



FORUM STATISTIKA DAN KOMPUTASI

Vol. 6 No. 1 April 2001

Implementing Bayesian Inference Using MCMC on MINITAB	Nur Iriawan	1 - 6
Perluasan Kueri Menggunakan Peluang Bersyarat (<i>Query Expansion using Conditional Probability</i>)	Julio Adisantoso	7 - 13
Analisis Data Longitudinal dengan Metode Regresi Berstruktur Pohon (Khusus Penyakit Kencing Manis)	Hazmira Yozza Siswadi Budi Suharjo	14 - 21
Transformasi BOX-COX untuk Kenormalan Komponen Utama (Kasus Beberapa Data Pertanian)	Hanny A.H. Komalig Siswadi Budi Suharjo Aji Hamim Wigena	22 - 26
Metode AMMI Pada Model Campuran	Suwardi Ahmad Ansori Mattjik Budi Susetyo	27 - 34

Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Petanian Bogor
Bogor, Indonesia

FORUM STATISTIKA DAN KOMPUTASI

PENANGGUNG JAWAB

Ketua Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Pertanian Bogor

EDITOR PELAKSANA

Ir. Hari Wijayanto, MS
Ir. Aji Hamim Wigena, MSc
Ir. Anang Kurnia, MSi
Anwar Fitrianto, SSI
Bagus Sartono, SSI

TIM PENELAAH

Dr. Ir. Khairil A. Notodiputro
Prof. Dr. Barizi, MES
Dr. Ir. Aunuddin
Dr. Ir. Budi Susetyo
Dr. Ir. Bambang Juanda
Dr. Ir. Asep Saefuddin, MSc

PEMASARAN

Dra. Itasia Dina Sulvianti, MSi
Ir. Erfiani, MSi

ADMINISTRASI DAN KEUANGAN

Ir. I Made Sumertajaya, MSi

ISSN

0853-8115

ALAMAT

Divisi Pengembangan Statistika dan Komputasi
Jurusan Statistika FMIPA-IPB
Jalan Raya Pajajaran, Bogor
Telp. (0251) 313023, 379830
Fax. (0251) 381807
e-mail:statistika@fmipa.ipb.ac.id

Forum Statistika dan Komputasi diterbitkan berkala dua kali setahun yang memuat tulisan ilmiah yang berhubungan dengan bidang Statistika yang berasal dari dalam dan luar lingkungan Institut Pertanian Bogor. Tulisan ilmiah dapat berupa hasil penelitian, bahasan tentang metodologi, komputasi, tulisan populer dan tinjauan buku.

TRANSFORMASI BOX-COX UNTUK KENORMALAN KOMPONEN UTAMA (Kasus Beberapa Data Pertanian)¹⁾

Hanny A.H. Komalig²⁾, Siswadi³⁾, Budi Suharjo³⁾, Aji Hamim Wigena³⁾

Abstrak

Untuk memperbaiki kenormalan data peubah ganda bagi analisis komponen utama (AKU) dapat dilakukan transformasi Box-Cox. Penelitian ini bertujuan untuk menelusuri perubahan konfigurasi sebagai akibat perbaikan asumsi kenormalan beberapa komponen utama dalam AKU.

Penggunaan transformasi Box-Cox dapat memperbaiki kenormalan sejumlah data yang tidak berdistribusi normal. Perbaikan kenormalan dengan transformasi Box-Cox dalam dimensi rendah cenderung merubah konfigurasi dan perubahan ini berkaitan dengan struktur data yang ada gugus data tersebut.

Penggunaan transformasi Box-Cox untuk menormalkan komponen utama sangat berkaitan dengan tujuan penggunaan hasil komponen utama tersebut. Apabila tujuannya untuk inferensia mengenai struktur komponen utama, maka transformasi Box-Cox dapat digunakan untuk menormalkan komponen-komponen utama. Tetapi bila komponen utama ditujukan hanya untuk mendapatkan suatu deskripsi sederhana mengenai pengamatan-pengamatan, maka penggunaan transformasi Box-Cox tidak penting untuk dilakukan.

