S1 Sains Data
FMIPA UNESA

Analisis Multivariat – Pertemuan 9

MDS - Biplot

Prodi S1 Sains Data
Universitas Negeri Surabaya
8 April 2025

MULTIDIMENSIONAL SCALING

Tujuan

Membuat peta/konfigurasi posisi objek dalam ruang berdimensi rendah (umumnya 2 dimensi) berdasarkan data jarak antar objek atau data multivariate yang sebelumnya diubah dulu menjadi matriks jarak

Kegunaan Analisis

- Mendapatkan posisi relatif suatu objek dibandingkan objek lain. Dalam banyak kasus strategi bisnis, digunakan untuk menentukan pesaing dan benchmarking.
- Melakukan pengelompokan objek, salah satu alternatif untuk cluster analisis

MDS

Informasi ordinal (peringkat)	Informasi dari jarak
Non-metric	Metric

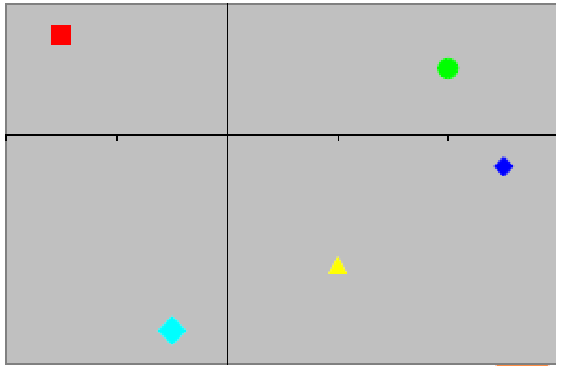
Sekilas MDS

Matriks Jarak

	A	B	C	D	E
A	0	d_{AB}	d_{AC}	d_{AD}	d_{AE}
B	d_{BA}	0	d_{BC}	d_{BD}	d_{BE}
C	d_{CA}	d_{CB}	0	d_{CD}	d_{CE}
D	d_{DA}	d_{DB}	d_{DC}	0	d_{DE}
E	d_{EA}	d_{EB}	d_{EC}	d_{ED}	0

Data multivariate

	X_1	X_2	...	X_p
A	X_{11}	X_{12}	...	X_{1p}
B	X_{21}	X_{22}	...	X_{2p}
C	X_{31}	X_{32}	...	X_{3p}
D	X_{41}	X_{42}	...	X_{4p}
E	X_{51}	X_{52}	...	X_{5p}



Konfigurasi Obyek

Perhitungan Jarak

JARAK EUCLIDEAN

Adalah jarak antara dua objek yang dibandingkan.

Jika dimisalkan objek 1 adalah $x' = (x_1, x_2, \ldots, x_p)'$ dan objek 2 adalah $y' = (y_1, y_2, \ldots, y_p)'$

Maka **jarak Euclidean-nya** adalah:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_p - y_p)^2}$$

$$Dn = \begin{bmatrix} d_{11} . d_{12} \ldots d_{1n} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ d_{n1} . d_{n2} \ldots d_{nn} \end{bmatrix}$$

Dimana : x_i= objek ke-1 pada pengamatan ke-i

y_i= objek ke-2 pada pengamatan ke-i

p= banyaknya pengamatan

Atau dalam notasi matrik, rumus jarak Euclidean-nya menjadi:

$$d(x, y) = \sqrt{(x - y)'(x - y)}$$

Langkah-Langkah MDS Metrik

1. start with distances d_{ij}
2. define $\mathcal{A} = -\frac{1}{2}d_{ij}^2$
3. put $\mathcal{B} = (a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet})$

$$a_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n} a_{ij} \quad a_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij} \quad a_{\bullet\bullet} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$$

4. find the eigenvalues $\lambda_1, \ldots, \lambda_p$ and the associated eigenvectors $\gamma_1, \ldots, \gamma_p$ where the eigenvectors are normalized so that $\gamma_i^\top \gamma_i = 1$.
5. Choose an appropriate number of dimensions p (ideally $p = 2$)
6. The coordinates of the n points in the Euclidean space are given by $x_{ij} = \gamma_{ij}\lambda_j^{1/2}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

Kebaikan Hasil MDS

$$\text{STRESS} = \left(\frac{\sum_{1<j} \left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i<j} d_{ij}^2} \right)^{1/2}$$

- Ide untuk menemukan tampilan dari objek-objek dalam titik pada ruang dimensi akhir sedemikian sehingga **nilai STRESS sekecil mungkin**.
- Nilai STRESS yang minimum merujuk pada kebaikan model, yakni hubungan monotonik antara kemiripan dan jarak akhir.

