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

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

Sample Covariance

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

Correlation

$$Correlation = \frac{Cov\,(x,y)}{\sigma x * \sigma y}$$

Matrix Variance Covariance ($\boldsymbol{\Sigma}$)

$$\Sigma = \begin{pmatrix} \sigma_{X_1X_1} & \cdots & \sigma_{X_1X_p} \\ \vdots & \ddots & \vdots \\ \sigma_{X_pX_1} & \cdots & \sigma_{X_pX_p} \end{pmatrix}.$$

Dua Variabel

$$\begin{array}{ccc} & x & y \\ \begin{array}{c} x \\ y \end{array} & \multicolumn{2}{c}{\begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix}} \end{array}$$

Tiga Variabel

$$\begin{array}{cccc} & x & y & z \\ \begin{array}{c} x \\ y \\ z \end{array} & \multicolumn{3}{c}{\begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix}} \end{array}$$

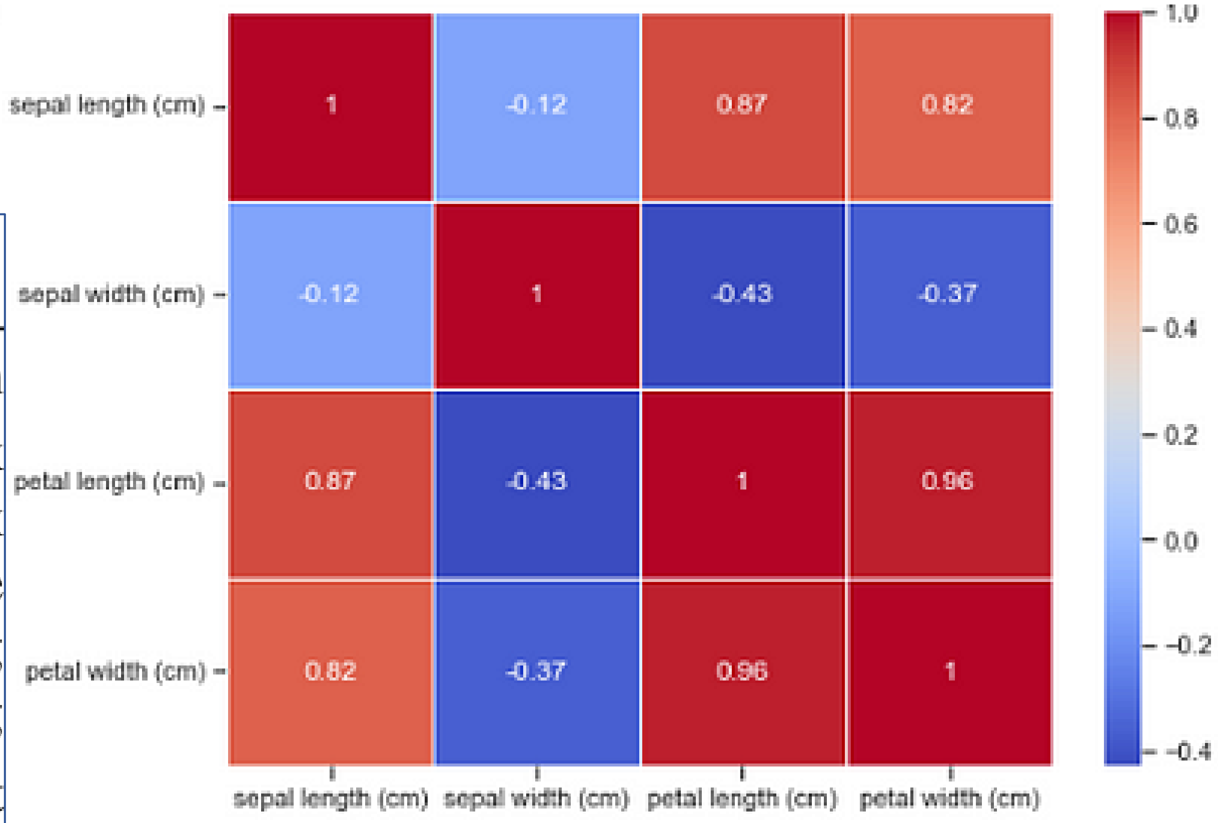
- How about invers of variance covariance matrix ($\boldsymbol{\Sigma}^{-1}$)?