PENDAHULUAN

Menurut Jolliffe (1986) dalam penerapan AKU, yaitu mereduksi dimensi data untuk mendapat beberapa komponen utama, asumsi utama peubah acak X berdistribusi normal peubah ganda sering tidak terpenuhi, dan berakibat hasil penggunaan AKU menjadi terbatas dalam mendapatkan inferensia yang akurat.

Dalam memperbaiki kenormalan data dapat dilakukan dengan transformasi Box-Cox. Selain menormalkan data yang mendasari penggunaan AKU, suatu transformasi, seperti transformasi Box-Cox, dikatakan baik jika dengan transformasi itu konfigurasi skor-skor komponen utama data hasil transformasi cukup dekat kesesuaiannya dengan konfigurasi skor komponen utama data asal baik dengan atau tanpa standarisasi.

$$g_j(\mathbf{Y}'; \lambda) = Y_j^{(\lambda_j)} = \begin{cases} \frac{(Y_j^{\lambda_j} - 1)}{\lambda_j} & , \lambda_j \neq 0 \\ \ln(Y_j) & , \lambda_j = 0 \end{cases} \dots(1)$$

Tujuan penelitian ini ialah menelusuri sejauhmana perubahan konfigurasi sebagai akibat perbaikan asumsi kenormalan beberapa komponen utama pertama dalam AKU. Sedangkan manfaat penelitian ini diharapkan

memberikan informasi dari hasil transformasi Box-Cox terhadap AKU, khususnya keperluan dalam data-data pertanian apakah diperlukan suatu transformasi Box-Cox untuk perbaikan asumsi kenormalan.

TINJAUAN PUSTAKA

Analisis Komponen Utama.

AKU biasanya digunakan untuk : (1) identifikasi peubah baru yang mendasari data peubah ganda, (2) mengurangi banyaknya dimensi himpunan peubah yang biasanya terdiri atas peubah yang banyak dan saling berkorelasi menjadi peubah-peubah baru yang tidak berkorelasi, dan (3) menghilangkan peubah-peubah asal yang mempunyai sumbangan informasi yang relatif kecil (Siswadi & Suharjo, 1997).

Peubah baru yang dimaksudkan sebagai komponen utama tersebut mempunyai ciri: (1) kombinasi linear peubah-peubah asal, (2) jumlah kuadrat koefisien dalam kombinasi linear tersebut bernilai satu, (3) tidak berkorelasi, dan (4) mempunyai ragam yang terurut dari terbesar ke terkecil.

Perlunya Asumsi Kenormalan dalam AKU dan Perbaikannya.

Menurut Morrison (1990), bila menggunakan AKU dengan contoh yang diambil dari populasi normal, maka dapat dibuat inferensia mengenai koefisien-koefisien komponen dan akar-akar ciri. Inferensia

¹⁾ Sebagian dari tesis S2 penulis pertama.

²⁾ Dosen pada Fakultas Pertanian Universitas Sam Ratulangi Manado.

³⁾ Dosen pada Fakultas MIPA, Institut Pertanian Bogor.

Transformasi Box-Cox untuk Kenormalan Komponen Utama (Kasus Beberapa Data Pertanian)

transformasi yang tepat digunakan untuk tujuan tersebut adalah transformasi Box-Cox.

Transformasi Box-Cox

Misalkan $Y'=(Y_1, Y_2, \dots, Y_p)$ merupakan p-peubah, maka transformasi Box-Cox didefinisikan pada rumus (1), dengan fungsi kemungkinan maksimum pada rumus (2).

$$L_{\max}(\lambda_1, \lambda_2, \dots, \lambda_p) = -\frac{n}{2} \ln \left| \hat{\Sigma} \right| - (np/2) \cdot \ln(2\pi) + \sum_{j=1}^p \left[(\lambda_j - 1) \sum_{i=1}^n \ln(y_{ij}) \right] \dots (2)$$

Komputasi Vektor Parameter Transformasi Box-Cox.

Umumnya metode untuk menduga parameter transformasi Box-Cox peubah ganda merupakan modifikasi metode pendugaan kasus peubah tunggal.

