MODUL 1 ANALISIS MULTIVARIAT
SEBARAN NORMAL MULTIVARIAT

Mika Alvionita Sitinjak

11 September 2025

Analisis multivariat adalah cabang statistika yang mempelajari data dengan lebih dari satu variabel secara bersamaan untuk memahami hubungan, pola, maupun struktur yang tersembunyi di dalamnya. Data multivariat biasanya direpresentasikan dalam bentuk matriks data, di mana baris menunjukkan observasi dan kolom menunjukkan variabel. Melalui representasi matriks ini, hubungan antar variabel dihitung menggunakan matriks kovarians atau matriks korelasi. Selanjutnya, teknik-teknik analisis multivariat seperti analisis faktor, atau analisis korelasi kanonik sangat bergantung pada konsep aljabar linear, khususnya nilai eigen dan vektor eigen. Nilai eigen digunakan untuk mengukur besarnya keragaman data yang dapat dijelaskan oleh suatu komponen, sedangkan vektor eigen menunjukkan arah atau kombinasi linier variabel yang membentuk sumbu baru dalam ruang data. Dengan demikian, matriks berperan sebagai dasar matematis yang memungkinkan data multivariat diproses, direduksi dimensinya, serta diinterpretasikan dalam bentuk yang lebih sederhana namun tetap informatif.

Salah satu penerapan analisis multivariat banyak ditemukan di berbagai bidang, misalnya di bidang kesehatan. Bayangkan sebuah penelitian ingin memahami faktor-faktor yang memengaruhi risiko penyakit jantung pada sekelompok pasien. Data yang dikumpulkan mencakup berbagai variabel: tekanan darah, kadar kolesterol, indeks massa tubuh (IMT), kadar gula darah, kebiasaan merokok, dan tingkat aktivitas fisik. Karena jumlah variabel yang banyak, peneliti menggunakan analisis faktor untuk menyederhanakan data. Dari hasil perhitungan matriks korelasi dan ekstraksi eigenvalue, ternyata variabel-variabel tersebut dapat dipadatkan menjadi dua faktor utama: faktor metabolik (tekanan darah, gula darah, dan kolesterol) dan faktor gaya hidup (merokok, IMT, dan aktivitas fisik). Dengan cara ini, analisis multivariat membantu peneliti memahami pola yang tersembunyi, sehingga lebih mudah mengidentifikasi kelompok pasien dengan risiko tinggi dan menyusun strategi pencegahan yang tepat

Secara konsep, analisis multivariat berasal dari lebih dari satu variabel acak pada saat yang bersamaan. Ketika hanya ada satu variabel, kita cukup melihat rataan (mean) dan varians untuk menggambarkan pusat dan penyebaran data. Namun, dalam

konteks multivariat dengan p variabel, kita membutuhkan generalisasi konsep tersebut.
Pertama, rataan tidak lagi berupa satu angka, melainkan berbentuk vektor mean $\mu=(\mu_1,\mu_2,\dots,\mu_p)'$ yang merupakan rata dari masing-masing variabel. Vektor ini memberi gambaran di titik mana data multivariat "berpusat" dalam ruang berdimensi p.
Kedua, varians juga tidak lagi cukup untuk menggambarkan penyebaran, karena variabel-variabel bisa saling berhubungan. Untuk itu, digunakan matriks kovarians Σ, yaitu matriks berukuran p×p yang berisi varians dari masing-masing variabel pada diagonal utama, serta kovarians antar variabel pada elemen di luar diagonal. Matriks kovarians inilah yang menggambarkan distribusi, arah sebaran, dan keeratan hubungan antar variable (korelasi).

ARRAYS OF BASIC DESCRIPTIVE STATISTICS

Sample means
$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Sample variances and covariances
$$\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \qquad (1\text{-}8)$$

Sample correlations
$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

Dengan masing – masing nilainya diperoleh sebagai berikut.
Sample means
$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad j = 1,2,\dots,p$$
Sample variances and covariances

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)\left(x_{ik} - \overline{x_k}\right)$$

Sample correlations

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj} \cdot s_{kk}}}$$

Contoh.

