

1805

Konstantin Momirović i Janez Štalec
Katedra za kineziološko psihologiju i sociologiju
Egon Zakrajšek
Inštitut za matematiko, fiziko in mehaniko
Univerze v Ljubljani

**ODREĐIVANJE BROJA ZNAČAJNIH GLAVNIH
KOMPONENATA NA TEMELJU REALNE VARI-
JANCE MATRICE INTERKORELACIJA REAL-
NIH I IMAGE VARIJABLI**

LK KRITERIJ

DETERMINING THE NUMBER OF SIGNIFICANT PRINCIPAL COMPONENTS ON THE BASIS OF REAL VARIANCE OF REAL AND IMAGE VARIABLES INTERCORRELATION MATRIX

It was suggested that a number of significant factors of complete intercorrelation matrixes could be determined to be just sufficient for explaining real variance of real and image variables intercorrelation matrix. The results of some empirical studies showed that such criterion was generally more superior than standard Guttman-Kaiser's criterion. In analysing matrixes of great hypothetical range the proposed criterion was also more superior than PB criterion by Momirović and Štalec.

ОПРЕДЕЛЕНИЕ ЧИСЛА ЗНАЧИМЫХ ГЛАВНЫХ КОМПОНЕНТОВ НА ОСНОВАНИИ РЕАЛЬНОЙ ВАРИАНЦЫ МАТРИЦЫ ИНТЕРКОРРЕЛЯЦИИ РЕАЛЬНЫХ И ИМАЖ ПЕРЕМЕННЫХ

Предлагается определение числа значимых факторов полных корреляционных матриц таким способом, чтобы это число было как раз достаточным для объяснения реальной вариации матрицы ковариации реальных и имаж переменных. Результаты определенных эмпирических исследований показали, что такой критерий обыкновенно лучше принятого Guttman-Kaiser критерия.

В анализе матриц, обладающих большим гипотетическим рангом, предлагаемый критерий является лучше, чем ПБ критерий Шталеца и Момировича.

1. UVOD

Guttmanova donja granica broja značajnih glavnih komponenata sigurno nije najpogodnija mjera broja psihologijski ili kineziologijski značajnih latentnih dimenzija. Browne je pokazao (Browne, 1968) da je to, u stvari, gornja granica psihologijski značajnih latentnih dimenzija, a Štalec i Momirović su utvrdili ne samo slabu konzistentnost Guttman-Kaiserova kriterija za određivanje broja faktora kompletne korelacijske matrice, već i to da ovaj kriterij i suviše često dovodi do hiperfaktorizacije, efekti koje se ne mogu uvijek neutralizirati u toku transformacijskih procedura (Štalec i Momirović, 1971).

No njihov kriterij broja faktora, utemeljen na donjoj granici ukupne zajedničke varijance, iako se ponaša vrlo konzistentno i gotovo nikada ne dovodi do hiperfaktorizacije, ipak nije bez nedostataka. Najveći od njih je svakako opasnost od hipofaktorizacije. Zaista, empirijska su istraživanja pokazala* da je ta opasnost realna; u nevelikom, ali ipak značajnom postotku slučajeva ovaj je kriterij dovodio do odbacivanja ne samo statistički značajnih latentnih dimenzija.

Kriterij broja značajnih glavnih komponenata koji je ovdje predložen temelji se na veličini realne varijance matrice interkorelacija realnih i image varijabli, odnosno, što je naravno isto, matrice interkorelacija realnih varijabli reducirane tako da su u dijagonalu stavljeni koeficijenti determinacije svake varijable na temelju skupa preostalih. Takva matrica u pravilu nije pozitivno definitna. Kako je zbroj karakterističnih korjenova te matrice ravan zbroju koeficijenata determinacije, tj. veličini na kojoj se temelji PB kriterij, zbroj nenegativnih korjenova, dakle onih koji pripadaju realnim karakterističnim vektorima, veći je ili jednak od zbroja koeficijenata determinacije. Otuda, ako se broj značajnih glavnih komponenata definira tako da njihova varijanca dostigne veličinu realne varijance reducirane matrice interkorelacija, opasnost od hipofaktorizacije znatno će se smanjiti, jer je taj broj veći (ili jednak) od broja faktora određenih PB kriterijem, a u pravilu manji (ili jednak) od broja faktora određenih Guttman-Kaiserovim kriterijem, pa je i opasnost od hiperfaktorizacije manja.

