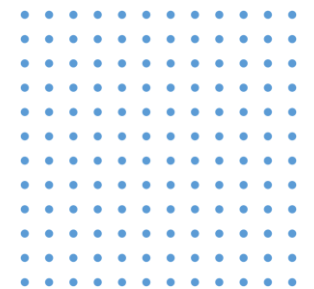
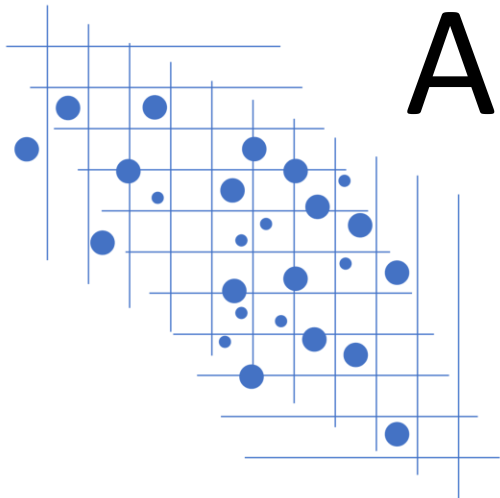




S1 Sains Data
FMIPA UNESA

Analisis Multivariat – Materi 03

Distribusi Normal Multivariat dan Analisis Korelasi Kanonik



Prodi S1 Sains Data

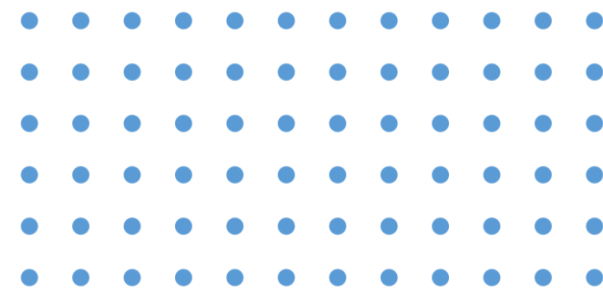
Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Negeri Surabaya

2025

Outline

- Review Matriks Varians Kovarian dan Korelasi
- Distribusi Normal Multivariat
- Review Nilai dan Vector Eigen + Kombinasi Linier
- Analisis Korelasi Kanonik
- Korelasi Kanonik dengan R



Review Matriks Varians Kovarian dan Korelasi

Kovarians dan Korelasi

Covariance

Population Covariance Formula

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

Sample Covariance

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

Correlation

$$\text{Correlation} = \frac{\text{Cov}(x, y)}{\sigma_x * \sigma_y}$$

Matrix Variance Covariance (Σ)

$$\Sigma = \begin{pmatrix} \sigma_{X_1 X_1} & \dots & \sigma_{X_1 X_p} \\ \vdots & \ddots & \vdots \\ \sigma_{X_p X_1} & \dots & \sigma_{X_p X_p} \end{pmatrix}.$$

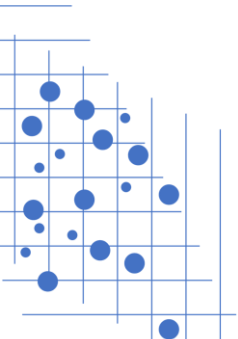
Dua Variabel

$$\begin{matrix} & x & y \\ \begin{matrix} x \\ y \end{matrix} & \begin{bmatrix} var(x) & cov(x, y) \\ cov(x, y) & var(y) \end{bmatrix} \end{matrix}$$

Tiga Variabel

$$\begin{matrix} & x & y & z \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} var(x) & cov(x, y) & cov(x, z) \\ cov(x, y) & var(y) & cov(y, z) \\ cov(x, z) & cov(y, z) & var(z) \end{bmatrix} \end{matrix}$$

- How about invers of variance covariance matrix (Σ^{-1})?

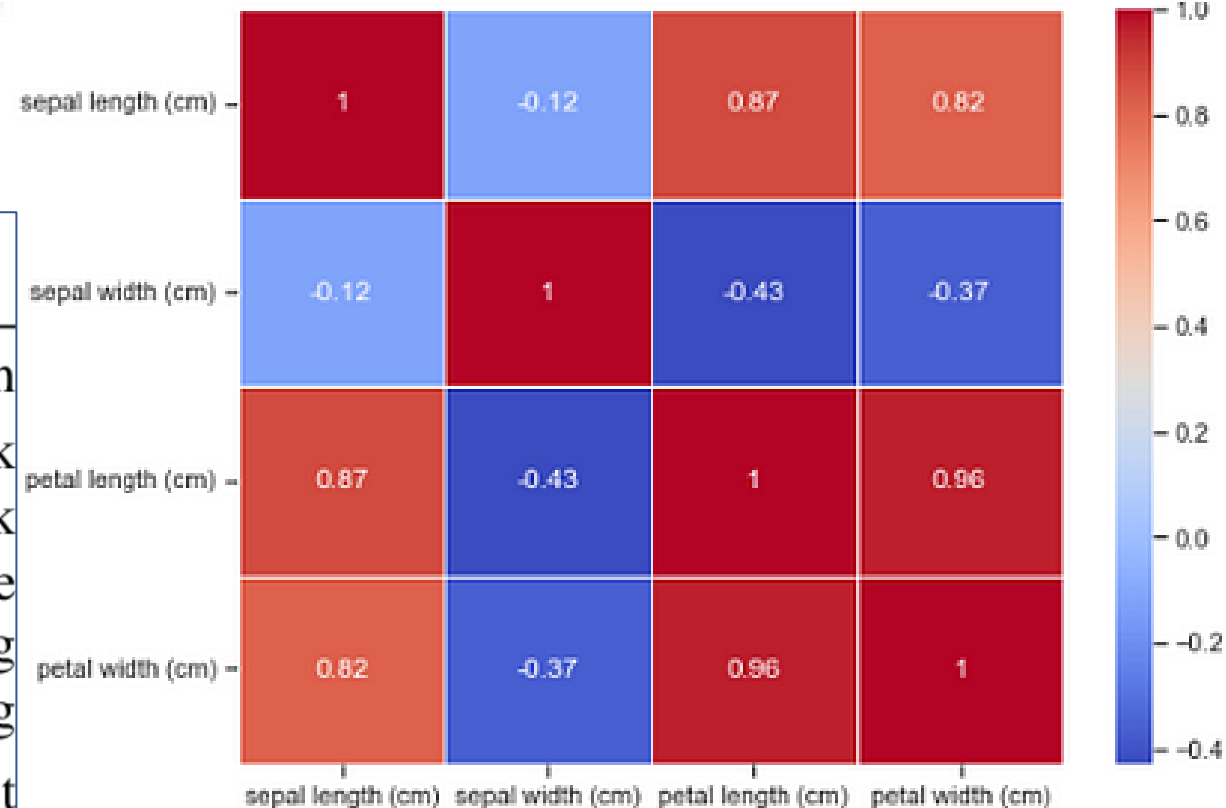


Matrix Correlation and Heatmap

$$\mathcal{P} = \begin{pmatrix} \rho_{X_1 X_1} & \cdots & \rho_{X_1 X_p} \\ \vdots & \ddots & \vdots \\ \rho_{X_p X_1} & \cdots & \rho_{X_p X_p} \end{pmatrix},$$