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)

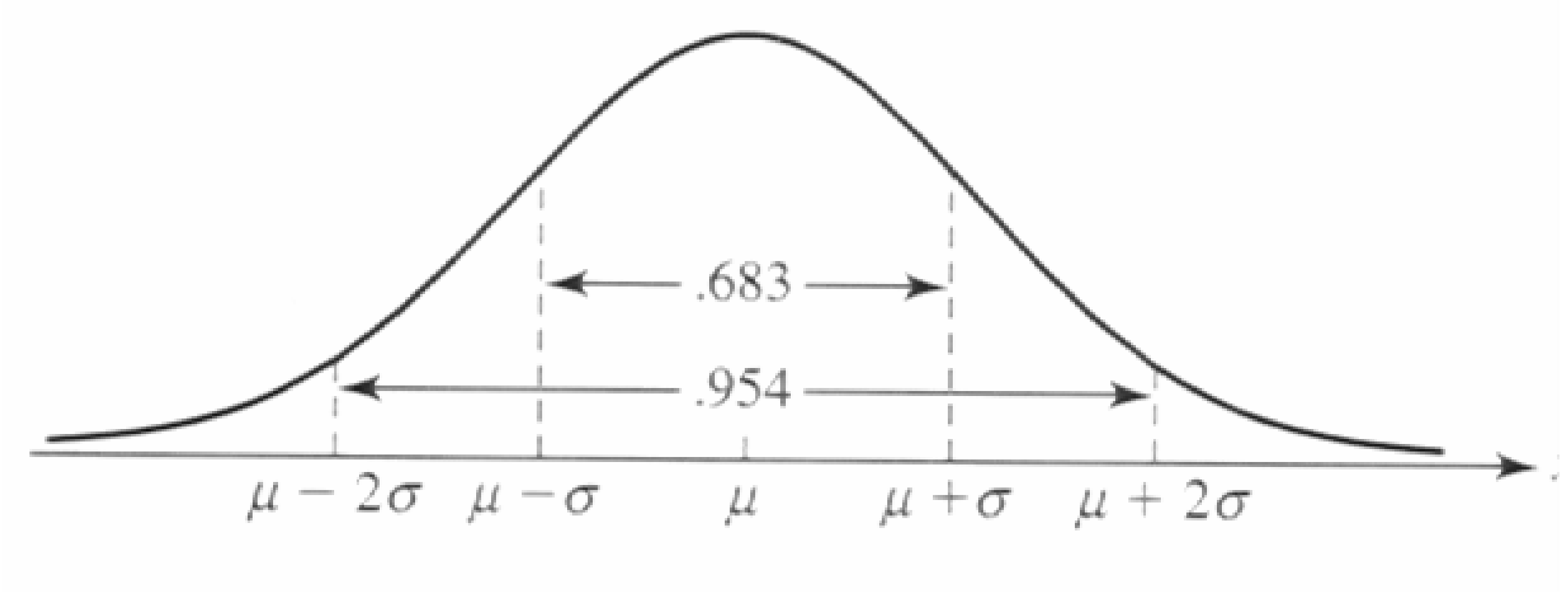


Distribusi Normal Multivariat

Distribusi Normal Univariat

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

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



Fungsi Kepadatan Normal *p*-dimensi

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

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

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

$\mathbf{x}$ is a sample from random vector

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

Kuadrat Jarak (Jarak Mahalanobis)

