

Penggunaan Analisis Komponen Utama Dalam Penggabungan Data Peubah Ganda pada Kasus Produksi Pertanian dan Perkebunan Di Wilayah Bolaang Mongondow Tahun 2008

¹Sunarsi Habib Abdurachman, ²Hanny Komalig, ³Nelson Nainggolan

¹Program Studi Matematika, F-MIPA, UNSRAT, sunarsiabdurachman@yahoo.co.id

²Program Studi Matematika, F-MIPA, UNSRAT, hanoy07@yahoo.com

³Program Studi Matematika, F-MIPA, UNSRAT, bapaivana@yahoo.co.id

Abstract

The objective of this research is to study the combine the two groups of data with multivariate variables using Principal Component Analysis. The data used in this study is a secondary data drawn from the North Sulawesi BPS data in Production Agriculture and Plantation Bolaang Mongondow region in 2008. The results show that PCA can be used to combining two separate groups multivariate data and the correlation between the Principal Components of the data are combined with the Principal Component of the overall initial data (intact) is relatively high with correlation between PC1 and PC1AB as big 0,987 and correlation between PC2 and PC2AB as big 0,916.

Keywords : Principal Component Analysis, Agriculture Production and Plantation

Abstrak

Tujuan penelitian ini adalah menggabungkan dua gugus data peubah ganda dengan menggunakan Analisis Komponen Utama. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari BPS Sulawesi Utara yakni Data Produksi Pertanian Dan Perkebunan Di Wilayah Bolaang Mongondow Tahun 2008. Hasilnya menunjukkan bahwa AKU dapat digunakan untuk menggabungkan dua gugus data peubah ganda yang terpisah dan korelasi antara komponen utama dari data yang digabungkan dengan komponen utama dari keseluruhan data awal (utuh) relatif tinggi yakni dengan nilai korelasi PC1 dan PC1AB sebesar 0,987 dan PC2 dan PC2AB sebesar 0,916.

Kata kunci : Analisis Komponen Utama, Produksi Pertanian dan Perkebunan

1. Pendahuluan

Menurut ilmu pengetahuan dan teknologi (IPTEK) saat ini, bahwa ilmu statistika telah mempengaruhi hampir seluruh aspek kehidupan manusia, hampir semua kebijakan publik dan keputusan-keputusan yang diambil oleh para pakar ilmu pengetahuan didasarkan dengan metode statistika serta hasil analisis dan interpretasi data baik secara kuantitatif maupun kualitatif. Analisis Peubah Ganda (APG) merupakan salah satu metode analisis dalam statistika yang dapat digunakan untuk menganalisis beberapa variabel secara bersamaan.

Salah satu metode yang digunakan dalam APG adalah pereduksian dimensi data peubah ganda [1] dan Analisis Komponen Utama (AKU) merupakan salah satu metode yang digunakan tersebut, yakni untuk mereduksi beberapa variabel menjadi beberapa variabel baru yang jumlahnya lebih sedikit tanpa menghilangkan karakteristik dari variabel asal, sehingga lebih mudah untuk menginterpretasikan data-data tersebut [2].

Semua data peubah ganda dapat digabungkan dengan AKU asalkan memenuhi kriteria atau syarat data yang ditentukan yakni data yang bisa dibentuk menjadi dua bagian, data berskala rasio, memiliki korelasi antar komponen, misalnya data perikanan, pertanian, perkebunan dan lain-lain. Adapun data peubah ganda yang perlu digabungkan yaitu data tahunan produksi pertanian dan perkebunan, sehingga ide untuk menggabungkan dua gugus data dengan menggunakan Analisis Komponen Utama dapat dicoba pada kasus data tahunan produksi pertanian dan perkebunan wilayah Bolaang Mongondow, Sulawesi Utara.

Dalam penelitian ini AKU digunakan untuk menggabungkan dua gugus data peubah ganda yang terpisah dan mempelajari AKU dalam menggabungkan dua gugus data peubah ganda menjadi dua komponen utama yang mewakili kedua gugus data tersebut.

