

Review Jurnal

Variable Reduction Schemes Based on Principal Component, Canonical Correlation and Procrustes Analyses

Latar Belakang

Korelasi kanonik dikembangkan oleh Hotelling pada tahun 1936. Analisis korelasi kanonik adalah analisis peubah ganda yang sering digunakan untuk menguji hubungan secara linear antara dua kelompok peubah (Rencher dan Christensen 2012).

Salah satu masalah kompleks yang seringkali terjadi adalah analisis peubah ganda yang mempunyai peubah yang sangat banyak sehingga sulit untuk merepresentasikan hasil penelitian. Kesulitan lainnya adalah hasil penelitian sulit untuk disimpulkan secara global. Selain itu, dengan banyaknya peubah memperbesar kemungkinan tidak akuratnya koefisien dari penduga parameter. Terjadi juga kondisi saat data yang diperlukan dalam satu peubah memerlukan biaya yang besar. Permasalahan ini dapat diatasi dengan mengurangi peubah antara lain dengan menggunakan analisis korelasi kanonik dengan menghilangkan satu per satu peubah secara langsung dan dengan peubah kanonik. Analisis lainnya dapat dilakukan dengan analisis komponen utama dan analisis Procrustes.

Dalam analisis korelasi kanonik, beberapa peubah dalam dua kelompok peubah dapat digunakan untuk mempelajari korelasi antara dua kelompok peubah (Timm 2002). Oleh karena itu, peubah yang dihilangkan adalah yang memberikan pengurangan nilai korelasi kanonik paling sedikit atau yang mempunyai koefisien peubah dalam peubah kanonik paling dekat dengan nol.

Analisis komponen utama merupakan analisis peubah ganda yang digunakan untuk mengurangi dimensi data yang berukuran sangat besar dengan mempertahankan sebanyak mungkin informasi yang terkandung pada data asalnya. Analisis komponen utama mentransformasi peubah-peubah asli yang masih saling berkorelasi satu dengan yang lain menjadi satu himpunan peubah baru yang tidak berkorelasi lagi. Peubah-peubah baru tersebut disebut dengan komponen utama. Peubah yang dihilangkan adalah peubah yang mempunyai korelasi paling besar dengan komponen utama yang bersesuaian dengan nilai eigen terkecil.

Analisis Procrustes adalah teknik yang mengacu pada perbandingan dua objek dengan berbagai kondisi dan menghasilkan ukuran kesesuaian. Analisis Procrustes dapat menghilangkan kemungkinan peubah yang tidak dapat dibandingkan dalam kelompok data individu dan perbedaan ukuran antara kelompok data dengan mengubah skala data dan menghitung jarak antar data. Perhitungan dengan jarak Euclid dan transformasi dengan cara translasi-rotasi dan dilasi adalah langkah yang efektif dalam analisis Procrustes (Bakhtiar dan Siswadi 2015). Pengurangan peubah dalam analisis Procrustes dilakukan dengan membandingkan jarak antara matriks awal yang ingin dihilangkan peubahnya dengan matriks itu sendiri setelah pengurangan peubah satu per satu.

Metode

1.1. Sumber Data

Data yang digunakan dalam karya ilmiah ini adalah data peubah ganda dua kelompok tentang jumlah ikan jenis tertentu dan lingkungan laut di sekitar Laut Barents (Anonim 1997). Kelompok pertama berkaitan dengan lingkungan terdapat empat peubah yaitu garis lintang, garis bujur, kedalaman dan suhu di 89 teritorial Laut Barents. Kelompok kedua berkaitan dengan jenis ikan yang berada di 89 teritorial Laut Barents. Terdapat 30 jenis ikan yang setiap jenisnya mewakili satu peubah. Data ini adalah data yang diunduh dari internet.

1.2. Proses Analisis Data

Pengurangan peubah dilakukan dengan tiga jenis analisis, yaitu analisis korelasi kanonik, analisis komponen utama, dan analisis Procrustes. Proses analisis data pada setiap metode adalah sebagai berikut:

Analisis korelasi kanonik

Terdapat dua cara mengurangi peubah dengan analisis korelasi kanonik.

