MODUL 2 ANALISIS MULTIVARIAT
Inferensia Vektor Nilai Tengah
Mika Alvionita Sitinjak
16 September 2025

Konsep dasar tentang inferensia vektor nilai tengah berhubungan dengan pengujian atau pendugaan rata-rata dari suatu vektor peubah acak multivariat. Dalam statistik multivariat, setiap individu atau pengamatan biasanya diwakili oleh lebih dari satu variabel sehingga rata-rata yang dibahas bukan lagi berupa angka tunggal, melainkan sebuah vektor yang mencerminkan nilai tengah dari seluruh variabel tersebut secara simultan. Inferensia ini dilakukan untuk mengetahui apakah rata-rata populasi (vektor mean μ) sama dengan suatu nilai hipotesis tertentu, atau untuk membandingkan rata-rata dari dua atau lebih kelompok multivariat. Salah satu metode yang sering digunakan adalah uji Hotelling's T^2, yang merupakan generalisasi dari uji-t pada kasus univariat. Konsep ini penting karena mempertimbangkan tidak hanya nilai rata-rata tiap variabel, tetapi juga korelasi antarvariabel, sehingga kesimpulan yang diambil lebih akurat dan komprehensif dalam memahami karakteristik populasi multivariat.

Uji Hotelling's T^2 adalah metode statistik dalam analisis multivariat yang digunakan untuk menguji hipotesis mengenai rata-rata vektor populasi. Uji ini merupakan generalisasi dari uji-t pada kasus univariat, tetapi diperluas untuk situasi di mana terdapat lebih dari satu variabel yang diamati secara bersamaan. Inti dari uji Hotelling's T^2 adalah menguji apakah vektor nilai tengah populasi sama dengan vektor nilai tengah yang dihipotesiskan, atau membandingkan dua vektor rata-rata dari dua kelompok. Statistik uji T^2 dibentuk dengan mempertimbangkan baik perbedaan rata-rata maupun struktur kovariansi antarvariabel, sehingga informasi tentang hubungan antarvariabel tidak diabaikan. Nilai statistik T^2 ini kemudian dapat dikonversi menjadi distribusi F untuk menentukan signifikansi pengujian. Uji ini sangat bermanfaat dalam penelitian yang melibatkan banyak variabel terukur, seperti psikologi, biologi, atau ilmu sosial, karena memberikan gambaran menyeluruh tentang perbedaan rata-rata tanpa harus menguji setiap variabel secara terpisah.Tabel F dan Tabel t dapat diakses disini. Berikut ditampilkan untuk Persamaan Uji Hotelling's T^2

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then with $\overline{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_j$ and $\mathbf{S} = \frac{1}{(n-1)}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$,

$$\alpha = P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right]$$

$$= P\left[n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] \quad (5\text{-}6)$$

whatever the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Here $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of the $F_{p,n-p}$ distribution.

Contoh :

Suatu penelitian dilakukan terhadap 20 perempuan sehat untuk mengetahui bagaimana tubuh mereka mengeluarkan keringat. Para peneliti tertarik pada tiga komponen penting dari keringat, yaitu:

1. Laju keringat (X_1),
2. Kandungan natrium (X_2), dan
3. Kandungan kalium (X_3).

Mereka ingin membuktikan apakah rata-rata ketiga komponen tersebut pada perempuan sehat benar-benar sesuai dengan nilai standar yang diduga, yaitu:

- Laju keringat sekitar 4 unit,
- Kandungan natrium sekitar 50 unit,
- Kandungan kalium sekitar 10 unit.

Dengan kata lain, peneliti menguji hipotesis:

H_0: Rata-rata ketiga komponen sama dengan [4,50,10]

H_1: Rata-rata ketiga komponen tidak sama dengan [4,50,10]

Ditunjukkan pada tabel dibawah ini.

TABLE 5.1 SWEAT DATA

Individual	X_1 (Sweat rate)	X_2 (Sodium)	X_3 (Potassium)
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Uji ini dilakukan dengan tingkat signifikansi α = 0,10, artinya peneliti memberi toleransi 10% kemungkinan kesalahan dalam menyimpulkan hasil penelitian.