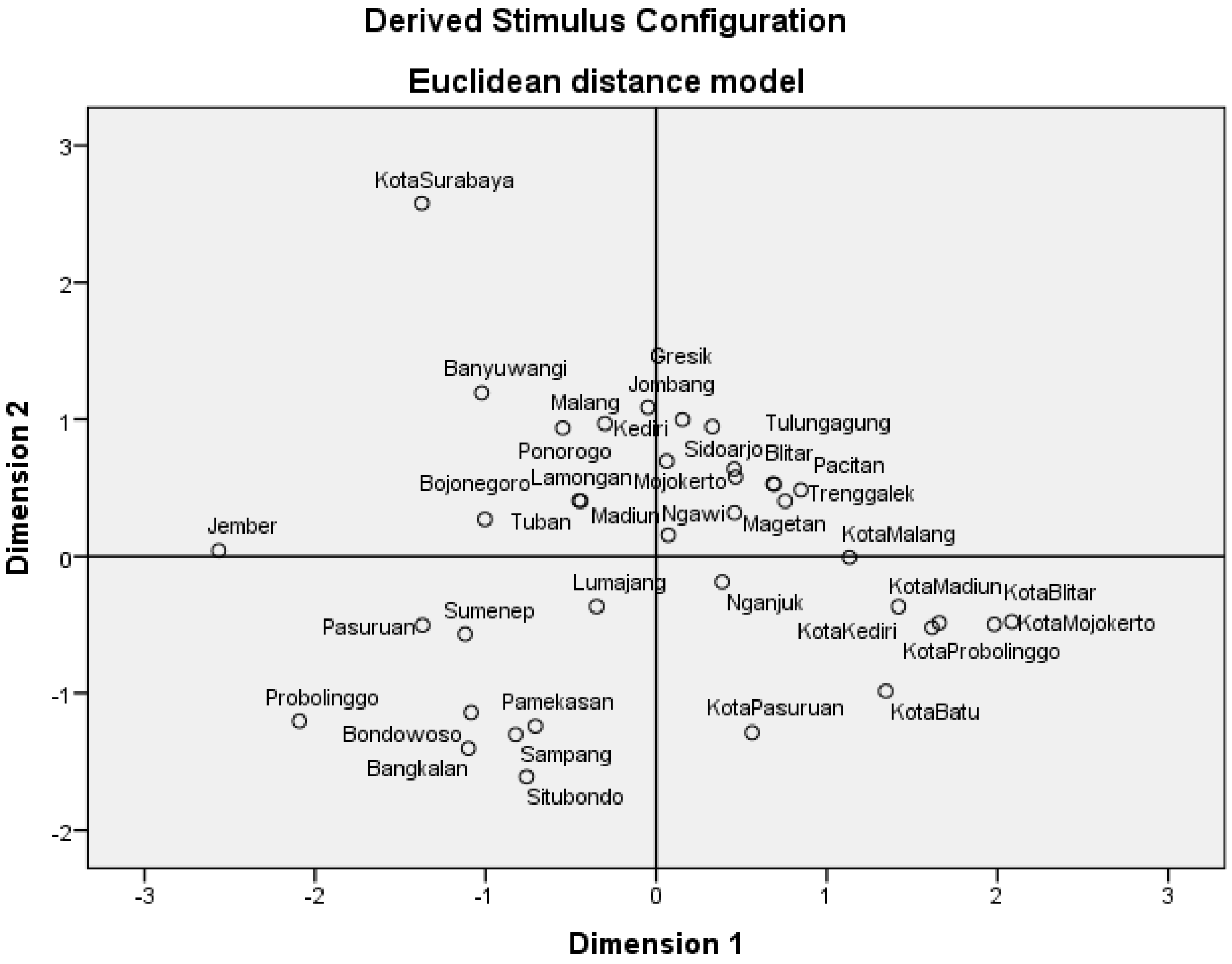
Nilai STRESS (%)	Kesesuaian
≥ 20	Buruk
10 – 20	Cukup baik
5 – 10	Baik
2.5 – 5	Sangat baik
< **2.5**	Sempurna

Contoh Data Metrik

Kabupaten/Kota	Keluhan	AKB	AHH	Jumlah Puskesmas
Pacitan	30	22.12	71.9	24
Ponorogo	26.02	25.83	70.49	31
Trenggalek	25.05	20.8	72.3	22
Tulungagung	27.23	21.4	72.09	31
Blitar	34.31	23.12	71.46	24
Kediri	27.51	26.83	70.34	37
Malang	28.17	29.46	69.69	39
Lumajang	19.07	36.92	67.93	25
Jember	23.51	55.42	63.39	49
Banyuwangi	34.7	32.56	68.66	45
Bondowoso	34.38	52.28	64.13	25
Situbondo	26.19	53.82	63.65	17
Probolinggo	25.39	62.45	61.87	33
Pasuruan	30.59	49.74	64.8	33
Sidoarjo	23.34	23.36	71.27	26
Mojokerto	33.74	23.99	70.82	27
Jombang	39.92	27.05	70.38	34
Nganjuk	25.16	30.46	69.48	20
Madiun	27.79	30.64	69.39	26

