sekitar 8,53 persen. Secara umum, angka buta huruf laki-laki lebih rendah dibanding angka buta huruf perempuan, yaitu 4,80 persen dibanding 12,12 persen. Selama kurun waktu 2009-2012 terjadi peningkatan angka melek huruf dari 87,80 di tahun 2009 dan menjadi

88,34 persen di tahun 2010 dan 88,79 persen di tahun 2011. Pada tahun 2012 angka melek huruf 89,00 persen pada penduduk berusia 15 tahun ke atas di Jawa Timur. Capaian indikator ini pada tahun 2012, terpaut sebesar 6 persen di bawah target Pendidikan Untuk Semua (PUS) Tahun 2014. Sementara itu perlu upaya keras untuk mencapai target yang terdapat dalam RPJMN 2010-2014 Kemdiknas, mengingat capaian Jawa Timur pada tahun 2012 terpaut jauh yaitu sebesar 6,40 persen.

Menurut jenis kelamin capaian melek huruf penduduk usia 15 tahun ke atas secara umum laki-laki lebih tinggi dibandingkan perempuan. Capaian melek huruf antara laki-laki dan perempuan kalau dilihat menurut kelompok umur antara laki-laki dan perempuan semakin tinggi kelompok umur semakin besar perbedaan capaian melek huruf. Capaian melek huruf laki-laki diatas 90 persen mulai kelompok umur 15-19 tahun hingga 45-49 tahun kelompok umur 15-19 tahun hingga 35-39 tahun keatas. Kondisi ini memberikan gambaran bahwa penduduk laki-laki yang buta huruf lebih sedikit dibanding penduduk perempuan.

Berdasarkan Laporan Eksekutif Pendidikan tahun 2013 bahwa penduduk usia 15 tahun ke atas di Jawa Timur pada tahun 2013 memiliki rata-rata lama untuk mengenyam pendidikan formal adalah 7,53 tahun. Hal ini sejalan dengan banyaknya penduduk usia 15 tahun ke atas di Jawa Timur yang menamatkan pendidikannya hanya sebatas SD sederajat (pendidikan dasar) apabila dihubungkan dengan tingkat pendidikan tertinggi yang ditamatkan. Rata-rata lama sekolah penduduk usia 15 tahun ke atas terjadi peningkatan yaitu dari 7,20 tahun di tahun 2009 meningkat menjadi 7,53 tahun di tahun 2013 di Jawa Timur selama 2009-2013.

Rata-rata lama sekolah penduduk di Jawa Timur terus mengalami peningkatan mulai pada kelompok usia 7-12 tahun sebesar 2,75 tahun hingga tertinggi pada kelompok usia 19-24 tahun, yaitu rata – rata lama sekolahnya mencapai 10,35 tahun atau setara dengan kelas 1 SLTA dan terus menurun hingga pada kelompok usia 65 tahun ke atas, rata-rata lama sekolah sekitar 1,89 tahun saja. Menurut kelompok usia laki-laki lebih tinggi rata lama sekolahnya dibandingkan perempuan. Rata lama sekolah laki-laki dan perempuan capaiannya hampir sama mulai kelompok usia 7-12 tahun hingga 19-24 tahun, namun semakin bertambah usia, capaian rata rata lama sekolah laki-laki lebih tinggi dibandingkan perempuan. Kondisi ini memberikan gambaran bahwa banyak penduduk perempuan untuk tidak meneruskan pendidikannya ke jenjang yang lebih tinggi pada usia di atas 24 tahun.

Secara umum rata lama sekolah pada daerah-daerah kabupaten wilayah Madura dan sekitar kawasan tapal kuda (kab Bondowoso, kab Probolinggo, kab, Situbondo, kab, Jember, dll) lebih rendah berkisar hanya 5–6 tahun saja jika dibandingkan dengan daerah kota dengan rata rata lama sekolah yang lebih tinggi.

Berdasarkan Laporan Eksekutif Kesehatan tahun 2013 bahwa angka harapan hidup penduduk Jawa Timur mengalami peningkatan yaitu dari 69,15 tahun pada tahun 2009 menjadi 70,19 tahun pada tahun 2013. Namun pencapaian kinerja peningkatan AHH Jawa Timur menjadi satu perhatian tersendiri karena masih lebih rendah, AHH pada tahun 2013 sebesar 70,19 tahun lebih rendah dibanding rata-rata nasional Tahun 2011 sebesar 70,74.

