Metode Statistika Multivariat

Dr. Moch. Fandi Ansori

Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro

August 25, 2023

Daftar Isi

Kenapa Harus Belajar Metode Statistika Multivariat?

Sebagai seorang matematikawan yang akan bekerja di berbagai bidang, mengapa kita perlu belajar Metode Statistika Multivariat?

1) Di kehidupan nyata, sebagian besar berbagai hal dipengaruhi oleh lebih dari satu variabel.
2) Analisis multivariat adalah sekumpulan metode untuk menginvestigasi data dengan banyak variabel (≥ 2).

Pengetahuan dasar yang harus dikuasai sebelum kuliah ini:

- Statistik univariat maupun bivariat, dan dasar-dasar multidimensi linier ($Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$), serta fungsi taklinier seperti kuadrat, eksponensial, dan logaritma natural.
- Vektor, matriks dan operasinya, serta dasar-dasar aljabar linier lainnya.

Objektif dari Investigasi Data

Objektif dari investigasi sains yang melibatkan analisis multivariat biasanya memuat hal-hal berikut:

(1) **Reduksi data dan penyederhanaan struktur**. Fenomena yang dikaji direpresentasikan sesederhana mungkin tanpa membuang informasi yang berharga.

(2) **Menyortir dan mengelompokkan**. Menggolongkan objek atau variabel yang "serupa" berdasarkan karakteristik terukur.

(3) **Investigasi ketergantungan antar variabel**. Kita tertarik untuk mengetahui hubungan antar variabel.

(4) **Prediksi**. Hubungan antar variabel ditentukan dengan tujuan untuk memprediksi nilai satu atau lebih variabel.

(5) **Konstruksi dan uji hipotesis**. Untuk validasi asumsi atau menguatkan keyakinan awal.

Metode Statistika Multivariat **Minggu ke-1**

Aspek Analisis Multivariat

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Data Multivariat

Data multivariat muncul manakala seorang peneliti memilih sejumlah p variabel, dan nilai dari semua variabel dicatat sebanyak n item, individu, atau unit percobaan yang berbeda.

Misal x_{jk} menyatakan pengukuran variabel ke-k pada item ke-j. Maka, secara keseluruhan, data multivariat tersebut dapat ditulis sebagai:

	Variable 1	Variable 2	$\cdots$	Variable k	$\cdots$	Variable p
Item 1:	x_{11}	x_{12}	$\cdots$	x_{1k}	$\cdots$	x_{1p}
Item 2:	x_{21}	x_{22}	$\cdots$	x_{2k}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item j:	x_{j1}	x_{j2}	$\cdots$	x_{jk}	$\cdots$	x_{jp}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Item n:	x_{n1}	x_{n2}	$\cdots$	x_{nk}	$\cdots$	x_{np}

Atau dalam bentuk **matriks**:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Contoh

Example 1

Sebanyak empat kuitansi dari toko buku universitas dipilih untuk menyelidiki tren penjualan buku. Setiap kuitansi yang diberikan antara lain jumlah buku yang terjual dan jumlah total setiap penjualan. Misalkan variabel pertama adalah total penjualan (dolar) dan variabel kedua adalah jumlah buku yang terjual. Kemudian kita dapat menganggap angka yang sesuai pada kuitansi sebagai empat pengukuran pada dua variabel. Misalkan datanya, dalam bentuk tabel, adalah:

Variable 1 (dollar sales):	42	52	48	58
Variable 2 (number of books):	4	5	4	3

Matriks datanya adalah $X = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$.

Statistik Deskriptif

Rata-rata sampel ($\overline{x}_k$) dan **variansi sampel** (s_{kk}) untuk variabel ke-k dihitung sebagai

$$\overline{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk}, \qquad s_{kk} = \frac{1}{n}\sum_{j=1}^{n}\left(x_{jk} - \overline{x}_k\right)^2.$$

Nilai $\sqrt{s_{kk}}$ disebut **simpangan baku**.
Covariance sampel (s_{ik}) antara variabel ke-i dan ke-k dihitung sebagai

$$s_{ik} = \frac{1}{n}\sum_{j=1}^{n}\left(x_{ji} - \overline{x}_i\right)\left(x_{jk} - \overline{x}_k\right).$$

Sifat *covariance*: $s_{ik} = s_{ki}$.

Koefisien korelasi sampel (r_{ik}) untuk variabel ke-i dan ke-k dihitung sebagai

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}}.$$

Sifat koefisien korelasi: $|r_{ik}| \leq 1$.