Berikutnya tahap komputasi untuk kasus peubah tunggal tersebut:

- Pilihlah beberapa nilai $\lambda \in [-2, +2]$, dalam hal ini : -2, -1.9, -1.8, ..., 0, 0.1, ..., 1.9, 2.
- Tentukan fungsi kemungkinan $L(\lambda)$ untuk masing-masing nilai λ , dengan rumus (3).
- Selanjutnya dibuat grafik antara nilai-nilai $L(\lambda)$ yang diperoleh dengan nilai-nilai λ yang berpadanan.
- Tentukan suatu nilai λ sehingga nilai $L(\lambda)$ merupakan nilai yang maksimum (λ_{\max}). Nilai λ_{\max} ini merupakan penduga titik untuk pangkat yang diperlukan dalam transformasi data.

Plot Chi-Square.

Metode plot *chi-square* digunakan untuk memeriksa kenormalan gabungan dari suatu data yang didasarkan pada jarak Mahalanobis kuadrat. Prosedur ini mengukur jarak setiap pengamatan dari sentroid contoh, dalam suatu kontur, dan membandingkan dengan peluang teoritisnya. Johnson & Wichern (1988) menyatakan bila populasi induk bersifat normal peubah ganda dan n lebih besar dari 25, maka setiap jarak kuadrat $d_1^2, d_2^2, \dots, d_n^2$ akan merupakan peubah acak *chi-square*.

METODE PENELITIAN.

Data.

Data yang dipakai dalam penelitian ini merupakan

$$L(\lambda) = -\frac{n}{2} \ln \left[\frac{1}{n} \sum_{i=1}^n \left(x_i^{(\lambda)} - \overline{x^{(\lambda)}} \right)^2 \right] + (\lambda - 1) \sum_{i=1}^n \ln(x_i) \dots (3)$$

set data peubah ganda yang merupakan hasil-hasil penelitian di bidang pertanian yang menggunakan data peubah ganda yaitu data dari penelitian: a) Herawati (1997), b) Wasrdani (1995), c) Trisnawati (1997), dan d) Khristiningrum (1997). Data ini mewakili data yang bersifat tidak normal tapi bisa dinormalkan dengan

transformasi Box-Cox (data a dan b), data yang bersifat normal (data c), dan data tidak normal juga tidak bisa dinormalkan dengan transformasi (data d).

Analisis Data

Tahap analisis dalam penelitian ini dimulai dengan mentransformasikan data, yang meliputi: a) Standarisasi dan b) Transformasi Box-Cox.

Mendapatkan Beberapa Komponen Utama Pertama.

Setelah matriks standarisasi dan koefisien-koefisien pangkat transformasi Box-Cox serta matriks transformasi diperoleh, dilakukan AKU berdasarkan matriks-matriks tersebut untuk mendapatkan beberapa komponen utama pertama beserta skor-skor komponen-nya. Adapun batasan jumlah komponen utama tersebut digunakan total ke-ragaman dibatasi dua komponen utama pertama atau minimal 65 % total keragaman.

Metode Plot Chi-Square.

1. Urutkan jarak-jarak kuadrat tersebut dari nilai terkecil ke terbesar, yaitu $d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(n)}^2$ dimana n adalah banyaknya pengamatan.
2. Untuk setiap $d_{(i)}^2$, hitung nilai-nilai $(i-0.5)/n$.
3. Hitung $\chi_p^2((i-0.5)/n)$ yang merupakan presentil $100(i-0.5)/n$ dari distribusi *chi-square* dengan derajat bebas p.
4. Selanjutnya buat plot antara pasangan-pasangan ($d_{(i)}^2; \chi_p^2((i-0.5)/n)$).
5. Hitung korelasi antara nilai-nilai $d_{(i)}^2$ dan $\chi_p^2((i-0.5)/n)$.

Analisis Procrustes.

Prosedur ini membandingkan dua konfigurasi skor komponen. Misalkan X dan Y merupakan matriks-matriks berukuran nxk yang berisi koordinat konfigurasi dari titik-titik dalam ruang berdimensi-k. Kedua konfigurasi tersebut dilakukan penyesuaian translasi, rotasi dan dilasi secara seragam untuk mendapatkan sepasang konfigurasi yang sedekat mungkin.