Berdasarkan Susenas tahun 2013, bahwa sebagian besar pengeluaran penduduk sudah bergeser ke arah untuk memenuhi kebutuhan non makanan, yaitu mencapai 50,52 persen pada tahun 2012. Pengeluaran makanan kembali meningkat persentasenya di tahun 2013, yaitu sebesar 50,54 persen. Hal ini terjadi sebagai dampak dari kenaikan harga BBM sehingga mengakibatkan kenaikan harga komoditi makanan di tahun 2013.

Menurut Mudrajad (2006:115), Badan Pusat Statistik (BPS) Indonesia menggunakan konsep kemampuan memenuhi kebutuhan dasar (basic needs approach) untuk mengukur banyaknya jumlah penduduk miskin dalam menentukan indeks pembangunan manusia. Pengukuran ini dilakukan dengan melihat besarnya rupiah yang dibelanjakan per kapita per bulan untuk memenuhi kebutuhan minimum makanan dan non-makanan. Untuk kebutuhan minimum makanan digunakan patokan 2.100 kalori per hari. Sementara itu, untuk kebutuhan non-makanan meliputi pengeluaran untuk perumahan, sandang, serta aneka barang dan jasa.

Oleh karena itu penelitian ini ingin meneliti indeks pembangunan manusia pada tahun 2013 di Jawa Timur dan ingin melihat seberapa besar pengaruh faktor yang mempengaruhi indeks pembangunan manusia di Jawa Timur pada tahun 2013.

Berdasarkan penjelasan diatas Penelitian ini ingin menggunakan data tentang indeks pembangunan manusia untuk mengaplikasikan metode PCA (Principal Component Analysis) dalam mengatasi multikolinearitas. Indeks pembangunan manusia dipengaruhi oleh faktor angka melek huruf, rata lama sekolah, pengeluaran per kapita dan angka harapan hidup.

METODE PENELITIAN

Ada beberapa cara untuk mengetahui keberadaan multikolinearitas dalam suatu model regresi, dan untuk penelitian ini dengan melihat Nilai Tolerance atau VIF. Melihat nilai Tolerance atau VIF pada tabel "coefficient" dengan menggunakan bantuan software SPSS. Dikatakan terjadi multikolinearitas jika nilai korelasi antar variabel independen pada tabel output Coefficient Correlations > 95 persen. Sedangkan hasil perhitungan nilai tolerance dan VIF pada tabel output Coefficient dapat dikatakan terjadi multikolinearitas, jika nilai Tolerance < 0,1 atau nilai VIF > 10.

Diadakan perbaikan agar data agar tidak mengalami multikolinearitas, setelah di deteksi adanya multikolinearitas pada data. Di dalam penelitian ini untuk mengatasi multikolinearitas menggunakan metode PCA (Principal Component Analysis). Menyederhanakan variabel yang diamati dengan cara menyusutkan (transformasi) dimensinya pada dasarnya merupakan tujuan dari prosedur PCA. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel bebas baru yang tidak berkorelasi sama sekali. Setelah beberapa komponen hasil PCA yang bebas multikolinearitas diperoleh, maka beberapa komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisis pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi.

Metode PCA terdiri dari :

1. Menghitung Covariance-Matrix atau Bentuk Matrik Korelasi
2. Menentukan Banyak Factor / Mencari Eigenvalue dan Eigenvector
3. Rotasi factor dan Urutkan principal components secara menurun
4. Transformasi Data ke Sumbu Principal Components
5. Interpretasi

Aplikasi PCA juga dapat dilakukan dengan bantuan software SPSS. Jika menggunakan SPSS akan memunculkan output yang akan menuntun terbentuknya variabel bebas yang baru seperti

a. Uji barlett, uji Kaiser Mayer Olkin (KMO)

Uji Kaiser-Meyer-Olkin (KMO) dan Barlett Test untuk menguji layak atau tidaknya analisis faktor dilakukan. Jika nilai KMO berkisar antara 0,5 sampai dengan 1 maka analisis faktor layak digunakan. Namun, jika nilai KMO kurang dari 0,5 maka analisis faktor tidak layak dilakukan. Sedangkan Barlett Test digunakan untuk menguji apakah benar beberapa variabel yang dilibatkan berkorelasi.

b. Anti-Image Matric

Bagian Anti Image Correlation, dapat dilihat 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 sebagai berikut:

- MSA = 1, variabel tersebut dapat diprediksi tanpa kesalahan oleh variabel lain.
- MSA > 0,5, variabel masih bisa diprediksi dan bisa dianalisis lebih lanjut.
- MSA < 0,5, variabel tidak bisa diprediksi dan tidak bisa dianalisis lebih lanjut, atau dikeluarkan dari variabel lainnya.

