ANALISIS FAKTOR (FACTOR ANALYSIS)

Bahan Kuliah Secara Daring
Mahasiswa Departemen Statistika-FMIPA-IPB
Oleh: Dr. Ir. Budi Susetyo

Latar Belakang

- Dalam bidang penelitian tertentu, misalnya psikometri, sering kali ingin menggambarkan karakteristik individu tetapi tidak dapat diukur secara langsung (unobservable), misalnya intelegensi seseorang, prestasi siswa, bentuk ideal tubuh, dls.
- Karakteristik individu tersebut, yang selanjutnya disebut faktor, kemungkinan dapat dicirikan oleh segugus peubah yang dapat diukur (observable).
- Analisis Faktor merupakan suatu metode untuk menggambarkan (jika ada) pola hubungan internal banyak peubah sehingga membentuk beberapa kelompok unobservable faktor yang memiliki makna.
- Peubah-peubah yang membentuk suatu faktor tersebut memiliki korelasi tinggi didalam faktor itu sendiri dan berkorelasi rendah dengan faktor lainnya.
- Analisis faktor ini sering dikatakan sebagai pengembangan dari AKU

Struktur Data Amatan

Individu	Peubah				
	X1	X2	X3	…	Xp
1	x11	x12	x13		x1p
2	x21	x22	x23		x2p
3	x31	x32	x33		x3p
4	x41	x42	x43		x4p
5	x51	x52	x53		x5p
…	…	…	…	…	…
…	…	…	…	…	…
n	xn1	xn2	xn3		xnp

Model Faktor Ortogonal (1)

- Didefinisikan vektor peubah acak **observable X** dengan p komponen memiliki nilai tengah **μ** dan matriks peragam **Σ**.
- Model factor mendefinisikan bahwa vector X merupakan fungsi linear dari beberapa peubah acak unobservable F1, F2,…Fm (disebut factor umum/common factor) dan p sumber keragaman lainnya (disebut error).
- Model faktor dapat dituliskan dalam bentuk:

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \ldots + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \ldots + l_{2m}F_m + \varepsilon_2$$

:

:

:

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pm}F_m + \varepsilon_p$$

Model Faktor Ortogonal (2)

- Model factor dapat ditulis dalam bentuk matriks:

$$\underset{(px1)}{(\underline{X} - \underline{\mu})} = \underset{(pxm)}{L}\ \underset{(mx1)}{\underline{F}} + \underset{(px1)}{\underline{\varepsilon}}$$

-

Dimana koefisien l_{ij} dikatakan sebagai loading dari peubah ke-j pada factor ke-i sehingga matriks L adalah matriks dari loading faktor.

F adalah vektor acak dari F_1, F_2,… F_m. dan ε vektor galat/error dari ε_1, ε_2 …ε_p dimana kedua vector tersebut unobservable.

Model Faktor Ortogonal (3)

- Yang membedakan antara model factor dan regresi linear berganda adalah bahwa dalam regresi vector peubah F observable sehingga koefisien L dapat dengan mudah diduga.
- Meskipun vector peubah F dalam model factor unobservable, melalui beberapa asumsi tambahan terhadap vector acak F dan **ε** maka dapat dilakukan pendugaan terhadap model factor

Asumsi-Asumsi Model Faktor (1)

- $E(\underline{F}) = 0$, $E(\underline{\varepsilon}) = 0$
- $Cov(\underline{F}) = E(\underline{F}\underline{F}') = I$
- $Cov(\underline{\varepsilon}) = E(\underline{\varepsilon}\,\underline{\varepsilon}') = \psi = \text{diag}(\psi_1, \psi_2, \ldots, \psi_p)$
- $\underline{F}$ dan $\underline{\varepsilon}$ saling bebas
- $Cov(\underline{\varepsilon},\underline{F}) = E(\underline{\varepsilon},\underline{F}) = 0$

Berdasarkan asumsi dan model factor diatas maka struktur peragam model factor dapat dinyatakan:

1. $Cov(X) = LL' + \psi$ atau

$$Var(X_i) = l_{i1}^2 + \ldots + l_{im}^2 + \psi_i$$

$$Cov(X_i,X_k) = l_{i1}\, l_{k1} + \ldots + l_{im}\, l_{km}$$

2. $Cov(X,F) = L$ atau

$$Cov(X_i,F_j) = l_{ij}$$

Asumsi-Asumsi Model Faktor (2)