2. PRINCIPI I ALOGARITAM ZA ODREĐIVANJE BROJA KOMPONENATA NA TEMELJU REALNE VARIJANCE REDUCIRANOG SISTEMA VARIJABLI

Neka je z n -dimenzionalni vektor varijabli, standardiziranih i normaliziranih, i neka je R matrica interkorelacija tih varijabli. Koeficijenti determinacije svake od tih varijabli na temelju preostalih $(n-1)$ varijabli bit će

* Kriterij Štaleca i Momirovića, poznat pod imenom PB kriterij, primjenjen je, od dana kada je predložen, u više od 200 analiza kognitivnih, konativnih, motoričkih, antropometrijskih i socioloških varijabli.

$$d_j^2 = 1 - 1/r^{jj} \quad j = 1, \dots, n$$

gdje su r^{jj} dijagonalne vrijednosti matrice R^{-1} . Neka je

$$D^2 = (d_j^2) = I - dg^{-1} R^{-1}$$

dijagonalna matrica koeficijenata determinacije, i neka je

$$U^2 = dg^{-1} R^{-1}$$

dijagonalna matrica unikateta u_j^2 . Nazovimo reduciranom matricom korelacija matricu

$$P = R - U^2 = R - I + D^2$$

Lako se može pokazati da je P matrica kovarijanci realnih i image varijabli. Naime, ako je $Z = (z_{ij})$, $i = 1, \dots, N$; $j = 1, \dots, n$ matrica standardiziranih vrijednosti N entiteta u n varijabli, matrica

$$\psi = Z(I - R^{-1}U^2) = (\psi_{ij})$$

matrice image vrijednosti, kovarijance realnih i image varijabli bit će

$$Z'N^{-1}\psi = Z'N^{-1}Z(I - R^{-1}U^2) = R - U^2 = P$$

Kako je dobro poznato (vidi, npr. Harman, 1970) matrica P u pravilu nije pozitivno definitna. Neka su L_j , $j = 1, \dots, t$; $t \leq n$ nenegativni karakteristični korjenovi matrice P . Za sve korjenove te matrice, naravno, vrijedi

$$v = \sum_{i=1}^n L_i = \text{tr } P = \sum_{i=1}^n d_i^2$$

Otuda, za t negativnih korjenova

$$w = \sum_{j=1}^t L_j \geq \sum_{j=1}^n d_j^2$$

Nazovimo w realnom varijancom sistema analiziranih varijabli.

Neka su λ_j karakteristični korjenovi matrice R . Neka je k broj korjenova, koji zadovoljavaju uvjete

$$\sum_{j=1}^k \lambda_j \geq w$$

$$\sum_{j=1}^{k-1} \lambda_j < w$$

Prema tome, k je upravo onoliki broj, koji omogućava da varijanca prvih k glavnih komponenata dostigne realnu varijancu w . Zbog toga što je $w \geq v$, vrijedi $k \geq p$, gdje je p broj značajnih komponenata određenih na temelju PB kriterija, tj.:

$$\sum_{j=1}^p \lambda_j \geq v$$

$$\sum_{j=1}^{p-1} \lambda_j < v$$

Ovo je istraživanje provedeno u okviru teme »Metode za transformaciju i kondenzaciju kinezioloških informacija« koju je financiralo Sveučilište u Zagrebu kao dio istraživačkog programa »Utjecaj tjelesne aktivnosti na psihosomatski status«.

Općenito, vrijedi $k \leq q$, gdje je q broj značajnih komponenata određen Guttman-Kaiserovim postupkom, tj.

$$q = \text{NUM} (\lambda_j \geq 1.)$$

Zbog toga što je q empirijski određen parametar relaciji $k \leq q$ nije moguće na pristojan način dokazati; no serija empirijskih istraživanja je pokazala (vidi odjeljak 2. ovog članka) da ta relacija vrijedi uvijek ako je $N \gg n$.

Algoritam za određivanje faktora sukladno LK* kriteriju, zajedno sa odredbama programskog sistema SS+ koji omogućava određivanje glavnih komponenata korelacijskih matrica reda 200 (na računalu CDC CYBER), odnosno 300 (na računalu UNIVAC 1110 ili 1106) može se napisati na slijedeći način:

Operacija	Programska naredba
(1) $B = (b_{ij})$ $i = 1, N$ $j = 1, n$	INPUT.
(2) $B \rightarrow Z = (z_{ij})$	STATISTICS.
(3) $R = Z' Z N^{-1}$	CORRELATION.
(4) R^{-1} $U^2 = dg^{-1} R^{-1}$	INVERSION.
(5) $P = R - U^2$	MATRIX REDUCTION ($R=R$, $RINV=RINV$, $RRED=P$)
(6) $Y' P Y = L$	DIAGONALISATION ($R = P$, $X = Y$, $LAMBDA = L$)
(7) $X' R X = \Lambda$	DIAGONALISATION ($R = P$, $X = Y$, $LAMBDA = LAMBDA$)
(8) $X \Lambda^{1/2} = H$ (samo za komponente značajne po LK kriteriju)	HOTELLING ($F = F$, $LAMBDA=LAMBDA$, $X = X$, LK , $KRIT = L$)
(9) (štampanje kompletne korelacijske matrice)	PRINT (MATRIX = R, TEXT = KORELACIJSKA MATRICA)
(10) (štampanje $\lambda_j, \lambda_j/n$ $\sum_{j=1}^a \lambda_j/n$ ($a=j, j+1 \dots$) H)	PRINT (MATRIX = F, T, TEXT = GLAVNE KOMPONENTE PO LK KRITERIJU)

Matrica U^2 štampa se kao vektor automatski uz naredbu INVERSION; ta naredba također vodi

* L(udi) K(onj). Ime kriterija motivirano je ponašanjem karakterističnih korjenova bliskih nuli matrica koje nisu pozitivno definitne, i nije bez korelacije sa stavom autora prema metodi koju predlažu.

† S(tistical) S(ystem) je programski sistem za multivarijantnu analizu podataka, razvijen najprije na Institutu za matematiku, fiziku i mehaniku Univerze v Ljubljani a zatim na Računskom centru Instituta za kineziologiju za računala IMB 1130. Kasnije je proširen i adaptiran za računala CDC CYBER i UNIVAC, serija 1100.

do izračunavanja i štampanja $tr(I-U^2) = \text{SUMA SMC}$ i postotka zajedničke varijance.

Očito, primjena LK kriterija nije odveć ekonomična; no problem ekonomičnosti nije više ozbiljan sa računalima na kojima je SS implementiran. Uostalom, operacija (6) može biti iskorištena za određivanje glavnih osovina po nekom od uobičajenih kriterija, tj. za određivanje latentnih dimenzija sukladno faktorskom, a ne komponentnom modelu. Operacija (7) može se upotrijebiti i za određivanje glavnih komponenata po nekom drugom kriteriju. U oba je slučaja, naravno, potrebno napisati onoliko naredaba HOTELLING (uz navođenje odgovarajućih datoteka i kriterija za određivanje broja faktora) koliko je različitih solucija potrebno. Npr., da bi se dobile

- (1) značajne komponente po Guttman-Kaiserovom kriteriju
- (2) značajne komponente po PB kriteriju
- (3) k značajnih glavnih osovina reducirane matrice interkorelacija
dovoljno je, uz navedene naredbe, napisati
- (11) HOTELLING ($X = X$, $LAMBDA = LAMBDA$,
 $F = FK$)
- (12) PRINT (MATRIX = FK, T, TEXT = GLAVNE KOMPONENTE PO GK KRITERIJU)
- (13) HOTELLING ($X = X$, $LAMBDA = LAMBDA$,
 $F = FB$, $RINV$, PBC)
- (14) PRINT (MATRIX = FB, T, TEXT = GLAVNE KOMPONENTE PO PB KRITERIJU)
- (15) HOTELLING ($X = Y$, $LAMBDA = L$, $NUM = k$, $F = FG$)
- (16) PRINT (MATRIX = FG, T, TEXT = GLAVNE OSOVINE REDUCIRANE MATRICE KORELACIJA)

Naravno, bazična struktura dobijena u svakoj soluciji bit će u pravilu transformirana u neku parsimoničnu poziciju. U SS sistemu u tu je svrhu dovoljno napisati naredbe (uz navođenje odgovarajućih datoteka) VARIMAX (za ortogonalne rotacije sukladno Kaiserovom normal-varimax kriteriju), odnosno DIRECT OBLIMIN (za kose transformacije sukladno Zakrajšekovoj modifikaciji Jenrich-Sampsonovog postupka). Naredbama PRINT mogu se odštampati sve matrice dobijene ovim operacijama.

3. REZULTATI NEKIH EMPIRIJSKIH STUDIJA PONAŠANJA LK KRITERIJA

Ponašanje LK kriterija ispitano je na određenom broju empirijskih matrica. Neke od njih (navedene u Tabeli 1 pod $RB = 1,9$) dobijene su na temelju rezultata pokusa modeliranih tako da proizvedu samo jednu značajnu latentnu dimenziju. Hipotetski, matrice dobijene u studijama SOCVAR

A i SOCVAR B morale bi imati jednaki broj značajnih faktora; na temelju prethodnih istraživanja taj bi broj morao iznositi 6.