No	Coefficient	Correlation Coefficient Classification
1	0	No correlation
2	0-0.2	Very weak
3	0.21-0.40	Weak
4	0.41-0.60	Moderate
5	0.61-0.80	Strong
6	0.81-0.99	Very strong
7	1	Perfect

Source: [Roflin & Zulvia \(2021\)](#)



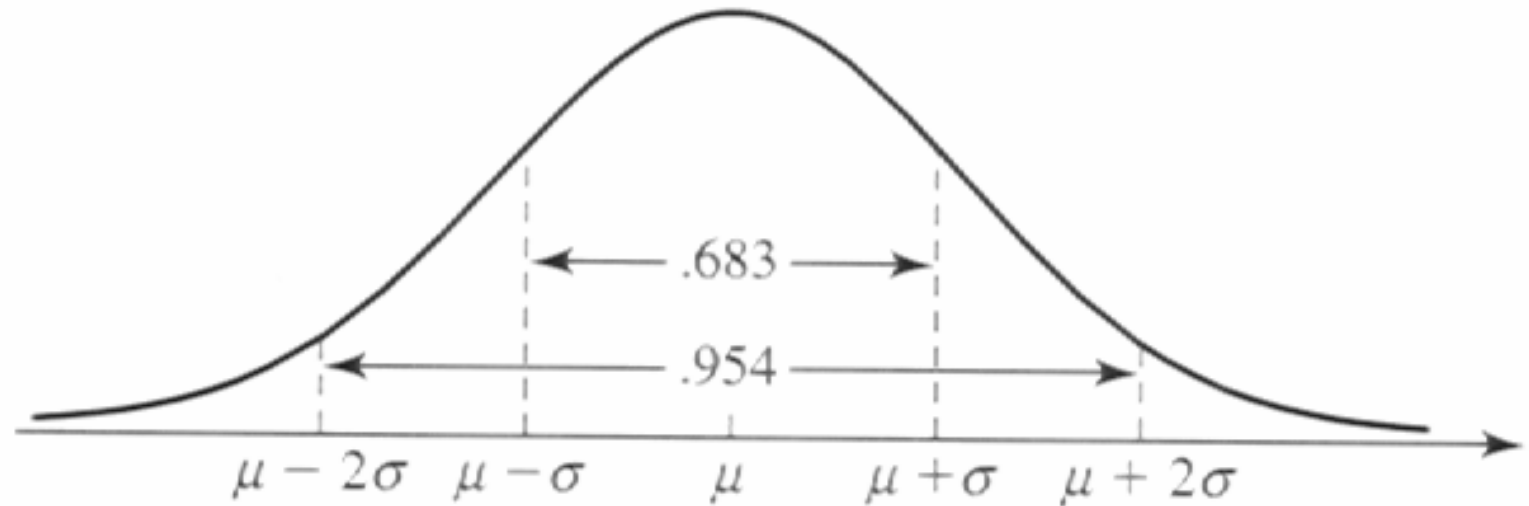


Distribusi Normal Multivariat

Distribusi Normal Univariat

$$N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \quad -\infty < x < \infty$$



Fungsi Kepadatan Normal p -dimensi

$$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

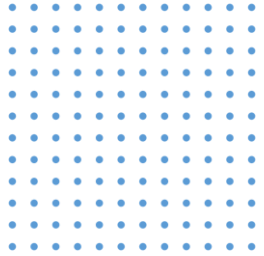
$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}}$$

$$-\infty < x_i < \infty, \quad i = 1, 2, \dots, p$$

\mathbf{x} is a sample from random vector

$$\mathbf{X}' = [X_1, X_2, \dots, X_p]$$

Bivariate Normal



Kurva Bivariate Normal dengan
 $\sigma_{11} = \sigma_{22}, \rho_{12} = 0$

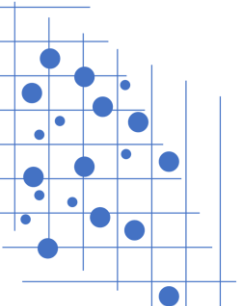
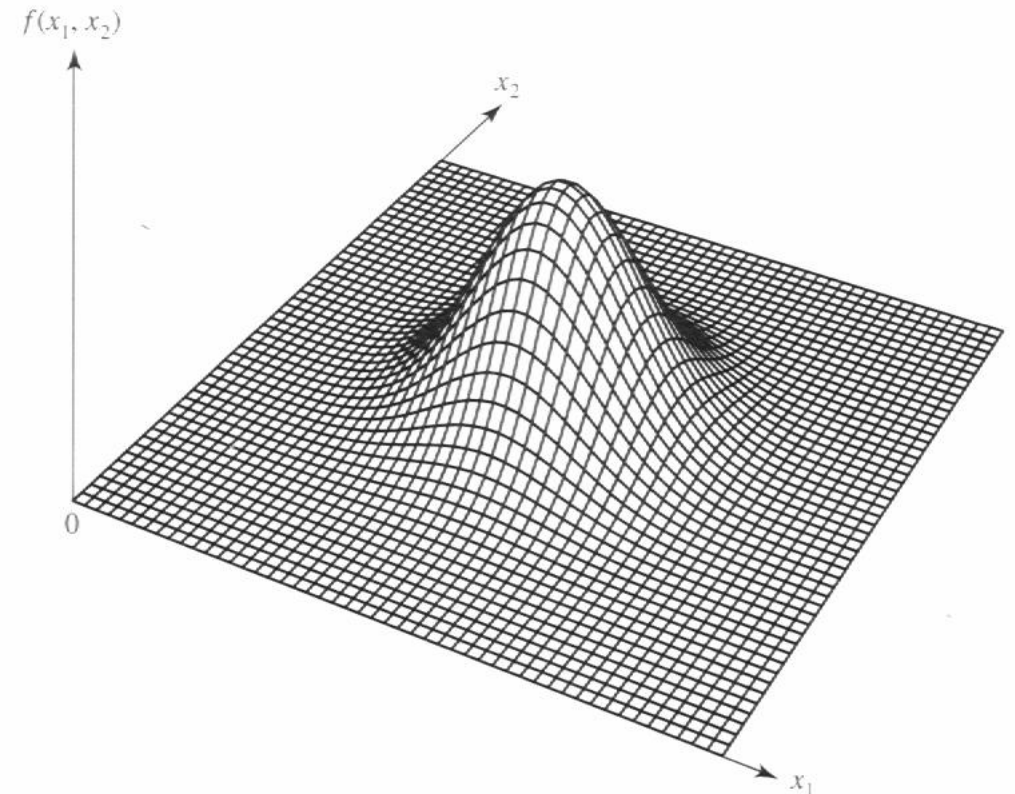
$$\mu_1 = E(X_1), \mu_2 = E(X_2)$$

$$\sigma_{11} = \text{Var}(X_1), \sigma_{22} = \text{Var}(X_2)$$

$$\rho_{12} = \sigma_{12} / (\sqrt{\sigma_{11}} \sqrt{\sigma_{22}}) = \text{Corr}(X_1, X_2)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho_{12}^2)$$



