H. Maman Suherman,Drs.,M.Si

PROSEDUR PENGUJIAN HIPOTESIS SEHUBUNGAN DENGAN AKAR-AKAR LATEN DARI MATRIKS KOVARIANS

(Dalam Analisis Komponen Utama)

Abstrak

Untuk membuat kesimpulan tentang karakteristik populasi multivariat khusunya populasi m variat sehubungan dengan analisis komponen utama, maka perlu adanya suatu prosedur atau langkah-langkah untuk menyelesaikan permasalahan ini.

Masalahnya adalah bagaimana karakteristik dari komponen utama populasi (ruang individu) apakah dapat direduksi ke dalam ruang yang lebih kecil dari m, idealnya ruang k = 1, atau k = 2 atau tidak dapat direduksi?

Untuk menjawab ini diperlukan teknik sampling dengan cara mengambil sampel acak m variat berukuran n. Dari sampel ini dicari akar-akar laten dari matriks kovariansnya, sebab permasalahan mereduksi ruang dimensi identik dengan masalah nilai-nilai dari akar laten matriks kovarians. Nilai ini natinya digunakan dan termuat dalam statistik uji sehubungan dengan pengujian hipotesis null H_k tentang akar-akar laten matriks kovarians populasi.

Ternyata statistik ujinya (statistik hitung) adalah berdistribusi chikuadrat dengan derajat kebebasan ½ (q + 2) (q – 1), dengan q = m-k

Kata kunci : Matriks Kovarians, akar laten, hipotesis null H_k

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Analisis statistik Multivariat atau Teori Statistik Multivariat merupakan bagian atau topik dalam statistika. Topik dalam statistika ini mempelajari variabel banyak (m, variabel, $m \geq 1$), yang mana m variabel ini menyatakan karakteristik dari sejumlah individu atau objek. Artinya setiap individu memiliki ciri yang dinyatakan dengan m variabel. Salah satu bagian dari topik analisis multivariat ini adalah menyajikan individu (dan juga variabel) pada ruang yang berdimensi k, dengan $k \leq m$, dan kalau mungkin $k = 1$, $k = 2$, atau $k = 3$. Jadi kita berusaha mereduksi ruang individu yang tadinya berdimensi m menjadi ruang yang berdimensi lebih kecil dari m, dan kita cukup menganalisis individu di ruang yang lebih kecil (kalau bisa $k = 2$) tanpa banyak kehilangan informasi. Timbul sebuah pertanyaan, yaitu bagaimana prosedur mereduksi dari ruang berdimensi besar ke ruang berdimensi kecil?

Beberapa metoda multivariat harus digunakan untuk menjawab masalah ini baik secara verbal maupun secara visual (grafik). Salah satu metoda dari analisis multivariat adalah analisis komponen utama. Persoalan berikutnya muncul apabila data multivariat merupakan sampel yang diambil dari sebuah populasi yang berdistribusi m variat, dalam hal ini berdistribusi normal multi variat. Kita menginginkan penegsan (konfirmasi) bahwa sampel yang kita ambil berasal dari populasi yang diharapkan, yaitu populasi dimana individu-individu cukup hanya dipelajari dengan k komponen utama pertama saja.

B. Rumusan dan Pembatasan Masalah

Secara umum masalahnya adalah “Bagaimanakah prosedur atau metoda inferensi sehubungan dengan komponen utama populasi?” secara rinci (khusus) masalahnya adalah:

1. Bagaimanakah kriteria penentuan statistik uji yang digunakan sehubungan dengan pengujian hipotesis kesamaan akar-akar laten terkecil?
2. Bagaimana distribusi dari statistik uji dalam pengujian hipotesis tentang akar-akar laten terkecil?

Catatan :

Materi penunjang (prasyarat) untuk pembahasan masalah ini antara lain adalah komponen utama populasi, komponen utama sampel, dan distribusi gabungan akar-akar laten dari matriks kovarians sampel.

