MODUL 3 ANALISIS MULTIVARIAT
MANOVA SATU ARAH
Mika Alvionita Sitinjak
16 September 2025

Sering kali, lebih dari dua populasi perlu dibandingkan. Sampel acak, yang dikumpulkan dari masing-masing dari ggg populasi, disusun sebagai berikut:

Populasi 1: $X_{11}, X_{12}, \ldots, X_{1n1}$

.

.

.

Populasi g: $X_{g1}, X_{g2}, \ldots, X_{gn}$

MANOVA digunakan pertama kali untuk menyelidiki apakah vektor rata-rata populasi sama, dan jika tidak, komponen rata-rata mana yang berbeda secara signifikan.

Asumsi mengenai Struktur Data untuk One-way MANOVA

1. $X_{l1}, X_{l2}, \ldots, X_{ln}$ adalah sampel acak berukuran n_l dari suatu populasi dengan rata-rata μ_l, untuk $l=1,2,\ldots,g$. Sampel acak dari populasi yang berbeda adalah saling independen.
2. Semua populasi memiliki matriks kovarians yang sama Σ.
3. Setiap populasi berdistribusi normal multivariat.

Kondisi (3) dapat dilonggarkan dengan merujuk pada Teorema Limit Tengah (Result 4.13) ketika ukuran sampel n_l besar.

Sebuah tinjauan mengenai analisis varians univariat (ANOVA) akan memfasilitasi pembahasan kita mengenai asumsi multivariat dan metode penyelesaiannya.

Dalam statistika multivariat, inferensia vektor nilai tengah berfokus pada pengujian hipotesis mengenai rata-rata multivariat dari suatu populasi atau perbandingan rata-rata antara beberapa kelompok dengan mempertimbangkan korelasi antar variabel respon. Hal ini menjadi dasar bagi MANOVA (Multivariate Analysis of Variance), di mana uji dilakukan terhadap kesamaan vektor rata-rata antar kelompok dengan melibatkan lebih dari satu variabel dependen secara bersamaan. Dengan kata lain, MANOVA dapat dipandang sebagai perluasan dari analisis varian (ANOVA) yang berbasis pada inferensia vektor nilai tengah, sehingga tidak hanya menilai perbedaan rata-rata pada satu variabel, tetapi juga memperhitungkan hubungan antar variabel respon dalam membedakan kelompok yang dibandingkan.

Model MANOVA sebagai berikut.

MANOVA MODEL FOR COMPARING *g* POPULATION MEAN VECTORS

$$\mathbf{X}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \mathbf{e}_{\ell j}, \quad j = 1, 2, \ldots, n_\ell \quad \text{and} \quad \ell = 1, 2, \ldots, g \qquad (6\text{-}34)$$

where $\mathbf{e}_{\ell j}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ variables. Here the parameter vector $\boldsymbol{\mu}$ is an overall mean (level), and $\boldsymbol{\tau}_\ell$ represents the ℓth treatment effect with $\sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \mathbf{0}$.

Model MANOVA di atas merupakan perluasan dari ANOVA ke kasus multivariat, yaitu ketika variabel respon lebih dari satu dan saling berkorelasi. Secara umum, model dituliskan sebagai $X_{\ell j}=\mu+\tau_\ell+e_{\ell j}$ adalah vektor pengamatan ke-j pada kelompok ke-ℓ. Parameter μ merepresentasikan rata-rata umum (grand mean vector), sedangkan τ_ℓ menunjukkan efek perlakuan (treatment effect) dari kelompok ke-ℓ yang menyatakan perbedaan rata-rata kelompok tersebut terhadap rata-rata keseluruhan. Komponen $e_{\ell j}$ merupakan vektor galat yang berdistribusi normal multivariat dengan rata-rata nol dan matriks kovarians Σ, serta dianggap saling bebas antar individu. Agar model dapat diidentifikasi dengan baik, efek perlakuan harus memenuhi syarat keseimbangan $\sum_{l=1}^{g} n_l \tau_l = 0$. Dengan kerangka ini, MANOVA digunakan untuk menguji hipotesis nol bahwa semua vektor rata-rata populasi dari *g* kelompok adalah sama, dibandingkan dengan hipotesis alternatif bahwa setidaknya ada satu kelompok yang berbeda, sehingga memungkinkan peneliti mengevaluasi perbedaan antar kelompok secara simultan pada lebih dari satu variabel respon.