Distribusi Normal Multivariat

- Generalisasi dari fungsi kepadatan peluang normal univariat
- Sebagai dasar banyak analisis multivariat
- Berguna untuk menghampiri distribusi populasi yang sebenarnya
- Central limit distribution (Distribusi limit pusat) untuk banyak statistik multivariat
- Dapat dijelaskan secara matematis

$$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})/2}$$

$$-\infty < x_i < \infty, \quad i = 1, 2, \dots, p$$

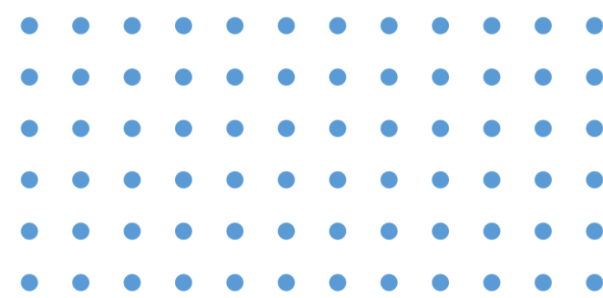
Kasus Multivariat

$\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ independent observations from population (may not be multivariate normal)

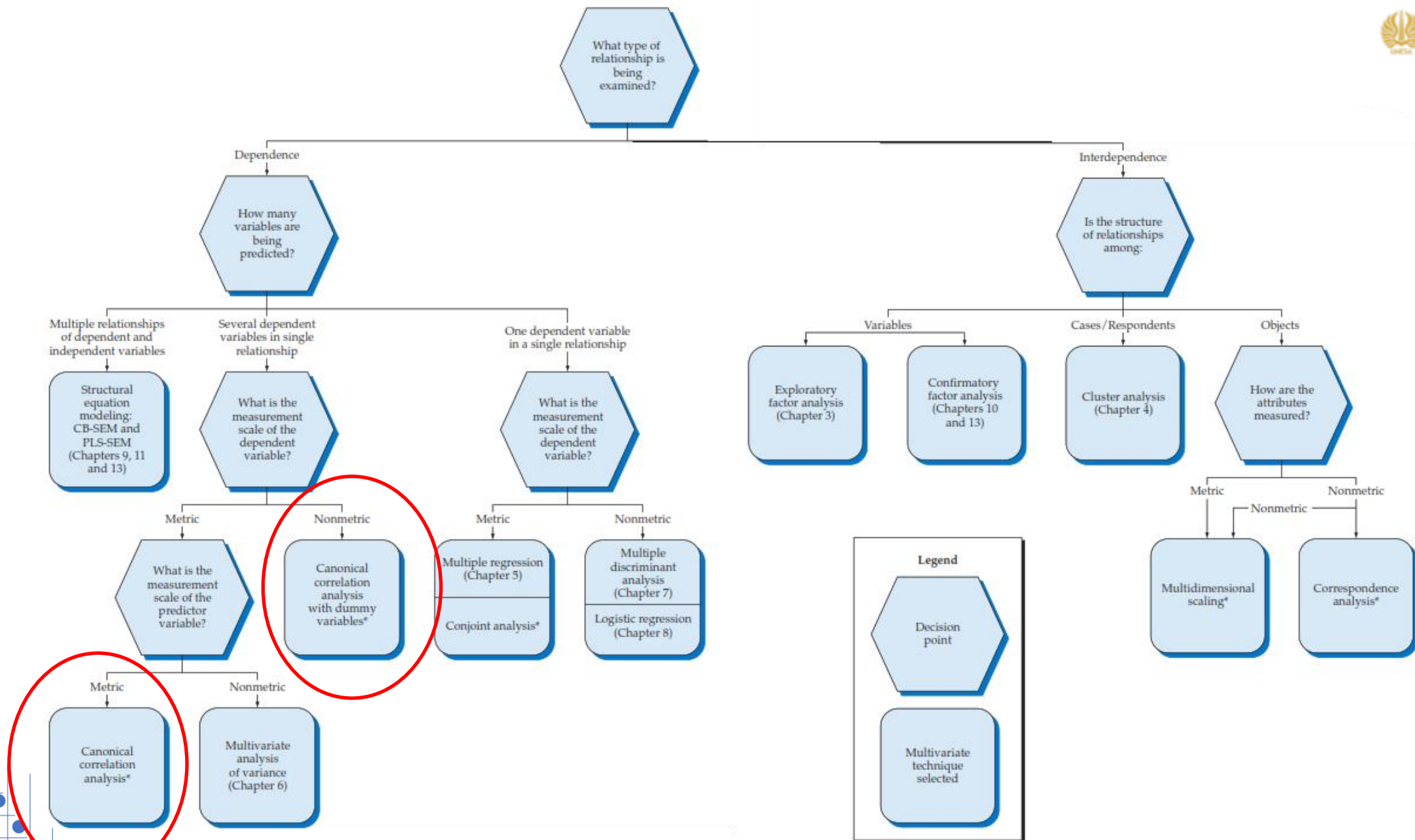
with mean $E(\mathbf{X}_i) = \boldsymbol{\mu} \Rightarrow$

$\bar{\mathbf{X}}$ converges in probability to $\boldsymbol{\mu}$

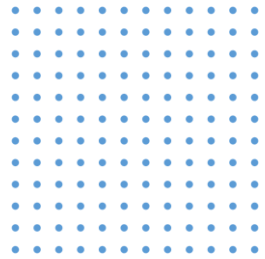
\mathbf{S} converges in probability to $\boldsymbol{\Sigma}$



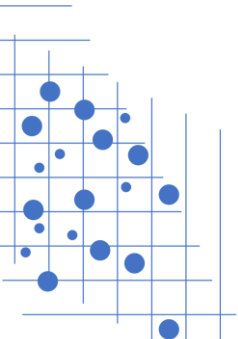
Analisis Korelasi Kanonik



Korelasi Kanonik



- Analisis korelasi kanonik (*canonical correlation*) adalah salah satu teknik analisis multivariat untuk **mengidentifikasi dan mengukur hubungan atau asosiasi antara dua kelompok/himpunan variabel.**
- Analisis kanonik ini difokuskan pada korelasi antara kombinasi linear dari variabel dalam satu kelompok dan kombinasi linear dari variabel pada kelompok yang lain.
 - Pasangan kombinasi linear dinamakan **variabel kanonik**
 - Nilai korelasinya disebut disebut sebagai **korelasi kanonik**



Asumsi dalam Analisis Kanonik

1. **Linieritas**, yaitu keadaan dimana hubungan antar variabel bersifat linier.
2. **Normalitas multivariat**, yaitu menguji signifikansi setiap fungsi kanonik. Namun, pengujian normalitas secara multivariat sulit dilakukan, maka cukup dilakukan uji normalitas untuk setiap variabel. Asumsi yang digunakan adalah jika secara individu sebuah variabel memenuhi kriteria normalitas, maka secara keseluruhan juga akan memenuhi asumsi normalitas.
3. **Tidak ada multikolinieritas** antar anggota kelompok variabel.