2. Tinjauan Pustaka

2.1. Analisis Komponen Utama

Analisis Komponen Utama (AKU) dipelopori oleh Karl Pearson pada tahun 1901 untuk peubah nonstokastik, AKU merupakan suatu teknik untuk membentuk variabel baru yang merupakan kombinasi linear dari variabel asal. Jumlah maksimum dari variabel baru yang dapat dibentuk adalah sama dengan jumlah variabel asal, dan antara variabel baru tidak berkorelasi [3]. Menurut [5] AKU adalah teknik yang digunakan untuk menyederhanakan suatu data dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum.

AKU terkonsentrasi pada penjelasan struktur variansi dan kovariansi melalui suatu kombinasi linear variabel-variabel asal, dengan tujuan utama melakukan reduksi data dan membuat interpretasi. Analisis komponen utama lebih baik digunakan jika variabel-variabel asal saling berkorelasi [2].

AKU dimulai dengan data pada variabel ke p dari banyaknya n data. Seperti yang ditunjukkan pada tabel, maka kombinasi linear dari variabel-variabel X_1, X_2, \dots, X_p , diperoleh komponen utama yaitu :

$$\begin{aligned} PC_1 &= \alpha_{11}X_1 + \alpha_{12}X_2 + \dots + \alpha_{1p}X_p \\ PC_2 &= \alpha_{21}X_1 + \alpha_{22}X_2 + \dots + \alpha_{2p}X_p \\ PC_n &= \alpha_{n1}X_1 + \alpha_{n2}X_2 + \dots + \alpha_{np}X_p \end{aligned} \quad (1)$$

Menurut [1] secara aljabar, komponen utama adalah kombinasi linear khusus dari p variabel acak X_1, X_2, \dots, X_p . Secara geometris, kombinasi linear ini menggambarkan pemilihan dari sistem koordinat yang diperoleh dengan merotasikan sistem awal dengan X_1, X_2, \dots, X_p sebagai sumbu koordinat. Seperti yang kita lihat, komponen utama semata-mata bergantung pada matriks kovarians Σ (matriks korelasi ρ) dari X_1, X_2, \dots, X_p dalam perkembangannya tidak membutuhkan asumsi multivariat normal. Di sisi lain, komponen utama yang berasal dari populasi multivariate normal mempunyai interpretasi yang berguna dalam kepadatan ellipsoid konstan. Misalkan vektor acak $X' = [X_1, X_2, \dots, X_p]$ memiliki matriks kovarians Σ dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \lambda_p \geq \dots \geq 0$. Maka komponen utama pertama yang dilambangkan oleh Y_1 didefinisikan sebagai:

$$Y_1 = \Sigma \alpha_1 X_j = \alpha_1 X$$

yang memaksimalkan ragam Y_1 , yaitu $\alpha_1' \Sigma \alpha_1$, dengan kendala $\alpha_1' \alpha_1 = 1$. Komponen utama kedua, dilambangkan dengan Y_2 , dengan kendala $\alpha_2' \alpha_2 = 1$, dan tidak ada korelasi antara Y_1 dan Y_2 (koragam Y_1 dan Y_2 yaitu $\alpha_1' \Sigma \alpha_2 = 0$ yang nantinya akan berarti $\alpha_1' \alpha_2 = 0$). Untuk Y_3 sama halnya dengan Y_1, Y_2 . Demikian seterusnya sampai komponen utama ke - 4 sampai ke - p .

Dengan menggunakan pengganda Lagrange diperoleh $\alpha_1, \alpha_2, \dots, \alpha_p$ sebagai vektor ciri yang berpadanan dengan akar ciri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ dari matriks koragam Σ . Nilai akar ciri ke- i merupakan ragam komponen utama ke- i . Karena solusi bagi vektor α merupakan vektor ciri maka vektor ini tidak bersifat khas, misalnya penggandaannya dengan -1 juga akan merupakan solusi.