Diberikan dua variabel dengan 7 pengamatan masing-masing:

Variabel 1	5	4	6	2	2	8	3
Variabel 2	5	5.5	4	7	10	5	7.5

Tentukan **Sample means , Sample variances and covariances, Sample correlations**

Jawaban.

Sample means

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\,, \quad j = 1,2,\ldots,p$$

$$\overline{x_1} \quad = \frac{5+4+6+2+2+8+3}{7} = \frac{30}{7} \approx 4.285714$$

$$\overline{x_2} \quad = \frac{5+5.5+4+7+10+5+7.5}{7} = \frac{44}{7} \approx 6.285714$$

Sample variances and covariances

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)\left(x_{ik} - \overline{x_k}\right)$$

$$s_{11} = \frac{1}{6}\sum_{i=1}^{7}(x_{1i} - \overline{x_1})^2 \approx 4.904762$$

$$s_{22} = \frac{1}{6}\sum_{i=1}^{7}(x_{2i} - \overline{x_2})^2 \approx 4.154762$$

$$s_{12} = \frac{1}{6}\sum_{i=1}^{7}(x_{1i} - \overline{x_1})(x_{2i} - \overline{x_2}) \approx -3.511905$$

Sehingga,

$$s = \begin{bmatrix} 4.9 & -3.5 \\ 4.1 & -3.5 \end{bmatrix}$$

Sample correlations

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj} \cdot s_{kk}}}$$

$$s_1 = \sqrt{s_{11}} \approx 2.21467,$$

$$s_2 = \sqrt{s_{22}} \approx 2.03832$$

$$r_{12} = \frac{s_{12}}{s_1 \cdot s_2} = \frac{-3.511905}{2.21467 \times 2.03832} \approx -0.7780$$

Sehingga,

$$r = \begin{bmatrix} 1 & -0.77 \\ -0.77 & 1 \end{bmatrix}$$

Latihan. (Buku Johnson hal 38 no. 1.4)

Diketahui data dari 10 perusahaan industri terbesar di Amerika Serikat (dalam jutaan dolar):

Perusahaan	Penjualan (x_1)	Laba (x_2)	Aset (x_3)
General Motors	126,974	4,224	173,297
Ford	96,933	3,835	160,893
Exxon	86,656	3,51	83,219
IBM	63,438	3,758	77,734
General Electric	55,264	3,939	128,344
Mobil	50,976	1,809	39,08
Philip Morris	39,069	2,946	38,528
Chrysler	36,156	359	51,038
Du Pont	35,209	2,48	34,715
Texaco	32,416	2,413	25,636

a. Buatlah diagram pencar (scatter diagram) antara variabel penjualan x_1d an laba x_2. Selain itu, gambarkan juga diagram titik marginal (marginal dot diagrams) untuk masing-masing variabel x_1 dan x_2. Berikan komentar mengenai pola atau bentuk diagram tersebut.
b. Hitunglah Sample means, Sample variances and covariances dan Sample correlations. Berikan interpretasi dari nilai korelasi r_{12} yang diperoleh, khususnya hubungan antara penjualan dan laba perusahaan.

Sebaran Normal Ganda

Dalam analisis multivariat, salah satu konsep paling mendasar adalah sebaran normal ganda (multivariate normal distribution). Sebaran ini merupakan perluasan dari sebaran normal univariat ke dimensi yang lebih tinggi. Jika pada sebaran normal univariat data hanya memiliki satu rata-rata (μ) dan satu varians (σ^2), maka pada sebaran normal ganda, data dengan p variabel acak mempunyai vektor mean $\mu=(\mu_1,\mu_2,\dots,\mu_p)^t$ dan matriks kovarians Σ berukuran p×p, yang tidak hanya menggambarkan penyebaran tiap variabel, tetapi juga hubungan antarvariabel. Fungsi kepadatan probabilitas dari sebaran normal ganda diberikan oleh:

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

dengan x adalah vektor pengamatan berukuran p |Σ| adalah determinan dari matriks kovarians, dan Σ^{-1} adalah invers dari matriks kovarians.