Cara pertama (skema 1) ialah dengan mencari korelasi kanonik dari dua kelompok peubah, dengan data kelompok kedua dikurangi satu peubah. Lakukan berulang kali hingga semua peubah pernah dikurangi. Peubah yang dihilangkan ialah peubah yang memberikan pengurangan korelasi kanonik pertama yang paling sedikit. Setelah mengetahui peubah yang dihilangkan, tentukan korelasi kanonik dari kelompok peubah yang baru. Ulangi langkah ini sampai nilai korelasi kanonik berkurang cukup banyak dari korelasi kanonik semula.

Cara kedua (skema 2) ialah dari dua kelompok data dicari korelasi kanonik dan peubah kanoniknya. Peubah kanonik yang dicari adalah kombinasi linear dari konstanta dengan peubah pada kelompok yang ingin dikurangi peubahnya, dalam hal ini adalah data kelompok dua dengan 30 peubah. Kombinasi linear tersebut didapat dengan cara mencari vektor eigen yang bersesuaian dari nilai eigen terbesar dari invers dari matriks koragam data kelompok dua dikalikan matriks koragam data kelompok dua dengan kelompok satu dikalikan dengan matriks koragam data kelompok satu dikalikan matriks koragam data kelompok satu dengan kelompok dua. Nilai eigen dari matriks tersebut merupakan korelasi kanonik antar dua kelompok peubah. Peubah dengan koefisien peubah dalam peubah kanonik yang mempunyai nilai paling dekat dengan nol adalah peubah yang dihilangkan. Setelah menentukan peubah yang dihilangkan, tentukan kembali korelasi kanonik antara kelompok satu dan kelompok dua dengan 29 peubah. Tentukan juga kombinasinya dan hilangkan peubah dengan koefisien peubah dalam peubah kanonik paling dekat dengan nol. Ulangi tahap ini hingga korelasi kanonik berkurang cukup banyak.

Analisis komponen utama

Pada analisis skema ke-3, terdapat matriks data dua kelompok. Kelompok pertama berdimensi 89×4 . Kelompok kedua berdimensi 89×30 . Kedua kelompok data ini distandardisasi agar tidak terdapat peubah yang mendominasi. Kedua kelompok data yang sudah terstandardisasi ini dicari korelasinya dengan analisis korelasi kanonik. Terdapat empat korelasi kanonik. Korelasi pertama merupakan korelasi terbesar pertama, korelasi kedua merupakan korelasi terbesar kedua, korelasi ketiga merupakan korelasi terbesar ketiga dan korelasi keempat merupakan korelasi terbesar keempat.

Kelompok data kedua yang mempunyai 30 peubah dicari komponen utamanya. Komponen utama yang digunakan adalah vektor eigen yang bersesuaian dengan nilai eigen terkecil dikalikan dengan data asli. Kemudian cari korelasi antara peubah asal ke- i dengan komponen utama ke- j . Peubah yang mempunyai korelasi terbesar adalah peubah yang dihilangkan.

Setelah menghilangkan satu peubah, tentukan kembali korelasi kanonik dari data kelompok pertama dan kelompok kedua. Jika selisih korelasi kanonik sebelum dan sesudah pengurangan peubah tidak besar, lakukan pengurangan peubah kembali dengan mencari komponen utama yang bersesuaian dengan nilai eigen terkecil dan tentukan kembali korelasinya. Proses analisis ini terus berlanjut hingga nilai korelasi kanonik dari kedua kelompok peubah turun cukup jauh dari korelasi kanonik dengan peubah lengkap.

Analisis Procrustes

Langkah pertama pada analisis skema ke-empat adalah menentukan korelasi kanonik data lengkap. Setelah itu lakukan pengurangan peubah.

