LIVIA PUSPA T
0607007

9.3 METODE KOMPONEN UTAMA

Informasi yang dibutuhkan dalam teknik komponen utama suatu data didapat dari matriks variansi kovariansi, atau matriks korelasinya. Metode komponen utama ini, bertujuan untuk menaksir parameter pada analisis faktor, yaitu varians spesifik $\left(\Psi_{(pxp)}\right)$, komunalitas (h), dan matriks faktor loading $\left(L_{(pxm)}\right)$. Matriks varians kovarians dari sampel yaitu S yang merupakan estimator (penduga) bagi matriks varians kovarians populasi yang tidak diketahui yaitu Σ. Komponen utama analisis faktor pada matriks varians kovarians populasi Σ memiliki pasangan nilai eigen dan vektor eigen $\left(\lambda_{i,} e_i\right)$ dimana $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$ dengan $\sum = LL' + \psi$

$$= \left[\sqrt{\lambda_1} e_1 \vdots \sqrt{\lambda_2} e_2 \vdots \cdots \vdots \sqrt{\lambda_p} e_p\right] \begin{bmatrix} \overline{\sqrt{\lambda_1} e'_1} \\ \sqrt{\lambda_2} e'_2 \\ \vdots \\ \sqrt{\lambda_p} e'_p \end{bmatrix} + 0$$

Untuk mencari penaksir $\widehat{L}$, substitusikan Σ dengan S pada persamaan $\sum = LL' + \psi$ sehingga diperoleh: $S \cong \widehat{L}\widehat{L}^t + \widehat{\Psi}$ (9-3-1)

Dalam pendekatan analisis komponen utama, $\widehat{\Psi}$ diabaikan dan S dapat difaktorkan menjadi $S = \widehat{L}\widehat{L}^t$.

Solusi komponen utama pada model faktor adalah sebagai berikut :

Komponen utama analisis faktor pada matriks varians kovarians sampel S memiliki pasangan nilai eigen dan vektor eigen $\left(\widehat{\lambda}_1, \widehat{e}_1\right), \left(\widehat{\lambda}_2, \widehat{e}_2\right), ..., \left(\widehat{\lambda}_p, \widehat{e}_p\right)$ dimana $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Jika $m < p$ merupakan banyaknya faktor umum, maka matriks yang dihasilkan dari *estimasi faktor loading* $\left\{\widehat{\ell}_{ij}\right\}$ didefinisikan sebagai:

$$\widehat{L} = \left[\sqrt{\lambda_1} e_1 \vdots \sqrt{\lambda_2} e_2 \vdots \ldots \vdots \sqrt{\lambda_m} e_m\right] \qquad \text{(9-3-2)}$$

Estimasi varians spesifik diberikan oleh elemen diagonal dari matriks $S-\widehat{L}\widehat{L}^t$ adalah

$$\widehat{\psi}=\begin{bmatrix}\widehat{\psi}_1 & 0 & \cdots & 0\\ 0 & \widehat{\psi}_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \widehat{\psi}_p\end{bmatrix} \text{ dengan } \widehat{\psi}_i=s_{ii}-\sum_{j=1}^{m}\widehat{\ell}_{ij}^{\,2} \qquad (9\text{-}3\text{-}3)$$

Sedangkan *estimasi varians spesifik* oleh elemen diagonal dari matriks $R-\widehat{L}\widehat{L}^t$ adalah

$$\widehat{\psi}_i=1-\sum_{j=1}^{m}\widehat{\ell}_{ij}^{\,2} \qquad (9\text{-}3\text{-}4)$$

Komunalitas diestimasi sebagai:

$$\hat{h}_i^2=\hat{\ell}_{i1}^2+\hat{\ell}_{i2}^2+\cdots+\hat{\ell}_{im}^2=\sum_{j=1}^{m}\hat{\ell}_{ij}^2 \qquad (9\text{-}3\text{-}5)$$