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

$$\Downarrow$$

$$(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

Bivariate Normal

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

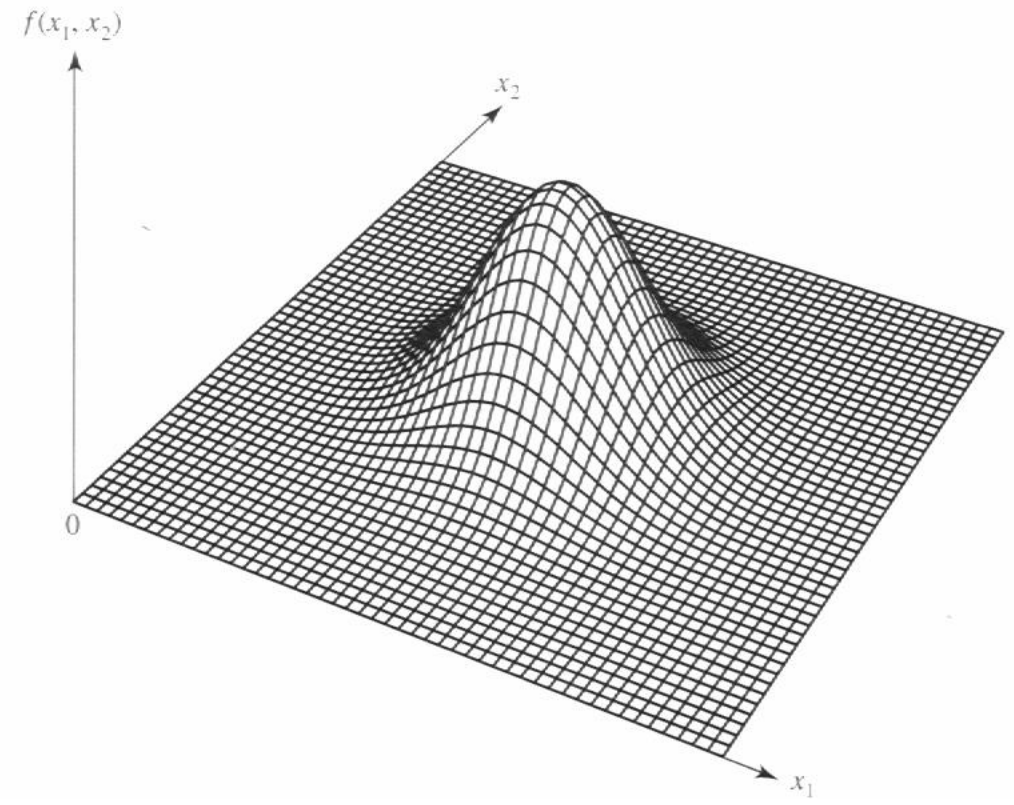
$$\sigma_{11} = \mathrm{Var}(X_1),\ \sigma_{22} = \mathrm{Var}(X_2)$$

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

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

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

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



Squared Distance

$$
(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})
$$

$$
= [x_1 - \mu_1,\, x_2 - \mu_2] \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}
\begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}
\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}
$$

$$
= \frac{1}{1-\rho_{12}^2} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) \right]
$$

Distribusi Normal Multivariat

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

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

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

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

Kasus Multivariat

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ independent observations from population (may not be multivariate normal) with mean $E(\mathbf{X}_\mathrm{i}) = \boldsymbol{\mu} \Rightarrow$

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

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

Central Limit Theorem Distribusi Multivariat

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$: independent observation from a population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$

$\Rightarrow \sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu})$ is approximately $N_p(\mathbf{0}, \boldsymbol{\Sigma})$

for large sample size $n >> p$

(quite good approximation for moderate n when the parent population is nearly normal)

Distribusi Hampiran Jarak Statistik

$$\overline{\mathbf{X}} : \text{nearly } N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}) \text{ for large sample size } n >> p$$

$$n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) : \text{approximately } \chi_p^2$$

for large n-p

$\mathbf{S}$ close to $\boldsymbol{\Sigma}$ with high probability when

n is large

$$\therefore n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) : \text{approximately } \chi_p^2$$

for large n-p

Review Nilai dan Vector Eigen + Kombinasi Linier

Nilai dan Vektor Eigen

DEFINITION

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.[1]

Contoh: Apakah u dan v adalah eigenvectors dari A.

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \text{and } \mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Solusi:

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

u merupakan eigenvector dari eigenvalue -4
namun v bukan eigenvector karena Av bukan perkalian dari v

Nilai dan Vektor Eigen

- Sebuah matrik **A**, **x** merupakan eigenvector dan λ adalah eigenvalue terkait jika $\mathbf{Ax} = \lambda\mathbf{x}$
  - **A** merupakan matrik *square* dan determinant dari $\mathbf{A} - \lambda\mathbf{I}$ harus sama dengan nol

$$\mathbf{Ax} - \lambda\mathbf{x} = 0 \text{ iff } (\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = 0$$