Matriks Statistik Deskriptif

Statistik deskriptif yang dihitung dari n pengukuran terhadap p variabel dapat dituliskan dalam bentuk vektor/matriks.

$$\text{Matriks rata-rata sampel} \quad \bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}, \quad \text{Matriks variansi sampel} \quad \mathbf{V} = \begin{bmatrix} s_{11} \\ s_{22} \\ \vdots \\ s_{pp} \end{bmatrix}$$

$$\text{Matriks } \textit{covariance} \text{ sampel} \quad \mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix},$$

$$\text{Matriks korelasi sampel} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Contoh

Example 2

Pandang Example 1. Diperoleh rata-rata sampel:

$$\overline{x}_1 = \frac{1}{4}\sum_{j=1}^{4} x_{j1} = \frac{1}{4}(42+52+48+58) = 50, \quad \overline{x}_2 = \frac{1}{4}\sum_{j=1}^{4} x_{j2} = \frac{1}{4}(4+5+4+3) = 4, \quad \bar{\mathbf{x}} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix} = \begin{bmatrix} 50 \\ 4 \end{bmatrix}.$$

Sedangkan untuk variansi dan covariance sampel:

$$s_{11} = \frac{1}{4}\sum_{j=1}^{4}(x_{j1}-\overline{x}_1)^2 = \frac{1}{4}\left[(42-50)^2+(52-50)^2+(48-50)^2+(58-50)^2\right] = 34$$

$$s_{22} = \frac{1}{4}\sum_{j=1}^{4}(x_{j2}-\overline{x}_2)^2 = \frac{1}{4}\left[(4-4)^2+(5-4)^2+(4-4)^2+(3-4)^2\right] = 0.5$$

$$s_{12} = \frac{1}{4}\sum_{j=1}^{4}(x_{j1}-\overline{x}_1)(x_{j2}-\overline{x}_2) = \frac{1}{4}\left[(42-50)(4-4)+\cdots+(58-50)(3-4)\right] = -1.5, \quad s_{21} = s_{12} = -1.5.$$

Contoh (cont.)

Sehingga, matriks covariance-nya adalah;

$$\mathbf{S} = \begin{bmatrix} 34 & -1.5 \\ -1.5 & 0.5 \end{bmatrix}$$

Untuk korelasi, kita peroleh

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} = \frac{-1.5}{\sqrt{34}\sqrt{0.5}} = -0.36, \quad r_{21} = r_{12} = -0.36,$$
$$\mathbf{R} = \begin{bmatrix} 1 & -0.36 \\ -0.36 & 1 \end{bmatrix}.$$

Jarak Statistik

Kita ingin menghitung jarak dari titik $P = (x_1, x_2)$ ke titik pusat $Q = (0, 0)$. Variabel x_1 dan x_2 mungkin saja memiliki rentang yang berbeda, agar mereka "setara", keduanya harus dibagi dengan **simpangan baku**-nya. Sehingga, kita bisa mendefinisikan **jarak statistik** dari P ke Q sebagai

$$d(P, Q) = \sqrt{\left(\frac{x_1}{\sqrt{s_{11}}}\right)^2 + \left(\frac{x_2}{\sqrt{s_{22}}}\right)^2}.$$

Secara khusus, ketika $d(P, Q) = c$ konstan, maka kita punyai persamaan elips

$$\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2.$$

Secara umum, **jarak statistik** antara titik $P = (x_1, x_2, \ldots, x_p)$ dengan suatu titik tetap $Q = (y_1, y_2, \ldots, y_p)$ didefinisikan sebagai

$$d(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \cdots + \frac{(x_p - y_p)^2}{s_{pp}}}.$$

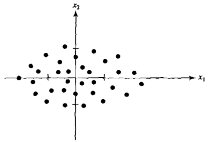


Fig. 1. Diagram pencar (*scatter plot*).

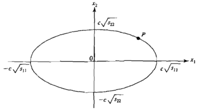


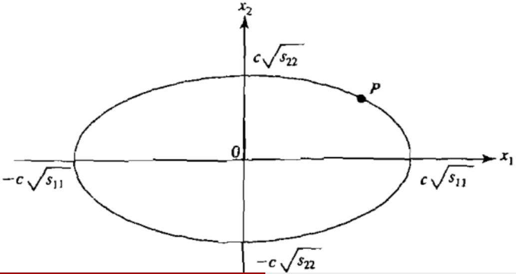
Fig. 2. Elips.