Kabupaten/Kota	Keluhan	AKB	AHH	Jumlah Puskesmas
Magetan	24.39	22.29	71.81	22
Ngawi	28.61	25.83	70.81	24
Bojonegoro	20.31	38.24	67.53	36
Tuban	25.89	32.86	68.37	33
Lamongan	28.54	33.25	68.68	33
Gresik	21.08	22.65	71.7	32
Bangkalan	19.11	53.69	63.81	22
Sampang	32.39	51.72	64.39	21
Pamekasan	20.87	49	65.05	20
Sumenep	23.7	47.48	65.25	30
Kota Kediri	31.36	23.3	71.08	9
Kota Blitar	23.9	18.71	72.99	3
Kota Malang	31.68	22.84	71.21	15
Kota Probolinggo	33.44	23.13	71.01	6
Kota Pasuruan	25.93	38.38	66.5	8
Kota Mojokerto	44.5	21.38	72.13	5
Kota Madiun	33.98	22.62	71.55	6
Kota Surabaya	27.79	22.48	71.72	62
Kota Batu	19.73	27.91	70.18	5

Output MDS



Pengelompokkan dengan MDS bukan dilihat melalui observasi tersebut masuk ke dalam kuadran mana, melainkan melalui kedekatan antar observasi.

Contoh

No	Site	Bur Oak	Black Oak	White Oak	Red Oak	American Elm	Basswood	Ironwood	Sugar Maple
1	S1	9	8	5	3	2	0	0	0
2	S2	8	9	4	4	2	0	0	0
3	S3	3	8	9	0	4	0	0	0
4	S4	5	7	9	6	5	0	0	0
5	S5	6	0	7	9	6	2	0	0
6	S6	0	0	7	8	0	7	0	5
7	S7	5	0	4	7	5	6	7	4
8	S8	0	0	6	6	0	6	4	8
9	S9	0	0	0	4	2	7	6	8
10	S10	0	0	2	3	5	6	5	9

Contoh Data Non-Metrik

	Mercedes	Jaguar	Ferrari	VW
Mercedes	–			
Jaguar	3	–		
Ferrari	2	1	–	
VW	5	4	6	–

Langkah-Langkah MDS Non Metrik

1. Menentukan koordinat awal masing-masing objek
2. Menghitung jarak Euclidean antar objek dij
3. Menentukan δij yaitu jarak actual titik objek. Pada MDS non metrik berupa peringkat.
4. Menentukan $\hat{d}_{ij}$ (disparitas), dengan algoritma PAV
5. Menghitung koordinat baru

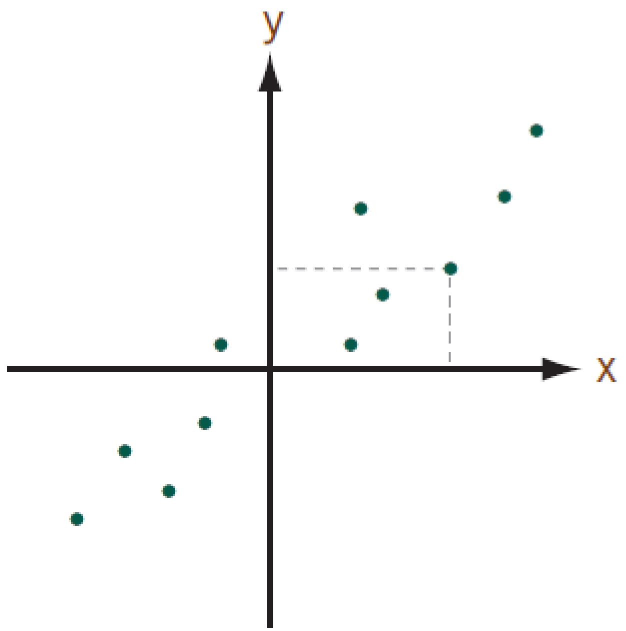
$$x_{il}^{NEW} = x_{il} + \frac{\alpha}{n-1} \sum_{j=1, j\neq i}^{n} \left(1 - \frac{\widehat{d}_{ij}}{d_{ij}}\right)(x_{jl} - x_{il}), \qquad l = 1, \ldots, p^*.$$