Ukuran simetrik dari kesamaan antara dua

konfigurasi ialah (Cox & Cox, 1994): $\gamma = 1 - M^2_{\min} / \text{tr}(X'X')$, dimana minimum M^2 diperoleh dengan $M^2 = \text{tr}[XX' + c^2 YY' - 2c XQ'Y']$, untuk $Q = AU'$, dan ULA' adalah hasil penguraian nilai singular dari $X'Y$, serta $c = \text{tr}(XQ'Y') / \text{tr}(YY')$.

Dengan nilai $\gamma=1$ dapat dinyatakan bahwa kedua konfigurasi identik.

HASIL DAN PEMBAHASAN.

Analisis Komponen Utama.

Secara keseluruhan pada tabel di bawah menunjukkan bahwa untuk hasil AKU dari data asal memperlihatkan dominasi dua Komponen Utama (KU) pertama, dengan kumulatif proporsi keragaman di atas 85 %. Penggunaan standarisasi terhadap AKU, diperoleh nilai-nilai kumulatif yang mengecil. Kecuali pada kasus data Trisnawati menunjukkan penurunan yang relatif besar untuk penggunaan standarisasi dan standarisasi setelah transformasi, sehingga memerlukan tiga KU pertama untuk mendapatkan total keragaman 65 %.

Tabel 1. Daftar nilai kumulatif proporsi komponen utama.

Sumber Data	Data asal	Standarisasi	Trans. Box-Cox
Herawati	0.995	0.411	0.956
	0.998	0.665	0.987
	0.999	0.865	1.000
	1.000	0.939	1.000
	1.000	0.983	1.000
Wardani	0.975	0.490	0.986
	0.989	0.720	0.999
	0.998	0.893	0.999
Trisnawati	0.789	0.300	0.625
	0.880	0.510	0.887
	0.930	0.667	0.997
	0.965	0.800	0.999
	0.986	0.906	1.000
Khristiningrum	0.971	0.979	0.983
	0.995	0.993	0.999
	0.999	0.999	1.000

Transformasi Box-Cox dari ketiga metode untuk seluruh data terlihat hasil AKU yang agak mirip dengan yang diperoleh dari data asal. Ini menyatakan transformasi Box-Cox tidak mengurangi pengaruh perbedaan keragaman antar peubah pada data asal. Sebaliknya transformasi ini cenderung memperkuat dominasi beberapa KU pertama tersebut.

Untuk kombinasi standarisasi setelah transformasi Box-Cox menghasilkan dominasi dua KU pertama yang lebih kecil dibandingkan dengan penggunaan standarisasi sebelum transformasi Box-Cox. Ini karena standarisasi setelah transformasi Box-Cox cenderung mengurangi pengaruh peubah yang dominan dari hasil transformasi Box-Cox terhadap beberapa KU pertama tersebut.

Analisis Plot Chi-Square untuk Beberapa Skor Komponen Utama.

Hasil analisis plot *chi-square* dari keempat gugus data tersebut ditampilkan dalam Tabel 2 berikut ini.

Tabel 2. Daftar nilai koefisien korelasi dari plot chi-square.

Sumber Data	Kenormalan awal		Kenormalan Skor Komponen Utama			Nilai kritis 5%
	Data Asal	Box-Cox	Data asal	Standarisasi	Box-Cox	
Herawati	0.971	0.992	0.971	0.958	0.989	0.983
Wardani	0.955	0.980	0.985	0.978	0.980	0.971
Trisnawati	0.985	0.991	0.986	0.968	0.977	0.974
Khristiningrum	0.935	0.936	0.760	0.762	0.773	0.970

a). Kasus Data Herawati.

Dengan penerapan AKU untuk dua KU pertama dari data asal, nilai korelasinya lebih kecil dari r_{tabel} , demikian juga dengan standarisasi memberikan hasil yang tidak normal. Dibandingkan dengan data asal, tampaknya standarisasi cenderung menurunkan nilai kenormalan skor KU tersebut. Kemudian untuk transformasi Box-Cox, terlihat transformasi ini dapat menormalkan kedua KU pertamanya.

b). Kasus Data Wardani.