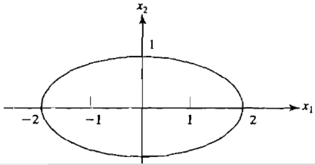
Contoh

Sejumlah pengukuran dalam variabel (x_1, x_2) memiliki statistik deskriptif: $\overline{x}_1 = \overline{x}_2 = 0$, $s_{11} = 4$, dan $s_{22} = 1$. Misalkan pengukuran x_1 tidak berkorelasi dengan x_2. Maka, jarak statistik dari sembarang titik $P(x_1, x_2)$ ke pusat $O(0, 0)$ adalah

$$d(O, P) = \sqrt{\frac{x_1^2}{4} + \frac{x_2^2}{1}}.$$

Semua titik (x_1, x_2) yang memiliki jarak statistik konstan 1 dari pusat memenuhi persamaan elips $\frac{x_1^2}{4} + \frac{x_2^2}{1} = 1$.





Transformasi Jarak Statistik

Definisi jarak statistik di atas mengasumsikan semua variabel saling independen. Terkadang, kita temui keadaan dimana ada beberapa variabel yang berkorelasi. Untuk itu, definisi jarak statistik sebelumnya perlu ditransformasi menjadi

$$d(P,Q) = \sqrt{\sum_{j=1}^{p} a_{jj}(x_j - y_j)^2 + \sum_{i=1}^{p} \sum_{j \neq i} a_{ij}(x_i - y_i)(x_j - y_j)}.$$

Koefisien-koefisien a_{ij} di atas dapat dirangkum dalam bentuk matriks

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{12} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{bmatrix}.$$

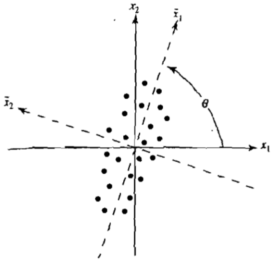


Fig. 3. Diagram pencar hasil transformasi

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Soal Latihan

Misalkan tujuh **sampel acak** diambil dengan variabel x_1 dan x_2. Datanya disajikan sebagai berikut

x_1	3	4	2	6	8	2	5
x_2	5	5.5	4	7	10	5	7.5

(1) Tentukan vektor rata-rata sampel.

(2) Tentukan vektor variansi sampel.

(3) Tentukan matriks *covariance* sampel.

(4) Tentukan matriks koefisien korelasi sampel.

(5) Tentukan jarak statistik ke titik $O(0, 0)$.

(6) Tentukan jarak statistik ke titik rata-rata $Q(\overline{x}_1, \overline{x}_2)$.

Metode Statistika Multivariat **Minggu ke-2**

Distribusi Normal Multivariat

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Aljabar Matriks

- **Transpose**, **determinan**, dan **inverse** dari matriks A berturut-turut dinotasikan dengan A', $|A|$, danA^{-1}.

Example 3

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 7 & -4 & 6 \end{bmatrix} \Rightarrow A' = \begin{bmatrix} 2 & 7 \\ 1 & -4 \\ 3 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 3 \\ 6 & 4 \end{bmatrix} \Rightarrow |B| = -14,\ B^{-1} = \begin{bmatrix} -\frac{4}{14} & \frac{3}{14} \\ \frac{6}{14} & -\frac{1}{14} \end{bmatrix}.$$

- Matriks **simetri** adalah matriks yang memenuhi $A' = A$. Matriks simetri dikatakan **definit positif** jika $\forall x \neq \mathbf{0}$ berlaku $\mathbf{x}' A \mathbf{x} > 0$.

Example 4

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \Rightarrow A' = A \text{ dan } \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 - 2x_1x_2 + x_2^2 = (x_1 - x_2)^2.$$

Aljabar Matriks (cont.)

- **Nilai eigen** dan **vektor eigen**. Nilai eigen λ pada persamaan $A\mathbf{x} = \lambda\mathbf{x}$ dicari dengan cara $|A - \lambda I| = 0$ atau $|\lambda I - A| = 0$. Nilai eigen yang diperoleh disubstitusikan ke persamaan awal $A\mathbf{x} = \lambda\mathbf{x}$ untuk mendapatkan vektor eigen $\mathbf{x}$. Bentuk **normalisasi** dari vektor $\mathbf{x}$ adalah $\mathbf{e} = \frac{1}{\|\mathbf{x}\|}\mathbf{x}$.

Example 5

$A = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$ memiliki nilai eigen $\lambda_1 = 6$ dan $\lambda_2 = -4$. Vektor eigen normalisasi yang bersesuaian $\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ dan $\mathbf{e}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

Aljabar Matriks (cont.)

- Matriks definit positif A dapat di**dekomposisi** menjadi $A = P\Lambda P'$, dengan $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ adalah matriks diagonal nilai eigen dan $P = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$ adalah matriks vektor eigen yang dinormalisasi dan bersifat $PP' = P'P = I$.