Jawaban:

1. Hipotesis
   H_0: Rata-rata ketiga komponen sama dengan [4,50,10]'
   H_1: Rata-rata ketiga komponen tidak sama dengan [4,50,10]'
2. Taraf Signifikansi : α =10%
3. Statistika uji
   Rata-rata

$$\overline{X_J} = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$$

$$\overline{X_1} = \frac{3.7 + 5.7 + 3.8 + \ldots + 6.5 + 4.1 + 5.5}{20} = 4.61$$

$$\overline{X_2} = \frac{48.5 + 65.1 + 47.2 + \ldots + 44.1 + 40.9}{20} \approx 45.27$$

$$\overline{X_3} = \frac{9.3 + 8.0 + \ldots + 11.2 + 9.4}{20} \approx 10.00$$

$$\bar{X} = \begin{bmatrix} 4.61 \\ 45.27 \\ 10 \end{bmatrix}$$

Matriks kovarians

$$S_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i1} - \overline{X_1})^2$$

X_i1	rataan	X_{i1}-\bar{X_1}	(X_{i1}-\bar{X_1})^2
3,7	4,61	-0,91	0,83
5,7	4,61	1,09	1,19
3,8	4,61	-0,81	0,66
3,2	4,61	-1,41	1,99
3,1	4,61	-1,51	2,28
4,6	4,61	-0,01	0,00
2,4	4,61	-2,21	4,88
7,2	4,61	2,59	6,71
6,7	4,61	2,09	4,37
5,4	4,61	0,79	0,62
3,9	4,61	-0,71	0,50
4,5	4,61	-0,11	0,01
3,5	4,61	-1,11	1,23
4,5	4,61	-0,11	0,01
1,5	4,61	-3,11	9,67
8,5	4,61	3,89	15,13
4,5	4,61	-0,11	0,01
6,5	4,61	1,89	3,57
4,1	4,61	-0,51	0,26
5,5	4,61	0,89	0,79
		Jumlah	**54,73**

$$S_{11} = \frac{1}{19}(54{,}71) = 2.8$$

Kemudian lanjutkan sampai S_{33}

Sehingga, matriks kovarian yang diperoleh,

$$S = \left[\begin{bmatrix} 2.87 & 10.01 & -1.810 \\ 10.01 & 199.78 & -5.64 \\ -1.81 & -5.64 & 3.62 \end{bmatrix}\right]$$

Kemudian di cari matriks S^{-1} nya

$$\det(S) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

Substitusi nilai:
$a_{11} = 2.87, \quad a_{12} = 10.01, \quad a_{13} = -1.81$
$a_{21} = 10.01, \quad a_{22} = 199.78, \quad a_{23} = -5.64$
$a_{31} = -1.81, \quad a_{32} = -5.64, \quad a_{33} = 3.62$

Hitung sub-determinannya :
$t_1 = a_{22}a_{33} - a_{23}a_{32} = 199.78 \times 3.62 - (-5.64) \times (-5.64)$
$= 722.9916 - 31.5976 = 691.394$
$t_2 = a_{21}a_{33} - a_{23}a_{31} = 10.01 \times 3.62 - (-5.64) \times (-1.81) = 36.2362 - 10.2084$
$= 26.0278$
$t_3 = a_{21}a_{32} - a_{22}a_{31} = 10.01 \times (-5.64) - 199.78 \times (-1.81)$
$= -56.4564 + 361.6018 = 305.1454$

Sekarang determinan:
$\det(S) = (a_{11}t_1) - (a_{12}t_2) + (a_{13}t_3)$
$= (2.87 \times 691.394) - (10.01 \times 26.0278) + (-1.81 \times 305.1454)$
$= 1984.30078 - 260.538278 - 552.313174$
$= 1171.449328$