Sumber : Mattjik dan Sumertajaya (2011)

Pendugaan Koefisien Kanonik

Misalkan ingin dibuat hubungan antara dua kelompok variabel. Variabel pertama terdiri dari p variabel berukuran ($p \times 1$) yang dinotasikan dengan vektor variabel random $\mathbf{X}^{(1)}$. Variabel kedua terdiri dari q variabel berukuran ($q \times 1$) yang dinotasikan dengan vektor variabel random $\mathbf{X}^{(2)}$. dimana $p \leq q$. Misalkan karakteristik dari vektor variabel random $\mathbf{X}^{(1)}$ dan $\mathbf{X}^{(2)}$ adalah sebagai berikut:

$$\begin{aligned} E(\mathbf{X}^{(1)}) &= \boldsymbol{\mu}^{(1)}; & \text{Cov}(\mathbf{X}^{(1)}) &= \boldsymbol{\Sigma}_{11} \\ E(\mathbf{X}^{(2)}) &= \boldsymbol{\mu}^{(2)}; & \text{Cov}(\mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{22} \\ \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21} \end{aligned} \quad (1)$$

Kombinasi linear dari kedua kelompok variabel dapat dituliskan sebagai berikut :

$$U = \mathbf{a}'\mathbf{X}^{(1)} = a_1X_1^{(1)} + a_2X_2^{(1)} + \dots + a_pX_p^{(1)} \quad (2)$$

$$V = \mathbf{b}'\mathbf{X}^{(2)} = b_1X_1^{(2)} + b_2X_2^{(2)} + \dots + b_qX_q^{(2)}$$

Sehingga diperoleh

$$\begin{aligned} \text{Var}(U) &= \mathbf{a}' \text{Cov}(\mathbf{X}^{(1)}) \mathbf{a} = \mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a} \\ \text{Var}(V) &= \mathbf{b}' \text{Cov}(\mathbf{X}^{(2)}) \mathbf{b} = \mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b} \\ \text{Cov}(U, V) &= \mathbf{a}' \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \mathbf{b} = \mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b} \end{aligned} \quad (3)$$

Korelasi kanonik diperoleh dengan menghitung

$$\text{Corr}(U, V) = \frac{\mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b}}{\sqrt{\mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a}} \sqrt{\mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b}}} \quad (4)$$

Pendugaan Koefisien Kanonik

- Pasangan variabel kanonik pertama (U_1, V_1) adalah :

$$U_1 = \underbrace{\mathbf{e}_1' \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)}}_{\mathbf{a}_1'} \quad V_1 = \underbrace{\mathbf{f}_1' \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}}_{\mathbf{b}_1'}$$

- Pasangan variabel kanonik ke- k , $k = 2, 3, \dots, p$ adalah :

$$U_k = \mathbf{e}_k' \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)} \quad V_k = \mathbf{f}_k' \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}$$

maximizes

$$\text{Corr}(U_k, V_k) = \rho_k^*$$

Dimana $\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \rho_p^{*2}$ adalah eigenvalue dari matriks $\boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2}$ yang berpadanan dengan eigenvector $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$.

$\rho_1^{*2} \geq \rho_2^{*2} \geq \dots \rho_p^{*2}$ juga merupakan eigenvalue dari matriks $\boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2}$ yang berpadanan dengan eigenvector $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p$.

Pendugaan Koefisien Kanonik (lanjutan)

Definisi :

- Variabel kanonik pertama memiliki korelasi terbesar pertama

$$U_1 = a_1' \mathbf{X}^{(1)} \quad \text{Var}(U_1) = 1$$

$$V_1 = b_1' \mathbf{X}^{(2)} \quad \text{Var}(V_1) = 1$$

$$\text{maksimum corr}(U_1, V_1) = \rho_1$$

- Variabel kanonik kedua memiliki korelasi terbesar kedua

$$U_2 = a_2' \mathbf{X}^{(1)} \quad \text{Var}(U_2) = 1 \quad \text{Cov}(U_1, U_2) = 0 \quad \text{Cov}(U_1, V_2) = \text{Cov}(U_2, V_1) = 0$$

$$V_2 = b_2' \mathbf{X}^{(2)} \quad \text{Var}(V_2) = 1 \quad \text{Cov}(V_1, V_2) = 0$$

$$\text{maksimum corr}(U_2, V_2) = \rho_2$$

- Variabel kanonik ke-k memiliki korelasi terbesar ke-k

$$U_k = a_k' \mathbf{X}^{(1)} \quad \text{Var}(U_k) = 1 \quad \text{Cov}(U_k, U_l) = 0 \quad \text{Cov}(U_k, V_l) = 0 ; k \neq l$$

$$V_k = b_k' \mathbf{X}^{(2)} \quad \text{Var}(V_k) = 1 \quad \text{Cov}(V_k, V_l) = 0$$

$$\text{maksimum corr}(U_k, V_k) = \rho_k$$

Pasangan variabel kanonik memiliki sifat :

$$\text{Var}(U_k) = \text{Var}(V_k) = 1$$

$$\text{Cov}(U_k, U_\ell) = \text{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell$$

$$\text{Cov}(V_k, V_\ell) = \text{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell$$

$$\text{Cov}(U_k, V_\ell) = \text{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell$$

Matriks varians kovarians

$$\mathbf{S}_{p \times p} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{11} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

where

$$s_{ik} = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)$$

Kombinasi Linear

$$U = \mathbf{a}' \mathbf{X}^{(1)} = a_1 X_1^{(1)} + a_2 X_2^{(1)} + \cdots + a_p X_p^{(1)}$$

$$V = \mathbf{b}' \mathbf{X}^{(2)} = b_1 X_1^{(2)} + b_2 X_2^{(2)} + \cdots + b_q X_q^{(2)}$$

Matriks varians kovarians

$$\text{Var}(U) = \mathbf{a}' \text{Cov}(\mathbf{X}^{(1)}) \mathbf{a} = \mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a}$$

$$\text{Var}(V) = \mathbf{b}' \text{Cov}(\mathbf{X}^{(2)}) \mathbf{b} = \mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b}$$

$$\text{Cov}(U, V) = \mathbf{a}' \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \mathbf{b} = \mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b}$$

$$\text{Corr}(U, V) = \frac{\mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b}}{\sqrt{\mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a}} \sqrt{\mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b}}}$$

$$\rho_k = \text{Corr}(U_k, V_k) = \sqrt{\lambda_k}.$$

We are looking for vectors $\mathbf{a} \in \mathbb{R}^p, \mathbf{b} \in \mathbb{R}^q$ such that $\text{Corr}(U, V)$ is maximised.