2.2. Analisis Korelasi

Menurut [4], analisis korelasi berusaha mengukur eratnya hubungan antara dua peubah dengan menggunakan suatu bilangan yang disebut koefisien korelasi. Koefisien korelasi (r) adalah nilai yang menunjukkan kuat atau lemah hubungan antar x dan y . Nilai r terletak antara -1 dan 1 ($-1 \leq r \leq 1$) Nilai r dapat dihitung dengan persamaan :

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

3. Metodologi Penelitian

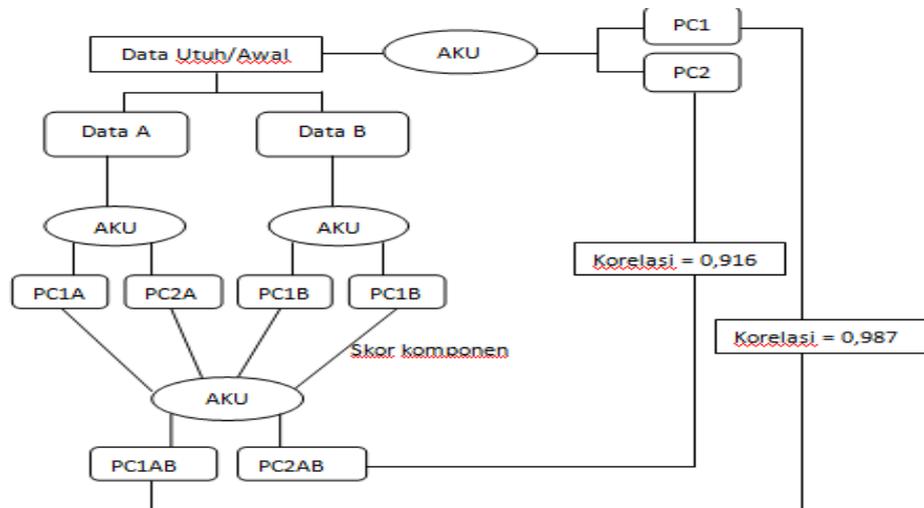
Data yang dipakai dalam penelitian ini merupakan data sekunder yakni data tahunan Produksi Pertanian dan Perkebunan Wilayah Bolaang Mongondow, Sulawesi Utara yang diterbitkan oleh Badan Pusat Statistik (BPS) Propinsi Sulawesi Utara. Fokus pembahasan hanya dilakukan pada data Produksi pertanian dan perkebunan di Wilayah Bolaang Mongondow setiap kecamatan pada tahun 2008.

3.1. Analisis Komponen Utama

- A. Dilakukan Analisis Komponen Utama Untuk Data utuh, langkah – langkahnya sebagai berikut:
 - 1) Pemasukkan Data Standarisasi Data (dengan *software* Minitab 14)
 - 2) Pembentukan dua Komponen Utama (dengan *software* Minitab 14)
- B. Penggabungan dua gugus data dengan Analisis Komponen Utama, langkah-langkah sebagai berikut :
 - 1) Pemasukkan Data
 - 2) Standarisasi Data (dengan *software* Minitab 14)
 - 3) Data dipilah menjadi dua gugus data (data A dan data B)
 - 4) Pembentukan dua Komponen Utama dari Data A dan dua Komponen Utama dari data B (dengan *software* Minitab 14)
 - 5) Komponen Utama dari data A dan Komponen Utama dari data B di gabungkan (dengan *software* Minitab 14)

3.2. Analisis Korelasi

Data asal setelah dilakukan Analisis komponen Utama, analisis terakhir yang digunakan yaitu analisis korelasi guna mencari tahu apakah ada hubungan antara KU dari Data Utuh dan KU dari Data yang telah digabungkan (Data A dan Data B) sehingga jika terdapat korelasi maka dapat dikatakan bahwa penggabungan data dapat dilakukan. (dengan *software* Minitab 14)



Gambar 1. Diagram Alir Metodologi Penelitian

4. Hasil Dan Pembahasan

4.1. Analisis Komponen Utama

Hasil analisis terhadap ketiga gugus data yang ada yakni Data Utuh (Awal), Data A (Perkebunan) dan Data B (Pertanian) dengan menggunakan Analisis Komponen Utama pada *software minitab 14* maka dibuat tabel hasil AKU yang meliputi nilai eigenvalue, proporsi keragaman dan kumulaif proporsi keragaman serta skor komponen utama yang dapat dilihat pada pembahasan dibawah ini.