6. Ulangi langkah 2, hingga konvergen (perubahan konfigurasi sangat kecil atau stress minimum tercapai).

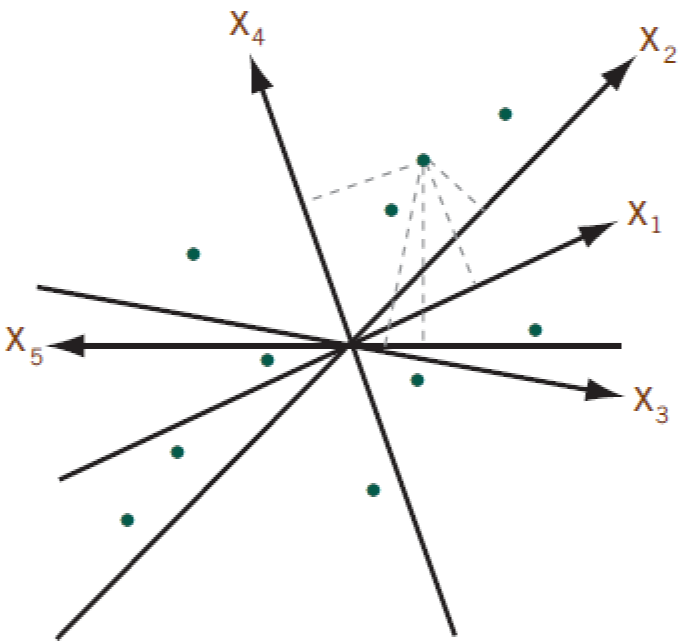
ANALISIS BIPLOT

Analisis Biplot

Analisis eksplorasi yang menyajikan data peubah ganda dalam bentuk dua dimensi.



Scatterplot



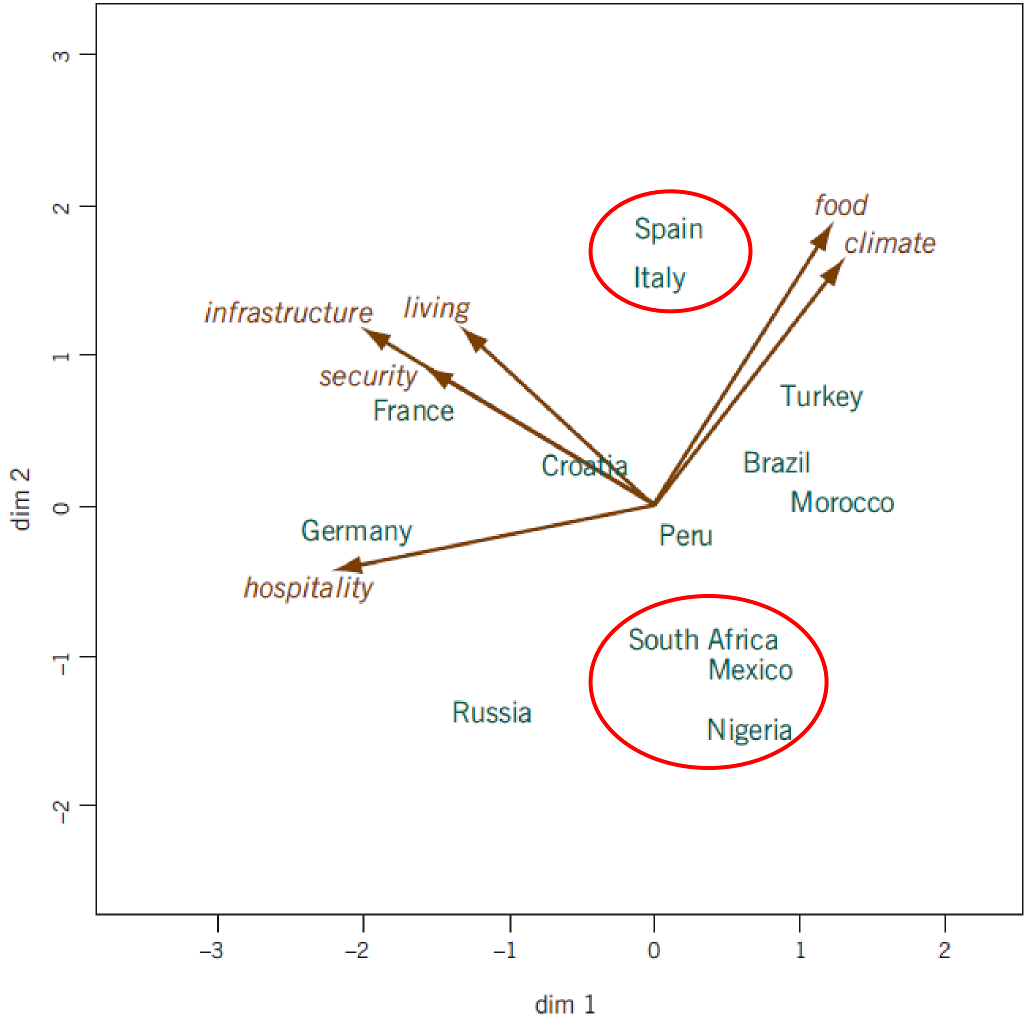
Biplot

Analisis Biplot

- Pertama kali diperkenalkan oleh Gabriel (1971) yang memungkinkan representasi visual dari hubungan antar variabel dan individu dalam dataset.
- Empat hal penting yang bisa didapatkan dari tampilan biplot, yaitu:
  1. Kedekatan antar objek yang diamati
  2. Keragaman peubah
  3. Korelasi antar peubah
  4. Nilai peubah pada suatu objek