Persamaan kanonikal

$$|\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} - \rho^{*2} \mathbf{I}| = 0$$

$$\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \mathbf{a} = \rho^{*2} \mathbf{a}$$

$$\boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \mathbf{b} = \rho^{*2} \mathbf{b}$$

$$(\mathbf{A} - \rho^{*2} \mathbf{I}) \mathbf{a} = \mathbf{0}$$

$$(\mathbf{B} - \rho^{*2} \mathbf{I}) \mathbf{b} = \mathbf{0}$$

$$\mathbf{A} = \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$$

$$\mathbf{B} = \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$$

Canonical correlation coefficient from eigenvalue of a symmetric matrix

The relationship between the canonical coefficients of the standardized variables and the canonical coefficients of the original variables:

$$\mathbf{A}^* = \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2}$$

$$\mathbf{B}^* = \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1/2}$$

Kombinasi Linear

$$\mathbf{a}_i = \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{e}_i, \text{ and } \mathbf{b}_i = \boldsymbol{\Sigma}_{22}^{-1/2} \mathbf{f}_i.$$

Contoh

	X_1	X_2	Y_1	Y_2
	Sistolik	Diastolik	Tinggi	Berat
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Tekanan Darah
Blood Pressure (BP)

Ukuran Badan
Body Size (BS)

Bagaimana korelasi kanonik antara BP dan BS?

1. Hitung matriks kovarians

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	128,567	37,267	24,167	48,333
Diastolik	37,267	17,467	19,267	31,933
Tinggi	24,167	19,267	74,167	91,333
Berat	48,333	31,933	91,333	132,667

$$\begin{aligned}
 E(\mathbf{X}^{(1)}) &= \boldsymbol{\mu}^{(1)}; & \text{Cov}(\mathbf{X}^{(1)}) &= \boldsymbol{\Sigma}_{11} \\
 E(\mathbf{X}^{(2)}) &= \boldsymbol{\mu}^{(2)}; & \text{Cov}(\mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{22} \\
 \text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) &= \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'
 \end{aligned}
 \quad
 \Sigma = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

$$S_{xx} = \begin{bmatrix} 128,567 & 37,267 \\ 37,267 & 17,467 \end{bmatrix} \quad S_{xy} = \begin{bmatrix} 24,167 & 48,333 \\ 19,267 & 31,933 \end{bmatrix}$$

$$S_{yx} = \begin{bmatrix} 24,167 & 19,267 \\ 48,333 & 31,933 \end{bmatrix} \quad S_{yy} = \begin{bmatrix} 74,167 & 91,333 \\ 91,333 & 132,667 \end{bmatrix}$$

Contoh

	X_1	X_2	Y_1	Y_2
	Sistolik	Diastolik	Tinggi	Berat
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Tekanan Darah
Blood Pressure (BP)

Ukuran Badan
Body Size (BS)

Bagaimana korelasi kanonik antara BP dan BS?

2. Hitung matriks korelasi kanonik

$$R_x = S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2}$$

$$S_{xx}^{-1/2} = P D^{-1/2} P^{-1}$$

P adalah matriks vektor eigen dari S_{xx}

D adalah matriks diagonal nilai eigen dari S_{xx}

$$S_{xx} = \begin{bmatrix} 128,567 & 37,267 \\ 37,267 & 17,467 \end{bmatrix}$$

```
> eigen_Sxx
eigen() decomposition
$values
[1] 139.909170  6.124164

$vectors
      [,1]      [,2]
[1,] -0.9566705  0.2911728
[2,] -0.2911728 -0.9566705
```

$$\begin{aligned}
 S_{xx}^{-1/2} &= P D^{-1/2} P' \\
 &= \begin{pmatrix} -0.9567 & 0.2912 \\ -0.2912 & -0.9567 \end{pmatrix} \begin{pmatrix} 1/\sqrt{139.909} & 0 \\ 0 & 1/\sqrt{6.124} \end{pmatrix} \begin{pmatrix} -0.9567 & 0.2912 \\ -0.2912 & -0.9567 \end{pmatrix}' \\
 &= \begin{pmatrix} 0.1116 & -0.0890 \\ -0.0890 & 0.3770 \end{pmatrix}
 \end{aligned}$$

Contoh

	X_1	X_2	Y_1	Y_2
	Sistolik	Diastolik	Tinggi	Berat
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Matriks Kovarian

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	128,567	37,267	24,167	48,333
Diastolik	37,267	17,467	19,267	31,933
Tinggi	24,167	19,267	74,167	91,333
Berat	48,333	31,933	91,333	132,667

Bagaimana korelasi kanonik antara BP dan BS?

2. Hitung matriks korelasi kanonik

$$R_x = S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2}$$

$$S_{xx}^{-1/2} = \begin{pmatrix} 0.1116 & -0.0890 \\ -0.0890 & 0.3770 \end{pmatrix}$$

$$R_x = \begin{pmatrix} 0.1116 & -0.0890 \\ -0.0890 & 0.3770 \end{pmatrix} \begin{pmatrix} 24,167 & 48,333 \\ 19,267 & 31,933 \end{pmatrix} \begin{pmatrix} 74,167 & 91,333 \\ 91,333 & 132,667 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 0.1023 & 0.1636 \\ 0.1636 & 0.4552 \end{pmatrix}$$

3. Hitung nilai eigen dan vector eigen dari R_x

\$values

[1] 0.51936537 0.03817159

\$vectors

[,1] [,2]
[1,] 0.3651336 -0.9309551
[2,] 0.9309551 0.3651336

Contoh

	X_1	X_2	Y_1	Y_2
	Sistolik	Diastolik	Tinggi	Berat
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	128,567	37,267	24,167	48,333
Diastolik	37,267	17,467	19,267	31,933
Tinggi	24,167	19,267	74,167	91,333
Berat	48,333	31,933	91,333	132,667

Bagaimana korelasi kanonik antara BP dan BS?

