KORELASI ANTARA DUA KELOMPOK VARIABEL KUANTITATIF DALAM ANALISIS KANONIK

Oleh :

Dewi Rachmatin, S.Si., M.Si.

Jurusan Pendidikan Matematika
FPMIPA Universitas Pendidikan Indonesia

ABSTRAK

Gagasan dasar dari Analisis Kanonik adalah mengembangkan pengertian koefisien korelasi antara dua kelompok variabel kuantitatif menjadi pengertian "korelasi" antara dua kelompok variabel kuantitatif. Melalui teknik Analisis Kanonik dapat dipelajari kemiripan antara kedua kelompok variabel kuantitatif. Berdasarkan hasil ini, penyajian data variabel maupun individu dapat dilakukan pada ruang bagian berdimensi kecil yang optimal seperti halnya pada Analisis Komponen Utama.

Kata Kunci : Analisis Kanonik, korelasi antara dua kelompok variabel kuantitatif, penyajian data.

PENDAHULUAN

Gagasan dasar dari Analisis Kanonik adalah mengembangkan pengertian koefisien korelasi antara dua kelompok variabel kuantitatif menjadi pengertian "korelasi" antara dua kelompok variabel kuantitatif. Jika ada dua kelompok variabel X dan Y, melalui teknik Analisis Kanonik dapat diselidiki derajat ekivalensi antara X dan Y atau seberapa jauh keduanya mendekati keadaan ekivalen.

Misalkan ada dua kelompok variabel kuantitatif $\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^p\}$ dimana $\bar{x}^j = (x_1^j, x_2^j, \ldots, x_n^j)$ untuk setiap j = 1, 2, …, p dan $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^q\}$ dimana $\bar{y}^j = (y_1^j, y_2^j, \ldots, y_n^j)$ untuk setiap j=1, 2, …, q. Pengertian ekivalensi antara kedua kelompok tersebut didefinisikan sebagai berikut.

Definisi

Dua kelompok variabel kuantitatif $\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^p\}$ dan $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^q\}$ dikatakan ekivalen, jika himpunan semua kombinasi linier dari $\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^p\}$ berimpit dengan himpunan kombinasi linier dari $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^q\}$.

Untuk memperjelas definisi tersebut, pandang matriks-matriks data berikut.. Misalkan I = {1, 2, …, n} himpunan individu X(pxn) dan Y (qxn) matriks-matriks data berturut-turut hasil pengukuran kelompok variabel pertama dan kedua pada I. Diagram dual yang sesuai dengan X, Y, dan I tersebut adalah:

$$\begin{tikzcd}
E_1=R^p \arrow[d, shift right, "M"'] & F^* \arrow[l, "X"'] \arrow[r, "Y"] & E_2^*=R^q \arrow[d, shift left, "N"] \\
E_1^* \arrow[u, shift right, "V_{11}"'] \arrow[r, "X^T"] & F=R^n \arrow[u, "D_p"] & E_2 \arrow[l, "Y^T"'] \arrow[u, shift left, "V_{22}"]
\end{tikzcd}$$

Misalkan himpunan semua kombinasi linier dari $\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^p\}$:

$$W_1 = \{\bar{\tau} \mid \bar{\tau} = X^T \bar{a}\,;\, \bar{a} \text{ di } E_1^*\} = X^T(E_1^*) \subset R^n \text{ dimana } \bar{a} = \sum_{j=1}^{p} a_j \bar{e}_j^*(1)\,,\ E_1 = R^p.$$

Sedangkan himpunan semua kombinasi linier dari $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^q\}$:

$$W_2 = \{\bar{\eta} \mid \bar{\eta} = Y^T\,\bar{b}\,;\, \bar{b} \text{ di } E_2^*\} = Y^T(E_2^q) \subset R^n \text{ dimana } \bar{b} = \sum_{j=1}^{q} b_j \bar{e}_j^*(2)\,,\ E_2 = R^q.$$

Kedua kelompok variabel $\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^p\}$ dan $\{\bar{y}^1, \bar{y}^2, \ldots, \bar{y}^q\}$ ekivalen jika $W_1 = W_2$.