c. Communalities

Communalities merupakan banyaknya varians yang dapat dijelaskan oleh faktor yang terbentuk.

d. Total Varian Explained

Setiap faktor mewakili variabel yang dianalisis. Eigenvalue merupakan kemampuan setiap faktor mewakili variabel yang dianalisis ditunjukkan oleh besarnya varians yang dijelaskan. Eigenvalue

menunjukkan kepentingan relatif beberapa faktor dalam menghitung varians ketiga variabel yang dianalisis. Susunan eigenvalue selalu diurutkan dari yang terbesar hingga yang terkecil, dengan syarat bahwa angka eigenvalue dibawah 1 tidak digunakan dalam menghitung jumlah faktor yang terbentuk.

e. Component Matrix

Componen Matriks merupakan tabel yang berisikan factor loading (nilai korelasi) beberapa variabel analisis dengan faktor yang terbentuk.

f. Component Score Coefficient Matriks

Setelah didapatkan faktor yang terbentuk melalui metode Principal Component Analysis, maka perlu dicari persamaan sehingga dapat dihitung skor setiap faktor secara manual. Persamaan yang dibuat hampir sama dengan regresi linear berganda, hanya dalam persamaan faktornya tidak terdapat konstanta. Setelah komponen hasil PCA yang bebas multikolinearitas diperoleh maka beberapa komponen tersebut diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi linear.

HASIL

1. Metode Varians Inflation Factor (VIF)

Metode Varians Inflation Factor (VIF) merupakan analisis dan pengolahan data untuk mendeteksi multikolinearitas. Data mengalami multikolinearitas, tergantung pada nilai tolerance dan VIF, jika nilai Tolerance < 0,1 atau nilai VIF >10 maka dapat dikatakan terjadi multikolinearitas.

No	Variabel Bebas	Kolinearitas	
		Tolerance	VIF
1	Angka Melek Huruf	0,124	8,070
2	Lama Sekolah	0,061	16,304
3	Pengeluaran Per Kapita	0,148	6,779
4	Angka Harapan Hidup	0,375	2,664

Berdasarkan tabel 5.2 diatas menunjukkan bahwa data yang dianalisis dengan metode VIF mengalami multikolinearitas karena dari empat variabel bebas yang di analisis, variabel lama sekolah memiliki nilai toleransi kurang dari 0,1 yaitu sebesar 0,061 dan memiliki nilai VIF lebih dari 10 yaitu sebesar 16,304. Oleh karena itu perlu diadakan perbaikan agar data tersebut tidak mengalami multikolinearitas.

2. Metode Principal Component Analisis (PCA)

Pengujian sebelumnya telah menunjukkan bahwa terdapat permasalahan multikolinearitas dalam data tersebut, maka dilakukan penanggulangan untuk mengatasi masalah multikolinearitas tersebut. Penelitian ini menggunakan prosedur Principal Component Analysis (PCA) untuk mengatasi multikolinearitas. Prosedur PCA bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan

(transformasi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali.

Analisis menggunakan metode Principal Component Analysis (PCA) akan menghasilkan beberapa pengujian yang penting dan output yang menjelaskan terbentuknya faktor baru. Pengujian pertama, mengenai uji Kaiser-Meyer-Olkin (KMO) Measure Of Adequacy dan Barlett Test of Sphericity. Uji KMO digunakan untuk menentukan layak atau tidak analisis factor dilakukan, pada tabel 5.3 terlihat bahwa nilai KMO adalah 0,729, maka analisis factor layak dilakukan.

(Uji Kaiser Mayer Olkin) KMO	Barlet		
	Chi-Square	Signifikan	Desibel
0,729	161,608	0,000	6

Barlett Test merupakan uji statistik untuk melihat adanya korelasi antara variabel bebas. Barlett tes akan menampilkan nilai Chi-Square dan nilai signifikan dalam menentukan adanya korelasi di antara variabel bebas. Pada tabel 5.3 dapat dilihat bahwa nilai Chi-Square adalah 161,608, dengan derajat bebas sebesar 6, dan p-value (sig) sebesar 0,000. Karena p-value (0,000) < 0,05 maka Ho di tolak. Artinya, benar-benar terdapat korelasi antar variabel bebas.