Pengurangan peubah diawali dengan melakukan konfigurasi matriks baru dari data kelompok dua dengan salah satu setiap unsur kolomnya diubah menjadi nol. Misalkan matriks pertama adalah matriks dengan unsur setiap kolom pertama nol dan kolom lainnya merupakan data dari kelompok dua, matriks kedua adalah matriks dengan semua unsur kolom kedua nol dan kolom lainnya tetap, seterusnya sampai 30 peubah dengan 30 matriks baru. Setiap matriks baru ditentukan ukuran Procrustesnya. Ukuran Procrustes menyatakan selisih jarak antara dua matriks. Semakin kecil ukuran Procrustes, semakin kecil pula jaraknya. Peubah yang dihilangkan ialah peubah yang memiliki ukuran Procrustes terkecil antara matriks yang terkonfigurasi dari kolom setiap unsur peubahnya nol dengan matriks data asal. Ukuran Procrustes yang kecil ini menyatakan peubah yang telah dibuat nol tidak banyak berpengaruh terhadap jarak antar matriks konfigurasi dan matriks asal.

Setelah melakukan pengurangan peubah hitung kembali korelasi kanoniknya. Jika korelasi tidak turun secara drastis, lakukan kembali pengurangan peubah dengan cara yang sama. Matriks yang

digunakan adalah matriks dengan 29 peubah. Konfigurasi kembali matriks baru dan hitung ukuran Procrustesnya. Tahap ini berlanjut hingga korelasi kanonik setelah peubah yang dihilangkan mengalami banyak penurunan.

Hasil

Pengurangan peubah pada masing-masing skema terdapat pada tabel berikut:

Number of removed variables	Scheme-1		Scheme-2		Scheme-3		Scheme-4	
	Removed variables	First CC	Removed variables	First CC	Removed variables	First CC	Removed variables	First CC
0		0.9264		0.9264		0.9264		0.9264
1	Y_2	0.9264	Y_{25}	0.9253	Y_{26}	0.9264	Y_{13}	0.9264
2	Y_{13}	0.9264	Y_{14}	0.9209	Y_{28}	0.9251	Y_{18}	0.9249
3	Y_{26}	0.9263	Y_5	0.9197	Y_{22}	0.9222	Y_6	0.9195
4	Y_{10}	0.9263	Y_{22}	0.9158	Y_{25}	0.9210	Y_{21}	0.9146
5	Y_{19}	0.9262	Y_{11}	0.9077	Y_{14}	0.9155	Y_4	0.9097
6	Y_8	0.9262	Y_{24}	0.9040	Y_3	0.9132	Y_{12}	0.8972
7	Y_{17}	0.9260	Y_{29}	0.8981	Y_{23}	0.9126		
8	Y_{23}	0.9255			Y_{20}	0.8987		
9	Y_{30}	0.9243			Y_5	0.8985		
10	Y_1	0.9231						
11	Y_{18}	0.9213						
12	Y_{28}	0.9193						
13	Y_{24}	0.9177						
14	Y_5	0.9156						
15	Y_{16}	0.9141						
16	Y_{22}	0.9112						
17	Y_3	0.9079						
18	Y_4	0.9016						
19	Y_9	0.8929						
Percent CC deterioration		3.61		3.05		3.01		3.15

Skema yang paling berhasil menghilangkan peubah adalah skema 1 sebanyak 19 peubah sedangkan skema 2 hanya 7 peubah, skema 3 sebanyak 9 peubah sedangkan skema 4 sebanyak 6 peubah.

Kesimpulan

Terdapat 4 analisis yang dapat digunakan dalam mengurangi suatu peubah, yaitu analisis korelasi kanonik (langsung dan tidak langsung), analisis komponen utama, dan analisis Procrustes. Prosedur melalui proses iterasi telah dilaksanakan, dengan menghilangkan satu per satu peubah yang mempunyai kontribusi paling sedikit pada data. Maksimum 3% toleransi kekurangan data, maka didapatkan analisis korelasi kanonik langsung dapat mengurangi 18 peubah dengan tetap mempertahankan data sebanyak 90,16%.