Tabel MANOVA, yang digunakan untuk mendapatkan model diatas.

MANOVA TABLE FOR COMPARING POPULATION MEAN VECTORS

Source of variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})'$	$g - 1$
Residual (Error)	$\mathbf{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)'$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (corrected for the mean)	$\mathbf{B} + \mathbf{W} = \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})'$	$\sum_{\ell=1}^{g} n_\ell - 1$

Berdasarkan Tabel Manova diatas yang digunakan **untuk** digunakan untuk menguji apakah terdapat perbedaan yang signifikan antara rata-rata vektor dari beberapa grup (misalnya perlakuan atau kelompok) dan tidak hanya membandingkan satu variabel respon, tetapi beberapa variabel respon sekaligus. Tabel diatas menjelaskan tentang

1. Sumber variasi

   Tabel ini memecah total variasi dalam data menjadi dua komponen:
   - Treatment (B) ➔ variasi antar kelompok (perbedaan rata-rata grup).
   - Residual/Error (W) ➔ variasi dalam kelompok (perbedaan individu terhadap rata-rata grup).
   - Total (B+W) ➔ gabungan keduanya, total variasi dari data.
2. Matrix of Sum of Squares and Cross Products (SSP)
   - B disebut *between-group sum of squares and cross products matrix* ➔ mengukur variasi antar kelompok.
   - W disebut *within-group sum of squares and cross products matrix* ➔ mengukur variasi dalam kelompok.
   - B + W = total variasi (setelah dikoreksi dengan rata-rata keseluruhan).

   Ini adalah bentuk multivariat dari *Sum of Squares (SS)* dalam ANOVA, tapi dalam bentuk matriks, bukan sekadar angka.
3. Derajat kebebasan (degrees of freedom, d.f.)
   - Untuk Treatment (B): g - 1, di mana g adalah jumlah grup.
   - Untuk Error (W): Σ n_{ℓ} - g, di mana n_l adalah jumlah sampel tiap grup.
   - Untuk Total: Σ n_{ℓ} - 1, jumlah total sampel dikurangi 1.

Jadi tabel ini menjelaskan bagaimana variasi data dipisahkan (antara grup dan dalam grup) dalam MANOVA, dan digunakan sebagai dasar untuk menguji hipotesis apakah vektor rata-rata grup berbeda secara signifikan.

Selain itu, tabel ini memiliki bentuk yang persis sama, komponen demi komponen, dengan tabel ANOVA, kecuali bahwa kuadrat dari skalar digantikan oleh padanannya dalam bentuk vektor. Sebagai contoh, $(x_e - \bar{x})^2$ menjadi $(x_e - \bar{x})(x_e - \bar{x})'$.
Derajat kebebasan sesuai dengan geometri univariat dan juga dengan beberapa teori distribusi multivariat yang melibatkan distribusi Wishart.
Salah satu uji untuk hipotesis nol $H_0: \tau_1 = \tau_2 = \cdots = \tau_g = 0$ melibatkan generalized variances Kita menolak H_0 jika rasio generalized variances:

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{\left| \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)' \right|}{\left| \sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})' \right|}$$

adalah terlalu kecil. Besaran $\Lambda^* = \frac{|W|}{|B+W|}$, yang awalnya diajukan oleh Wilks.
Tabel dibawah ini adalah Tabel Distribusi Wilks' Lambda (Λ*), yang digunakan dalam MANOVA (Multivariate Analysis of Variance) untuk menguji hipotesis apakah rata-rata vektor dari beberapa populasi berbeda secara signifikan.

TABLE 6.3 DISTRIBUTION OF WILKS' LAMBDA, $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}|$

No. of variables	No. of groups	Sampling distribution for multivariate normal data
$p = 1$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g}{g - 1}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g - 1}{g - 1}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\Sigma n_\ell - p - 1}{p}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$
$p \geq 1$	$g = 3$	$\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\Sigma n_\ell - p - 2)}$

Contoh:

Suatu percobaan dilakukan untuk mengetahui perbedaan tiga varietas jagung. Data respon yang diambil antara lain Y1 = Produksi per hektar, dan Y2 = bobot/1000 butir. Rancangan lingkungan yang digunakan adalah rancangan acak lengkap. Datanya diperoleh sebagai berikut :