Dalam praktek boleh dikatakan tidak pernah kita jumpai X ekivalen dengan Y. Jadi $W_1 \neq W_2$. Yang ingin diselidiki adalah derajat ekivalensi antara X dan Y atau seberapa jauh keduanya mendekati keadaan ekivalen. Berdasarkan hal ini, penyajian data akan kita lakukan pada ruang bagian berdimensi kecil yang optimal seperti halnya pada analisis komponen utama.

Jika $W_1 \neq W_2$, ada lima ruang bagian yang harus diperhatikan, yaitu : $W_1 \cap W_2$, $W_1 \cap W_2^{\perp}$, $W_1^{\perp} \cap W_2$, $(W_1 \cap W_2 \oplus W_1 \cap W_2^{\perp})$, dan $(W_1 \cap W_2 \oplus W_1^{\perp} \cap W_2)^{\perp}$ (Djauhari; 1988).

Masalah selanjutnya yang dihadapi adalah bagaimana mencari pasangan variabel kanonik $(\bar{\tau}^j, \bar{\eta}^j)$, untuk semua j = 1,2,…,k. Cara yang sederhana untuk mencari pasangan variabel kanonik, diberikan pada dalil di bawah ini. Sebelumnya jika X dan Y terbakukan, kita lukis :

V_{11} = X Dp X^T adalah matriks variansi-kovariansi kelompok variabel yang pertama.

V_{22} = Y Dp Y^T adalah matriks variansi-kovariansi kelompok variabel yang kedua.

V_{12} = X Dp $Y^T = V_{21}^T$ adalah matriks variansi-kovariansi kelompok variabel yang pertama dan kedua.

Dalil

Misalkan $A_1 = X^T(X\ D_p\ X^T)^{-1}\ X\ D_p$ dan $A_2 = Y^T(Y\ D_p\ Y^T)^{-1}\ Y\ D_p$.

A_1A_2 dan $V_{11}^{-1}\ V_{12}\ \ V_{22}^{-1}\ V_{22}$ memiliki nilai-nilai karakteristik positif yang sama.

Jika $V_{11}^{-1}\ V_{12}\ \ V_{22}^{-1}\ V_{22}\ \bar{a}_j = \lambda_j\ \bar{a}_j$ dan $\left\|\bar{a}_j\right\|_{V_{11}}$ = 1, maka $\bar{\tau}^j = X^T\bar{a}_j$; j = 1, 2,…, p.

Bukti dalil di atas dapat dilihat pada referensi [1]. Akibat dalil di atas adalah sebagai berikut.

Akibat

Bila matriks data X dan Y terpusat, maka :

a) $\bar{\tau}^j$ dan $\bar{\eta}^j$ juga terbakukan sebab keduanya adalah kombinasi linier dari variabel-variabel terbakukan . $\left(\bar{\tau}^j, \bar{\eta}^j\right)$adalah pasangan kanonik ke-j, j =1, 2, …, k. dimana : $\bar{\tau}^j = X^T\bar{a}_j$ dan $\bar{\eta}^j = Y^T\bar{b}_j$. Pasangan $\left(\bar{a}_j, \bar{b}_j\right)$ dimana $\left\|\bar{a}_j\right\|_{V_{11}}$ = 1 dan $\left\|\bar{b}_j\right\|_{V_{22}}$ disebut pasangan faktor kanonik ke-j.

b) $D_p\left(\bar{\tau}^j, \bar{\eta}^j\right) = \mathrm{Cov}\left(\bar{\tau}^j, \bar{\eta}^j\right) = \sqrt{\lambda_j}$.

c) Karena $\bar{\tau}^j$ dan $\bar{\eta}^j$ juga terbakukan, maka $D_p\left(\bar{\tau}^j, \bar{\eta}^j\right) = r\left(\bar{\tau}^j, \bar{\eta}^j\right) = \sqrt{\lambda_j}$

$r\left(\bar{\tau}^j, \bar{\eta}^j\right)$ disebut koefisien korelasi kanonik ke-j, untuk setiap j =1, 2, …, k.