Jika banyaknya faktor umum tidak ditentukan atau tidak diketahui, maka m dapat dicari berdasarkan pada estimasi nilai eigen. Seharusnya kontribusi beberapa faktor awal pada varians sampel cukup besar. Kontribusi dari faktor ke-1 pada varians sampel total $tr(S)=s_{11}+s_{22}+...+s_{pp}$ adalah:

$$\widehat{\ell}_{i1}^{\,2}+\widehat{\ell}_{i2}^{\,2}+...+\widehat{\ell}_{i1}^{\,2}=\sum_{i=1}^{p}\widehat{\ell}_{i1}^{\,2}=\sum_{i=1}^{p}\left(\sqrt{\hat{\lambda}_1}\hat{e}_{i1}\right)^2=\hat{\lambda}_1\sum_{i=1}^{p}\hat{e}_{i1}^{\,2}=\hat{\lambda}_1$$

Dimana $\hat{e}_1$ adalah vektor eigen satuan, yang artinya memiliki panjang 1. Sehingga secara umum dapat juga ditulis kontribusi dari faktor ke-*j* pada varians sampel total adalah :

$$\sum_{i=1}^{p}\widehat{\ell}_{i1}^{\,2}=\sum_{i=1}^{p}\left(\sqrt{\hat{\lambda}_1}\hat{e}_{i1}\right)^2=\hat{\lambda}_j\sum_{i=1}^{p}\hat{e}_{ij}^{\,2}=\hat{\lambda}_j \qquad (9\text{-}3\text{-}6)$$

Secara umum, ***proporsi dari varians sampel total yang berasal dari faktor umum ke-j*** adalah : $\dfrac{\hat{\lambda}_j}{s_{11}+s_{22}+...+s_{pp}}$ untuk analisis faktor pada S

$$\frac{\hat{\lambda}_j}{p} \quad \text{untuk analisi faktor pada R} \qquad (9\text{-}3\text{-}7)$$

Kriteria pada persamaan (9-3-7) diatas seringkali digunakan dalam menentukan banyaknya faktor umum. Dengan menggunakan program komputer kita juga dapat menentukan banyaknya faktor umum berdasarkan pada banyaknya nilai eigen dari matriks korelasi R atau dari matriks varians kovarians S yang lebih dari rata-rata nilai eigen.

Setelah seluruh nilai taksiran parameter didapatkan kemudian dihitung matriks sisa. Matriks sisa didefinisikan sebagai selisih dari korelasi sampel R atau dari matriks S dengan nilai-nilai taksiran yang didapat.

$$Matriks\, sisa = R - \hat{L}\hat{L}^t + \hat{\Psi} \text{ atau } S - \hat{L}\hat{L}^t + \hat{\Psi} \qquad (9\text{-}3\text{-}8)$$

Contoh 9.3

dalam study pilihan konsumen., suatu sampel random dari konsumen untuk menilai atribut suatu produk baru. Tanggapan pada 7 titik skala berbeda membentuk matriks korelasi. Matriks korelasi disajikan dalam bentuk sebagai berikut.

$$\begin{array}{ll|l} \text{Attribute (Variabel)} & & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \text{Taste} & 1 & \\ \text{Good buy for money} & 2 & \\ \text{Flavour} & 3 & \\ \text{Suitable for snack} & 4 & \\ \text{Provides lots of energy} & 5 & \end{array}$$

Attribute (Variabel)		1	2	3	4	5
Taste	1	1	.02	.96	.42	.01
Good buy for money	2	.02	1	.13	.71	.85
Flavour	3	.96	.13	1	.5	.11
Suitable for snack	4	.42	.71	,5	1	.79
Provides lots of energy	5	.01	.85	.11	.79	1

Telah jelas dari yang dilingkari anggota matriks korelasi bahwa variabel 1 dan 3 dan variabel 2 dan5 membentuk grup. Variabel 4 lebih mendekati grup(2,5) daripada grup(1,3).diberikan hasil dan jumlah dari variabel, kita mengira bahwa hubungan linier anta variabel dapat dijelaskan dalam bentuk satu, dua atau tiga factor umum.