- Porsi ragam peubah X ke-i yang dapat dijelaskan oleh m faktor umum disebut dengan komunalitas ke-i sedangkan porsi yang dijelaskan oleh factor spesifik disebut ragam spesifik. Struktur ragam peubah X dapat ditulis sbb:

$$\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i \text{ ;}$$

$Var(X_i)$=komunalitas+ragam spesifik

Atau dapat juga ditulis $\sigma_{ii} = h_i^2 + \psi_i \text{ ; } i = 1,2, \dots, p$

dengan $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$

Pendugaan Parameter

- Ada beberapa metode pendugaan parameter model factor, yang dapat dikelompokkan dalam metode non-iterati dan metode iterative
- Metode non-iteraif yang paling banyak digunakan adalah metode komponen utama
- Metode iterative yang banyak digunakan adalah metode kemungkinan maksimum

Metode

Metode non-iteratif

- Metode komponen utama
- Metode faktor utama
- Analisis Citra
- Analisis faktor kanonik non-iteratif

Metode iteratif

- Metode kemungkinan maksimum
- Metode kuadrat terkecil tak terboboti
- Metode komponen utama iteratif Harris
- Metode analisis faktor alpha

Metode Komponen Utama

- Misal **Σ** merupakan matriks peragam dari matriks pengamatan **X** yang memiliki pasangan nilai akar ciri (eigenvalue) dan vektor cirinya (λ_i , $\underline{e}_i$) dengan $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Misalkan m< p merupakan jumlah peubah dari faktor umum (*common factor*), maka penduga parameter sebagai berikut:
- Matriks penduga faktor loadingnya {l_{ij}} yaitu:

  $\acute{L} = [\sqrt{\lambda_1}\, \underline{e}_1 \mid \sqrt{\lambda_2}\, \underline{e}_2 \mid \sqrt{\lambda_3}\, \underline{e}_3 \mid \ldots \mid \sqrt{\lambda_m}\, \underline{e}_m]$
- Penduga ragam spesifik adalah $\Psi = S - \acute{L}\acute{L}'$
- Nilai komunalitas untuk peubah ke-i: $h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$

Seperti pada Analisis Komponen Utama, dalam analisis factor juga dapat menggunakan matriks korelasi R

Ilustrasi 1

Tersedia data harga saham 100 mingguan (n=100) dari 5 jenis saham (p=5). Dari data 5 jenis saham tersebut ingin factor yang mencirikan kondisi ekonomi. Analisis dilakukan dengan menggunakan matriks korelasi R

Tabel pendugaan loading faktor,komunalitas dan total proporsi keragaman yang dijelaskan dari setiap faktor untuk m=1 dan m=2

Jenis Saham (X)	**Solusi satu faktor**		**Solusi dua faktor**		
	F_1	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
Allied Chemical .1	0.783	0.39	0.783	0.217-	0.34
DuPont	0.773	0.40	0.773	0.458-	0.19
Union Carbide .2	0.794	0.37	0.794	0.234-	0.31
Exxon .3	0.713	0.49	0.713	0.472	0.27
Texaco .4	0.712	0.49	0.712	0.524	0.22
Total proporsi kumulatif keragaman yang dapat dijelaskan	0.571		0.571	0.733	

Penjelasan Hasil Analisis

- Jika menggunakan 1 factor maka terdapat 57,1% keragaman X yang dapat dijelaskan Faktor 1, sedangkan jika menggunakan 2 factor sebesar 73,3%
- Faktor pertama merepresentasikan kondisi ekonomi secara umum dan dapat disebut faktor pasar.
- Faktor kedua merupakan kontras antara saham perusahaan kimia dengan saham perusahaan minyak (pada faktor perusahaan kimia memiliki loading negatif yang relatif besar dan perusahaan minyak memiliki loading positif yang relatif besar).
- Dengan demikian faktor kedua dapat disebut faktor industri karena sebagai pembeda harga saham di industri yang berbeda.

Metode Kemungkinan Maksimum

- Metode kemungkinan maksimum (MKM) mengasumsikan bahwa matriks ragam-peragam atau matriks korelasi semua peubah bersifat non-singular.
- Fungsi kepekatan peluang bagi S adalah L(S) yaitu:

$$L(S) = c.|\mathbf{\Sigma}|^{-\frac{n-1}{2}}|\mathbf{S}|^{\frac{n-1}{2}-\frac{p+1}{2}} e^{\frac{n-1}{2}tr(\mathbf{\Sigma}^{-1}\mathbf{S})}$$

dengan *c* adalah konstanta. Sehingga log-likelihood dari **L** dan **ψ** , jika **Σ** = **LL'** + **ψ** adalah:

$$\ln c - \left(\frac{n-1}{2}\right)\left\{tr[(\mathbf{LL'}+\boldsymbol{\psi})^{-1}\mathbf{S}] - ln\,|\,\mathbf{LL'}+\boldsymbol{\psi})^{-1}\mathbf{S}\,|\right\}$$

Penduga kemungkinan maksimum bagi **L** dan **ψ** diperoleh dengan memaksimumkan diatas dengan kendala *k*(*k*-1)/2 persyaratan kenunikan (Johnson & Wichern, 1998).