Example 6

Berdasarkan Example 5, diperoleh

$$P\Lambda P' = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & -4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} = A.$$

Distribusi Normal Univariat

Peubah acak X ber**distribusi normal** dengan parameter rata-rata μ dan variansi σ^2 (ditulis $X \sim N(\mu, \sigma^2)$) memiliki **fungsi kepadatan peluang**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty.$$

Sifat:

1) Ekspektasi $E(X) = \mu$.
2) Variansi $Var(X) = \sigma^2$.

Distribusi normal juga biasa disebut sebagai distribusi Gauss

Karl Fiedrich Gauss (1777-1855)

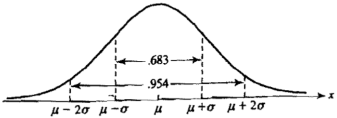


Fig. 4. Grafik $f(x)$ (kurva lonceng/kurva Gauss).

Distribusi Normal Bivariat

Peubah acak-peubah acak $X_1 \sim N(\mu_1, \sigma_1^2)$ dan $X_2 \sim N(\mu_2, \sigma_2^2)$ ber**distribusi normal bivariat** memiliki fungsi kepadatan peluang

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

dengan

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad -\infty < x_1 < \infty, \quad -\infty < x_2 < \infty,$$

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

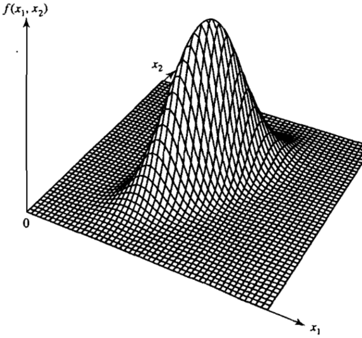


Fig. 5. Grafik $f(x_1, x_2)$.

Distribusi Normal Multivariat

Peubah acak-peubah acak $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\ldots$, $X_p \sim N(\mu_p, \sigma_p^2)$ ber**distribusi normal multivariat** memiliki fungsi kepadatan peluang

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

dengan

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \quad -\infty < x_i < \infty, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix},$$

$|\Sigma| = \text{determinan}, \quad \Sigma^{-1} = \text{inverse}.$

Soal Latihan

Misalkan peubah acak $X_1 \sim N(\mu_1, \sigma_1^2)$ dan $X_2 \sim N(\mu_2, \sigma_2^2)$, dengan $\mu_1 = \mu_2 = 0$ dan matriks covariance Σ dengan $\sigma_{11} = \sigma_{22}$.

1. Tentukan nilai eigen dan vektor eigen dari matriks Σ.
2. Misalkan $\sigma_{11} = \sigma_{22} = 0.25$ dan $\sigma_{12} = -0.1$. Uraikan ekspresi berikut $-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$.
3. Tuliskan fungsi kepadatan peluang dari X_1 dan X_2, yaitu $f(x_1, x_2)$.

Metode Statistika Multivariat **Minggu ke-3**

Kontur Densitas Elips pada Distribusi Normal Multivariat

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Matriks Definit Positif dan Elips

Misalkan matriks definit positif $A_{2\times 2}$ memiliki nilai eigen λ_1 dan λ_2 dengan $0 < \lambda_1 < \lambda_2$ serta vektor eigen dalam bentuk normalisasi $\mathbf{e}_1$ dan $\mathbf{e}_2$. Misalkan matriks A dapat didekomposisi menjadi $A = P\Lambda P'$, dengan $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, dan $P = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}$. Perhatikan bahwa

$$A = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2'.$$

Ambil sembarang $\mathbf{x} \in \mathbb{R}^2$. Maka

$$\mathbf{x}' A \mathbf{x} = \lambda_1 (\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2 (\mathbf{x}'\mathbf{e}_2)^2.$$

Jika $\mathbf{x}' A \mathbf{x} = c^2$ (c konstan) dan dimisalkan $y_1 = \mathbf{x}'\mathbf{e}_1$ dan $y_2 = \mathbf{x}'\mathbf{e}_2$, maka diperoleh persamaan elips (dengan pusat di $(0,0)$)

$$c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2.$$

Matriks Definit Positif dan Elips (cont.)

Perhatikan bahwa, $\mathbf{x} = \frac{c}{\sqrt{\lambda_1}}\mathbf{e}_1$ memenuhi persamaan

$$\mathbf{x}'A\mathbf{x} = (\frac{c}{\sqrt{\lambda_1}}\mathbf{e}_1)'(\lambda_1)(\frac{c}{\sqrt{\lambda_1}}\mathbf{e}_1) = \lambda_1(\frac{c^2}{\lambda_1})\mathbf{e}_1'\mathbf{e}_1$$
$$= \lambda_1(\frac{c^2}{\lambda_1})\|\mathbf{e}_1\|^2 = c^2.$$

Hal ini dikarenakan $\mathbf{e}_1$ adalah vektor normalisasi. Hal diatas berlaku juga untuk $\mathbf{x} = \frac{c}{\sqrt{\lambda_2}}\mathbf{e}_2$.