Contoh.

Diberikan dua variabel dengan 7 pengamatan masing-masing:

Variabel 1	5	4	6	2	2	8	3
Variabel 2	5	5.5	4	7	10	5	7.5

Tentukan fungsi kepadatan probabilitasnya.

Jawaban.

Nilai mean

$$\overline{x_1} = \frac{5+4+6+2+2+8+3}{7} = \frac{30}{7} \approx 4.2857$$

$$\overline{x_2} = \frac{5+5.5+4+7+10+5+7.5}{7} = \frac{44}{7} \approx 6.2857$$

$$\mu = \begin{bmatrix} 4.2857 \\ 6.2857 \end{bmatrix}$$

Matriks Kovariannya

$$s_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \overline{x_1})^2 \approx 4.9048$$

$$s_{22} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{2i} - \overline{x_2})^2 \approx 4.1548$$

$$s_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \overline{x_1})(x_{2i} - \overline{x_2}) \approx -3.5119$$

$$\Sigma = \begin{bmatrix} 4.9 & -3.5 \\ -3.5 & 4.1 \end{bmatrix}$$

Nilai dari (x-μ) yaitu

$$(x-\mu)' = \begin{bmatrix} 0.7143 & -1.2857 \\ -0.2857 & -0.7857 \\ 1.7143 & -2.2857 \\ -2.2857 & 0.7143 \\ -2.2857 & 3.7143 \\ 3.7143 & -1.2857 \\ -1.2857 & 1.2143 \end{bmatrix}$$

$$(x-\mu) = \begin{bmatrix} 0.7143 & -0.2857 & 1.7143 & -2.2857 & -2.2857 & 3.7143 & -1.2857 \\ -1.2857 & -0.7857 & -2.2857 & 0.7143 & 3.7143 & -1.2857 & 1.2143 \end{bmatrix}$$

Nilai dari $|\boldsymbol{\Sigma}|$

$$|\Sigma| = 4.904762 \cdot 4.154762 - (-3.511905) \cdot (-3.511905) = 8.044642047619$$

Nilai dari $\boldsymbol{\Sigma^{-1}}$

$$\Sigma^{-1} = \begin{bmatrix} 0.516 & 0.437 \\ 0.437 & 0.61 \end{bmatrix}$$

Fungsi Kepadatan probabilitasnya yaitu

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

$f(x)$

$$= \frac{1}{(2\pi^{2/2})\sqrt{8.04}} e^{-\begin{bmatrix} 0.7143 & -1.2857 \\ -0.2857 & -0.7857 \\ 1.7143 & -2.2857 \\ -2.2857 & 0.7143 \\ -2.2857 & 3.7143 \\ 3.7143 & -1.2857 \\ -1.2857 & 1.2143 \end{bmatrix} \begin{bmatrix} 0.516 & 0.437 \\ 0.437 & 0.61 \end{bmatrix} \begin{bmatrix} 0.7143 & -0.2857 & 1.7143 & -2.2857 & -2.2857 & 3.7143 & -1.2857 \\ -1.2857 & -0.7857 & -2.2857 & 0.7143 & 3.7143 & -1.2857 & 1.2143 \end{bmatrix}/2}$$

Latihan. (Buku Johnson hal 38 no. 1.4)
Berdasarkan Latihan diatas carilah fungsi kepadatan probabilitasnya.

