ANALISIS KORESPONDENSI UNTUK PEMETAAN PERSEPSI

Agus Rusgiyono[1]
[1]Staf Pengajar Program Studi Statistika FMIPA UNDIP

Abstract

Correspondence analysis used to investigate the relationship between two or more qualitative variables. This technique could shrink the dimensions of variables and describe the profile vector of rows and columns of a matrix vector data from the contingency table. Target correspondence analysis is to show the relationship variables rows and columns as well as visualization variables in R2-dimensional space, using the Chi square of the distance definition in sub-Euclidean space.

Keywords: Profile of Row and Column Vectors, Chi Square Distance, Euclidean Subset

1. Pendahuluan

Ketersediaan barang atau suatu merk produk sejenis di pasaran menyebabkan konsumen mempunyai kesempatan untuk membandingkan produk yang satu dengan yang lainnya. Hal ini menyebabkan produsen memerlukan peta persepsi konsumen terhadap produknya untuk mengetahui posisi di tengah persaingan sehingga dapat disusun strategi pemasaran yang tepat. Untuk keperluan pembuatan peta persepsi ini kemiripan antar produk beserta atribut yang menjadi focus penelitian diterjemahkan dalam pengertian jarak antara dua titik pada bidang atau ruang. Sehingga diperlukan transformasi dari persepsi konsumen ke dalam bentuk vector dan matrik.

Misalnya X dan Y adalah peubah kategorik dengan masing-masing peubah mempunyai a dan b kategori. Hasil pengamatan disajikan dalam tabel kontingensi a x b dengan $n_{ij} \geq 0$ menyatakan frekuensi dari sel ke (i, j). Matriks dari frekuensi relatif dinyatakan sebagai berikut

$$M_{axb} = [m_{ij}] = \left[\frac{n_{ij}}{n}\right]$$

dengan $n = \sum_i \sum_j n_{ij}$

M_{axb} disebut sebagai matriks korespondensi, untuk $i = 1,2,3,\ldots,a$ dan $j = 1,2,3,\ldots,b$

Misal vektor jumlahan baris dari matriks M_{axb} adalah

$$r = M_{+j} = (m_{1+}, \cdots, m_{a+})' = \left(\frac{n_{1+}}{n}, \ldots, \frac{n_{a+}}{n}\right)'$$

dan vektor jumlahan kolom dari matriks M_{axb} adalah

$$c = M'1 = (m_{+1}, \cdots, m_{+b})' = \left(\frac{n_{+1}}{n}, \ldots, \frac{n_{+a}}{n}\right)'$$

dengan **1** = (1,…,1)' adalah sebuah vektor satuan, suatu vektor yang semua unsurnya bernilai 1, dan $n_i = \sum_{j=1}^{b} n_{ij}, i = 1, 2, \ldots, a, n_j = \sum_{j=1}^{a} n_{ij}, j = 1, 2, \ldots, b.$

Misal $D_r = \begin{bmatrix} m_{1+} & 0 & . & . & 0 \\ 0 & m_{2+} & . & & 0. \\ . & . & . & . & m_{i+} & 0 \\ 0 & & & 0 & m_{a+} \end{bmatrix}$ dan $D_c = \begin{bmatrix} m_{+1} & 0 & . & . & 0 \\ 0 & m_{+2} & . & & 0. \\ . & . & . & . & m_{+j} & 0 \\ 0 & & & 0 & m_{+b} \end{bmatrix}$

dengan D_r = diag(r) dan D_c = diag (c) merupakan sebuah matriks diagonal yang masing-masing berukuran a x a dan b x b[1].

Selanjutnya R didefinisikan sebagai :

$$R = D_r^{-1} M = \begin{bmatrix} \frac{m_{11}}{m_{1+}} & \cdots\cdots & \frac{m_{1b}}{m_{1+}} \\ \vdots & \ddots & \vdots \\ \frac{m_{a1}}{m_{a+}} & \cdots & \frac{m_{ab}}{m_{a+}} \\ . \end{bmatrix}$$

maka a baris dari matriks R $_{a \times b}$, disebut profil baris-profil baris (*row profiles*) dalam ruang berdimensi b. Jumlah dari unsur-unsur profil baris (*row profiles)* adalah 1.