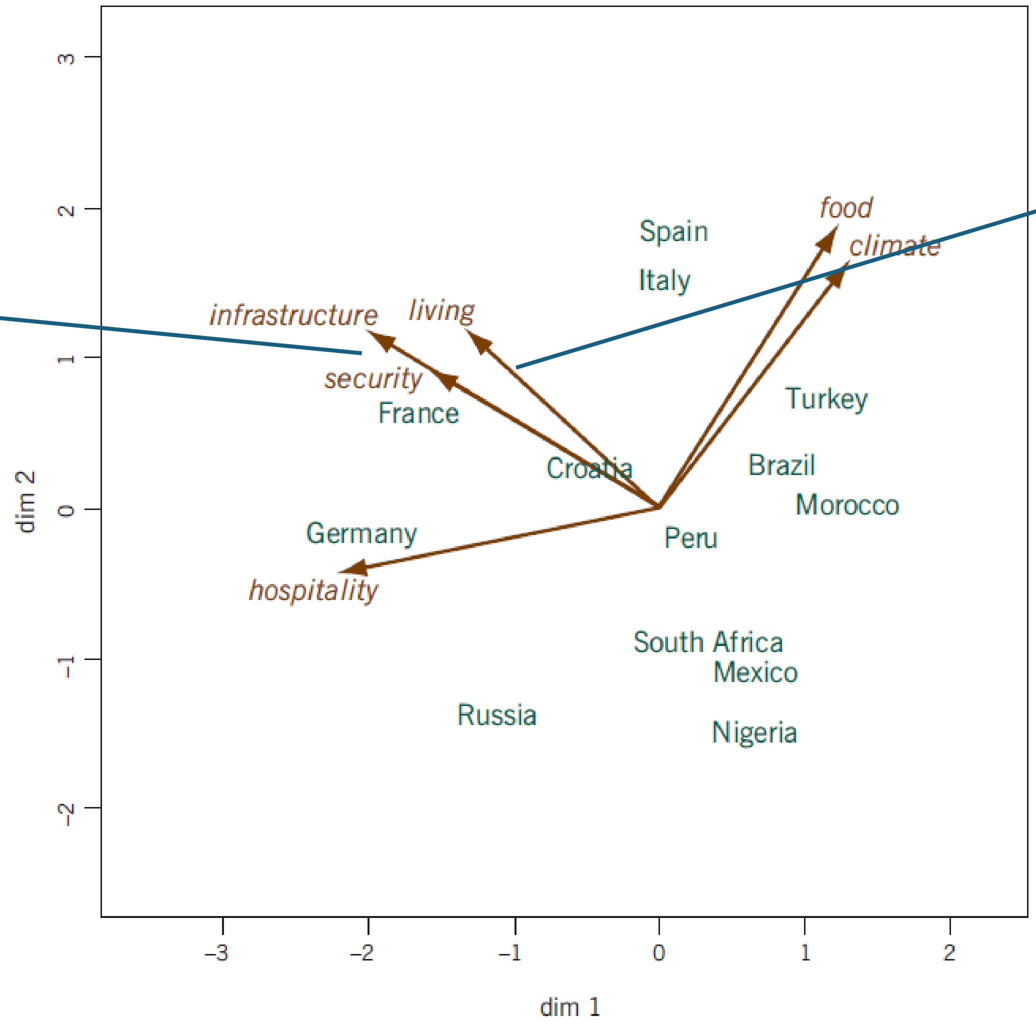
Kedekatan antar objek yang diamati

Dua objek yang memiliki karakteristik sama akan digambarkan sebagai dua titik dengan posisi yang berdekatan