4.1.1. Analisis Komponen Utama untuk Data Utuh

Tabel 1. Hasil AKU pada Data Utuh

Eigenvalue	6.3493	4.0439
Proportion	0.373	0.238
Cumulative	0.373	0.611

Selain nilai eigenvalue seperti pada Tabel di atas, pada output juga terlihat skor Komponen Utama dari 17 variabel, akan tetapi tidak dibahas pada pembahasan kali ini, dikarenakan kurang dibutuhkan dalam penelitian ini. yang akan dibahas hanyalah Skor Komponen Utama dari 12 Kecamatan yang nantinya akan digunakan untuk analisis lanjutan. Adapun Skor 2 Komponen Utama dari 12 Kecamatan dapat dilihat pada Tabel dibawah ini.

Tabel 2. Skor Dua Komponen

Kecamatan	PC1	PC2
Dumoga Barat	-0.82020	2.39745
Dumoga Utara	-1.00107	1.70746
Dumoga Timur	-1.18077	2.29801
Lolayan	0.94893	3.84775
Passi Barat	2.92859	-0.11589
Passi Timur	3.81887	-0.96274
Bilalang	4.83718	-1.70331
Potger	-1.15157	-1.06659
Bolaang	-2.37371	-1.61283
Bolaang Timur	-1.82576	-1.12778
Lolak	-2.46568	-2.12896
Sang Tumbolang	-1.71481	-1.53256

4.1.2. Analisis Komponen Utama untuk Data A

Setelah melakukan analisis Komponen Utama pada Data Utuh (Produksi Pertanian dan Perkebunan di Wilayah Bolaang Mongondow, 2008) untuk Analisis selanjutnya terlebih dahulu dilakukan pengelompokkan Data, Data Utuh dibagi menjadi dua gugus data yakni menjadi Data A dan Data B, Data A adalah Data Perkebunan yang terdiri dari 10 variabel dan Data B adalah Data Pertanian yang terdiri dari 7 variabel, cara membedakan antara Data Perkebunan dan Pertanian adalah Data Perkebunan merupakan Data jenis tanaman industri sedangkan Data pertanian adalah Data jenis tanaman pangan, dan masing-masing Data akan dilakukan Analisis Komponen Utama dengan menggunakan *software Minitab 14*. Di bawah ini adalah hasil Analisis Komponen Utama dari Data A.

Tabel 3. Hasil Analisis Komponen Utama untuk Data A

Eigenvalue	3.9833	2.7619
Proportion	0.398	0.276
Cumulative	0.398	0.675

Sama halnya dengan analisis pada Data Utuh, Analisis pada Data A juga menghasilkan Skor dua Komponen Utama dari ke 12 Kecamatan yang akan digunakan untuk Analisis lanjutan pada *Software Minitab 14*.

Tabel 4. Skor Dua Komponen Utama Untuk Analisis Lanjutan

kecamatan (Y)	PC1A	PC2A
Dumoga Barat	-0.36091	1.43098
Dumoga Utara	0.08263	1.88177
Dumoga Timur	-0.35906	2.16941
Lolayan	-3.49222	1.89249
Passi Barat	-1.36533	-0.10397
Passi Timur	-2.18242	-2.71793
Bilalang	-2.18242	-2.71793
Poigar	1.45688	0.04533
Bolaang	1.93880	-0.25875
Bolaang Timur	1.93880	-0.25875
Lolak	2.46885	-1.25332
Sang Tombolang	2.05640	-0.10933

4.1.3. Analisis Komponen Utama untuk Data B

Data B merupakan Data Pertanian yang terdiri dari tujuh variabel, Sama halnya dengan Analisis pada Data Utuh (Awal) dan Data A, Data B juga dianalisis dengan Analisis Komponen Utama pada software minitab 14, dan menghasilkan nilai Eigenvalue, Proporsi Keragaman dan Proporsi Kumulatif Keragaman seperti pada Tabel di bawah ini.