Jadi, titik-titik (x_1, x_2) yang memiliki jarak c akan berada di elips yang axis-nya diberikan oleh vektor eigen $\mathbf{e}$ dengan panjang yang proporsional terhadap $\frac{1}{\sqrt{\lambda}}$. Ilustrasinya diberikan pada Fig. 1.

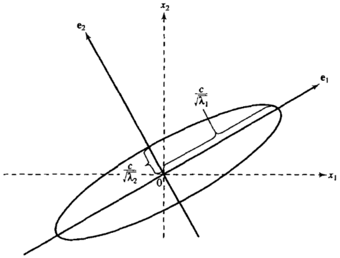


Fig. 1. Grafik elips berdasarkan matriks definit positif.

Kontur Densitas Elips pada Distribusi Normal Bivariat

Perhatikan kembali fungsi kepadatan peluang dari distribusi normal bivariat

$$f(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2 |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Kontur dari $f(x_1, x_2)$ diperoleh dengan membuat $f(x_1, x_2) =$ konstan. Karena $\frac{1}{(\sqrt{2\pi})^2 |\Sigma|^{1/2}}$ konstan, maka untuk mendapatkan kontur dari f haruslah $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) =$ konstan.

Misalkan ditulis

$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = c^2. \tag{1}$$

Lemma 7

Misal Σ adalah matriks definit positif. Jika Σ^{-1} ada dan λ adalah nilai eigen dari Σ, maka $\frac{1}{\lambda}$ adalah nilai eigen dari Σ^{-1}.

Buktikan!

Kontur Densitas Elips pada Distribusi Normal Bivariat (cont.)

Misal λ_1 dan λ_2 adalah nilai eigen dari matriks Σ dan vektor eigen normalisasi yang bersesuaian $\mathbf{e}_1$ dan $\mathbf{e}_2$.

Karena Σ^{-1} memiliki nilai eigen $\frac{1}{\lambda}$, maka perhatikan bahwa Pers. (1) merupakan persamaan elips dengan pusat (μ_1, μ_2) dan memiliki axis dengan arah $\mathbf{e}_1$ (dengan panjang $\frac{c}{1/\sqrt{\lambda_1}} = c\sqrt{\lambda_1}$) dan $\mathbf{e}_2$ (dengan panjang $\frac{c}{1/\sqrt{\lambda_2}} = c\sqrt{\lambda_2}$).

Pers. (1) disebut **kontur densitas elips** dari distribusi normal bivariat.

Contoh Kontur Densitas Elips pada Distribusi Normal Bivariat

Example 8

Perhatikan kembali soal latihan pada bab 2. Misalkan peubah acak $X_1 \sim N(\mu_1, \sigma_{11})$ dan $X_2 \sim N(\mu_2, \sigma_{22})$ berdistribusi normal bivariat dengan matriks *covariance* $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ dengan $\sigma_{11} = \sigma_{22}$.
Nilai eigennya adalah

$$\lambda_1 = \sigma_{11} + \sigma_{12} \text{ dan } \lambda_2 = \sigma_{11} - \sigma_{12},$$

dengan vektor eigen normalisasi $\mathbf{e}_1' = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ dan $\mathbf{e}_2' = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$.
Panjang dan lebar dari kontur densitas elips-nya adalah $c\sqrt{\sigma_{11} + \sigma_{12}}$ dan $c\sqrt{\sigma_{11} - \sigma_{12}}$.

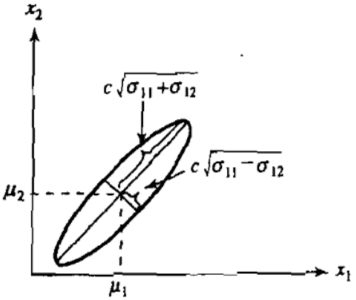


Fig. 2. Kontur densitas elips pada distribusi normal bivariat.

Kontur Densitas Elips pada Distribusi Normal Multivariat

Dengan cara serupa, untuk distribusi normal multivariat dengan

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix},$$

diperoleh **kontur densitas elips** dalam bentuk persamaan

$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = c^2,$$

dengan pusat elips $\mu = (\mu_1, \ldots, \mu_p)$ dan ukuran elips $c\sqrt{\lambda_i}$, dengan λ_i adalah nilai eigen matriks Σ dengan vektor eigen normalisasi $\mathbf{e}_i$.

Soal Latihan

Misal suatu distribusi normal bivariat dengan $\mu_1 = 1$, $\mu_2 = 3$, $\sigma_{11} = 2$, $\sigma_{22} = 1$, $\rho_{12} = -0.8$.