Model po kome su izvedene studije VAREFK I i VAREFK II pretpostavljao je da sistem varijabli ima samo dvije značajne latentne dimenzije.

Broj faktora i količina objašnjene varijance po LK kriteriju uspoređen je sa brojem faktora i količinom objašnjene varijance po Guttman-Kaiserovom i PB kriteriju.

Vidi se, da se i PB kriterij i LK kriterij ponašaju bolje od standardnog Guttman-Kaiserovog kriterija. PB kriterij je očito superiorniji od LK kriterija u analizi matrica sa hipotetskim rangom 1. Ovaj je kriterij općenito superiorniji od LK kriterija u analizi matrica sa malim hipotetskim rangom, ali osjetljivo inferiorniji u analizi matrica sa relativno velikim hipotetskim rangom. Izuzev LK kriterija u jednom slučaju, sva tri analizirana kriterija ponašaju se konzistentno.

Tabela 1

USPOREDNA ANALIZA GUTTMAN-KAISEROVOG, PB I LK KRITERIJA

RB	JOB	n	N	MV	RV	GK	VGK	PBC	VPB	LK	VLK
1	STRUP	6	30	40.66	49.92	2	64.44	2	64.44	2	64.44
2	KOORNOG	6	46	36.81	44.80	1	51.67	1	51.67	1	51.67
3	KOORRUK	6	46	38.90	46.65	2	67.07	1	47.55	1	47.55
4	REORGN	7	42	17.29	28.12	3	62.40	1	24.58	2	44.97
5	EXSNAG	6	36	55.99	65.27	2	75.30	1	56.39	2	75.30
6	REPTRUP	6	32	23.42	32.28	3	68.83	1	25.92	2	48.61
7	AGILITY	6	47	49.72	56.78	2	70.79	2	70.79	2	70.79
8	PRECIZ	6	72	9.70	18.34	2	47.82	1	27.78	1	27.78
9	KRIT	6	41	39.76	47.27	2	67.66	1	50.71	1	50.71
10	SOCVAR A	34	483	33.53	41.66	11	60.57	4	34.41	6	43.16
11	SOCVAR B	31	543	35.41	43.73	11	65.04	4	37.71	6	47.29
12	VAREFK I	20	467	28.37	35.36	6	54.99	2	35.39	3	38.74
13	VAREFK II	18	467	21.10	28.93	6	55.97	2	29.53	2	29.53

n = broj varijabli

N = broj entiteta

MV = minimalni postotak zajedničke varijance

RV = postotak realne varijance

GK = broj faktora po Guttman-Kaiserovom kriteriju

VGK = postotak objašnjene varijance po Guttman-Kaiserovom kriteriju

PBC = broj faktora po PB kriteriju

VPB = postotak objašnjene varijance po PB kriteriju

LK = broj faktora po LK kriteriju


VLK = postotak objašnjene varijance po LK kriteriju

Premda nije vjerojatno da će LK ikada zamijeniti PB kriterij, nema sumnje da je u nekim slučajevima opravdana njegova primjena, najčešće u komparativne svrhe. Posebno je umjesno primijeniti LK kriterij ako je razlika između broja faktora dobijenih po Guttman-Kaiserovom i PB kriteriju vrlo velika. Čini se da je u takvim slučajevima LK superiorniji nad PB kriterijem.

4. ZAKLJUČAK

Predloženo je da se broj značajnih faktora kompletnih korelacijskih matrica odredi tako, da bude upravo dovoljan za objašnjenje realne varijance matrice kovarijanci realnih i image varijabli. Rezultati nekih empirijskih studija pokazali su da je takav kriterij redovito superiorniji od standardnog Guttman-Kaiserovog kriterija. U analizi matrica sa velikim hipotetskim rangom predloženi je kriterij superiorniji i od PB kriterija Štaleca i Momirovića.

5. LITERATURA

- Browne, M. W. A Note on Lower Bounds for the Number of Common Factors. *Psychometrika*, 1968, 33, No, 2 pp. 233-236.
- Guttman, L. Image Theory for the Structure of Quantitative Variates. *Psychometrika*, 1953, 18, pp. 277-296.
- Harman, H. H. Modern Factor Analysis. 2. ed. The University of Chicago Press. Chicago, 1970.
- Horst, P. Factor Analysis of Data Matrices. Holt, Rinehart and Winston, New York, 1965.
- Štalec, J. i K. Momirović  Ukupna količina valjane varijance kao osnov kriterija za određivanje broja značajnih glavnih komponenata. *Kineziologija*, 1971, 1, br. 1, str. 79-81.