Dengan beberapa KU pertama untuk data asal dan yang distandarisasi memberikan hasil yang normal, padahal pemeriksaan awal terhadap seluruh peubah diperoleh bahwa data ini tidak normal. Hasil ini berkaitan dengan yang diperoleh Johnson & Wichern (1988), dimana data yang direpresentasikan dimensi rendah menampilkan keadaan yang bersifat normal tapi bila direpresentasikan dalam dimensi yang lebih tinggi dapat bersifat tidak normal. Jadi dapat dinyatakan bahwa tidak perlu menggunakan transformasi menormalkan untuk representasi dimensi rendah tapi cukup dengan menggunakan AKU.

Untuk standarisasi nampaknya hasil yang sama seperti data Herawati, dimana standarisasi ini menekan nilai kenormalan skor-skor beberapa KU pertama. Demikian juga dengan transformasi Box-Cox diperoleh hasil yang berdistribusi normal.

c). Kasus Data Trisnawati.

Dari pemeriksaan kenormalan beberapa komponen utama (KU) pertama data asal diperoleh hasil yang normal. Sedangkan untuk standarisasi, nampaknya mirip dengan hasil pada data-data sebelumnya, dimana dengan standarisasi maka nilai kenormalan diturunkan, sehingga beberapa KU pertama tersebut cenderung menjadi bersifat tidak normal lagi. Sebaliknya untuk transformasi Box-Cox, terlihat hasil yang cukup normal.

d). Kasus Data Khristiningrum.

Dari pemeriksaan kenormalan beberapa KU pertama dari data asal, terlihat hasil tersebut bersifat tidak normal. Hasil yang sama juga diperoleh pada hasil-hasil yang menggunakan standarisasi dan yang melakukan transformasi Box-Cox.

Analisis Procrustes untuk Konfigurasi Beberapa Skor KU pertama.

Hasil analisis Procrustes dari keempat gugus data tersebut ditampilkan dalam Tabel 3 berikut ini.

Tabel 3. Daftar nilai ukuran kesesuaian koefisien Procrustes.

	Konfigurasi yang Dibandingkan	
	(Za,Z)	(Za,Zb)
Herawati	48.42	42.25
Wardani	64.86	62.14
Trisnawati	44.84	46.08
Khristiningrum	99.75	92.82
		(Z,Zb)
Herawati		5.01
Wardani		99.03
Trisnawati		81.16
Khristiningrum		94.41

Keterangan :

- Z = Skor beberapa KU pertama dari data asal sebagai salah satu pembandingan,
- Za = Skor beberapa KU pertama dari data standarisasi sebagai suatu pembandingan,
- Zb = Skor beberapa KU pertama dari transformasi Box-Cox.

a). Perbandingan Konfigurasi dengan Data yang Di standarisasi.

Berdasarkan ukuran kesesuaian procrustes untuk keempat data yang distandarisasi dengan data-data yang ditransformasi Box-Cox, terlihat bahwa data Wardani dan Khristiningrum mempunyai nilai-nilai kesesuaian di atas 60 %, yang relatif lebih besar dibanding hasil data Herawati dan Trisnawati. Ini berkaitan dengan jumlah pengamatan dalam setiap data, dimana semakin banyak pengamatan dari konfigurasi yang dibandingkan maka semakin besar kemungkinan mempunyai ukuran kesesuaian yang kecil.

Pada data Khristiningrum terlihat nilai kesesuaian di atas 90 %. Ini jelas berhubungan dengan struktur data itu sendiri, dimana data ini tidak normal dan tidak dapat dinormalkan dengan transformasi tersebut. Sehingga konfigurasi transformasi tidak jauh berbeda dengan konfigurasi standarisasi tersebut.

b). Perbandingan Konfigurasi dengan Data Asal.

Hasil perbandingan konfigurasi beberapa KU pertama data asal dengan hasil-hasil transformasi Box-Cox, terlihat perbedaan antara kasus data Herawati dengan data-data lain. Ada dua alasan yang mungkin; pertama, data tersebut tidak berdistribusi normal peubah ganda, dan kedua, jumlah pengamatan yang jelas lebih besar dari data-data lain. Sehingga dengan menormalkan data tersebut akan mengubah konfigurasi dan semakin banyak pengamatan maka

semakin banyak titik-titik pasangan perbandingan konfigurasi yang berubah.