Pengujian selanjutnya adalah MSA (Measure Of Sampling Adequacy), nilai MSA terletak pada diagonal "anti image correlation", yang diatas nilainya ada tanda 'a'. Berdasarkan hasil analisis yang telah dilakukan bahwa semua angka MSA memiliki nilai di atas 0,5 artinya analisis dapat dilanjutkan.

Pada tabel 5.4 menjelaskan tentang komunalitas atau berapa varians yang dapat dijelaskan oleh faktor yang diekstrak (faktor yang terbentuk). Setiap variabel berkorelasi dengan beberapa faktor yang terbentuk. Pada variabel angka melek huruf, diperoleh nilai sebesar 0,906. Hal ini berarti sekitar 90,6 persen variabel angka melek huruf dapat dijelaskan oleh faktor yang terbentuk. Variabel lama sekolah, diperoleh nilai sebesar 0,945. Hal ini berarti sekitar 94,5 persen variabel lama sekolah dapat dijelaskan oleh faktor yang terbentuk. Pada variabel pengeluaran per kapita, diperoleh nilai sebesar 0,792. Hal ini berarti sekitar 79,2 persen variabel lama sekolah dapat dijelaskan oleh faktor yang terbentuk. Variabel angka harapan hidup, diperoleh nilai sebesar 0,687. Hal ini berarti sekitar 68,7 persen variabel angka harapan hidup dapat dijelaskan oleh faktor yang terbentuk.

Tabel 5.3 Komunalitas

No	Variabel bebas	Proporsi varian yang dijelaskan
1	Angka Melek Huruf	0,906
2	Lama Sekolah	0,945
3	Pengeluaran Per Kapita	0,792
4	Angka Harapan Hidup	0,687

Pada hasil analisis juga menunjukkan bahwa dari ke empat variabel bebas yang telah diuji hanya ada satu faktor yang terbentuk. Faktor 1 (satu) ini memiliki eigenvalue sebesar 3.330, artinya faktor 1 (satu) ini dapat menjelaskan 3.330 atau 83.251 persen dari total komunalitas. Sedangkan untuk 3 faktor, angka eigenvalues sudah di bawah 1. Sehingga proses faktoring seharusnya berhenti pada satu faktor saja. Faktor yang terbentuk dari ke empat variabel yaitu variabel bebas F1. Hal ini menunjukkan bahwa variabel bebas F1 adalah jumlah yang paling optimal untuk menggantikan empat variabel bebas tersebut.

Setelah mendapatkan faktor yang terbentuk melalui proses Principal Component Analysis (PCA), maka perlu menarik persamaannya. Persamaan tersebut, digunakan untuk mencari skor setiap faktor secara manual. Persamaan yang dibuat mirip dengan regresi linier berganda, hanya dalam persamaan faktornya tidak terdapat konstanta.

Tabel 5.4 Komponen Nilai Koefisien Matrik

No	Variabel bebas	Nilai faktor yang terbentuk
1	Angka Melek Huruf	0,286
2	Lama Sekolah	0,292
3	Pengeluaran Per Kapita	0,267
4	Angka Harapan Hidup	0,249

Berdasarkan hasil tabel 5.5, maka persamaan untuk faktor baru yang terbentuk adalah sebagai berikut :

$$F1 = 0,292 (\text{Lama Sekolah}) + 0,286 (\text{Angka Melek Huruf}) + 0,267 (\text{Pengeluaran Per Kapita}) + 0,249 (\text{Angka Harapan Hidup})$$

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

3 Model Regresi yang Cocok

Setelah mendapatkan variabel bebas baru (F1) yang bebas multikolinearitas melalui teknik Principal Component Analysis (PCA), maka selanjutnya meregresikan variabel bebas yang baru (F1) tersebut terhadap variabel tak bebas (Y). Karena variabel bebas baru (F1) 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 = Indeks Pembangunan Manusia

$$F1 = 0,292 (\text{Lama Sekolah}) + 0,286 (\text{Angka Melek Huruf}) + 0,267 (\text{Pengeluaran Per Kapita}) + 0,249 (\text{Angka Harapan Hidup})$$

Tabel 5.5 Koefisien

Model	Koefisien Tidak Standar		Signifikan
	Beta	Standar Error	
(Konstan)	72,440	0,114	0,000
F1	4,470	0,116	0,000