Misalkan didefinisikan profil baris ke-*i* sebagai r_i', dengan $r_i = \left(\frac{m_{i1}}{m_{i+}}, \frac{m_{i2}}{m_{i+}} \ldots, \frac{m_{ib}}{m_{i+}} \right)'$

Bila C didefinisikan sebagai :

$$C = D_c^{-1} M' = \begin{bmatrix} \frac{m_{11}}{m_{+1}} & \cdots\cdots & \frac{m_{1b}}{m_{+1}} \\ \vdots & \ddots & \vdots \\ \frac{m_{1b}}{m_{+b}} & \cdots & \frac{m_{ab}}{m_{+b}} \\ . \end{bmatrix}$$

maka b baris dari matriks C $_{b \times a}$, disebut profil kolom-profil kolom (*column profiles*) dalam ruang berdimensi a.

Sebagai catatan jumlah unsur-unsur dari profil kolom (*column profiles*) adalah sama dengan 1. Jika didefinisikan profil kolom ke-j dengan c_j', maka,

$$c_j = \left(\frac{m_{1j}}{m_{+j}}, \frac{m_{2j}}{m_{+j}} \ldots, \frac{m_{aj}}{m_{+j}} \right)'$$

Seperti pada kasus profil baris, jumlah unsur-unsur pada masing-masing profil kolom, sesuai harapan adalah 1.

Vektor $c = (m_{+1}, \cdots, m_{+b})'$ disebut sebagai rataan profil baris atau pusat baris atau vektor dari massa baris. Sedangkan vektor $r = (m_{1+}, \cdots, m_{a+})'$ disebut sebagai rataan profil kolom atau pusat kolom atau vektor dari massa baris. Rataan profil baris dan rataan profil kolom ini merupakan rata-rata pembobot atau dengan kata lain rataan profil baris dan rataan profil, kolom merupakan rata-rata pembobot dari profil kolom dan profil baris. Lebih khusus, rataan profil baris adalah $c = \sum_{i=1}^{n} m_{i+} r_i^{'}$ dengan r_i' adalah profil baris ke-*i*. sedangkan rataan profil kolom adalah $r = \sum_{i=1}^{b} m_{+j} c_{j+}^{'}$ [2].

Dalam analisis korespondensi, a baris dari matriks yang dibentuk dari dua kolom pertama F dan b baris dari matriks yang dibentuk dari dua kolom pertama G secara umum ditampilkan dalam satu grafik. Dalam plot, jarak antar titik berhubungan dengan profil-profil baris atau antar titik berhubungan dengan profil-profil kolom merupakan pendekatan terhadap jarak *chi square* antar masing-masing profil yang mencerminkan kemiripan antar produk.

2. Rumusan Masalah

1. Bagaimana cara menampilkan profil baris-profil baris tersebut ke dalam ruang dimensi *euclid* yang lebih rendah. Cara yang sama digunakan juga untuk menampilkan profil kolom-profil kolom dalam ruang dimensi *euclid* yang lebih rendah.
2. Pendekatan jarak apa yang digunakan?

3. Bahan Dan Metode

Statistik uji *chi square* dari Pearson untuk menguji kebebasan antara X dan Y adalah[3]:

$$\chi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2}{\frac{n_{i+}n_{+j}}{n}} = n \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{\left(m_{ij} - m_{i+}m_{+j}\right)^2}{m_{i+}m_{+j}} = n\, tr(E) = n \sum_{i=1}^{p} \lambda_i^2$$

dengan $E = D_r^{-1}(M - rc)D_c^{-1}(M - rc)$

$\lambda_1^2 \geq \cdots \geq \lambda_p^2$ adalah akar karakteristik tak nol dari E

p = rank (E) = rank (M - rc') = rank (M) = min (a, b) - 1

$$\chi^2 = n \sum_i m_{i+} \left[\sum_j \frac{\left(\frac{m_{ij}}{m_{i+}} - m_{+j}\right)^2}{m_{+j}} \right] = \sum_i n m_{i+} [(r_i - c)' D_c^{-1}((r_i - c)] = n \sum_i m_i d_i^2$$

dengan $d_i^2 = (r_i - c)' D_c^{-1}(r_i - c)$

Besaran d_i^2 merepresentasikan kuadrat jarak antara profil baris ke-i dan rata-rata profil baris. Jarak ini disebut jarak *chi square*. Kenyataannya, d_i^2 mirip dengan jarak *euclid* $(r_i - c)'((r_i - c)$ antara vektor $r_i\ dan\ c,$ kecuali jarak *euclid* diboboti dengan unsur–unsur vektor c, rataan profil baris.