Tabel 5. Hasil Analisis Komponen Utama pada Data B

Eigenvalue	3.2720	1.4008
Proportion	0.467	0.200
Cumulative	0.467	0.668

Dari Tabel di atas diperoleh nilai kedua Komponen Utama menyatakan 66,8% dari total variabel, ini menunjukkan bahwa pembagian Data menjadi 2 Gugus Data sangat berpengaruh pada hasil AKU, dimana setelah dilakukan pembagian data diperoleh nilai-nilai kumulatif proporsi keragaman yang semakin besar. Dan dalam hasil juga menunjukkan skors tiap komponen, adapun skors dua komponen utama sebagai berikut;

Tabel 6. Skors Dua Komponen Utama untuk Analisis Lanjutan

kecamatan	PC1B	PC2B
Dumoga Barat	-1.42113	3.11991
Dumoga Utara	-0.69785	0.11183
Dumoga Timur	-0.99583	-0.14462
Lolayan	-0.82821	0.14811
Passi Barat	2.60784	-0.08148
Passi Timur	2.01260	1.20473
Bilalang	3.89405	-0.28448
Poigar	-0.16424	-0.86783
Bolaang	-1.46861	-0.95480
Bolaang Timur	-0.83833	-0.23067
Lolak	-1.64189	-0.60978
Sang Tombolang	-0.45840	-1.41092

4.1.4. Analisis Komponen Utama Untuk Data A dan Data B

Setelah dilakukan Analisis Komponen Utama untuk tiap-tiap Data yakni Data Utuh (Asal), Data A dan Data B maka dilanjutkan dengan Analisis Komponen Utama pada Data A dan Data B (penggabungan Data peubah ganda), yaitu dua Komponen Utama dari Data A dan dua Komponen Utama dari Data B digabungkan dengan Analisis Komponen Utama menggunakan software minitab 14 diperoleh hasil seperti pada Tabel di bawah ini.

Tabel 7. Hasil Analisis Komponen Utama untuk Data Gabungan (AB)

Eigenvalue	1.8689	1.4178
Proportion	0.467	0.354
Cumulative	0.467	0.822

Berdasarkan Tabel Hasil Analisis Komponen Utama untuk Data Gabungan (AB) di atas maka hasil AKU dari Data yang digabungkan memperlihatkan dominasi dua KU, dengan kumulatif proporsi keragaman yang dijelaskan oleh dua KU pada keseluruhan menunjukkan nilai sebesar 82,2%. Pada Data gabungan (AB) menunjukkan peningkatan nilai proporsi keragaman yang relatif cukup besar, sehingga dengan penggabungan Data Peubah Ganda dari pembagian Data Utuh (Awal) menjadi dua gugus Data menyebabkan pengaruh yang tinggi pada variabel yang mendominasi dua KU menjadi bertambah. Dalam hasil AKU dengan *software minitab 14* terlihat bahwa Data setelah dilakukan Penggabungan Data Peubah Ganda menghasilkan dominasi dua KU yang lebih besar dibandingkan dengan Data sebelum dilakukan penggabungan data atau Data Utuh (Awal) berdasarkan nilai kumulatif proporsi keragaman yang diperoleh. Hal ini karena Penggabungan Data cenderung menambah pengaruh variabel yang dominan dari hasil AKU terhadap KU tersebut. Skor dua Komponen Utama dapat dilihat pada Tabel dibawah ini;

Tabel 8. Skor Dua Komponen Utama Pada Data Gabungan (AB).

Variable	PC1	PC2
PC1A	-0.541	-0.456
PC2A	-0.456	0.541
PC1B	0.680	-0.193
PC2B	0.193	0.680

Dari Tabel 8 dapat dilihat bahwa skors untuk dua Komponen Utama yang ditampilkan sesuai dengan Persamaan (1) adalah :

$$PC1 = -0,541 PC1A - 0,456 PC2A + 0,680 PC1B + 0,193 PC2B$$

$$PC2 = -0,456 PC1A + 0,541 PC2A - 0,193 PC1B + 0,680 PC2B$$

Dan skor dua komponen untuk analisis lanjutan tertera pada Tabel 9.