Berdasarkan tabel 5.6 dapat dilihat bahwa nilai signifikan bernilai $0,000 < 0,05$, maka dapat disimpulkan bahwa variabel bebas baru (F1) memiliki pengaruh yang signifikan terhadap indeks pembangunan manusia di Jawa Timur. Selain itu pada tabel 5.8 hasil regresi antara variabel indeks pembangunan manusia (Y) dengan F1 didapatkan model persamaan sebagai berikut :

$$Y = 72,440 + 4,470 F1$$

$$\text{Dengan } F1 = 0,292 (\text{Lama Sekolah}) + 0,286 (\text{Angka Melek Huruf}) + 0,267$$

$$(\text{Pengeluaran Per Kapita}) + 0,249 (\text{Angka Harapan Hidup})$$

4. Faktor Dominan Yang Mempengaruhi Indeks Pembangunan Manusia

Indeks pembangunan manusia merupakan salah satu indikator untuk mengukur kesejahteraan suatu daerah atau wilayah. Indeks pembangunan manusia dipengaruhi oleh empat (4) faktor yaitu angka melek huruf, lama sekolah, pengeluaran per kapita dan angka harapan hidup. Namun dari ke empat faktor tersebut mengalami multikolinearitas setelah di analisis secara regresi linier berganda. Multikolinearitas terjadi karena adanya salah satu variabel bebas yang mempunyai korelasi sangat tinggi dibandingkan variabel lain.

Multikolinearitas dapat diatasi dengan banyak cara, tetapi dalam penelitian ini menggunakan metode principal component analysis (PCA). Hasil analisis PCA dapat digunakan untuk menentukan faktor yang dominan mempengaruhi indeks pembangunan manusia. Berdasarkan hasil analisis komponen matrik, terlihat bahwa hanya satu faktor yang terbentuk dari ketiga variabel yaitu variabel lama sekolah dengan nilai 0,972. Hal ini

menunjukkan bahwa variabel lama sekolah adalah jumlah yang paling optimal untuk mereduksi ketiga variabel bebas dan merupakan faktor yang dominan mempengaruhi indek pembangunan manusia.

PEMBAHASAN

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi yang kuat diantara beberapa variabel bebas (X) yang diikutsertakan dalam pembentukan model regresi linier. Multikolinieritas adalah suatu kondisi yang menyalahi asumsi regresi linier dan tidak mungkin terjadi apabila variabel bebas (X) yang diikutsertakan hanya satu. Hal ini didukung dengan data penelitian mengenai faktor yang mempengaruhi indek pembangunan manusia di Jawa Timur tahun 2013 yang terdiri dari beberapa variabel bebas yaitu angka melek huruf, lama sekolah, pengeluaran per kapita dan angka harapan hidup.

Menurut Montgomery, salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinieritas pada regresi linear berganda adalah Variance Inflation Factors (VIF). Adanya multikolinieritas dinilai dari nilai VIF yang dihasilkan. Besarnya nilai VIF ini bergantung pada nilai koefisien determinasi (R^2) yang dihasilkan. Jika nilai VIF melebihi 10 maka koefisien determinasi bernilai lebih besar dari 0,9. Nilai R^2 adalah koefisien determinasi ganda, jika R^2 besar maka VIF akan besar pula dan jika $VIF > 10$ maka terdapat multikolinieritas. Nilai VIF menunjukkan inflasi yang dialami oleh setiap koefisien regresi di atas nilai idealnya, yaitu di atas nilai yang dialami jika matriks korelasi adalah matriks identitas. Terdapat satu atau dua lebih nilai VIF yang besar menandakan adanya multikolinieritas.

Hal ini didukung dengan hasil penelitian bahwa nilai VIF pada variabel lama sekolah di atas 10 yaitu sebesar 16,304 sedangkan 3 (tiga) variabel lainnya tidak mengalami multikolinieritas karena nilai VIF-nya dibawah 10. Selain itu hasil penelitian juga menunjukkan bahwa koefisien determinasi lebih dari 0,9 yaitu sebesar 0,981. Hal ini menunjukkan adanya pengaruh nilai R^2 terhadap nilai VIF yang dihasilkan, yaitu semakin besar nilai R^2 maka semakin besar pula nilai VIF yang dihasilkan. VIF berkorespondensi dengan setiap estimasi kuadrat terkecil dari koefisien regresi.

Jika pada pengujian sebelumnya telah menunjukkan bahwa terdapat permasalahan multikolinieritas dalam data, maka dilakukan penanganan untuk mengatasi masalah multikolinieritas tersebut. Dalam hal ini menggunakan prosedur Principal Component Analysis (PCA). 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.