PENYAJIAN DATA

Analisis kanonik memungkinkan dilaksanakan penyajian variabel-variabel pada kedua kelompok secara global untuk kemudian mempelajari kemiripan variabel-variabel dalam kelompok maupun antar kelompok. Selain itu analisis kanonik memungkinkan untuk menyajikan individu-individu baik di $E_1 = R^p$ maupun di $E_2 = R^q$ (Djauhari; 1988).

PENYAJIAN VARIABEL

Penyajian variabel dapat dilakukan baik di W_1 dan W_2.

Penyajian variabel di W_1

Untuk memudahkan pembacaan hasil penyajian variabel, semua variabel dibakukan. Sebagai contoh penyajian variabel pada bidang, kita pandang bidang P yang dibangun oleh $\overline{\tau}^1$ dan $\overline{\tau}^2$. P adalah bidang di W_1 yang paling "dekat dengan W_2. Pada bidang inilah pendeteksian kemiripan antar variabel dapat dilakukan dengan hasil yang paling baik.

Misalkan $\overline{\alpha}^j$ adalah proyeksi $\overline{x}^j$ pada P, dan $\overline{\alpha}^j = r(\overline{x}^j, \overline{\tau}^1)\overline{\tau}^1 + r(\overline{x}^j, \overline{\tau}^2)\overline{\tau}^2$; untuk setiap j = 1, 2,…,p dan $\overline{\beta}^k$ adalah proyeksi $\overline{y}^k$ pada P, dan $\overline{\beta}^k = r(\overline{y}^k, \overline{\tau}^1)\overline{\tau}^1 + r(\overline{y}^k, \overline{\tau}^2)\overline{\tau}^2$; untuk setiap k = 1, 2,…, p.

Karena untuk penyajian ini semua variabel telah dibakukan, pada bidang P kita buat lingkaran L_1 berpusat di O dan berjari-jari 1. Lingkaran ini disebut lingkaran korelasi. Vektor $\overline{\tau}^1$ dan $\overline{\tau}^2$ diletakkan orthogonal, maka :

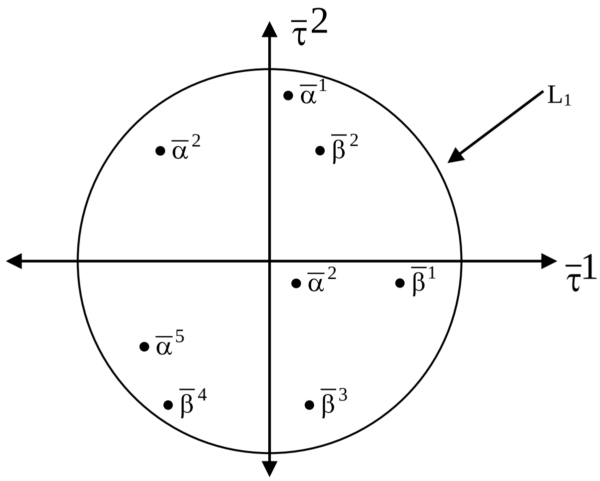
a) Variabel $\overline{x}^j$ atau $\overline{y}^k$ yang merupakan kombinasi linier dari $\overline{\tau}^1$ dan $\overline{\tau}^2$ akan terletak pada L_1. Dalam hal ini : $\|\overline{\alpha}^j\|^2_{Dp} = \|\overline{x}^j\|^2_{Dp} = 1$ dan $\|\overline{\beta}^k\|^2_{Dp} = \|\overline{y}^k\|^2_{Dp} = 1$.

b) Variabel $\overline{x}^j$ atau $\overline{y}^k$ yang tidak berkorelasi dengan $\overline{\tau}^1$ dan $\overline{\tau}^2$ akan terletak di titik pusat O dari L_1 .

Selanjutnya pengkajian variabel dilakukan pada lingkaran korelasi.