BAB II

PEMBAHASAN

(MASALAH PENGUJIAN HIPOTESIS SEHUBUNGAN DENGAN KOMPONEN UTAMA)

A. Rumusan Hipotesis mengenai Akar-Akar Laten

Seperti telah dijelaskan pada bab I, bahwa tujuan utama dari analisis komponen utama adalah menyajikan individu yang asalnya dalam ruang berdimensi besar (yaitu m, $k \geq 2$) ke dalam ruang berdimensi kecil (yaitu k, $k \leq m$), dan sedapat mungkin k = 2, atau k = 3.

Karena varians dari komponen utama adalah akar laten dari matriks kovarians, dan varian dari komponen utama pertama sampai dengan varians dari komponen utama ke m, berturut-turut mulai dari yang terbesar sampai dengan yang terkecil, maka analisis komponen utama identik dengan analisis akar-akar laten dari matriks kovarians.

Pada umumnya akar-akar laten dari matriks kovarians populasi tidak diketahui, hanya diasumsikan bahwa $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$ (> o). Sedangkan kita ingin mengetahui karakteristik populasi (dalam hal ini akar-akar laten), khususnya ingin mengetahui apakah k akar laten terbesar dari matriks kovarians populasi berbeda dan q = m-k akar laten terkecil lainnya sama? Untuk menjawab masalah ini kita perlu membuat rumusan hipotesis (hipotesis null), yaitu:

H_k : (m-k) akar-akar laten terkecil dari matriks kovarians populasi adalah sama (yaitu sama dengan λ).

Secara matematis (statistik) rumusan hipotesis tersebut adalah : $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m \ (=\lambda)$

Contoh :

Untuk k = o, maka $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$. Ini berarti semua akar laten dari matriks kovarians populasi adalah sama. Untuk k = 1, maka $H_1 : \lambda_2 = \lambda_3 = ... = \lambda_m$. Ini berarti (m-1) akar laten terkecil dari matriks kovarians populasi adalah sama, dan hanya akar laten terbesar λ_1 yang berbeda.

Catatan :

1. Jika $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$ diterima, maka disimpulkan bahwa komponen-komponen utama populasi memiliki varians yang sama, berarti kontribusinya sama untuk variasi total. Sehingga m variabel tidak direduksi ke dalam dimensi yang lebih kecil (tidak ada transformasi ke dalam komponen utama).
2. Jika $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$ ditolak, dan misalkan (m-1) akar laten terkecil adalah sama, berarti $H_1 : \lambda_2 = \lambda_3 = ... = \lambda_m$ diterima, dan bila nilai umumnya (perkiraan atau taksiran) dari akar-akar laten terkecil adalah relatif kecil, maka banyak variasi diterangkan oleh komponen utama pertama. Sehingga kita merasa layak untuk memperhatikan atau mempelajari individu hanya dengan komponen utama pertama saja (dimensi satu).

Jika $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$ ditolak, dan juga $H_1 : \lambda_2 = \lambda_3 = ... = \lambda_m$ ditolak, maka selanjutnya apakah (m-2) akar-akar laten terkecil sama? Yakni $H_2 : \lambda_3 = \lambda_4 = ... = \lambda_m$ diterima/ditolak? Dalam prakteknya, kita menguji barisan hipotesis null $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$ mulai dari k = 0, 1, 2, ..., (m-2).

B. Pengujian Hipotesis Mengenai Akar-Akar Laten

1. Pengujian untuk $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$

   Berdasarkan Muirhead (1982), bahwa uji ratio likelikood dari $H_O : \lambda_1 = \lambda_2 = ... = \lambda_m$ didasarkan pada statistik:

   $$V_O = \frac{\prod_{i=1}^{m} \ell_i}{\left(\frac{1}{m} \sum_{i=1}^{m} \ell_i\right)^m}$$

   Dengan $\ell_1 > \ell_2 > ... > \ell_m$ adalah akar-akar laten dari matriks kovarians sampel S. Kriteria pengujian dari taraf asimtotik α adalah:

   Menolak H_O, jika

   $$-\left(n - \frac{2m^2 + m + 2}{6m}\right) \ell_n V_O \le \chi^2_{\left[\alpha\, ;\, \frac{1}{2} (n+2)(m-1)\right]}$$ dan

   Menerima H_O, jika

   $$-\left(n - \frac{2m^2 + m + 2}{6m}\right) \ell_n V_O \le \chi^2_{\left[\alpha\, ;\, \frac{1}{2} (n+2)(m-1)\right]}$$ dengan

   $$\chi^2_{\frac{1}{2}(m = 2)(m-1)} \ge \chi^2_{\left[\alpha\, ;\, \frac{1}{2} (m+2)(m-1)\right]} = \alpha$$

2. Pengujian untuk $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$

   Bebebrapa teorema berikut dapat dibuktikan!