1) Tuliskan matriks *covariance* Σ.
2) Hitung $|\Sigma|$ dan Σ^{-1}.
3) Cari nilai eigen dari Σ.
4) Cari vektor eigen normalisasi dari Σ.
5) Tuliskan persamaan kontur densitas elips, serta tentukan panjang dan lebar ukuran elips tersebut. Gambarkan elipsnya juga.

Metode Statistika Multivariat **Minggu ke-5**

Uji Hipotesis Vektor Rata-rata untuk Satu Populasi

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Uji Hipotesis μ_0 sebagai Nilai dari rata-rata Distribusi Normal

Teori univariat untuk menentukan apakah suatu nilai μ_0 adalah nilai yang masuk akal untuk rata-rata populasi μ. Dari sudut pandang uji hipotesis,

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

Disini H_0 disebut **hipotesis nol**, dan H_1 disebut **hipotesis alternatif**.

Jika $X_1, X_2, \ldots, X_n$adalah sampel acak dari suatu populasi yang berdistribusi normal, **uji statistik** yang sesuai adalah

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}, \quad \text{dengan} \quad \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$$

Uji statistik ini memiliki **distribusi Student's** t (William Sealy Gosset, pen name *Student*, 1876-1937) dengan derajat kebebasan $t-1$. Kita **tolak** H_0 jika nilai pengamatan $|t|$ melebihi persentase tertentu dari distribusi t.

Hubungannya dengan Distribusi t

Menolak H_0 ketika $|t|$ nilainya besar ekuivalen dengan menolak H_0 jika

$$t^2 = \frac{(\overline{X} - \mu_0)^2}{s^2/n} = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0) \quad \text{nilainya besar.}$$

Ketika $\overline{X}$ dan s^2 nilainya diketahui, uji di atas menjadi:

$$\text{Tolak } H_0 \text{ jika } n(\overline{x} - \mu_0)(s^2)^{-1}(\overline{x} - \mu_0) > t_{n-1}^2(\alpha/2) \tag{2}$$

dengan $t_{n-1}(\alpha/2)$ menyatakan persentil atas ke-$100(\alpha/2)$ dari distribusi t, dan α disebut tingkat siginikan.

Dari (2), diperoleh

$$\text{Terima } H_0 \text{ jika } \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2)$$

yang memberikan selang kepercayaan

$$\overline{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \overline{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}$$

Generalisasi

Generalisasi

$$T^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)$$

dengan $\overline{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$, $\quad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X} - \overline{\mathbf{X}})'$, dan $\boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{10} \\ \vdots \\ \mu_{p0} \end{bmatrix}$. Statistik T^2 disebut Hotelling's T^2 (Harold Hotelling, 1895-1973).

Distribusi T^2 ekuivalen dengan

$$\text{distribusi } \frac{(n-1)p}{(n-p)} F_{p,n-p},$$

dengan $F_{p,n-p}$ menyatakan distribusi F dengan derajat kebebasan p dan $n-p$ (Dist. Fisher-Snedecor; Robert Fisher 1890-1962, George W. Snedecor 1881-1974)

Rangkuman

Uji hipotesis μ_0 sebagai nilai dari rata-rata distribusi normal:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

Pada tingkat signifikan α, kita **tolak** H_0 jika nilai observasi

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha),$$

dengan n adalah banyaknya pengamatan dan p adalah banyaknya variabel yang diamati.

Selang kepercayaan untuk $\boldsymbol{\mu}$ adalah

$$\bar{x}_j - t_{n-1}(\alpha/2)\sqrt{\frac{s_{jj}}{n}} \leq \mu_j \leq \bar{x}_j + t_{n-1}(\alpha/2)\sqrt{\frac{s_{jj}}{n}}$$

Catatan: Untuk menghitung $t_{n-1}(\alpha/2)$ dan $F_{p,n-p}(\alpha)$ kita membutuhkan tabel distribusi t dan F yang bersesuaian dengan tingkat signifikan α.

Contoh

Misal sampel acak berukuran $n = 3$ dari distribusi bivariat $\mathbf{X} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$. Uji hipotesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ untuk $\boldsymbol{\mu}_0 = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ dengan tingkat signifikan $\alpha = 5\%$. Dan tentukan selang kepercayaannya.