    - Trivial solution jika $\mathbf{x} = 0$
    - Nontrivial solution terjadi ketika $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- Apakah eigenvectors itu unique?
  - If **x** adalah eigenvector, maka $\beta\mathbf{x}$ juga eigenvector dan λ adalah eigenvalue

$$\mathbf{A}(\beta\mathbf{x}) = \beta(\mathbf{Ax}) = \beta(\lambda\mathbf{x}) = \lambda(\beta\mathbf{x})$$

Kombinasi Linier

Misalkan $\mathbf{X} : N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Jika

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{AX}$$

$$\mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

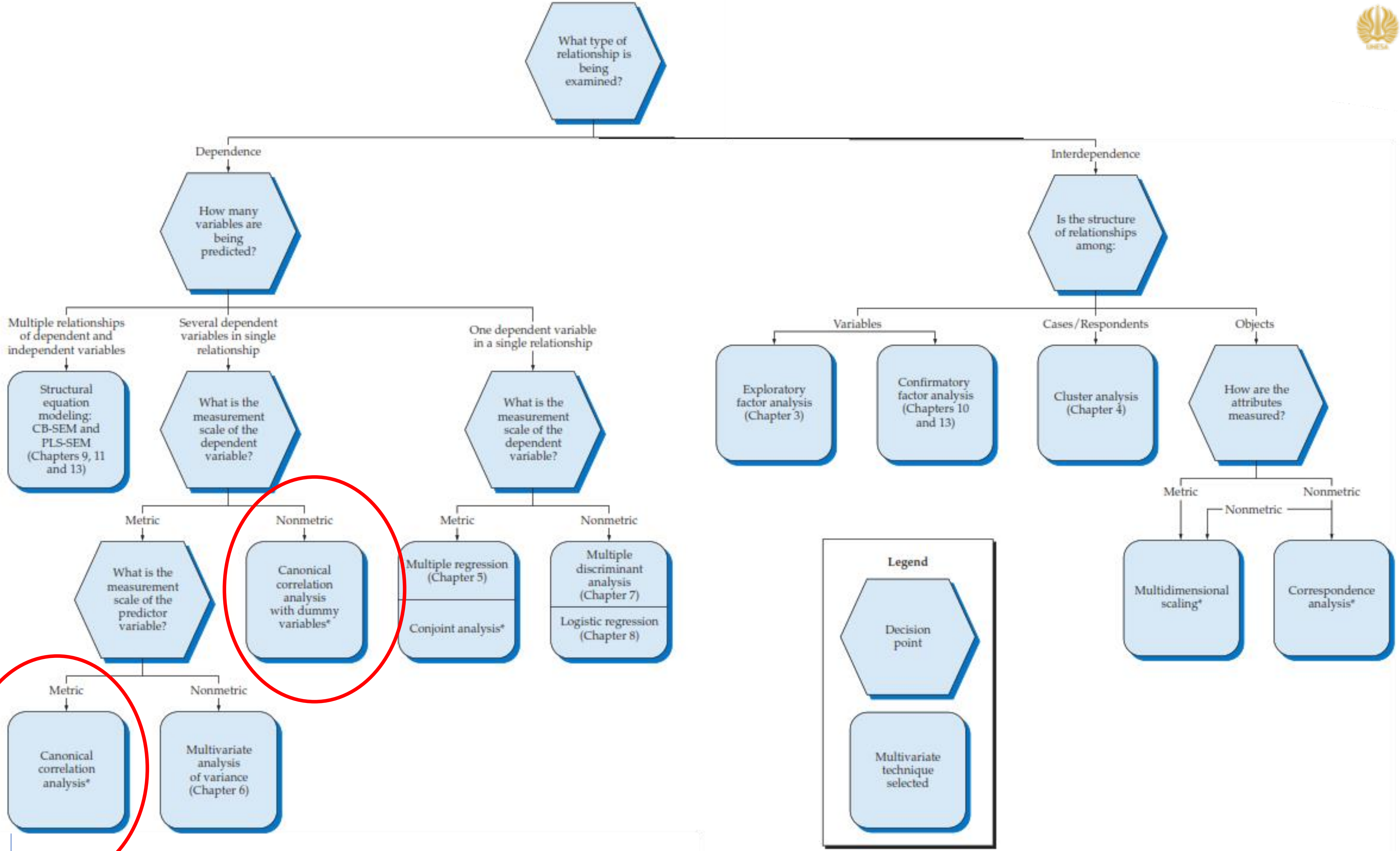
$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