4. Hitung a'_1

$$\hat{U}_k = \underbrace{\hat{\mathbf{e}}'_k \mathbf{S}_{11}^{-1/2}}_{\hat{\mathbf{a}}'_k} \mathbf{x}^{(1)}$$

$$\hat{V}_k = \underbrace{\hat{\mathbf{f}}'_k \mathbf{S}_{22}^{-1/2}}_{\hat{\mathbf{b}}'_k} \mathbf{x}^{(2)}$$

$$\mathbf{a}'_1 = (0.3651 \quad 0.9309) \begin{pmatrix} 0.1116 & -0.0890 \\ -0.0890 & 0.3770 \end{pmatrix} = (-0.0421 \quad 0.3184)$$

Hitung \mathbf{a}'_2

$$\mathbf{a}'_2 = (-0.9309 \quad 0.3651) \begin{pmatrix} 0.1116 & -0.0890 \\ -0.0890 & 0.3770 \end{pmatrix} = (-0.1364 \quad 0.2205)$$

$$U_1 = -0.0421 \text{ Systolik} + 0.3184 \text{ Diastolik}$$

$$U_2 = -0.1364 \text{ Systolik} + 0.2205 \text{ Diastolik}$$

5. Lakukan standarisasi pada U_1 dan U_2

6. Ulangi untuk R_y , \mathbf{b} , dan \mathbf{V} . Lakukan standarisasi pada V_1 dan V_2

7. Hitung Korelasi U_1 - V_1 , U_2 - V_2

Contoh

	X_1	X_2	Y_1	Y_2
	Sistolik	Diastolik	Tinggi	Berat
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Matriks Korelasi (Pearson) antar Fitur

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	1			
Diastolik	0,79	1,00		
Tinggi	0,25	0,54	1,00	
Berat	0,37	0,66	0,92	1

Tekanan Darah/
Blood Pressure (BP)

Ukuran Badan/
Body Size (BS)

Bagaimana korelasi kanonik antara BP dan BS?

Contoh

No	Sistolik X_1	Diastolik X_2	Tinggi Y_1	Berat Y_2
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Matriks Kovarians antar Fitur

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	128,567	37,267	24,167	48,333
Diastolik	37,267	17,467	19,267	31,933
Tinggi	24,167	19,267	74,167	91,333
Berat	48,333	31,933	91,333	132,667

Vectors

$$\begin{bmatrix} [1,] & 0.1310689 & -0.5261187 \\ [2,] & -0.9913733 & 0.8504112 \end{bmatrix}$$

$$S_{yy}^{-1} = \begin{bmatrix} 0,089 & -0,061 \\ -0,061 & 0,050 \end{bmatrix}$$

$$S_{xx} = \begin{bmatrix} 128,567 & 37,267 \\ 37,267 & 17,467 \end{bmatrix} \quad S_{xy} = \begin{bmatrix} 24,167 & 48,333 \\ 19,267 & 31,933 \end{bmatrix}$$

$$R_x = S_{xx}^{-1} S_{xy} \quad S_{yy}^{-1} S_{yx}$$

$$S_{yy} = \begin{bmatrix} 74,167 & 91,333 \\ 91,333 & 132,667 \end{bmatrix} \quad S_{yx} = \begin{bmatrix} 24,167 & 19,267 \\ 48,333 & 31,933 \end{bmatrix}$$

$$R_x = \begin{bmatrix} -0,093 & -0,081 \\ 0,989 & 0,650 \end{bmatrix}$$

Vektor eigen

$$v_x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0.131 & -0.526 \\ -0.991 & 0.850 \end{bmatrix}$$

Variabel kanonik pertama (U_1) adalah

$$BP = 0.131 X_1 - 0.991 X_2$$

Mencari Nilai eigen <https://www.youtube.com/watch?v=MYZLQVGiTjU>

Latihan

No	Sistolik X_1	Diastolik X_2	Tinggi Y_1	Berat Y_2
1	120	76	165	60
2	109	80	180	80
3	130	82	170	70
4	121	78	185	85
5	135	85	180	90
6	140	87	187	87

Matriks Kovarians antar Fitur

	Sistolik	Diastolik	Tinggi	Berat
Sistolik	128,567	37,267	24,167	48,333
Diastolik	37,267	17,467	19,267	31,933
Tinggi	24,167	19,267	74,167	91,333
Berat	48,333	31,933	91,333	132,667

$$S_{xx}^{-1} = \begin{bmatrix} 0,020 & -0,043 \\ -0,043 & 0,150 \end{bmatrix} \quad S_{yy}^{-1} = \begin{bmatrix} 0,089 & -0,061 \\ -0,061 & 0,050 \end{bmatrix}$$

$$S_{xx} = \begin{bmatrix} 128,567 & 37,267 \\ 37,267 & 17,467 \end{bmatrix} \quad S_{xy} = \begin{bmatrix} 24,167 & 48,333 \\ 19,267 & 31,933 \end{bmatrix}$$

$$R_y = S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy}$$

$$S_{yy} = \begin{bmatrix} 74,167 & 91,333 \\ 91,333 & 132,667 \end{bmatrix} \quad S_{yx} = \begin{bmatrix} 24,167 & 19,267 \\ 48,333 & 31,933 \end{bmatrix}$$

$$R_y = \begin{bmatrix} & \end{bmatrix}$$

$$v_y = \begin{bmatrix} & \end{bmatrix}$$

Variabel kanonik ke dua (V_1) adalah

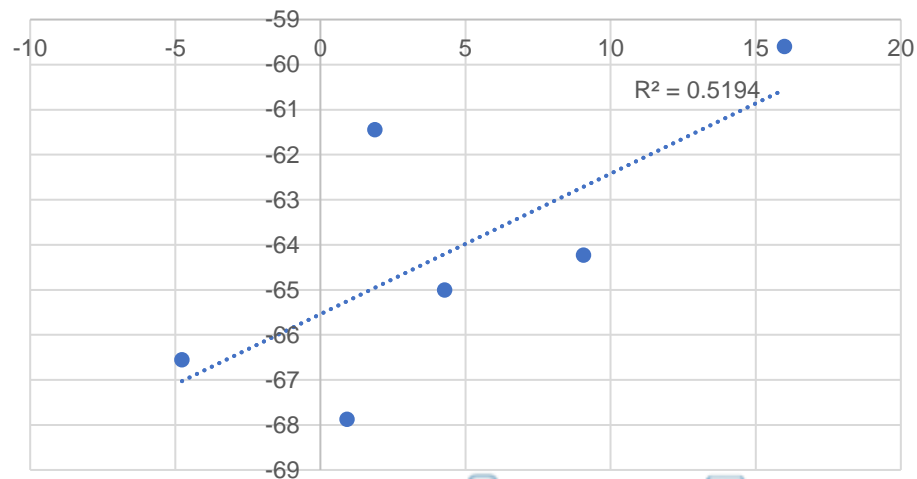
$$BS = \dots Y_1 - \dots Y_2$$

$$v_y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0.426 & -0.846 \\ -0.905 & 0.533 \end{bmatrix}$$