Contoh :

Misalkan kita mempunyai penyajian dua kelompok variabel $\{\overline{x}^1, \overline{x}^2,..., \overline{x}^5\}$ dan $\{\overline{y}^1, \overline{y}^2,..., \overline{y}^4\}$ sebagai berikut



Pada lingkaran korelasi itu dapat dibaca antara lain :

a) $\overline{x}^1$ atau $\overline{y}^2$ sangat berdekatan $r\left(\overline{x}^1, \overline{y}^2\right) \approx 1$.

b) $\overline{y}^2$ dan $\overline{y}^3$ saling orthogonal $r\left(\overline{x}^2, \overline{y}^3\right) \approx 0$

c) $\overline{x}^4$ orthogonal/tak berkorelasi dengan $\overline{\tau}^1$ dan $\overline{\tau}^2$. Juga dengan semua variabel yang dekat ke L_1.

d) walaupun $\overline{\alpha}^5$ dan $\overline{\beta}^4$ berdekatan, kita tak dapat menyatakan bahwa $\overline{x}^5$ dan $\overline{y}^4$ mirip satu sama lain, sebab mereka jauh dari L_1.

Penyajian variabel di W_2

Seperti pada bidang P, kita dapat membuat lingkaran korelasi L_2 pada bidang Q. Pada lingkaran ini variabel $\overline{x}^j$ disajikan oleh $\overline{\gamma}^j$; j=1,2,…,p dan variabel $\overline{y}^k$ disajikan oleh $\overline{\delta}^k$. k=1,2,…,q. $\overline{\gamma}^j$ adalah proyeksi $\overline{x}^j$ pada Q, dan $\overline{\gamma}^j = r\left(\overline{x}^j, \overline{\eta}^1\right)\overline{\eta}^1 + r\left(\overline{x}^j, \overline{\eta}^2\right)\overline{\eta}^2$; untuk setiap j = 1, 2,…,p dan $\overline{\delta}^k$ adalah proyeksi $\overline{y}^k$ pada Q, dan $\overline{\delta}^k = r\left(\overline{y}^k, \overline{\eta}^1\right)\overline{\eta}^1 + r\left(\overline{y}^k, \overline{\eta}^2\right)\overline{\eta}^2$; untuk setiap k = 1, 2,…, p. Selanjutnya pembacaan sama seperti pada L_1.

Catatan : Jika $\lambda_1 = \lambda_2 = 1$, maka P = Q. Dengan kata lain penyajian pada L_1 dan pada L_2 sama.

Untuk penyajian data variabel maupun individu dapat dibuat algoritma untuk mendapatkan koordinat proyeksi dua kelompok variabel tersebut. Sebagai contoh diberikan algoritma untuk mendapatkan proyeksi dua kelompok variabel pada bidang P berikut ini.

Algoritma

1. Diberikan $D_p = \begin{pmatrix} \frac{1}{n} & & O \\ & \frac{1}{n} & \\ O & & \frac{1}{n} \end{pmatrix}$ matriks bobot berukuran nxn, X_1 matriks berukuran pxn dan Y_1 matriks berukuran nxp. Bakukan matriks X_1 dan Y_1 dengan cara mengurangi matriks X_1 dan Y_1 dengan rataan baris masing-masing matriks; $X = X_1 - \overline{g}$ dengan elemen ke-j dari $\overline{g}$ adalah $\frac{1}{n}\sum_{i=1}^{n} x_i^j$.

2. Hitung V_{22} , V_{22}^{-1} , V_{21}, V_{11}, V_{22}^{-1} dan V_{12} .

3. Jika q < p,maka cari terlebih dahulu nilai-nilai eigen dari $V_{22}^{-1}\, V_{21}\, V_{11}^{-1}\, V_{12}$.

   Sebaliknya jika q > p cari terlebih dahulu nilai-nilai eigen dari $V_{11}^{-1}\, V_{12}\, V_{22}^{-1}\, V_{21}$.

Karena q = 2 < p = 3 maka terlebih dahulu akan dicari nilai-nilai eigen dari matriks $V_{22}^{-1}\ V_{21}\ V_{11}^{-1}\ V_{12}$ yang memenuhi persamaan :

det($V_{22}^{-1}\ V_{21}\ V_{11}^{-1}\ V_{12}\ -\ \lambda_j\ .\ I_2$) = 0 , untuk setiap j = 1,2.