Hal ini didukung dengan hasil penelitian yang menghasilkan satu faktor yang paling optimal dalam mereduksi variabel lain yaitu lama sekolah dengan nilai 0,972. Selain itu juga didapatkan model persamaan regresi yang baru setelah mendapatkan faktor yang terbentuk melalui proses reduksi, hanya saja dalam persamaan faktornya tidak terdapat konstanta. Berikut adalah persamaan untuk faktor baru yang terbentuk :

$$F1 = 0,292 (\text{Lama Sekolah}) + 0,286 (\text{Angka Melek Huruf}) + 0,267 (\text{Pengeluaran Per Kapita}) + 0,249 (\text{Angka Harapan Hidup})$$

Setelah beberapa komponen hasil PCA yang bebas multikolinieritas diperoleh, maka komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisis pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi. Hal ini didukung dengan hasil penelitian yang meregresikan variabel indek pembangunan manusia (Y) dengan variabel bebas baru yaitu F1. Karena variabel bebas baru (F1) yang terbentuk hanya satu, maka pada model tersebut digunakan analisis regresi linier sederhana. Berikut model persamaan hasil regresi antara variabel indek pembangunan manusia (Y) dengan F1:

$$Y = 72,440 + 4,470 F1$$

Dengan $F1 = 0,292 (\text{Lama Sekolah}) + 0,286 (\text{Angka Melek Huruf}) + 0,267 (\text{Pengeluaran Per Kapita}) + 0,249 (\text{Angka Harapan Hidup})$

Hasil regresi antara variabel indek pembangunan manusia dengan variabel bebas baru (F1) menunjukkan nilai signifikan $0,000 < 0,05$, maka dapat disimpulkan bahwa variabel bebas baru (F1) memiliki pengaruh yang signifikan terhadap indek pembangunan manusia di Jawa Timur. Variabel bebas baru (F1) terdiri dari lama sekolah, angka melek huruf, pengeluaran per kapita dan angka harapan hidup dengan skor koefisien baru yang memiliki pengaruh terhadap indek pembangunan manusia.

Hasil analisis dari metode PCA menunjukkan faktor yang dominan dari indek pembangunan manusia di Jawa Timur tahun 2013 yaitu lama sekolah, kemudian faktor berikutnya adalah angka melek huruf. Skor/nilai pada variabel lama sekolah sebesar 0,292 dan angka melek huruf sebesar 0,286.

Penelitian ini sejalan dengan penelitian yang dilakukan oleh (Pramudya dan Khoiriyah) dengan judul "Penerapan Analisis Regresi Ganda Untuk Mengidentifikasi Faktor-Faktor Yang Berasosiasi Terhadap Capaian IPM". Hasil penelitian tersebut menunjukkan bahwa faktor pendidikan sangat mempengaruhi indek pembangunan manusia. Hal itu ditunjukkan dengan telah direduksinya 75 variabel bebas menjadi tiga variabel penjelas yang berasosiasi dengan capaian IPM. Dengan demikian, permasalahan faktor-faktor yang terkait dengan capaian IPM telah dapat disederhanakan. Adapun, model linear terbaik yang menjelaskan pola hubungan antara variabel penjelas dengan IPM adalah persamaan

$y = 62,6448 - 0,399431X_1 + 0,182763X_{73} + 0,158506X_{74}$ dengan (X1) tidak dapat membaca dan menulis aksara latin, (X73) jumlah anggota rumah tangga yang tidak/belum pernah bersekolah, (X74) jumlah anggota rumah tangga yang tidak bersekolah lagi, dan (Y) IPM. Persamaan tersebut dengan sendirinya dapat digunakan untuk mengestimasi capaian IPM, khususnya bila ada kebutuhan informasi cepat yang terkait dengan capaian IPM.

Penelitian ini juga serupa dengan penelitian yang dilakukan oleh (Riani) dengan judul "Pembangunan Pendidikan Sebagai Motor Penggerak IPM Jawa Barat". Hasil analisis penelitian tersebut menunjukkan pada tahun 1996, kontribusi indeks pendidikan dalam membentuk IPM Jawa Barat adalah sebesar 36,173 persen. Dibandingkan dengan indeks kesehatan dan indeks daya beli, peranan hasil kinerja pembangunan pendidikan ini tidak begitu dominan. Tahun 2003, kontribusi indeks pendidikan dalam membentuk IPM Jawa Barat mengalami kenaikan yaitu menjadi sebesar 38,56 persen. Naiknya kontribusi indeks pendidikan tersebut menunjukkan hasil pembangunan di bidang pendidikan, yang antara lain tercermin dari meningkatnya Angka melek huruf sebesar 4,1 poin, dan meningkatnya Rata-rata lama sekolah sebesar 0,7 poin selama periode 1996-2003.