Perlakuan	Ulangan	Y1	Y2
Varietas 1	1	6	7
	2	5	9
Varietas 2	1	4	6
	2	6	6
	3	4	7
Varietas 3	1	5	4
	2	6	4

a. Tuliskan model liniernya beserta keterangan yang jelas.
b. Hitunglah vektor rataan untuk setiap perlakuan
c. Hitunglah matriks jumlah kuadrat dan hasil kali silang dari perlakuan (B), galat (W), dan Total (T)
d. Lakukan pengujian pada taraf nyata 5% untuk mengetahui apakah ketiga varietas memiliki respon yang berbeda. Gunakan Uji Wilks Lambda.
e. Apa kesimpulan anda?

Pembahasan.

a. Model linear:

$$\underline{y}_{lj} = \underline{\mu} + \underline{\tau}_l + \underline{\varepsilon}_{lj} \qquad ; l = 1,2,3 \; ; j = 1,2,\ldots,n_l$$

keterangan:

$\underline{y}_{lj}$ = respon varietas ke *l* ulangan ke *j*

$\underline{\mu}$ = vektor rataan umum

$\underline{\tau}_l$ = pengaruh varietas ke *l*

$\underline{\varepsilon}_{lj}$ = pengaruh acak varietas ke *l* ulangan ke *j*

b. Vektor Rataan

$$\bar{x}_1 = \begin{pmatrix} 5{,}5 \\ 8 \end{pmatrix};\ \bar{x}_2 = \begin{pmatrix} 4{,}67 \\ 6{,}33 \end{pmatrix};\ \bar{x}_3 = \begin{pmatrix} 5{,}5 \\ 4 \end{pmatrix}$$

$$\bar{x} = \begin{pmatrix} 5{,}14 \\ 6{,}14 \end{pmatrix}$$

c. Matriks Jumlah Kuadrat

$$\mathbf{B} = \sum_{l=1}^{3} n_l \left(\underline{\bar{\mathbf{X}}}_l - \underline{\bar{\mathbf{X}}}\right)\left(\underline{\bar{\mathbf{X}}}_l - \underline{\bar{\mathbf{X}}}\right)''$$

$$= 2\begin{pmatrix} 5{,}5 - 5{,}14 \\ 8 - 6{,}14 \end{pmatrix}(5{,}5 - 5{,}14 \quad 8 - 6{,}14) + 3\begin{pmatrix} 4{,}67 - 5{,}14 \\ 6{,}33 - 6{,}14 \end{pmatrix}(4{,}67 - 5{,}14 \quad 6{,}33 - 6{,}14) +$$

$$2\begin{pmatrix} 5{,}5 - 5{,}14 \\ 4 - 6{,}14 \end{pmatrix}(5{,}5 - 5{,}14 \quad 4 - 6{,}14)$$

$$= 2\begin{pmatrix} 0{,}36 \\ 1{,}86 \end{pmatrix}(0{,}36 \quad 1{,}86) + 3\begin{pmatrix} -0{,}47 \\ 0{,}19 \end{pmatrix}(-0{,}47 \quad 0{,}19) + 2\begin{pmatrix} 0{,}36 \\ -2{,}14 \end{pmatrix}(0{,}36 \quad -2{,}14)$$

$$= \begin{pmatrix} 1{,}1811 & -0{,}4695 \\ -0{,}4695 & 16{,}1867 \end{pmatrix}$$

$$\mathbf{W} = \sum_{l=1}^{3}\sum_{j=1}^{n_l}\left(\underline{\mathbf{X}}_{lj} - \underline{\bar{\mathbf{X}}}\right)\left(\underline{\mathbf{X}}_{lj} - \underline{\bar{\mathbf{X}}}\right)''$$

$$= \begin{pmatrix} 6 - 5{,}5 \\ 7 - 8 \end{pmatrix}(6 - 5{,}5 \quad 7 - 8) + \begin{pmatrix} 5 - 5{,}5 \\ 9 - 8 \end{pmatrix}(5 - 5{,}5 \quad 9 - 8) +$$

$$\begin{pmatrix} 4 - 4{,}67 \\ 6 - 6{,}33 \end{pmatrix}(4 - 4{,}67 \quad 6 - 6{,}33) +$$