   Teorema:

   Misalkan populasi $\vec{X} = (X_1, X_2, ..., X_m)1$ berdistribusi Nm ($\bar{\mu}$, $\sum$), dengan : $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_m$ akar-akar laten dari $\sum$. Maka statistik ratio likelihood untuk pengujian hipotesis null $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$ (=λ, λ tak diketahui) adalah:

   $$\Delta_k = \nabla_k^{N/2} = \left[\frac{\prod_{i=k+1}^{m} \ell_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^{m} \ell_i\right)^{m-k}}\right]^{N/2}$$

   dengan N = n + 1 ukuran sampel, dan $\ell_1 > \ell_{k+1}, \ell_{k+2}, ... \ell_m$ adalah akar-akar laten terkecil dari matriks kovarians sampel S.

   Teorema :

   Bila hipotesis null $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$ (=λ) benar, maka distribusi pendekatan ($n \to \sim$) dari statistik

   $$P_k = -\left[n - k - \frac{2q^2 + q + 2}{6q} + \sum_{i=k}^{k} \frac{\ell_q^{-2}}{(\ell_i - \ell_q)^2}\right] \ell_n \vee_k$$ adalah $\sum^2_{(q+2)\,(q-1)/_2}$, dan

   $$E_c\left[P_k\right] = \frac{1}{2}(q+2)\,(q-1) + \sigma(n^{-2})$$ dengan q =m – kdan $\bar{\ell}_2 = \frac{1}{2} \sum_{i=k+1}^{m} \ell_i$

3. Kriteria Penerimaan Hipotesis H_k

   Misalkan taraf signifikansi pengujian diambil sebesar α, maka kesimpulan yang dibuat adalah :

Tolak hipotesis $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$, jika P_k>c[α ; (q+2)(q + 1)
dan
Terima hipotesis $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m$, jika P_k>c[α ; (q+2)(q + 1) dengan c[α ; (q+2)(q + 1) adalah suatu titik pada sumbu datar dari grafik peubah acak χ^2 dengan derajat kebebasan (q+2)(q + 1)/$_2$ sedemikian sehingga luas daerah sebelah kanan titik ini adalah α.

Contoh :
Misalkan matriks kovarians sampel acak berukuran N = 101 dari populasi $\overline{\mathrm{X}}{:}n5(\overline{\mu},\Sigma)$ adalah :

$$S = \begin{bmatrix} 1.000 & 0,577 & 0,509 & 0,378 & 0,462 \\ 0,577 & 1,000 & 0,599 & 0,389 & 0,322 \\ 0,509 & 0,599 & 1.000 & 0,436 & 0,426 \\ 0,387 & 0,389 & 0,436 & 1,000 & 0,523 \\ 0,462 & 0,322 & 0,426 & 0,523 & 1,000 \end{bmatrix}$$

Dan akar-akar laten dari matriks kovarians ini adalah : $\ell_1 = 2{,}857$ $\ell_2 = 0{,}809$ $\ell_3 = 0{,}540$ $\ell_4 = 0{,}452$ dan $\ell_5 = 0{,}343$ Dapat ditunjukkan, bahwa dengan taraf arti α = 5% kedua hipotesis null $H_O : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$ dan $H_1 : \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$ ditolak.
Bagaimana untuk k = 2 atau untuk hipotesis null $H_2 : \lambda_3 = \lambda_4 = \lambda_5$?
Dalam hal ini m = 5 , n = 100 , k = 2 , q = 3 , $\overline{\ell}_3$ = 0,455 dan $\ell_n \vee_2$ = 0,051 , maka diperoleh statistik hitung P_2 = 5,011. Sedangkan untuk α = 5% diperoleh statisti tabel $\chi^2_{(0,05\,;5)} = 11{,}1$