Hitung determinan matriks S menggunakan kofaktor sebagai berikut.
Hitung minor untuk setiap entri (mis. minor pada posisi (1,1) = determinan submatriks terbentuk dari baris/kolom selain baris 1 kolom 1), lalu kofaktornya dengan tanda $(-1)^{i+j}$
Hasil minor (ringkas):

$$Minor = \begin{bmatrix} 691.39 & 26.02 & 305.14 \\ 26.02 & 7.11 & 1.93 \\ 305.14 & 1.93 & 473.16 \end{bmatrix}$$

Matriks kofaktor dikali dengan $(-1)^{i+j}$

$$Kofaktor = \begin{bmatrix} 691.39 & -26.02 & 305.14 \\ -26.02 & 7.11 & -1.93 \\ 305.14 & -1.93 & 473.16 \end{bmatrix}$$

Adjustate adalah transpose dari matriks kofaktor. Karena matriks S simetris, adjugate di sini sama dengan matriks kofaktor yang sudah ditranspos (nilai numerik sama tersusun serupa):

$$Adj(S) = \begin{bmatrix} 691.39 & -26.02 & 305.14 \\ -26.02 & 7.11 & -1.93 \\ 305.14 & -1.93 & 473.16 \end{bmatrix}$$

Sehingga invers dari matriks S

$$S^{-1} = \frac{1}{\det(S)}\, adj(S)$$

$$S^{-1} = \begin{bmatrix} 0.5 & -0.022 & 0.26 \\ -0.02 & 0.006 & -0.001 \\ 0.26 & -0.001 & 0.403 \end{bmatrix}$$

Menghitung T^2 Hotteling

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

$$T^2 = 20\left(\begin{bmatrix} 4.61 \\ 45.27 \\ 10 \end{bmatrix} - \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}\right)^T \begin{bmatrix} 0.5 & -0.022 & 0.26 \\ -0.02 & 0.006 & -0.001 \\ 0.26 & -0.001 & 0.403 \end{bmatrix}\left(\begin{bmatrix} 4.61 \\ 45.27 \\ 10 \end{bmatrix} - \begin{bmatrix} 4 \\ 50 \\ 10 \end{bmatrix}\right)$$

$$T^2 = 9.74$$

$$T_h^2 = \frac{(n-1)p}{(n-p)}F_{p,n-p} = \frac{(20-1)3}{(20-3)}F_{3,20-3} = 8.18$$

Karena $T^2 > T_h^2$ ➔ tolak Ho

4. Kesimpulan
   Tolak Ho, Vektor mean populasi μ berbeda secara signifikan dari μ_0. Dengan kata lain, minimal ada satu variabel (sweat rate, sodium, potassium) yang mean-nya berbeda dari hipotesis nol.

Penerapan dalam R Studio.

Penerapan kasus menggunakan dataset Boston Housing yang terdapat dalam R. Dataset Boston Housing adalah dataset klasik dalam statistik yang berisi informasi mengenai kondisi sosial-ekonomi, demografi, serta kualitas perumahan di berbagai wilayah di sekitar Boston, Massachusetts, Amerika Serikat pada tahun 1970-an. Dataset ini pertama kali dipublikasikan oleh Harrison & Rubinfeld (1978) dalam studi mengenai faktor-faktor yang memengaruhi harga rumah. Dataset ini memiliki 506 observasi (wilayah atau distrik) dengan 14 variabel, beberapa di antaranya adalah:

Variabel	Keterangan
crim	Tingkat kejahatan per kapita berdasarkan kota
zn	Persentase lahan pemukiman untuk rumah besar
indus	Persentase lahan untuk bisnis non-ritel
chas	Variabel dummy: apakah wilayah berbatasan dengan sungai Charles (1 = ya, 0 = tidak)
nox	Konsentrasi oksida nitrogen (indikator polusi udara)
rm	Rata-rata jumlah kamar per rumah
age	Persentase unit hunian yang dibangun sebelum tahun 1940
dis	Jarak rata-rata ke pusat kerja utama di Boston
rad	Indeks aksesibilitas ke jalan raya radial
tax	Tarif pajak properti (per $10,000)
ptratio	Rasio murid terhadap guru di tiap kota
lstat	Persentase populasi dengan status sosial ekonomi rendah
medv	Nilai median rumah yang ditempati pemilik (dalam ribuan dolar)