Dua nilai eigen pertama $\widehat{\lambda_1} = 2.85$ dan $\widehat{\lambda_2} = 1.81$dari R yang lebih baik dari lainnya. Lebih dari itu, m = 2 faktor umum akan dihitung untuk mencari proporsi cumulative

$$\frac{\lambda_1+\lambda_2}{p} = \frac{2.85+1.81}{5} = 0.93$$

Dari total sampel varian. Penaksiran factor pemuat, communality, dan factor khusus diberikan dalam table

Variabel	Penaksiran factor loading $\overline{l_{ij}} = \sqrt{\lambda_i} e_{ij}$		Communality	Factor khusus
	F_1	F_2	$\overline{h_i^2}$	$\overline{\psi_i} = 1 - \overline{h_i^2}$
1. Taste	0.56	0.82	0.98	0.02
2. Good buy for money	0.78	-0.53	0.88	0.12
3. Flavor	0.65	0.75	0.98	0.02
4. Suitable for snack	0.94	-0.11	0.89	0.11
5. Provides lots of energy	0.8	-0.54	0.93	0.07
Nilai Eigen	2.85	1.81		
Proporsi Communality dari variansi sampel total	0.571	0.932		

$$\tilde{L}\tilde{L}' + \tilde{\Psi} = \begin{bmatrix} 0.56 & 0.82 \\ 0.78 & -0.53 \\ 0.65 & 0.75 \\ 0.94 & -0.1 \\ 0.8 & -0.54 \end{bmatrix} \begin{bmatrix} 0.56 & 0.78 & 0.65 & 0.94 & 0.8 \\ 0.82 & -0.53 & 0.75 & -0.1 & -0.54 \end{bmatrix}$$

$$+ \begin{bmatrix} 0.02 & 0 & 0 & 0 & 0 \\ 0 & 0.12 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.11 & 0 \\ 0 & 0 & 0 & 0 & 0.07 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & .01 & .97 & .44 & 0 \\ & 1 & .11 & .79 & .91 \\ & & 1 & .53 & .11 \\ & & & 1 & .81 \\ & & & & 1 \end{bmatrix}$$

Hasilnya mendekati matriks korelasi R. kita akan menilai dua factor model dengan factor loading di atas yang cukup baik untuk data. The comunlity (.98,.88,.98,.89,.93) mengindikasikan perhitungan dua factor untuk percentasi besar dari variansi sampel masing-masing variabel.

PENAKSIRAN PARAMETER DENGAN METODE LIKELIHOOD

Jika faktor umum F dan faktor spesifik ε dapat diasumsikan berdistribusi normal, maka penaksiran maksimum likelihood dari faktor loading dan variansi spesifik dapat diperoleh. Ketika F_j dan ε_j berdistribusi normal gabungan, maka observasi $X_j - \mu = LF_j + \varepsilon_j$ adalah normal dan memiliki bentuk fungsi likelihood:

$$L(\mu,\Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left[-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^t + n(\bar{x}-\mu)(\bar{x}-\mu)^t\right)\right]\right]$$

$$= (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} \exp\left[-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^t\right)\right]\right]$$

$$\times\ (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}}\, 2\exp\left[-\left(\frac{1}{2}\right)\left(\bar{x}-\mu\right)^t \Sigma^{-1}\left(\bar{x}-\mu\right)\right]$$

(9-3-9)

Yang bergantung pada L dan Ψ dengan $\Sigma = LL^t + \Psi$. Model ini masih belum terdefinisi dengan baik karena banyaknya pilihan L. Agar L terdefinisi dengan baik, maka dengan cara melibatkan kondisi keunikan/ketunggalan pada proses perhitungan, yaitu :

$$L^t \Psi^{-1} L = \Delta \qquad (9\text{-}3\text{-}10)$$

Dimana Δ adalah matriks diagonal.