Besaran χ^2 / n merupakan *total inertia.* Sedangkan $n \sum_i m_i d_i^2$ menunjukan total inersia yang dinyatakan sebagai rata–rata terboboti dari kuadrat jarak *chi square* antara profil baris dengan rata–ratanya. Kuadrat Jarak *chi square* antara dua profil , misalkan $r_i\ dan\ r_j$ adalah

$$d_{ij}^2 = (r_i - r_j)' D_c^{-1}(r_i - r_j)$$

Hal tersebut, serupa dengan jarak kuadrat *euclid* $(r_i - r_j)'(r_i - r_j)$ antara dua vektor $r_i\ dan\ r_j$ kecuali rataan profil digunakan sebagai pembobot. Jarak *chi square* antara sebuah profil kolom dengan rataanya dan antara dua kolom profil mempunyai definisi yang sama.

Langkah selanjutnya adalah menentukan dua atau tiga subruang *euclid* dan memproyeksikan semua profil baris kedalam subruang *euclid* tersebut. Untuk mencari subruang *euclid* digunakan *generalized singular value decomposition* (GSVD). GSVD dari matriks (M - rc') adalah

$$(M - rc') = A\Lambda B'$$

dengan A adalah matriks berukuran a x p

B merupakan matriks berukuran b x p di mana berlaku A' $D_r^{-1}A = I_p$ dan B'$D_c^{-1}B = I_p$

Λ merupakan matriks diagonal yang mempunyai unsur–unsur diagonal nilai singular $\lambda_1, \ldots, \lambda_p$ dari (M - rc').

Matriks A dan B diperoleh dari penguraian nilai singular (*singular value decomposition)* dari $T = D_r^{-1/2}(M - rc)D_c^{-1/2}$

Sebagai catatan $\lambda_1^2 \geq \cdots \geq \lambda_p^2$ adalah akar karakteristik dari TT' sama dengan akar karakteristik dari E .

Misalkan akan direpresentasikan profil–profil baris dan profil–profil kolom ke dalam ruang berdimensi k (dengan k $\leq p$). Biasanya nilai k diambil 2 atau 3. Koordinat dari a profil baris adalah a buah baris dari matriks yang dibentuk dengan mengambil k kolom pertama dari F = $D_r^{-1}A\Lambda$. Dan koordinat dari b profil kolom adalah b buah baris dari matriks yang dibentuk dengan mengambil k kolom pertama dari G = $D_c^{-1}B\Lambda$. Karena total inersia adalah $\frac{\chi^2}{n} = tr(E) = \sum_{i=1}^{p} \lambda_i^2$ maka pendekatan ruang berdimensi p dengan ruang berdimensi k adalah bagus jika $\sum_{i=1}^{k} \lambda_i^2$ mendekati total inersia $\sum_{i=1}^{p} \lambda_i^2$, atau alternatifnya jika $\sum_{i=1}^{p} \lambda_i^2$ mendekati 0 . Besaran $\lambda_1^2, \lambda_2^2,$ dan seterusnya bisa diinterpretasikan sebagai besarnya kontribusi yang diberikan kepada total inersia oleh masing–masing dimensi pertama, dimensi kedua dan sebagainya.

Dalam analisis korespondensi, a baris dari matriks yang dibentuk dari dua kolom pertama F dan b baris matriks yang dibentuk dari dua kolom pertama G secara umum ditampilkan dalam satu grafik .Plot semacam ini disebut *symmetric plot* dari titik-titik yang berhubungan dengan profil-profil baris dan profil-profil kolom. Dalam plot, jarak antar titik berhubungan dengan profil-profil baris atau antar titik berhubungan dengan profil-profil kolom merupakan pendekatan terhadap jarak *chi square* antar masing-masing profil. Tidak ada interpretasi yang mengindikasikan antara dua titik, satu merupakan profil baris sedangkan yang lainnya merupakan profil kolom. Oleh sebab itu, hanya jarak antar titik yang berhubungan baik dengan dua baris atau dua kolom[3].