Karena P2 = 5,11 $\leq$ 11,1 = $\chi^2_{(0,05\,;5)}$, maka dengan derajat keyakinan 95% hipotesis null $H_2 : \lambda_3 = \lambda_4 = \lambda_5$ (= λ) diterima. Ini menyimpulkan bahwa, 3 akar laten terkecil dari matriks kovarians populasi adalah sama. Masalah berikutnya adalah, apakah kita dapat mengabaikan 3 akar laten terkecil ini? Dengan kata lain, apakah menyajikan data cukup hanya dengan dua komponen utama pertama saja? Untuk menjawab masalah ini kita perlu mencari interval kepercayaan satu sisi untuk λ, yakni

$$\lambda \leq \frac{\overline{\ell}_q}{1 - \left(\frac{2}{nq}\right)^{\frac{1}{2} z\alpha}}$$

Dengan q = 3 , $\overline{\ell}_2 = \overline{\ell}_3 = 0{,}445\,, n = 100\,, dan\ z_\alpha = z_{0,05} = 1{,}65$ maka $\lambda \leq$ 0,514. Karena batas atas terkecil dari interval kepercayaan 95% untuk λ adalah 0,514, yang dalam hal ini relatif kecil bila dibanding dengan dua akar laten terbesar lainnya, maka tiga akar laten terkecil dari matriks kovarians populasi dapat diabaikan. Hasil ini memperkuat dugaan bahwa, kita cukup mempelajari data hanya dengan dua komponen utama pertama saja.

BAB III

PENUTUP DAN KESIMPULAN

Pada bagian ini penulis mencoba membuat kesimpulan dan rangkuman yang merupakan jawaban atau masalah dan tujuan yang ingin dicapai.

1. Statistik uji (ratio likelikood) untuk pengujian hipotesis null $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m (= \lambda)$ adalah

$$\wedge_k = \vee_k^{N/2} = \left[\frac{\prod_{i=k+1}^{m} \ell_i}{\ell_q^z} \right]^{N/2}$$

2. Distribusi limit dari statisti

$$P_k = -\left[n - k - \frac{2q^2 = q + 2}{6q} + \sum_{i=1}^{k} \frac{\bar{\ell}_q^2}{(\ell_i - \bar{\ell}_q)^2} \right] \ell_n \vee_k \text{ adalah } \chi^2_{(q+2)\,(q-1)/2}$$

3. Kriteria penerimaan hipotesis null $H_k : \lambda_{k+1} = \lambda_{k+2} = ... = \lambda_m (= \lambda)$ dengan taraf signifikansi α adalah :

   Tolak H_k , jika $P_k > \chi^2_{[\alpha\,;\,(q+2)\,(q-1)/2]}$

   Terima H_k , jika $P_k \leq \chi^2_{[\alpha\,;\,(q+2)\,(q-1)/2]}$

4. Bila dengan taraf signifikansi α , dan untuk suatu k, k ≠ o hipotesis H_k diterima dan jika

$$\frac{\bar{\ell}_q}{1 - \left(\frac{2}{n\,q} \right)^{\frac{1}{2}} z_{\frac{1}{2}\alpha}}$$

   Relatif kecil, maka kita dapat memutuskan untuk mempelajari data hanya dengan k komponen utama pertama saja.

DAFTAR PUSTAKA

Anderson T.W. (1984), AN INTRODUCTION to MULTIVARIATE STATISTICAL ANALYSIS, 2nd Edition, John Wiley & Sons, Inc.

Djauhari, Maman A. (1988), MATERI POKOK STRUKTUR DATA STATISTIK, Karunia, Universitas Terbuka.

Everitt Brian S. (1994), A HANDBOOK of STATISTICAL ANALYSIS, New York, Macmillan Publishing Co, Inc.

Hogg R.V. & A.T. Craig. (1982), INTRODUCTION to MATHEMATICAL STATISTIES, New York, Macmillan Publishing Co, Inc.

Johson R.A (1982) APPLIED MULTIVARIATE STATISTICAL ANALYSIS, Prentice – Hall, Inc.

Muirhead R.J. (1982), ASPECT of MULTIVARIATE STATISTICAL THEORY, John Wiley & Sons, Inc.