<u>Teorema 9.1</u>

Misalkan $X_1, X_2, ..., X_n$ adalah sampel acak dari $N_p\left(\mu,\Sigma\right)$ dimana $\Sigma = LL^t + \Psi$ adalah matriks kovariansi dari faktor umum (m), sehingga diperoleh penaksir maksimum likelihood $\hat{L}, \hat{\psi}\, dan\, \hat{\mu} = \bar{x}$ dengan memaksimumkan persamaan (9-3-9) terhadap matriks diagonal $\hat{L}^t \hat{\Psi}^{-1} \hat{L} = \hat{\Delta}$.

Dengan demikian diperoleh penaksir Likelihood untuk komunalitas $\hat{h}_i^2$ adalah :

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2 \quad \text{untuk i=1,2,...,p} \qquad (9\text{-}3\text{-}11)$$

dengan ***proporsi ke-j terhadap variansi sampel total***

$$\xi^{(j)} = \frac{\hat{\ell}_{ij}^2 + \hat{\ell}_{2j}^2 + \ldots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \ldots + s_{pp}} \qquad (9\text{-}3\text{-}12)$$

Bukti:

Dengan menggunakan sifat invarian dari penaksir maksimum likelihood, fungsi L dan Ψ mempunyai penaksir dengan fungsi yang sama yaitu $\hat{L}$ dan $\hat{\Psi}$.

Sama halnya dengan komunalitas $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2$ mempunyai penaksir maksimum Likelihood $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2$.

Standarisasi vektor acak X

Matriks standar deviasi $V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$

$Z = V^{1/2}\left(X - \mu\right)$ dengan $V = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$ maka Z akan berdistribusi $N_p\left(0,1\right)$

$$\rho = V^{-\frac{1}{2}} \Sigma V^{-\frac{1}{2}} = (V^{-\frac{1}{2}}L)(V^{-\frac{1}{2}}L)' + V^{-\frac{1}{2}} \Psi V^{-\frac{1}{2}} \qquad (9\text{-}3\text{-}14)$$

maka penaksir likelihood dari ρ adalah

$$\hat{\rho} = (\hat{V}^{-\frac{1}{2}}\hat{L})(\hat{V}^{-\frac{1}{2}}\hat{L})' + \hat{V}^{-\frac{1}{2}} \hat{\Psi} \hat{V}^{-\frac{1}{2}} = \hat{L}_z \hat{L}_z' + \hat{\Psi}_z \qquad (9\text{-}3\text{-}15)$$

Dari proses standarisasi X diperoleh :

i. Matriks loading $L_z = V^{-\frac{1}{2}} L, L_z = (\ell_{ij}^*)$

ii. Matriks variansi khusus $\Psi_z = V^{-\frac{1}{2}} \Psi V^{-\frac{1}{2}}$

iii. Sedangkan penaksir Likelihood ρ adalah :

$$\hat{\rho} = (\hat{V}^{-\frac{1}{2}} \hat{L})(\hat{V}^{-\frac{1}{2}} \hat{L})^t + \hat{V}^{-\frac{1}{2}} \hat{\Psi} \hat{V}^{-\frac{1}{2}}$$

Atau : $\hat{\rho} = \hat{L}_z \hat{L}_z' + \hat{\psi}_z$

iv. Proporsi faktor ke-*j* terhadap variansi sampel total standarisasi

$$\xi^{(j)} = \frac{\hat{\ell}_{ij}^{*2} + \hat{\ell}_{2j}^{*\,2} + \ldots + \hat{\ell}_{pj}^{*\,2}}{p}$$