4. Hasil Dan Pembahasan

Penelitian di bidang sosial politik berikut ini ingin melihat hubungan antara wilayah tempat tinggal penduduk di Kota Semarang dengan pandangan mereka tentang calon walikota, serta ingin diketahui deskripsi pandangan calon walikota di Semarang dalam kaitannya dengan wilayah. Misalkan calon walikota dikategorikan dalam 4 calon , yaitu A,B,C,D. Sedangkan wilayah tempat tinggal dikategorikan menjadi 5 bagian, yaitu Semarang Utara,Tengah, Selatan, Timur dan Barat.

Disini, banyaknya kategori dari calon walikota adalah 4 dan banyaknya kategori Wilayah Tempat Tinggal adalah 5, sehingga pangkat dari matriks M = *min*(5,4)-1 = 3. Berarti tiga dimensi yang terbentuk nantinya mampu mempresentasikan data tanpa kehilangan informasi. Selanjutnya dilakukan pengolahan data dengan software SPSS.

a. Input data

Tabel 1. Hasil Survey Kepeminatan Responden Terhadap Calon Walikota

Correspondence Table

Wilayah	**Calon Walikota**				
	A	**B**	**C**	**D**	**Active Margin**
Barat	46	21	12	13	92
Utara	31	37	17	15	100
Timur	0	22	60	18	100
Selatan	42	19	18	21	100
Tengah	12	21	48	19	100
Active Margin	131	120	155	86	492

b. Prosedur Analisis

```
CORRESPONDENCE
TABLE=WILAYAH(1  5)   BY  CALONWALIKOTA (1   4)
/DIMENSIONS = 2
/MEASURE = CHISQ
/STANDARDIZE = RCMEAN
/NORMALIZATION = SYMMETRICAL
/PRINT = TABLE RPOINTS  RPROFILES
/PLOT = NDIM(1,MAX)  BIPLOT(20) RPOINTS(20) CPOINTS(20) TRROWS(20)
 TRCOLUMNS(20)
```

c. Output nya sebagai berikut

Tabel 2. Profile Baris

Row Profiles

Wilayah	**Calon Walikota**				
	A	**B**	**C**	**D**	**Active Margin**
Barat	.500	.228	.130	.141	1.000
Utara	.310	.370	.170	.150	1.000
Timur	.000	.220	.600	.180	1.000
Selatan	.420	.190	.180	.210	1.000
Tengah	.120	.210	.480	.190	1.000
Mass	.266	.244	.315	.175	

Tabel 3. Ringkasan Hasil

Summary

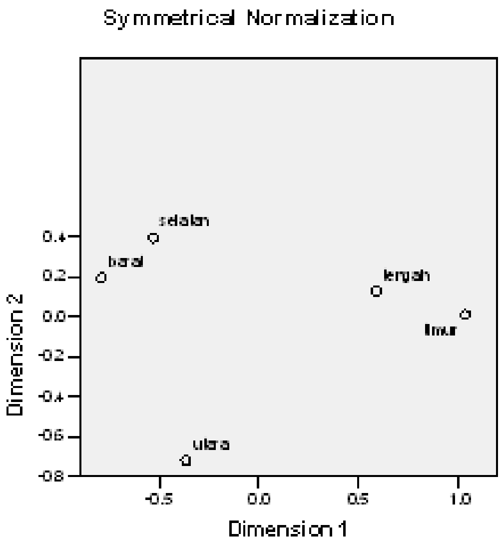
Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation
								2
1	.490	.240			.908	.908	.033	.079
2	.148	.022			.082	.990	.049	
3	.051	.003			.010	1.000		
Total		.265	130.271	.000[a]	1.000	1.000		

Intepretasi dari singular value yang merupakan akar kuadrat dari eigenvalue antar kategori dari variabel dalam analisis untuk setiap dimensi adalah 0,490 untuk dimensi pertama (terbesar), 0,148 untuk dimensi kedua (juga merupakan yang kedua terbesar) dan 0,051 dimensi ketiga. Dari yang diperoleh dalam analisis (proportion of inertia), dapat dinyatakan bahwa keragaman yang dapat diterangkan adalah sebesar 100% dengan rincian sebagai berikut:

1. Faktor pertama dengan eigen value sebesar 0,036 mampu menerangkan keragaman data sebesar 90.8%
2. Faktor kedua dengan eigen value sebesar 0,015 mampu menerangkan keragaman data sebesar 8.2% (total dengan figure pertama adalah 99%)
3. Faktor ketiga dengan eigen value sebesar 0,006 mampu menerangkan keragaman data sebesar 1% (total dengan sebelumnya menjadi 100%)

   Bila dilihat standar deviasinya terlihat bahwa untuk dimensi pertama ditemukan nilai sebesar 0.033. nilai ini juga menunjukkan presisi yang lebih baik dari dimensi kedua yang memiliki nilai standar deviasi sebesar 0.049.