Contoh

No	Sistolik X_1	Diastolik X_2	Tinggi Y_1	Berat Y_2	\widehat{BP} $= 0.131 X_1$ $- 0.991 X_2$	\widehat{BS} $= 0.426 Y_1$ $- 0.905 Y_2$	zBP	zBS
1	120	76	165	60	-59,596	15,99	1,45	1,59
2	109	80	180	80	-65,001	4,28	-0,28	-0,04
3	130	82	170	70	-64,232	9,07	-0,04	0,63
4	121	78	185	85	-61,447	1,885	0,86	-0,37
5	135	85	180	90	-66,55	-4,77	-0,78	-1,30
6	140	87	187	87	-67,877	0,927	-1,21	-0,51
Rataan					-64,117	4,564		
Stdev					3,112	7,190		

Scatter Plot Prediksi BP dan BS



Korelasi Kanonik

$$\text{corr}(BP, BS) = \sqrt{R^2}$$

$$= \sqrt{0,5194} = 0,72$$

Contoh

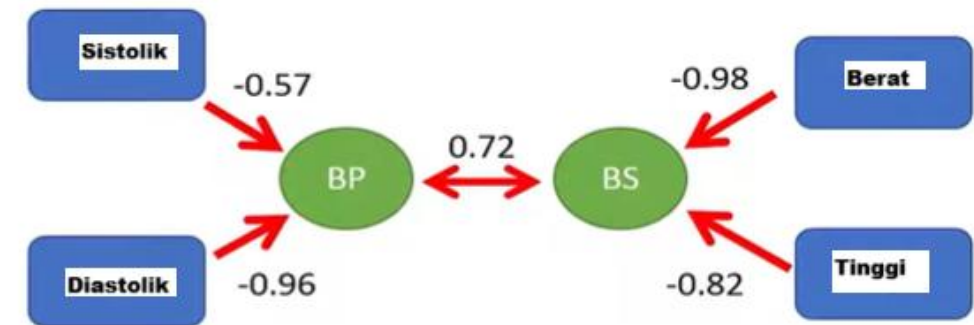
No	Sistolik X_1	Diastolik X_2	Tinggi Y_1	Berat Y_2	\widehat{BP} $= 0.131 X_1$ $- 0.991 X_2$	\widehat{BS} $= 0.426 Y_1$ $- 0.905 Y_2$	zBP	zBS
1	120	76	165	60	-59,596	15,99	1,45	1,59
2	109	80	180	80	-65,001	4,28	-0,28	-0,04
3	130	82	170	70	-64,232	9,07	-0,04	0,63
4	121	78	185	85	-61,447	1,885	0,86	-0,37
5	135	85	180	90	-66,55	-4,77	-0,78	-1,30
6	140	87	187	87	-67,877	0,927	-1,21	-0,51
Rataan					-64,117	4,564		
Stdev					3,112	7,190		

Korelasi Kanonik

$$\text{corr}(BP, BS) = \sqrt{R^2}$$

$$= \sqrt{0,5194} = 0,72$$

	Sistolik	Diastolik	zBP_topi
Sistolik	1,00		
Diastolik	0,79	1,00	
zBP_topi	-0,57	-0,96	1,00
	Tinggi	Berat	zBS_topi
Tinggi	1,00		
Berat	0,92	1,00	
zBS_topi	-0,82	-0,98	1,00



Uji Signifikansi Korelasi Kanonik

Uji korelasi kanonik secara keseluruhan

Hipotesis :

$$H_0: \Sigma_{12} = 0 \quad (\rho_1^* = \rho_2^* = \dots = \rho_p^* = 0)$$

$$H_1: \Sigma_{12} \neq 0 \quad (\rho_1^* \neq \rho_2^* \neq \dots \neq \rho_p^* \neq 0)$$

Statistik Uji :

$$-\left(n - 1 - \frac{1}{2}(p + q + 1)\right) \ln \prod_{i=1}^p (1 - \widehat{\rho}_i^{*2})$$

Daerah Penolakan :

Tolak H_0 jika statistik uji $> \chi_{pq}^2(\alpha)$

Uji korelasi kanonik secara sebagian

Hipotesis :

$$H_0^k: \rho_1^* \neq 0, \rho_2^* \neq 0, \dots, \rho_k^* \neq 0, \rho_{k+1}^* = \dots = \rho_p^* = 0$$

$$H_1^k: \rho_i^* \neq 0, \text{ untuk beberapa } i \geq k + 1$$

Statistik Uji :

$$-\left(n - 1 - \frac{1}{2}(p + q + 1)\right) \ln \prod_{i=k+1}^p (1 - \widehat{\rho}_i^{*2})$$

Daerah Penolakan :

Tolak H_0 jika statistik uji $> \chi_{(p-k)(q-k)}^2(\alpha)$

Contoh

Tujuan : menganalisis hubungan psikologi dengan pencapaian akademik

Data yang digunakan adalah data sekunder yang bersumber dari University of California (UCLA) yang berjumlah 600 observasi yang terdiri dari dua kelompok variabel yaitu variabel psikologi dan variabel akademik.

Variabel yang berupa indikator psikologi, diantaranya:

$Y_1 = \text{Locus of control}$

$Y_2 = \text{Self-concept}$

$Y_3 = \text{Motivation}$

Variabel yang berupa indikator akademik adalah tes standar dalam:

$X_1 = \text{Read}$

$X_2 = \text{Write}$

$X_3 = \text{Math}$

$X_4 = \text{Science}$

$X_5 = \text{Female}$

Uji Signifikansi Korelasi Kanonik

a. Uji korelasi kanonik secara keseluruhan

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 3	.75436	11.71573	15.00	1634.65	.000
2 TO 3	.96143	2.94446	8.00	1186.00	.003
3 TO 3	.98919	2.16461	3.00	594.00	.091

Statistik uji :

$$B = -(n - 1 - \frac{1}{2}(p + q + 1)) \ln \Lambda = 167.58$$

$$\chi^2_{0,05;15}(24.99)$$

Kesimpulan : Tolak H_0 , karena nilai $B = 167.58 > 24.99$

Hasil yang sama juga dapat dilihat dari *p-value* sebesar 0.000 kurang dari 0.05 yang menunjukkan tolak H_0 . Artinya, bahwa paling sedikit ada satu korelasi kanonik yang signifikan, sehingga fungsi kanonik dapat dilakukan uji lebih lanjut.

Uji Signifikansi Korelasi Kanonik

b. Uji korelasi kanonik secara sebagian

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 3	.75436	11.71573	15.00	1634.65	.000
2 TO 3	.96143	2.94446	8.00	1186.00	.003
3 TO 3	.98919	2.16461	3.00	594.00	.091

Fungsi Kanonik ke	Wilks L.	Statistik Uji	df	Chi square Tabel	P-value	Keputusan	Kesimpulan
1	0.75436	167.580972	15	24.99579014	0	Tolak H0	Fungsi kanonik ke-1 secara individu signifikan.
2	0.96143	23.3837774	8	15.50731306	0.003	Tolak H0	Fungsi kanonik ke-2 secara individu signifikan.
3	0.98919	6.46153285	3	7.814727903	0.91	Gagal Tolak H0	Fungsi kanonik ke-3 secara individu tidak signifikan.