Biasanya hasil penelitian-penelitian yang dilakukan datanya telah distandarisasi dan matriks korelasi sampel R dianalisis. Matriks sisa didefinisikan sebagai selisih dari korelasi sampel R dengan nilai-nilai taksiran yang diperoleh.

$$Matriks\, sisa = R - \hat{L}\hat{L}^t - \hat{\Psi} \qquad (9\text{-}3\text{-}16)$$

Example 9.5

Perhatikan example 8.5 dan 9.4. dengan model faktor m=2 dan gunakan metode maximum likelihood. Penaksir faktor loading, komunalitas, variansi spesifik dan Proporsi faktor terhadap variansi sampel total standarisasi dijelaskan pada tabel 9.3

$$R = \begin{bmatrix} 1 & 0.577 & 0.509 & 0.387 & 0.462 \\ 0.577 & 1 & 0.599 & 0.389 & 0.322 \\ 0.509 & 0.599 & 1 & 0.436 & 0.426 \\ 0.387 & 0.389 & 0.436 & 1 & 0.523 \\ 0.462 & 0.322 & 0.426 & 0.523 & 1 \end{bmatrix}$$

Tabel 9.3

Variabel	**Metode Maximum likelihood**			**Metode komponen utama**		
	Estimasi faktor loading		**Variansi spesifik**	**Estimasi faktor loading**		**Variansi spesifik**
	F_1	F_2	$\Psi_i = 1 - h_i^2$	F_1	F_2	$\Psi_i = 1 - h_i^2$
1.Allied Chemical	0.684	0.189	0.50	0.783	-0.217	0.34
2.du Pont	0.694	0.517	0.25	0.773	-0.458	0.19
3.Union Carbide	0.681	0.248	0.47	0.794	-0.234	0.31
4.Exxon	0.621	-0.073	0.61	0.713	0.412	0.27
5.Texaco	0.792	-0.442	0.18	0.712	0.524	0.22
Proporsi terhadap variansi sampel total standarisasi	0.485	0.598		0.517	0.733	

Matriks sisa maximum likelihood adalah

$$R - \hat{L}\hat{L}' - \hat{\Psi} = \begin{bmatrix} 0 & 0.005 & -0.004 & -0.024 & -0.004 \\ 0.005 & 0 & -0.003 & -0.004 & 0.000 \\ -0.004 & -0.003 & 0 & 0.031 & -0.004 \\ -0.024 & -0.004 & 0.031 & 0 & -0.000 \\ -0.004 & 0.000 & -0.004 & -0.000 & 0 \end{bmatrix}$$

Matriks sisa komponen utama adalah

$$R - \hat{L}\hat{L}' - \hat{\Psi} = \begin{bmatrix} 0 & -0.127 & -0.164 & -0.069 & 0.017 \\ -0.127 & 0 & -0.122 & 0.055 & 0.012 \\ -0.164 & -0.122 & 0 & -0.019 & -0.017 \\ -0.069 & 0.055 & -0.019 & 0 & -0.232 \\ 0.017 & 0.012 & -0.017 & -0.232 & 0 \end{bmatrix}$$

elemen-elemen dari matriks sisa maksimum likelihood lebih kecil dari matriks sisa pada metode komponen utama. oleh karena itu, pendekatan menggunakan metode maximum likelihood lebih baik. Namun, jika dilihat dari proporsi kumulatifnya maka metode komponen utama lebih diutamakan karena nilainya yang lebih besar dibandingkan dari metode maksimum likelihood.