Pada kasus ini ingin diuji apakah faktor-faktor multivariat (kriminalitas, ukuran rumah, dan status sosial ekonomi) memiliki perbedaan yang bermakna antara daerah dengan rumah lebih besar dan daerah dengan rumah lebih kecil. Dengan kata lain, ini memberikan gambaran apakah karakteristik sosial dan lingkungan secara bersama-sama berbeda secara signifikan antar kelompok hunian, bukan hanya dilihat dari satu variabel saja.

```r
if (!require(MASS)) install.packages("MASS")

if (!require(Hotelling)) install.packages("Hotelling")

library(MASS)
library(Hotelling)
data("Boston", package = "MASS")
vars <- c("crim", "rm", "lstat")
rm_med <- median(Boston$rm, na.rm = TRUE)
group_high <- Boston[Boston$rm > rm_med, vars]
group_low  <- Boston[Boston$rm <= rm_med, vars]
dim(group_high)   # observasi dan 3 variabel

## [1] 253   3

dim(group_low)

## [1] 253   3
hotres <- hotelling.test(group_high, group_low)
print(hotres)

## Test stat:  575.09
## Numerator df:  3
## Denominator df:  502
## P-value:  0
```

Uji Hotelling's T^2 pada dataset Boston Housing dari paket MASS. Paket MASS dan Hotelling dipanggil karena MASS menyediakan dataset *Boston*, sementara Hotelling berisi fungsi hotelling.test() untuk menjalankan uji. Dataset *Boston* kemudian dimuat, dan memilih tiga variabel penting: crim (tingkat kriminalitas), rm (rata-rata jumlah kamar per rumah), dan lstat (persentase penduduk berstatus sosial rendah). Selanjutnya, median dari variabel rm dihitung sebagai batas pemisah untuk membuat dua kelompok: group_high (wilayah dengan rata-rata jumlah kamar lebih tinggi dari median) dan group_low (wilayah dengan jumlah kamar kurang atau sama dengan median). Dengan cara ini, kita memiliki dua sampel multivariat yang masing-masing berisi observasi untuk tiga variabel. Setelah memeriksa jumlah data di kedua kelompok dengan dim(), selanjutnya menjalankan fungsi hotelling.test() untuk membandingkan apakah rata-rata vektor (crim, rm, lstat) berbeda signifikan antara kedua kelompok tersebut.

Interpretasi Hasil :

Nilai uji T^2=575.09T, dan *p-value* = 0 menunjukkan hasil sangat signifikan (lebih kecil dari tingkat signifikansi berapapun, misalnya 0.05 atau 0.01). Dengan demikian:

Hipotesis nol (H_0): rata-rata vektor (crim, rm, lstat) untuk kelompok rumah dengan jumlah kamar lebih tinggi dari median sama dengan kelompok rumah dengan jumlah kamar lebih rendah atau sama dengan median.

Tolak Ho, karena *p-value* ≈ 0 yang artinya, ada perbedaan yang sangat signifikan pada rata-rata gabungan ketiga variabel tersebut di antara kedua kelompok, dengan kata lain

- Wilayah dengan rumah lebih besar (rm > median) memiliki karakteristik yang sangat berbeda dalam hal tingkat kejahatan (crim), jumlah kamar (rm), dan status sosial ekonomi rendah (lstat) dibanding wilayah dengan rumah lebih kecil.
- Dengan kata lain, faktor sosial (status ekonomi rendah), faktor lingkungan (tingkat kriminalitas), dan karakteristik fisik (jumlah kamar) tidak bisa dianggap sama antara dua kelompok masyarakat ini.
- Hasil ini konsisten dengan intuisi bahwa daerah dengan rumah lebih besar biasanya punya status sosial-ekonomi lebih tinggi (lstat lebih rendah) dan sering kali tingkat kriminalitas lebih rendah (crim lebih kecil).

Latihan.

Buku Halaman 5.1 dan 5.11