Sebaliknya, hasil data Herawati dan Trisnawati sangat berbeda pada hasil transformasi Box-Cox, dimana untuk data Herawati hanya sekitar 5%. Ini berkaitan dengan struktur data yang ada; yaitu data tidak normal, sedangkan data Trisnawati normal. Jadi transformasi ini untuk menormalkan mempengaruhi konfigurasi. Sebagai pembandingan, perhatikan kasus data Khristiningrum yang tidak normal. Penggunaan transformasi ini tidak menormalkan, sehingga konfigurasi tidak jauh berbeda dengan konfigurasi data asalnya.

KESIMPULAN DAN SARAN.

Kesimpulan.

- Secara umum, penggunaan transformasi Box-Cox dapat meningkatkan nilai kenormalan dalam representasi pada dimensi rendah khususnya dengan penggunaan AKU, sedangkan pengaruh transformasi Box-Cox terhadap perubahan konfigurasi sangat berkaitan dengan keadaan struktur data.
- Penggunaan transformasi Box-Cox untuk menormalkan komponen utama sangat berkaitan dengan tujuan penggunaan hasil komponen utama tersebut. Apabila tujuannya untuk inferensia mengenai struktur komponen utama, maka transformasi Box-Cox dapat digunakan untuk menormalkan komponen-komponen utama. Tetapi bila komponen utama ditujukan hanya untuk mendapatkan suatu deskripsi sederhana mengenai pengamatan-pengamatan, maka penggunaan transformasi Box-Cox tidak penting untuk dilakukan.
- Perbaiki nilai kenormalan pada hasil-hasil AKU dengan transformasi Box-Cox yang direpresentasikan dalam dimensi rendah (dua atau tiga komponen utama) cenderung mengubah konfigurasi hasil transformasi tersebut apabila data asal bersifat tidak normal, tetapi perbaikan ini dapat menormalkan hasil-hasil AKU untuk beberapa komponen utama pertama.

Saran.

- Perlu adanya penelitian lanjutan mengenai keterkaitan antara jumlah pengamatan, standarisasi dengan perubahan konfigurasi sebagai akibat penggunaan transformasi Box-Cox.

DAFTAR PUSTAKA

Andrews, D., R. Gnanadesikan, & J. L. Warner. 1971. Transformations of Multivariate Data. *Biometrics* 27, pp 825-840.

Transformasi Box-Cox untuk Kenormalan Komponen Utama
(Kasus Beberapa Data Pertanian)

Cox, T. & M. A. A. Cox. 1994. *Multidimensional Scaling*. Chapman & Hall, London.

Gnanadesikan, R. 1977. *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York.

Herawatie, D. 1997. *Penggunaan Analisis Procrustes untuk Mengukur Kehilangan Informasi Akibat Reduksi Dimensi dengan Analisis Gradien Langsung*. Tesis. Pascasarjana IPB. Bogor. Tidak Dipublikasikan.

Johnson, R. A. & D. W. Wichern. 1988. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Jolliffe, I. T. 1986. *Principal Component Analysis*. Springer-Verlag, New York.

Kendall, S. M. 1975. *Multivariate Analysis*. Charles Griffin, London.

Khristiningrum, M. N. 1997. *Metode Regresi Komponen Utama untuk Kalibrasi Peubah Ganda*. Skripsi. Jurusan Statistika FMIPA-IPB. Bogor. Tidak Dipublikasikan.

Manly, BF 1986. *A Primer, Multivariate Statistical Methods*. Chapman & Hall, London.

Morrison, D.F. 1990. *Multivariate Statistical Methods*. 3rd ed. McGraw-Hill, Singapore.

Siswadi & B.Suharjo. 1997. *Analisis Eksplorasi Data Peubah Ganda*. FMIPA-IPB, Bogor.

Trisnawati, L. D. R. 1997. *Seleksi Klon Ubi Jalar dengan Metode Biplot*. Skripsi. Jurusan Statistika FMIPA-IPB. Bogor. Tidak Dipublikasikan.

Wardani, R. A. 1995. *Pengelompokan Sentra Produksi Padi Sawah di Jawa Berdasarkan Kemiripan Persentase Jenis Intensifikasi pada Periode 1989 dan 1993*. Skripsi. Jurusan Statistika FMIPA-IPB. Bogor. Tidak Dipublikasikan.

⊗