Jawab: Kita hitung

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{6+10+8}{3} \\ \frac{9+6+3}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \quad s_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{3-1} = -3$$

$$s_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{3-1} = 4, \quad s_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{3-1} = 9.$$

Sehingga $\mathbf{S} = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$ dan $\mathbf{S}^{-1} = \begin{bmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{bmatrix}$. Kemudian dihitung

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = 3 \begin{bmatrix} 8-9 & 6-5 \end{bmatrix} \begin{bmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{bmatrix} \begin{bmatrix} 8-9 \\ 6-5 \end{bmatrix} = 7/9 = 0.78.$$

Contoh (lanjutan)

Di sisi lain, dari tabel distribusi F diperoleh
$\frac{(3-1)2}{(3-2)}F_{2,3-2}(0.05) = 4F_{2,1}(0.05) = 4 \cdot 199.5 = 798.$

Karena $T^2 = 0.78 < 798$, maka kesimpulannya kita terima $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Jadi $\boldsymbol{\mu} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$.

Selanjutnya, selang kepercayaannya diberikan oleh

- $\overline{x}_1 - t_{3-1}(0.05/2)\sqrt{\frac{s_{11}}{3}} \le \mu_1 \le \overline{x}_1 + t_{3-1}(0.05/2)\sqrt{\frac{s_{11}}{3}}$
  $\Leftrightarrow 8 - 4.303\sqrt{4/3} \le \mu_1 \le 8 + 4.303\sqrt{4/3} \quad \Leftrightarrow \quad 3.03 \le \mu_1 \le 12.97$
- $\overline{x}_2 - t_{3-1}(0.05/2)\sqrt{\frac{s_{22}}{3}} \le \mu_2 \le \overline{x}_2 + t_{3-1}(0.05/2)\sqrt{\frac{s_{22}}{3}}$
  $\Leftrightarrow 6 - 4.303\sqrt{9/3} \le \mu_2 \le 6 + 4.303\sqrt{9/3} \quad \Leftrightarrow \quad -1.45 \le \mu_2 \le 13.45$

Metode Statistika Multivariat **Minggu ke-6**

Uji Hipotesis Vektor Rata-rata untuk Dua Populasi

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Rangkuman Statistik Dua Populasi

Misal sampel acak berukuran n_1 dari populasi 1 dan sampel acak berukuran n_2 dari populasi 2. Keduanya memiliki observasi p variabel.

	Sampel	Statistik	
Pop. 1	$\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$	$\bar{\mathbf{x}}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{x}_{1j}$	$\mathbf{S}_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)'$
Pop. 2	$\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$	$\bar{\mathbf{x}}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{x}_{2j}$	$\mathbf{S}_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)'$

Hipotesis yang ingin diajukan adalah:

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \mathbf{0}$$

Asumsi Terkait Struktur Data

Asumsi terkait struktur data:

1. Sampel $\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1}$ adalah sampel acak dari suatu populasi p-variat dengan vektor rata-rata $\boldsymbol{\mu}_1$ dan matriks *covariance* $\boldsymbol{\Sigma}_1$.
2. Sampel $\mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$ adalah sampel acak dari suatu populasi p-variat dengan vektor rata-rata $\boldsymbol{\mu}_2$ dan matriks *covariance* $\boldsymbol{\Sigma}_2$.
3. Sampel $\mathbf{X}_{11}, \mathbf{X}_{12}, \ldots, \mathbf{X}_{1n_1}$ independen terhadap $\mathbf{X}_{21}, \mathbf{X}_{22}, \ldots, \mathbf{X}_{2n_2}$.

Asumsi tambahan, ketika n_1 dan n_2 keduanya kecil:

1. Kedua populasi berdistribusi normal multivariat.
2. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

Ketika $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, maka untuk mengestimasi $\boldsymbol{\Sigma}$ dilakukan pembobotan

$$\mathbf{S}_{\text{pooled}} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}$$

Statistik dari Selisih rata-rata Dua Populasi

Untuk menguji hipotesis $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$, pertama-tama dilakukan

$$E(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = E(\overline{\mathbf{X}}_1) - E(\overline{\mathbf{X}}_2) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2.$$

Dari asumsi keindependenan, diperoleh $\overline{\mathbf{X}}_1$ dan $\overline{\mathbf{X}}_2$ saling bebas, dan $Cov(\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2) = \mathbf{0}$. Sehingga,

$$Cov(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) = Cov(\overline{\mathbf{X}}_1) - Cov(\overline{\mathbf{X}}_2) = \frac{1}{n_1}\boldsymbol{\Sigma} + \frac{1}{n_2}\boldsymbol{\Sigma} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}$$

Karena $\mathbf{S}_{\text{pooled}}$ mengestimasi $\boldsymbol{\Sigma}$, maka $\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}$ adalah estimator untuk $Cov(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$.