4. Tentukan $\bar{b}_j$ vektor eigen yang memenuhi kondisi $\left\|\bar{b}_j\right\|_{V_{22}} = 1$ atau $\bar{b}_j^{\,T}\, V_{22}\ \bar{b}_j = 1$

5. Tentukan $\bar{\tau}^j = \dfrac{1}{\sqrt{\lambda_j}} X^T\ V_{11}^{-1}\ V_{12}\ \bar{b}_j$, untuk setiap j =1,2.

6. Tentukan $\bar{\alpha}^j = r\left(\bar{x}^j, \bar{\tau}^1\right)\bar{\tau}^1 + r\left(\bar{x}^j, \bar{\tau}^2\right)\bar{\tau}^2$ atau $\bar{\alpha}^j = \begin{pmatrix} r\left(\bar{x}^j, \bar{\tau}^1\right) \\ r\left(\bar{x}^j, \bar{\tau}^2\right) \end{pmatrix}$, untuk j = 1, 2, 3.

7. Tentukan $\bar{\beta}^k = r\left(\bar{y}^k, \bar{\tau}^1\right)\bar{\tau}^1 + r\left(\bar{y}^k, \bar{\tau}^2\right)\bar{\tau}^2$ atau $\bar{\beta}^k = \begin{pmatrix} r\left(\bar{y}^k, \bar{\tau}^1\right) \\ r\left(\bar{y}^k, \bar{\tau}^2\right) \end{pmatrix}$, untuk k = 1, 2.

8. Plot $\bar{\alpha}^j$ dan $\bar{\beta}^k$ pada bidang P, untuk setiap j = 1,2,3, dan k = 1,2.

Pada langkah kelima, untuk penyajian data variabel di bidang Q penghitungan $\bar{\tau}^j$ diganti dengan $\bar{\eta}^j$, untuk j = 1,2. Langkah ketujuh dan kedelapan diganti dengan penghitungan $\bar{\gamma}^j$ dan $\bar{\delta}^k$. Dengan cara yang sama dapat dibuat algoritma untuk penyajian data individu baik di E_1 maupun di E_2.

Untuk data multivariat yang banyaknya variabel lebih dari tiga (p > 3 dan q > 3), diperlukan program khusus. Kendala dalam pembuatan program ini adalah : pada langkah keempat dan kelima menentukan solusi berupa nilai eigen dan vektor eigen yang memenuhi syarat keortonormalan yang dimaksudkan tidaklah mudah. Penentuan solusinya melibatkan teknik penentuan solusi dua sistem persamaan linier yang argumennya lebih dari tiga. Untuk kasus seperti ini diperlukan studi yang lebih lanjut.

Selain Analisis Kanonik, untuk melihat kemiripan antar variabel-variabel kuantitatif dalam kelompok dapat dilakukan teknik Analisis Komponen Utama. Analisis Komponen Utama akan mereduksi ruang individu ($E = R^p$) yang tadinya p menjadi k < p, kalau mungkin k = 2 atau 3. Jadi kita cukup menganalisis data pada ruang yang berdimensi 2 atau 3 untuk mempelajari kemiripan variabel-variabel maupun individu (Jackson; 1995 & Johnson; 1982).

Materi yang dibahas pada tulisan hanya membahas tentang Analisis Kanonik yang mempelajari dua kelompok variabel kuantitatif untuk individu yang sama. Jika individu-individu tersebut diklasifikasikan ke dalam kelompok-kelompok (misalkan ada g kelompok) maka Analisis Kanonik yang mempelajari struktur antar kelompok tersebut adalah Analisis Variat Kanonik. Metode ini berdasarkan pada dekomposisi nilai-nilai eigen (Lihat Canonical Analysis).

DAFTAR PUSTAKA

Djauhari, M. (1988). *Struktur Data Statistik*. Jakarta : Depdikbud.

Jackson, E. J. (1995). *Principal Component Analysis*. New York :Wiley & Johnson.

Johnson, R. A. (1982). *Applied Multivariate Statistical Analysis*. Prentice-Hall. Inc.

Canonical Analysis. [Online]. Tersedia :

http://www.nag.co.uk/numeric/FN/manual/pdf/c28/c28int_fn03.pdf

http://www.nag.co.uk/numeric/FN/manual/pdf/c28/c28m02_canon_analysis_fn03.pdf