$$\begin{pmatrix} 6 - 4{,}67 \\ 6 - 6{,}33 \end{pmatrix}(6 - 4{,}67 \quad 6 - 6{,}33) + \begin{pmatrix} 4 - 4{,}67 \\ 7 - 6{,}33 \end{pmatrix}(4 - 4{,}67 \quad 7 - 6{,}33) + \begin{pmatrix} 5 - 5{,}5 \\ 4 - 4 \end{pmatrix}(4 - 4) +$$

$$\begin{pmatrix} 6 - 5{,}5 \\ 4 - 4 \end{pmatrix}(6 - 5{,}5 \quad 4 - 4)$$

$$= \begin{pmatrix} 0{,}25 & -0{,}5 \\ -0{,}5 & 1 \end{pmatrix} + \begin{pmatrix} 0{,}25 & -0{,}5 \\ -0{,}5 & 1 \end{pmatrix} + \begin{pmatrix} 0{,}4489 & 0{,}2211 \\ 0{,}2211 & 0{,}1089 \end{pmatrix} + \begin{pmatrix} 1{,}7689 & -0{,}4389 \\ -0{,}4389 & 0{,}1089 \end{pmatrix} +$$

$$\begin{pmatrix} 0{,}4489 & -0{,}4489 \\ -0{,}4489 & 0{,}4489 \end{pmatrix} + \begin{pmatrix} 0{,}25 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0{,}25 & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 3{,}6667 & -1{,}6667 \\ -1{,}6667 & 2{,}6667 \end{pmatrix}$$

$$\mathbf{T} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}\left(\underline{\mathbf{X}}_{lj} - \underline{\bar{\mathbf{X}}}\right)\left(\underline{\mathbf{X}}_{lj} - \underline{\bar{\mathbf{X}}}\right)'$$

$$= \mathbf{W} + \mathbf{B}$$

$$= \begin{pmatrix} 3{,}6667 & -1{,}6667 \\ -1{,}6667 & 2{,}6667 \end{pmatrix} + \begin{pmatrix} 1{,}1811 & -0{,}4695 \\ -0{,}4695 & 16{,}1867 \end{pmatrix} = \begin{pmatrix} 4{,}8478 & -2{,}1362 \\ -2{,}1362 & 18{,}8534 \end{pmatrix}$$

d. **Hipotesis**

$H_0 : \underline{\tau}_1 = \underline{\tau}_2 = \underline{\tau}_3 = \underline{\tau}_4 = \underline{0}$

H_1 : minimal ada satu $\underline{\boldsymbol{\tau}}_l$ yang tidak sama dengan $\underline{\mathbf{0}}$

Statistik Uji :

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|}$$

$$\Lambda^* = \frac{\left|\begin{pmatrix} 3{,}6667 & {,}6667 \\ -1{,}6667 & 2{,}6667 \end{pmatrix}\right|}{\left|\begin{pmatrix} 4{,}8478 & -\ {,}1362 \\ -2{,}1362 & 18{,}8534 \end{pmatrix}\right|}$$

$$\Lambda^* = 0{,}081$$

Kriteria pengujian :

Tolak H_0 jika $\left[\frac{\sum n_l - p - 2}{p}\right]\left[\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right] \geq F_{2p;2(\sum n_l - p - 2)}(\alpha)$

Telah diketahui bahwa $\alpha = 5\%$, total perlakuan *(l)* = 3 dan total peubah *(p)* = 2. Sehingga diperoleh :

$$\left[\frac{7-2-2}{2}\right]\left[\frac{1-\sqrt{0{,}081}}{\sqrt{0{,}081}}\right] = 3{,}7704$$

dan

$$F_{2(2);2(7-2-2)}(0{,}05) = 4{,}5337$$

e. **Kesimpulan**

Karena $\left[\frac{\sum n_l - p - 2}{p}\right]\left[\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right] = 3{,}7704 < F_{2p;2(\sum n_l - p - 2)}(\alpha) = 4{,}5337$, maka dapat disimpulkan bahwa tidak cukup bukti untuk menolak H_0 (terima H_0) atau dengan kata lain dapat dikatakan bahwa varietas memberikan pengaruh yang nyata pada pengukuran secara bersama peubah produksi perhektar dan bobot/1000.

Latihan.

Buku dari exercise 6.8 halaman 333