Maka $\mathbf{AX} : N_2(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

can be verified with $Y_1 = X_1 - X_2, Y_2 = X_2 - X_3$

Analisis Korelasi Kanonik

Selecting a Multivariate Technique



- Joseph F. Hair, Multivariate Data Analysis, 8th Ed. Cengage, 2018

Korelasi Kanonik

- Analisis korelasi kanonik (*canonical correlation*) adalah salah satu teknik analisis multivariat untuk mengidentikasi 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, dan 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 x 1*) yang dinotasikan dengan vektor variabel random $\mathbf{X}^{(1)}$. Variabel kedua terdiri dari *q* variabel berukuran (*q x 1*) yang dinotasikan dengan vektor variabel random $\mathbf{X}^{(2)}$. dimana p≤q. Misalkan karakteristik dari vektor variabel random $\mathbf{X}^{(1)}$ dan $\mathbf{X}^{(2)}$ adalah sebagai berikut:

$$E(\mathbf{X}^{(1)}) = \boldsymbol{\mu}^{(1)}; \qquad \mathrm{Cov}(\mathbf{X}^{(1)}) = \boldsymbol{\Sigma}_{11}$$
$$E(\mathbf{X}^{(2)}) = \boldsymbol{\mu}^{(2)}; \qquad \mathrm{Cov}(\mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{22}$$
$$\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' \qquad (1)$$

Kombinasi linear dari kedua kelompok variabel dapat dituliskan sebagai berikut :

$$U = \boldsymbol{a}'\boldsymbol{X}^{(1)} = a_1X_1^{(1)} + a_2X_2^{(1)} + \cdots + a_pX_p^{(1)}$$
$$V = \boldsymbol{b}'\boldsymbol{X}^{(2)} = a_1X_1^{(2)} + a_2X_2^{(2)} + \cdots + a_qX_1^{(2)} \qquad (2)$$

Sehingga diperoleh

$$\mathrm{Var}(U) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)})\,\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}$$
$$\mathrm{Var}(V) = \mathbf{b}'\,\mathrm{Cov}(\mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b} \qquad (3)$$
$$\mathrm{Cov}(U,V) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\,\mathbf{b} = \mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}$$

Korelasi kanonik diperoleh dengan menghitung

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

Pendugaan Koefisien Kanonik

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

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

- Pasangan variabel kanonik ke-*k*, k = 2, 3, …, p adalah :

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

maximizes

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

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

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

Basu & Mandal (2010)

Pendugaan Koefisien Kanonik (lanjutan)

Definisi :

- Variabel kanonik pertama memiliki korelasi terbesar pertama

  $U_1 = a_1' \boldsymbol{X}^{(1)}$ $\qquad Var(U_1) = 1$

  $V_1 = b_1' \boldsymbol{X}^{(2)}$ $\qquad Var(V_1) = 1$

  $maksimum\ corr(U_1, V_1) = \rho_1$

- Variabel kanonik kedua memiliki korelasi terbesar kedua

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

  $V_2 = b_2' \boldsymbol{X}^{(2)}$ $\qquad Var(V_2) = 1$ $\qquad Cov(V_1, V_2) = 0$

  $maksimum\ corr(U_2, V_2) = \rho_2$

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

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

  $V_k = b_k' \boldsymbol{X}^{(2)}$ $\qquad Var(k) = 1$ $\qquad Cov(V_k, V_l) = 0$

  $maksimum\ corr(U_k, V_k) = \rho_k$

Pasangan variabel kanonik memiliki sifat :

$$\begin{aligned} \operatorname{Var}(U_k) &= \operatorname{Var}(V_k) = 1 \\ \operatorname{Cov}(U_k, U_\ell) &= \operatorname{Corr}(U_k, U_\ell) = 0 \quad k \neq \ell \\ \operatorname{Cov}(V_k, V_\ell) &= \operatorname{Corr}(V_k, V_\ell) = 0 \quad k \neq \ell \\ \operatorname{Cov}(U_k, V_\ell) &= \operatorname{Corr}(U_k, V_\ell) = 0 \quad k \neq \ell \end{aligned}$$