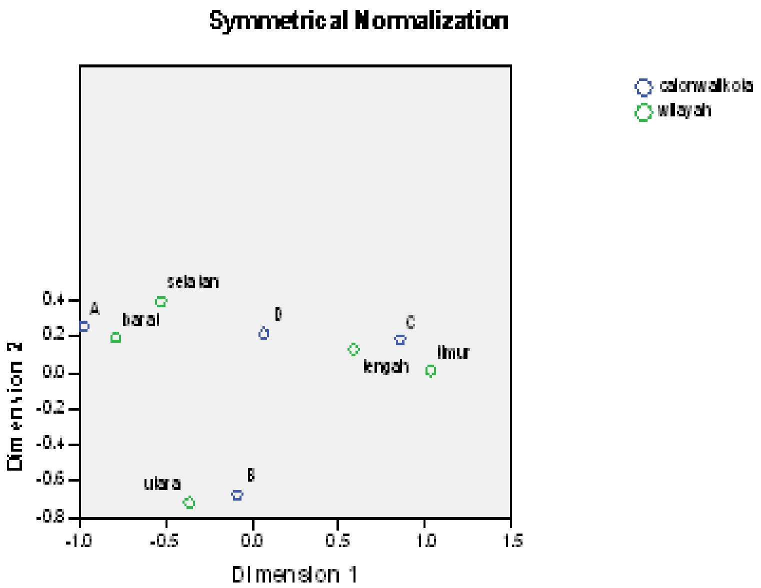
Total inersia diperoleh dari $\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 0,24^2 + 0,022^2 + 0,003^2 = 0,265$.



Gambar 1. Peta Kemiripan Antar Wilayah

Jika figure *Row Points* dan *Coloumn Points* digabungkan maka akan didapat figure akhir yang memperlihatkan pemetaan karakteristik pandangan politik penduduk pada keempat wilayah yang ada. Figur yang dimaksud ditampilkan pada Gambar 2.

Pada Gambar 2 tersebut dapat dilihat bahwa karakteristik pandangan calon walikota, penduduk Semarang yang bermukim di wilayah Utara pada umumnya adalah calon B, sementara yang bermukim di wilayah Selatan dan barat pada umumnya adalah dan cenderung ke calon A. Untuk wilayah Timur dan Tengah karakteristik umum adalah cenderung ke calon C sedangkan penduduk yang bermukim di wilayah Tengah berpandangan cenderung ke calon D.



Gambar 2. Peta Calon Walikota Dan Wilayah Pendukungnya

5. Kesimpulan

Berdasarkan analisis sebelumnya dapat diambil kesimpulan sebagai berikut:

1. Secara umum, analisis korespondensi adalah sebuah teknik multivariat secara grafik yang digunakan untuk eksplorasi data dari sebuah tabel kontingensi. Analisis korespondensi memiliki masukkan berupa tabel frekuensi, dan hasil keluarannya berupa peta (*mapping*) kategori dari variabel.
2. Analisis korespondensi dari contoh menunjukkan adanya perbedaan pandangan politik pada penduduk yang tinggal di wilayah tertentu. Hal ini terbukti dengan hasil pemetaan yang memperlihatkan bahwa pada Gambar 2, dapat dilihat bahwa karakteristik pandangan calon walikota, penduduk Semarang yang bermukim di wilayah Utara pada umumnya adalah calon B, sementara yang bermukim di wilayah Selatan dan barat pada umumnya adalah dan cenderung ke calon A. Untuk wilayah Timur dan Tengah karakteristik umum adalah cenderung ke calon C sedangkan penduduk yang bermukim di wilayah Tengah berpandangan cenderung ke calon D.

DAFTAR PUSTAKA

1. Lebart L., Moreneau A., and Warwick M., *Multivariate Descriptive Statistic Analysis Correspondence Analysis and Related Tecniques for Large Matrices*, John Wiley and Sons Inc, 1984.
2. Michael, J.G., *Theory and Aplications of Correspondence Analysis*, Academic Press, Inc , 1984
3. Pauls, G., *Multidimensional Scalling : Concepts and Aplications*, 1989.