Tabel 9. Skor Dua Komponen Utama untuk Analisis lanjutan.

Kecamatan	PC1AB	PC2AB
Dumoga Barat	-0.31953	2.49277
Dumoga Utara	-0.78261	0.73212
Dumoga Timur	-0.89578	0.81102
Lolayan	0.13971	1.58676
Passi Barat	1.36566	-0.04752
Passi Timur	2.29010	0.09152
Bilalang	2.75436	-0.96537
Poigar	-0.61056	-0.79914
Bolaang	-1.16238	-0.91871
Bolaang Timur	-0.80712	-0.56991
Lolak	-1.04201	-1.14650
Sang Tombolang	-0.92984	-1.26704

4.2. Analisis Korelasi

Berdasarkan hasil analisis dengan menggunakan *software minitab 14* diperoleh output sebagai berikut;

Tabel 10. korelasi PC1 dan PC1AB,PC2AB

	PC1
PC1AB	0.987
P-value	0.000
PC2AB	0.074
P-value	0.818

Tabel di atas menunjukkan bahwa PC1 dan PC1AB saling berkorelasi dengan nilai korelasi sebesar 0,987. Sedangkan PC1 dan PC2AB tidak berkorelasi karena hanya memiliki nilai korelasi sebesar 0,074.

Tabel 11. Korelasi PC2 dan PC1AB,PC2AB

	PC2
PC1AB	-0.119
p-value	0.713
PC2AB	0.916
p-value	0.000

Dari Tabel di atas dapat dilihat bahwa komponen yang saling berkorelasi adalah PC2 dan PC2AB yakni dengan korelasi sebesar 0,916, sedangkan antara PC2 dan PC1AB tidak saling berkorelasi. Dengan melihat hasil dari analisis terakhir yaitu analisis korelasi bahwa adanya korelasi antara komponen utama yaitu PC1 dan PC1AB serta PC2 dan PC2AB menandakan bahwa menganalisis Data secara keseluruhan sama halnya dengan menganalisis Data secara parsial dimana hasil Korelasi antara Data utuh dan Data gabungan menunjukkan hampir atau sama dengan 1, ini berarti bahwa hasil analisis komponen utama antara Data utuh dan Data gabungan adalah sama. Dibanding menganalisa Data keseluruhan, menganalisa Data secara parsial lebih memberikan hasil-hasil yang lebih baik berdasarkan nilai kumulatif proporsi keragaman yang diperoleh. Hal ini karena Penggabungan Data cenderung menambah pengaruh variabel yang dominan dari hasil AKU terhadap KU tersebut sehingga dengan penggabungan Data Peubah Ganda dari pembagian Data Utuh (Awal) menjadi dua gugus Data menyebabkan pengaruh yang tinggi pada variabel yang mendominasi dua KU menjadi bertambah.

5. Kesimpulan

Berdasarkan hasil analisa dan pembahasan, dapat disimpulkan bahwa :

1. Analisis Komponen Utama dapat digunakan untuk menggabungkan dua gugus data peubah ganda yang terpisah.
2. Tingkat korelasi antara komponen utama dari data utuh (awal) dengan komponen utama dari data yang digabungkan (data AB) sangat tinggi yang melebihi angka 90%.

6. Daftar Pustaka

- [1] Siswadi & B. Suharjo. 1997. Analisis Eksplorasi Data Peubah Ganda. FMIPA IPB, Bogor.
- [2] Johnson, R. A. dan D. W. Wichern. 1988. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.
- [3] Sharma, S. 1996. *Applied Multivariate Techniques*. Wiley, New York
- [4] Walpole, R. dan M, Raymond. 1995. Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan. ITB. Bandung.
- [5] Wikipedia. 2013. Analisis Komponen Utama from: http://id.wikipedia.org/wiki/analisis_komponen_utama. [13 februari 2013]