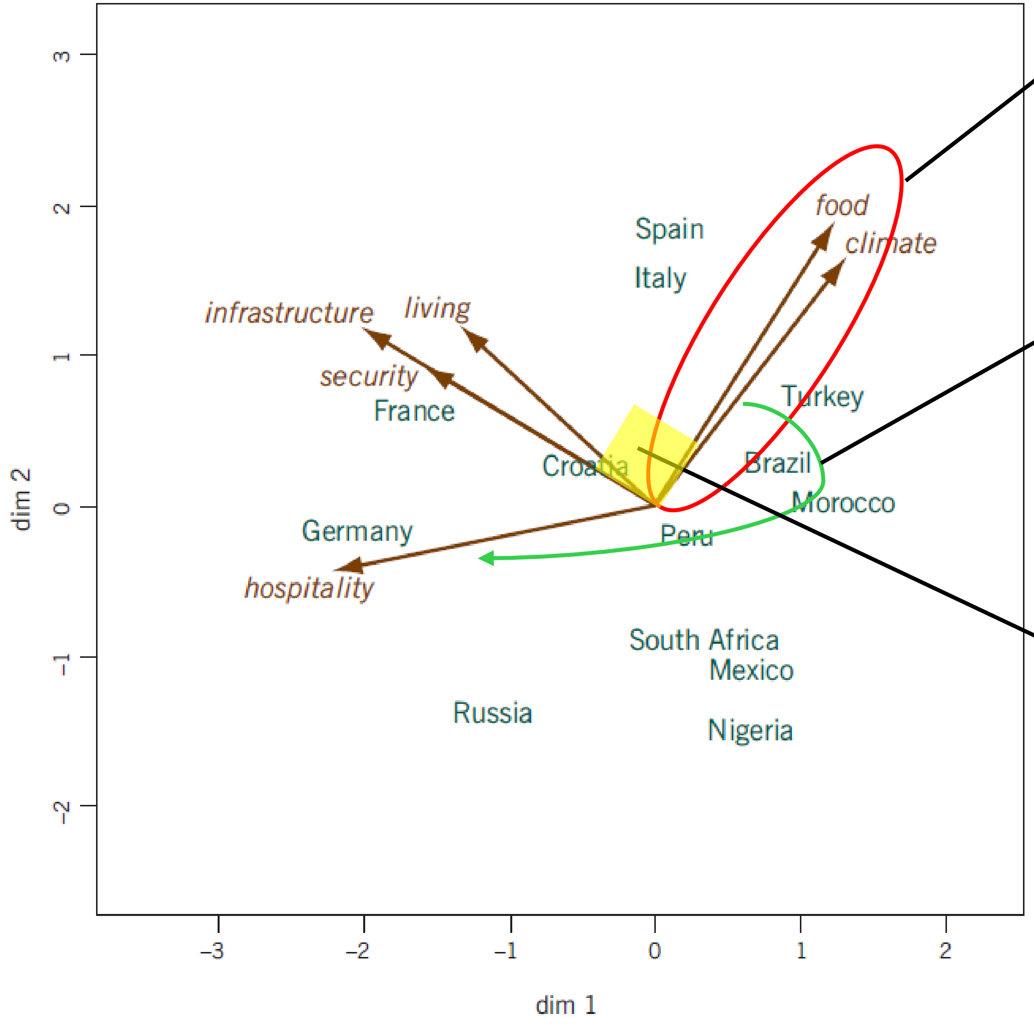
Keragaman peubah

Variabel dengan nilai keragaman besar digambarkan sebagai vector yang panjang



Variabel dengan nilai keragaman kecil digambarkan sebagai vector yang pendek

Korelasi antar peubah



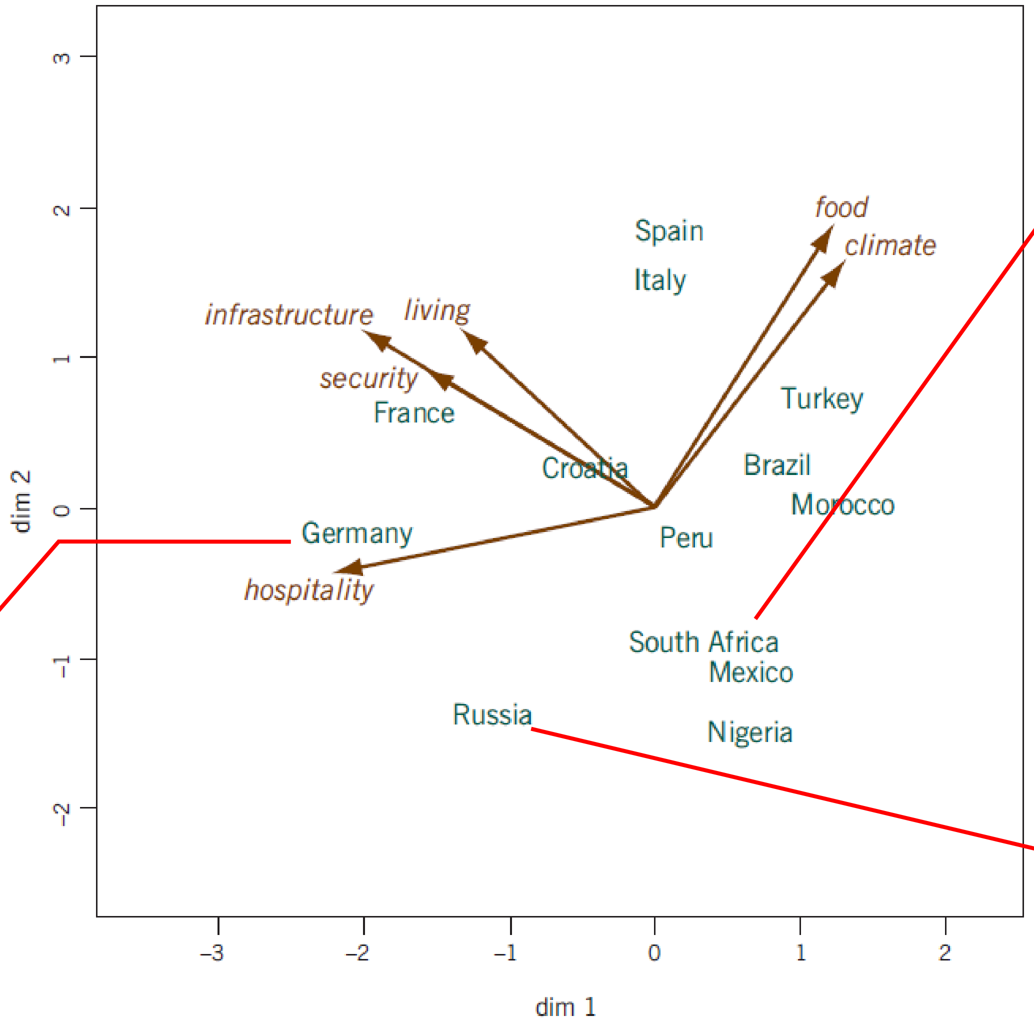
Korelasi positif: dua buah garis dengan arah yang sama atau membentuk sudut < 90