Disisi lain, naiknya kontribusi indeks pendidikan dalam membentuk IPM Jawa Barat juga disebabkan oleh turunnya pencapaian indeks Daya beli sebagai akibat merosotnya daya beli masyarakat. Tahun 1996, kemampuan konsumsi per kapita riil yang disesuaikan untuk masyarakat Jawa Barat mencapai 591,6 ribu rupiah. Terjadinya krisis moneter yang ditandai dengan melonjaknya harga barang-barang kebutuhan, menyebabkan terjadinya penurunan pada kemampuan konsumsi rata-rata masyarakat Jawa Barat menjadi sebesar 553,7 ribu rupiah pada tahun 2003. Hal ini menyebabkan merosotnya pencapaian indeks Daya beli sebesar 0,087 poin selama periode tersebut. Kontribusi indeks Daya beli terhadap IPM mengalami penurunan sebesar 4,17 persen.

Selain itu penelitian ini didukung dengan penelitian yang dilakukan oleh (Syamsuddin) dengan judul "Analisis Indeks Pembangunan Manusia Kabupaten Tanjung Jabung Barat Periode 2007-2011". Hasil analisis penelitian tersebut menunjukkan bahwa Kontribusi masing-masing komponen terhadap pembentukan IPM Kabupaten Tanjung Jabung Barat selama periode 2007-2011 menunjukkan bahwa dimensi pengetahuan selalu mendominasi dengan rata-rata 37,78 persen yang kemudian diikuti oleh dimensi hidup panjang rata-rata 34,16 persen. Sementara hisup layak berada pada posisi ketiga dengan rata-rata kontribusi sebesar 28,06 persen selama periode observasi. Dengan kondisi minimum 0 persen dan kondisi ideal 100persen, Kabupaten Tanjung Jabung Barat telah mampu mencapai angka melek huruf sebesar 97,93 persen. Capaian (indeks) tertinggi kedua adalah

untuk indikator angka harapan hidup (e0). Dengan nilai minimum 25 tahun dan ideal 85 tahun, Kabupaten Tanjung Jabung Barat telah mampu mencapai angka harapan hidup selama 69,70 tahun. Dengan kata lain persentase capaian pada indikator ini adalah sebesar 74,78 persen. Di posisi ketiga adalah paritas daya beli (PPP). Tingginya sumbangan dimensi pengetahuan ini terutama disebabkan oleh tingginya angka melek huruf di Kabupaten Tanjung Jabung Barat sebagai salah satu dari dua indikator dalam dimensi pengetahuan. Sebagaimana telah dikemukakan sebelumnya, angka melek huruf di Kabupaten Tanjung Jabung Barat menempati urutan kedua tertinggi setelah Kota Jambi.

PENUTUP

1. Kesimpulan

Berdasarkan uraian yang telah dijelaskan pada bab sebelumnya, maka dapat disimpulkan sebagai berikut:

1. Metode Principal Component Analysis (PCA) digunakan untuk mengatasi multikolinearitas yang bertujuan untuk menyederhanakan variabel yang diamati dengan cara 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. Setelah beberapa komponen hasil PCA yang bebas multikolinearitas diperoleh, maka komponen-komponen tersebut menjadi variabel bebas baru yang akan diregresikan atau dianalisis pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi.
2. Hasil regresi dari variabel bebas baru (F1) dengan variabel indeks pembangunan manusia (Y) menunjukkan adanya pengaruh yang kuat karena nilai signifikannya $0,000 < 0,05$.
3. Hasil analisis metode PCA (Principal Component Analysis) juga dapat digunakan untuk menentukan faktor dominan. Faktor dominan dari kasus indeks pembangunan manusia di Jawa Timur tahun 2013 yaitu lama sekolah dengan nilai koefisien sebesar 0,292.

2. Saran

Hasil penelitian ini diharapkan dapat memberikan sedikit sumbangan pemikiran sebagai usaha untuk mengkaji bidang ilmu pengetahuan khususnya bidang matematika. Saran yang dapat penyusun sumbangkan sehubungan dengan hasil penelitian ini adalah sebagai berikut:

1. Jika pada suatu model regresi terjadi penyimpangan asumsi multikolinearitas, maka harus dilakukan tindakan perbaikan untuk menghilangkan multikolinearitas tersebut.