Contoh

	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

$Matriks\ Korelasi\ (Pearson)\ antar\ Fitur$

	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

$$S_{xx}^{-1} = \begin{bmatrix} 0,020 & -0,043 \\ -0,043 & 0,150 \end{bmatrix} \qquad S_{yy}^{-1} = \begin{bmatrix} 0,089 & -0,061 \\ -0,061 & 0.050 \end{bmatrix}$$

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

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

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

$$R_x = \begin{bmatrix} -0,093 & -0,081 \\ 0,989 & 0,650 \end{bmatrix} \xrightarrow{\text{Eigen Vektor}} v_x = \begin{array}{c|cc} & 1 & 2 \\ \hline x_1 & 0.131 & -0.526 \\ x_2 & -0.991 & 0.850 \end{array}$$

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

$$S_{xx}^{-1} = \begin{bmatrix} 0,020 & -0,043 \\ -0,043 & 0,150 \end{bmatrix} \qquad S_{yy}^{-1} = \begin{bmatrix} 0,089 & -0,061 \\ -0,061 & 0.050 \end{bmatrix}$$

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

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

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

$$R_y = \begin{bmatrix} \quad \end{bmatrix} \qquad v_y = \begin{bmatrix} \quad \end{bmatrix}$$

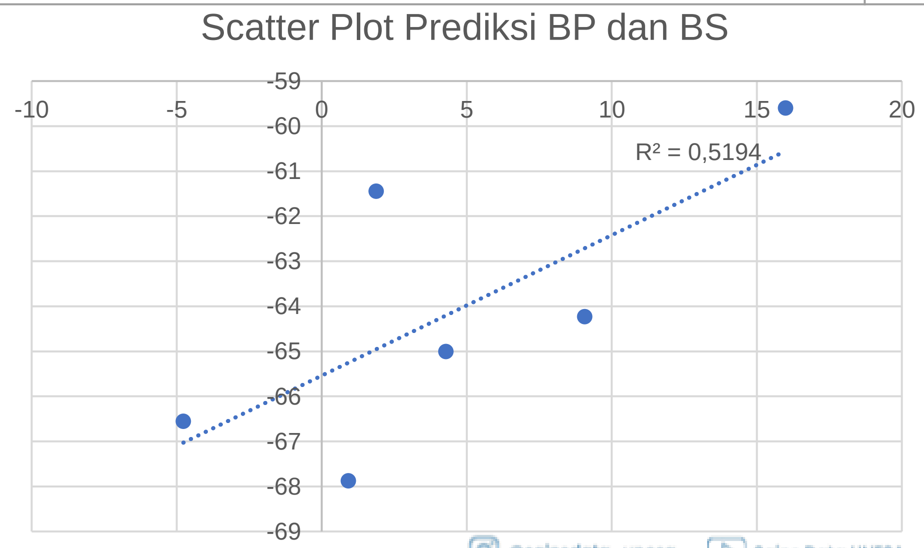
Variabel kanonik ke dua (V_1) adalah

$$BS = \cdots Y_1 - \cdots . Y_2$$

$$v_y = \begin{array}{c|cc} & 1 & 2 \\ \hline y_1 & \boxed{0.426} & -0.846 \\ y_2 & \boxed{-0.905} & 0.533 \end{array}$$

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		



$$Korelasi\ Kanonik$$
$$corr(BP, BS) = \sqrt{R^2} = \sqrt{0{,}5194} = 0{,}72$$

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		

$$Korelasi\ Kanonik$$
$$corr(BP, BS) = \sqrt{R^2}$$
$$= \sqrt{0{,}5194} = 0{,}72$$

	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 :

$H_0 : \boldsymbol{\Sigma}_{12} = 0 \; (\rho_1^* = \rho_2^* = \cdots = \rho_p^* = 0)$

$H_1 : \boldsymbol{\Sigma}_{12} \neq 0 \; (\rho_1^* \neq \rho_2^* \neq \cdots \neq \rho_p^* = 0)$

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 H0 jika statistik uji > $\chi^2_{pq}(\alpha)$

Uji korelasi kanonik secara sebagian

Hipotesis :

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

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

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 H0 jika statistik uji > $\chi^2_{(p-k)(q-k)}(\alpha)$

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 meunjukkan 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 https://www.youtube.com/watch?v=7TKvgpe3YOQ dan buat rangkuman untuk memahami CCA untuk data kuesioner

S1 Sains Data
FMIPA UNESA

Thank you