Langkah-Langkah Plot Uji Normalitas Multivariat

1. Hitung jarak Mahalanobis:
$$d_{ii}^2 = \left(x_{(i)} - \mu\right)' \Sigma^{-1} \left(x_{(i)} - \mu\right)$$

2. Beri peringkat nilai (d_{ii}^2).

3. Carilah nilai khi-kuadrat dari:
$$\chi_p^2\left(\frac{i - 1/2}{n}\right)$$
dengan derajat bebas p.

4. Buat plot $\chi_p^2\left(\frac{i-1/2}{n}\right)$ terhadap (d_{ii}^2)
   Jika pola hubungannya mengikuti garis lurus, maka data tersebut dapat dikatakan menyebar normal ganda.

Contoh.
Diberikan dua variabel dengan 7 pengamatan masing-masing:

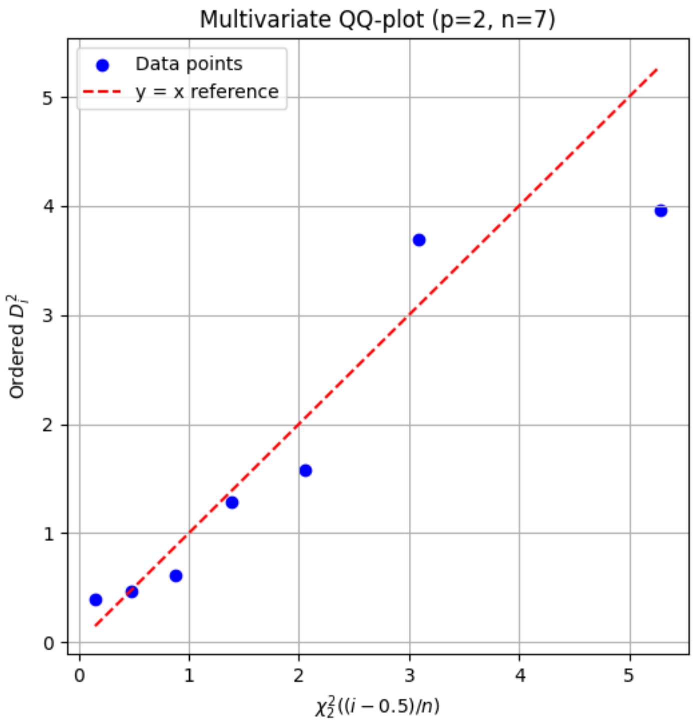
Variabel 1	5	4	6	2	2	8	3
Variabel 2	5	5.5	4	7	10	5	7.5

Plot QQ untuk dataset diatas.

1. Hitung jarak Mahalanobis:
$$d_{ii}^2 = \left(x_{(i)} - \mu\right)' \Sigma^{-1} \left(x_{(i)} - \mu\right)$$
   Untuk observasi 1 di titik (5,5)
$$\left(x_{(i)} - \mu\right) = \begin{bmatrix} 5 - 4.2857 \\ 5 - 6.2857 \end{bmatrix} = \begin{bmatrix} 0.7143 \\ -1.2857 \end{bmatrix}$$
$$d_{11}^2 = [0.7143 \quad -1.2857] \begin{bmatrix} 0.516 & 0.437 \\ 0.437 & 0.61 \end{bmatrix} \begin{bmatrix} 0.7143 \\ -1.2857 \end{bmatrix} = 0{,}46$$
   Lanjutkan sampai ke observasi ke-7
2. Urutkan nilai d_{11}^2

Variabel 1	Variabel 2	d_{ii}^2	Urutan d_{ii}^2	$\chi_p^2\left(\frac{i-1/2}{n}\right)$
5	5	0.46	0.38	0.14
4	5.5	0.61	0.46	0.48
6	4	1.2	0.61	0.88
2	7	1.5	1.28	1.38
2	10	3.6	1.58	2.05
8	5	3.9	3.69	3.08
3	7.5	0.3	3.96	5.2

3. Buat plot



Latihan. (Buku Johnson hal 38 no. 1.4)
Buatlah grafik QQ plot multivariat berdasarkan contoh diatas.