MENGUJI JUMLAH FAKTOR UMUM

Dengan asumsi populasi berdistribusi normal, maka akan dilakukan uji model yang sesuai.
Misalkan dalam hipotesis terdapat *m* buah faktor umum, maka dilakukan pengujian:

$$H_0 : \Sigma^* = \Sigma = LL' + \psi$$
$$H_1 : \Sigma^* \neq \Sigma$$

Dibawah H_0, Σ terbatas untuk mencapai bentuk H_0, dengan fungsi Likelihood yang dimaksimumkan diperoleh $\hat{\mu} = \bar{x}$ dan $\hat{\Sigma} = \hat{L}\hat{L}^t + \hat{\Psi}$ dimana $\hat{L}$ dan $\hat{\Psi}$ adalah penaksir Likelihood dari L dan. Ψ

Statistik uji untuk uji rasio Likelihood adalah:

$$-2\ln\Lambda = n\ln\left(\frac{\left|\hat{\Sigma}\right|}{\left|S_n\right|}\right). \qquad .....(9\text{-}3\text{-}17)$$

dengan derajat kebebasan

$$v - v_0 = \tfrac{1}{2}p(p+1) - \{p(m+1) - \tfrac{1}{2}m(m-1)\} = \tfrac{1}{2}\{(p-m)^2 - p - m\}$$

Bartlet menunjukkan bahwa dapat dilakukan aproksimasi chi-kuadrat terhadap distribusi sampel pada persamaan (9-3-17) dengan cara menggantikan nilai *n* dengan faktor koreksi

n-1-(2p+4m+5)/6.

Sehingga dengan menggunakan faktor koreksi Bartlett kita tolak H_0 dengan tingkat sinifikansi α, jika

$$\{n - 1 - (2p + 4m + 5)/6\}\ln\frac{\left|\hat{\Sigma}\right|}{\left|S_n\right|} > \chi^2_{(v-v_0);\alpha}$$

$$\{n-1-(2p+4m+5)/6\}\ln\frac{\left|\hat{L}\hat{L}'+\hat{\Psi}\right|}{|S_n|} > \chi^2_{(v-v_0);\alpha}$$

Dalam mengimplementasikan pengujian dengan persamaan di atas , untuk m factor umum didapat dengan cara membandingkan varian yang diperumum $|LL'+\psi|$ dengan $|S_n|$. jika n besar dan m relative kecil dari p, biasanya H_0 ditolak. Maka harus digunakan pertimbangan lain dalam memilih jumlah factor m.

Example 9.7

Pada exm.9.5 data stock-price telah dianalisis dengan maximum likelihood (m=2). misalkan matrix sisa dari solusi dua factor telah mencukupi.

Uji hipotesis $H_0 : \Sigma = LL'+\psi$, dengan m = 2 dan $\alpha = 0.05$

$$\frac{\left|\hat{L}_z\hat{L}'_z+\hat{\Psi}_z\right|}{|R|} = \frac{\begin{vmatrix} 1 & & & & \\ 0.572 & 1 & & & \\ 0.513 & 0.602 & 1 & & \\ 0.411 & 0.393 & 0.405 & 1 & \\ 0.458 & 0.322 & 0.430 & 0.523 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & & & & \\ 0.577 & 1 & & & \\ 0.509 & 0.599 & 1 & & \\ 0.387 & 0.389 & 0.436 & 1 & \\ 0.462 & 0.322 & 0.426 & 0.523 & 1 \end{vmatrix}} = \frac{0.194414}{0.193163} = 1.0065$$

dengan menggunakan Bartlett’s koreksi,

$$\{n-1-(2p+4m+5)/6\}\ln\frac{\left|\hat{L}\hat{L}'+\hat{\Psi}\right|}{|S_n|}=\left[100-1-\frac{10+8+5}{6}\right]\ln(1.0065)=0.62$$

$$\tfrac{1}{2}\left[(p-m)^2-p-m\right]=\tfrac{1}{2}\left[(5-2)^2-5-2\right]=1$$

$$\chi_1^2(0.05)=3.84$$

tolak H_0 dengan tingkat sinifikansi α, jika

$$\{n-1-(2p+4m+5)/6\}\ln\frac{\left|\hat{\Sigma}\right|}{|S_n|}>\chi^2_{(v-v_0);\alpha}$$

0.62<3.84 sehingga H_0 diterima.