- Penentuan banyaknya faktor bersama

Uji Nisbah Kemungkinan (likelihood ratio test)

Hipotesis nol yang diuji pada uji nisbah kemungkinan ini adalah:

$H_0 : \Sigma = LL' + \psi$, $r(L) = k$ diketahui

Misalkan $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\psi}}$, dan $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$ adalah penduga kemungkinan maksimum bagi **L**, **ψ** dan **Σ**, jika H_0 benar, maka nilai maksimum untuk log dari fungsi kemungkinannya adalah:

$$\ln L_{H0} = c^* - \left(\frac{n-1}{2}\right)\left\{tr[\mathbf{S}^{-1}\mathbf{S}] - ln\,|\,\mathbf{S}^{-1}\mathbf{S}\,|\right\}$$

$$= c^* - \frac{n-1}{2}p$$

Statistik uji nisbah kemungkinan, yaitu:

$$-2\ln\lambda = -2\ln\left(\frac{L_{\mathrm{H}_0}}{L}\right)$$

Menyebar khi-kuadrat dengan

db=½ [(p – k)^2 – (p + k)]. Jadi H0 ditolak jika,

$$-2\ln\left(\frac{L_{\mathrm{H}_0}}{L}\right) \geq \chi^2_{\alpha;db=\left[(p-k)^2-(p+k)\right]/2}$$

Akaike's information Criterion(AIC)

Statistik AIC untuk model dengan k faktor didefinisikan sebagai berikut:

$$AIC(k) = -2\ln L(k) + [2p(k+1) - k(k-1)]$$

Model berfaktor k dengan k adalah nilai yang berpadanan dengan AIC (k) yang paling kecil dianggap sebagai model terbaik

Data harga saham dianalisa kembali dengan menggunakan metode maksimum likelihood dengan tetap memakai dua model faktor

Variabel	Maksimum likelihood			Komponen utama		
	Penduga faktor		$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	Penduga faktor		$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
Allied Chemical .1	0.684	0.189	0.50	0.783	0.217-	0.34
DuPont .2	0.694	0.517	0.25	0.773	0.458-	0.19
Union Karbide .3	0.681	0.248	0.47	0.794	0.234-	0.31
Exxon .4	0.621	0.073-	0.61	0.713	0.412	0.27
Texaco .5	0.792	0.442-	0.18	0.712	0.524	0.22
Total proporsi kumulatif keragaman contoh yang dapat dijelaskan	0.485	0.598		0.571	0.733	

Dalam kasus data tersebut total proporsi kumulatif keragaman dengan metode komponen utama lebih besar dibandingkan dengan maximum likelihood.

Rotasi Faktor

- Dalam banyak kasus, hasil dari analisis factor sulit untuk diintepretasikan makna dari loading setiap factor
- Sebagai salah satu cara untuk membantu memudahkan intepretasi adalah melalui rotasi faktor
- Rotasi factor merupakan transformasi ortogonal dari loading factors dengan :

  L*= LT dimana **TT'=T'T=I**
- Beberapa jenis transformasi yaitu, varimax, oblique, quartimax, dan lain-lain

- Rotasi Varimax

Merupakan rotasi yang paling sering dipergunakan pada aplikasi yang merupakan transformasi ortogonal yang diperoleh dengan cara memaksimumkan:

$$\sum_{j=1}^{k}\left\{\frac{1}{p}\sum_{i=1}^{p}\left(\frac{{l^*_{ij}}^2}{h_i}\right)^2-\left[\frac{1}{p}\sum_{i=1}^{p}\frac{{l_{ij}}^2}{h_i}\right]^2\right\}$$

- Rotasi Oblique

Digunakan apabila transformasi ortogonal terhadap matriks loading faktor menghasilkan faktor yang masih sulit diinterpretasikan.

- Rotasi quartimax

Transformasi ortogonal dengan tujuan memperoleh Γ yang memaksimumkan

$$\sum_i \sum_j {l^*_{ij}}^4$$

L Adalah matriks loading faktor yang ingin ditransformasi menggunakan matriks ortogonal Γ menjadi $L^* = L\Gamma$

sehingga

$$\frac{1}{Pk}\sum_i \sum_j {l^*_{ij}}^4 - \left(\frac{1}{Pk}\sum_i \sum_j {l^*_{ij}}^2\right) = \frac{1}{Pk}\sum_i \sum_j {l^*_{ij}}^4 - \left(\frac{1}{Pk}\sum_i {l^*_{ij}}^2\right)$$

Mencapai maximum.

Terimakasih