Uji Hipotesis rata-rata dari Dua Populasi

Pada tingkat signifikan α, hipotesis $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$ kita **tolak** jika

$$T^2 = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - \boldsymbol{\delta}_0)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\text{pooled}} \right]^{-1} (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - \boldsymbol{\delta}_0)$$
$$> \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}(\alpha)$$

Contoh

Diketahui dua grup sampel acak dengan ukuran sama-sama 50, dan informasi $\bar{\mathbf{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}$, $\mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$, $\bar{\mathbf{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}$, $\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$. Ujilah hipotesis $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ untuk tingkat signifikan $\alpha = 5\%$.

Jawab: Diketahui $\boldsymbol{\delta}_0 = \mathbf{0}$. Pertama, dihitung

$$\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 = \begin{bmatrix} -1.9 \\ 0.2 \end{bmatrix}, \quad \mathbf{S}_{\text{pooled}} = \frac{49 \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix} + 49 \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}}{98} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

Kemudian, dihitung

$$T^2 = \begin{bmatrix} -1.9 - 0 & 0.2 - 0 \end{bmatrix} \left\{ \left(\frac{1}{50} + \frac{1}{50} \right) \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix} \right\}^{-1} \begin{bmatrix} -1.9 - 0 \\ 0.2 - 0 \end{bmatrix} = 52.47.$$

Sedangkan, dari tabel distribusi F, kita peroleh $\frac{(50+50-2)2}{(50+50-2-1)} F_{2,97}(0.05) = \frac{196}{97} 3.09 = 6.24$. Karena $T^2 = 52.47 > 6.24$, maka H_0 ditolak. Jadi $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.

Metode Statistika Multivariat **Minggu ke-6**

Uji Hipotesis terhadap Kesamaan Dua Matriks Covariance

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023

Uji Hipotesis Kesamaan Dua Matriks *Covariance*

Salah satu asumsi yang dipakai dalam uji membandingkan vektor rata-rata dari dua populasi adalah matriks *covariance* dari dua populasi tersebut sama.

Sekarang, akan dipelajari uji hipotesis kesamaan dua matriks *covariance*. Pandang hipotesis berikut

$$H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad \text{vs} \quad H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$$

Dengan mengasumsikan populasi normal bivariat, statistik rasio likelihood untuk menguji $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ adalah

$$\Lambda = \left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_{\text{pooled}}|} \right)^{\frac{n_1 - 1}{2}} \left(\frac{|\mathbf{S}_2|}{|\mathbf{S}_{\text{pooled}}|} \right)^{\frac{n_2 - 1}{2}}$$

Uji Box's M

Salah satu uji yang biasanya digunakan untuk menguji kesamaan dua matriks *covariance* adalah uji Box's M (George E. P. Box, 1919-2013). Uji Box's M berdasarkan aproksimasi χ^2 dari distribusi sampel $-2\ln\Lambda$.

Dengan mengambil $-2\ln\Lambda = M$, diperoleh

$$M = (n_1 + n_2 - 2)\ln|\mathbf{S}_{\text{pooled}}| - \left[(n_1 - 1)\ln|\mathbf{S}_1| + (n_2 - 1)\ln|\mathbf{S}_2|\right].$$

Hitung

$$u = \left[\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) - \frac{1}{n_1 + n_2 - 2}\right]\frac{2p^2 + 3p - 1}{6(p+1)}.$$

Pada tingkat signifikan α, hipotesis $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ **ditolak** jika

$$(1 - u)M > \chi^2_{p(p+1)/2}(\alpha)$$

Catatan: Jadi, untuk uji Box's M, kita memerlukan tabel distribusi χ^2.

Contoh

Diketahui dua grup sampel acak dengan ukuran sama-sama 50, dan informasi
$\mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$, $\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$. Ujilah hipotesis $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ untuk tingkat signifikan $\alpha = 5\%$.

Jawab: Terlebih dahulu dihitung

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}, \quad |\mathbf{S}_1| = 11, \quad |\mathbf{S}_2| = 7, \quad |\mathbf{S}_{\text{pooled}}| = 9.$$

Selanjutnya, dihitung

$$\begin{aligned} M &= 98 \ln 9 - [49 \ln 11 + 49 \ln 7] = 2.48, \\ u &= [(1/49 + 1/49) - 1/98] \frac{2(2)^2 + 3(2) - 1}{6(2+1)} = 0.02, \\ (1-u)M &= (1 - 0.02)2.48 = 2.43, \\ \chi^2_3(0.05) &= 7.81. \end{aligned}$$

Karena $(1-u)M = 2.43 < 7.81$, maka H_0 diterima. Jadi $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$.

Metode Statistika Multivariat **Minggu ke-7**

KUIS I

Dr. Moch. Fandi Ansori
Departemen Matematika
Fakultas Sains dan Matematika
Universitas Diponegoro
2023