2. Bila melakukan tindakan perbaikan untuk menghilangkan multikolinearitas sebaiknya menggunakan metode PCA (Principal Component Analysis) karena metode ini memiliki banyak kelebihan diantaranya dapat menghilangkan korelasi secara bersih (korelasi = 0) sehingga masalah multikolinearitas dapat benar-benar teratasi secara bersih, dapat digunakan untuk segala kondisi data/penelitian, dapat dipergunakan tanpa mengurangi jumlah variabel asal.

DAFTAR PUSTAKA

- Badan Pusat Statistik.(2013). Laporan Eksekutif Kesehatan 2013. http://jatim.bps.go.id/4dm!n/pdf_publicasi/Laporan-Eksekutif-Kesehatan-Provinsi-Jawa-Timur-2013.pdf. (diakses 27 Juni 2016)
- Badan Pusat Statistik.(2013). Indek Pembangunan Manusia tahun 2013. http://jatim.bps.go.id/4dm!n/pdf_publicasi/Indek-Pembangunan-Manusia-Provinsi-Jawa-Timur-2013.pdf. (diakses 27 Juni 2016)
- Badan Pusat Statistik.(2013). Survei Sosial Ekonomi Nasional 2013. http://jatim.bps.go.id/4dm!n/pdf_publicasi/Hasil-Survei-Sosial-Ekonomi-Nasional-Tahun-2013-Provinsi-Jawa-Timur.pdf. (diakses 27 Juni 2016)
- Badan Pusat Statistik.(2013). Laporan Eksekutif Pendidikan 2013. http://jatim.bps.go.id/4dm!n/pdf_publicasi/Laporan-Eksekutif-Pendidikan-Provinsi-Jawa-Timur-2013-.pdf. (diakses 27 Juni 2016)
- Badan Pusat Statistik.(2013). Jatim Dalam Angka 2013. http://jatim.bps.go.id/4dm!n/pdf_publicasi/Jatim-Dalam-Angka-Provinsi-Jawa-Timur-2013-.pdf. (diakses 27 Juni 2016)
- Damodar, Gujarati. 2006. Dasar-Dasar Ekonometrika Edisi ketiga Terjemahan: Mc. Graw Hill. Jakarta. Erlangga.
- Draper, N. dan Smith, H. 1992. Analisis Regresi Terapan. Edisi Kedua. Terjemahan Oleh Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.
- Draper dan Smith. 1992. Analisis Regresi Terapan. Jakarta: Gramedia Pustaka
- Gujarati, N.D. 2003. Basic Econometrics. 4th ed. New York: McGraw-Hill Companies, Inc.
- _____, D. N., dkk. 1995. Basics Econometrics, Mc Graw Hill, Inc. New York
- _____, Damodar dan zain, Sumarno. 1978. Ekonometrika Dasar. Jakarta: Erlangga.
- Hardius Usman, dkk. 2013. Aplikasi Teknik Multivariat Untuk Riset Pemasaran. Jakarta: PT Raja Grafindo Persada.
- HM Syamsudin. (2013). "Analisis Indeks Pembangunan Manusia Kabupaten Tanjung Jabung Barat Periode 2007-2011". Jurnal Paradigma Ekonomika, 7 (1): 9-1
- Kutner, M.H., C.J. Nachtsheim., dan J. Neter. 2004. Applied Linear Regression Models. 4th ed. New York: McGraw-Hill Companies, Inc.
- Montgomery, D.C & Peck. 1992. Introduction to linear Regression Analysis. New York. John Wiley & sons.
- Pramudya Ikrar. (2015). " Penerapan Analisis Regresi Ganda Untuk Mengidentifikasi Faktor-Faktor Yang Berasosiasi Terhadap Capaian IPM". Jurnal Statistik, 2(2): 764-776.
- Riani Westi. (2006). "Pembangunan Pendidikan Sebagai Motor Penggerak IPM Jawa Barat". Jurnal Ekonomi, 22 (3): 278-291.
- Sumodiningrat, Gunawan. 2007. Ekonometrika Pengantar. Yogyakarta: BPFEYOGYAKARTA
- Sembiring, R.K. 2003. Analisis Regresi. Edisi Kedua. Bandung: Institut Teknologi Bandung.
- Supranto. 2010. Analisis Multivariat Arti dan Interpretasi. Jakarta: PT RINEKA CIPTA
- Widarjono, A. 2007. Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis. Edisi Kedua. Yogyakarta: Ekonisia Fakultas Ekonomi Universitas Islam Indonesia.