Korelasi negatif: dua buah garis dengan arah yang berlawanan atau membentuk sudut > 90

Tidak ada korelasi: dua buah garis dengan sudut siku-siku atau 90

Nilai peubah pada suatu objek

Jika sebuah titik berada di arah yang sama dengan suatu vektor, maka objek tersebut memiliki nilai tinggi pada variabel itu



Jika sebuah titik berlawanan arah dengan suatu vektor, maka objek tersebut memiliki nilai rendah pada variabel itu.

Jika sebuah titik tegak lurus terhadap suatu vektor, maka variabel tersebut tidak terlalu berpengaruh terhadap objek tersebut.

Struktur Data

Atribut	X1	X2	X3	X4	X5	X6
Objek 1	8,21	5,21	5,21	4,18	3,90	4,23
Objek 2	7,13	6,44	6,17	5,33	5,38	5,84
Objek 3	7,35	6,87	7,02	6,12	6,56	6,06
Objek 4	4,34	5,40	3,99	3,62	3,75	3,79
Objek 5	5,43	6,35	4,85	4,23	4,53	4,56
Objek 6	7,98	6,67	7,12	5,19	4,35	3,60

Biplot

Biplot dibuat dengan **SVD (Singular Value Decomposition)**, yang merupakan teknik dekomposisi matriks untuk mereduksi dimensi data.

Misalkan X merupakan data yang berisi *n* objek dan *p* variable yang dikoreksi menjadi *r* dimensi, sehingga X dapat dituliskan menjadi :

$$\mathrm{X} = \mathbf{ULA}'$$

dimana

$\boldsymbol{L} = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_p})$ dari $X'X$ (atau XX')

U = matriks yang kolom-kolomnya merupakan eigenvector dari XX'

A = matriks yang baris-barisnya merupakan eigenvector dari $X'X$

Biplot

$$\mathrm{X} = \mathbf{ULA}'$$
$$X = \left(\boldsymbol{UL}^{\mathbf{1/2}}\right)\left(\boldsymbol{L}^{\mathbf{1/2}}\boldsymbol{A}'\right)$$

dimana

$$\boldsymbol{UL}^{\mathbf{1/2}} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1}u_{11} & \sqrt{\lambda_2}u_{12} \\ \sqrt{\lambda_1}u_{21} & \sqrt{\lambda_2}u_{22} \\ \vdots & \vdots \\ \sqrt{\lambda_1}u_{n1} & \sqrt{\lambda_2}u_{n2} \end{pmatrix}$$

sebagai koordinat setiap objek

$$\boldsymbol{L}^{\mathbf{1/2}}\boldsymbol{A}' = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{n1} & a_{n2} \end{pmatrix}' = \begin{pmatrix} \sqrt{\lambda_1}a_{11}, & \sqrt{\lambda_1}a_{21}, & \cdots & \sqrt{\lambda_1}a_{p1}, \\ \sqrt{\lambda_2}a_{12}, & \sqrt{\lambda_2}a_{22}, & \cdots & \sqrt{\lambda_2}a_{p2}, \end{pmatrix}$$