Hasil Analisis

Penentuan Koefisien Korelasi Kanonik dan Fungsi Kanonik

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	.27450	87.33628	87.33628	.46409	.21538
2	.02887	9.18537	96.52164	.16751	.02806
3	.01093	3.47836	100.00000	.10399	.01081

Pemilihan jumlah fungsi kanonik dapat dilihat dari nilai keragaman. Besarnya nilai proporsi keragaman menunjukkan baik tidaknya variabel kanonik yang dipilih untuk menerangkan keragaman asal. Semakin besar nilai proporsi keragaman maka semakin baik variabel-variabel kanonik yang dipilih menerangkan keragaman asal. Batasan yang digunakan untuk nilai proporsi bersifat relatif, sebagai acuan lebih besar dari 70%. Dari *output* di atas terlihat bahwa korelasi kanonik yang cukup besar yaitu fungsi ke-1 sebesar 0.46409 dengan keragaman sebesar 87.33628%. Sedangkan fungsi ke-2 dan ke-3 memiliki keragaman yang relatif kecil. Oleh karena itu, hanya nilai korelasi kanonik fungsi pertama saja yang digunakan untuk interpretasi.

Interpretasi Fungsi Kanonik

a. Bobot Kanonik

Standardized canonical coefficients for DEPENDENT variables
Function No.

Variable	1	2	3
Y1	.84042	-.41656	.44352
Y2	-.24788	-.83793	-.58326
Y3	.43267	.69480	-.68554

Bobot kanonik pada fungsi kanonik pertama urutan kontribusi terbesar sampai terkecil terhadap variabel kanonik Y adalah locus of control (Y1), motivation (Y3), dan self-concept (Y2). Sedangkan untuk fungsi kanonik kedua urutan kontribusinya adalah self-concept (Y2), motivation (Y3), dan locus of control (Y1).

Standardized canonical coefficients for COVARIATES
CAN. VAR.

COVARIATE	1	2	3
X1	.45080	-.04961	-.21601
X2	.34896	.40921	-.88810
X3	.22047	.03982	-.08848
X4	.04878	-.82660	1.06608
X5	.31504	.54057	.89443

Bobot kanonik untuk variabel X pada fungsi kanonik pertama urutan kontribusi terbesar sampai terkecil terhadap variabel X kanonik adalah read (X1), write (X2), female (X5), math (X3), dan science (X4). Untuk fungsi kanonik kedua urutan kontribusinya adalah science (X4), female (X5), write (X2), read (X1), dan math (X3).

Interpretasi Fungsi Kanonik

b. Loading Kanonik

Correlations between DEPENDENT and canonical variables
Function No.

Variable	1	2	3
Y1	.90405	-.38969	.17562
Y2	.02084	-.70874	-.70516
Y3	.56715	.35089	-.74513

Muatan kanonik variabel Y yang memiliki korelasi paling erat dengan fungsi kanonik pertama adalah locus of control (Y1) dengan nilai korelasi sebesar 0.90405. Sedangkan muatan kanonik variabel Y yang memiliki korelasi paling erat dengan fungsi kanonik kedua adalah self-concept (Y2) dengan nilai sebesar -0.70874.

Correlations between COVARIATES and canonical variables
CAN. VAR.

Covariate	1	2	3
X1	.84045	-.35883	-.13536
X2	.87654	.06484	-.25456
X3	.76395	-.29795	-.14776
X4	.65841	-.67680	.23036
X5	.36411	.75493	.54340

Muatan kanonik variabel X yang memiliki korelasi paling erat dengan fungsi kanonik pertama adalah write (X2), dengan nilai korelasi sebesar 0.87654, dan read (X1) dengan nilai korelasi sebesar 0.84045. Untuk muatan kanonik variabel X yang memiliki korelasi paling erat dengan fungsi kanonik kedua adalah female (X5) dengan nilai korelasi sebesar 0.75493.

Interpretasi Fungsi Kanonik

1. Bobot kanonik (*canonical weights*)

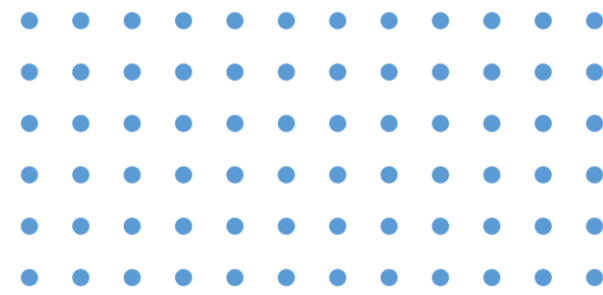
Bobot kanonik merupakan koefisien kanonik yang telah dibakukan, dapat diinterpretasikan sebagai besarnya kontribusi variabel asal terhadap variabel kanonik. Semakin besar nilai koefisien ini maka semakin besar kontribusi variabel yang bersangkutan terhadap variabel kanonik.

2. Muatan Kanonik (*canonical loadings*)

Muatan kanonik dapat dihitung dari korelasi antara variabel asal dengan masing-masing variabel kanoniknya. Semakin besar nilai muatan kanonik maka akan semakin penting peran variabel asal tersebut dalam kumpulan variabelnya. variabel asal yang memiliki nilai muatan kanonik besar (>0.5) akan dikatakan memiliki peran besar dalam kumpulan variabelnya, sedangkan tanda muatan kanonik menunjukkan arah hubungannya (Hair, dkk., 1998).

3. Muatan silang kanonik (*canonical cross-loadings*)

Muatan-silang kanonik dapat dihitung dari korelasi antara variabel asal dengan bukan variabel kanoniknya. Semakin besar nilai muatan silang mencerminkan semakin dekat hubungan fungsi kanonik yang bersangkutan dengan variabel asal. variabel asal yang memiliki nilai muatan silang kanonik besar (>0.5) akan dikatakan memiliki peranan besar dalam kumpulan variabelnya sedangkan tanda muatan silang kanonik menunjukkan arah hubungannya.



Korelasi Kanonik dengan R

CCA (Canonical Correlation Analysis) dengan R

1. Running ulang code berikut <https://zia207.quarto.pub/canonical-correlation%20-analysis.html>
2. Ganti data dengan data pada contoh soal dan lakukan analisis. Bagaimanakah hasilnya?
3. Tonton video berikut dan buat rangkuman untuk memahami CCA untuk data kuesioner



S1 Sains Data
FMIPA UNESA

Thank you