sebagai koordinat setiap variabel

- Objek digambarkan sebagai titik
- Variabel digambarkan sebagai tanda panah

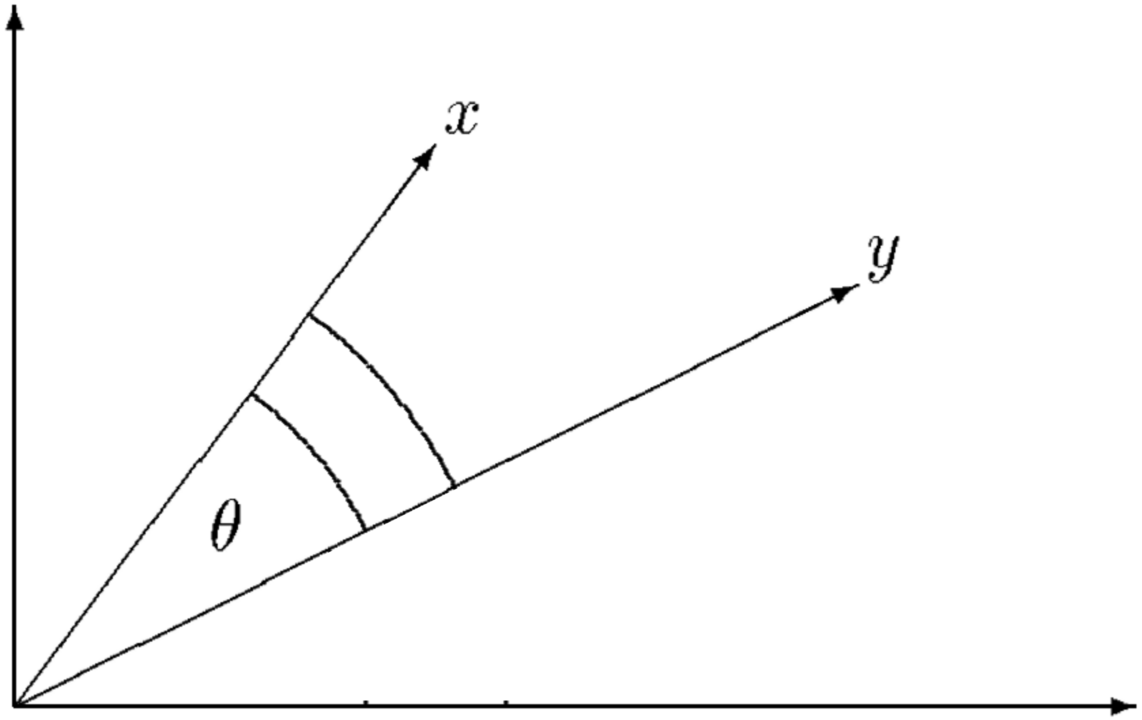
Ukuran Keragaman Biplot

Performa dengan menggunakan dua eigenvalue pertama

$$\frac{\lambda_1 + \lambda_2}{\sum_{i=1}^{p} \lambda_i}$$

Sudut antar variabel

- Cosinus dari sudut antara panah (garis) antar pasangan variable menunjukkan hubungan antara dua variabel korespondensi.



$$\cos\theta = \frac{x_1 y_1 + x_2 y_2}{\|x\| \; \|y\|} = \frac{x^\top y}{\|x\| \; \|y\|}$$

Langkah-Langkah Biplot

Menyusun matriks X (data yang telah distandarisasi)

Membuat matriks L, A, dan matriks U

Membuat matriks $(UL^{1/2})$ dan $(L^{1/2}A')$

Mengambil 2 kolom pertama dari matriks $(UL^{1/2})$ dan 2 baris pertama dari matriks $(L^{1/2}A')$

Membuat plot koordinat X dan Y untuk setiap objek dan variable hasil dari Langkah 4

Interpretasi plot

Contoh

No	Site	Bur Oak	Black Oak	White Oak	Red Oak	American Elm	Basswood	Ironwood	Sugar Maple
1	S1	9	8	5	3	2	0	0	0
2	S2	8	9	4	4	2	0	0	0
3	S3	3	8	9	0	4	0	0	0
4	S4	5	7	9	6	5	0	0	0
5	S5	6	0	7	9	6	2	0	0
6	S6	0	0	7	8	0	7	0	5
7	S7	5	0	4	7	5	6	7	4
8	S8	0	0	6	6	0	6	4	8
9	S9	0	0	0	4	2	7	6	8
10	S10	0	0	2	3	5	6	5	9

Referensi

Johnson, R. W., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis Sixth Edition. New York: Prentice Hall Inc.

Greenacre, Michael, 2010. Biplots in Practice. Spain: Fundacion BBVA / BBVA Foundation.

Rencher, A. C., 2002. Methods of Multivariate Analysis. 2nd ed. United States of America: John Wiley & Sons, Inc.

Manual Calculation

https://docs.google.com/spreadsheets/d/1gFrQMzc6wjTpDe98cFhzC0XjmfiSUWIH/edit?usp=sharing&ouid=103490478088003213054&rtpof=true&sd=true