

P. MATHAN D
PROFESOR

**ODREĐIVANJE BROJA ZNAČAJNIH
FAKTORA MATRICA
INTERKORELACIJA IMAGE VARIJABLI**

Egon Zakrajšek
Inštitut za matematiko, fiziko in mehaniko
Univerze v Ljubljani

Konstantin Momirović
Računski centar Inštituta za kineziologiju

Determining the Number of Significant Factors of Intercorrelation Matrixes of Image Variables

In analysing intercorrelation matrix of image variables it was proposed to consider being significant just that number of factors which is sufficient to explain that part of the matrix trace equivalent to the trace of the image covariance matrix.

In that case, other principal axes limit the space in which random oscillations of real latent dimensions are possible.

Определение числа значимых факторов матриц интеркорреляций иماج изменяемых

Предлагается в анализе матрицы корреляций иماج изменяемых считать значимым именно такое число факторов, которое как раз достаточно для объяснения той части следа матрицы, которая равняется следу матрицы ковариации иماج изменяемых.

Остальные главные оси в таком случае ограничивают пространство, в котором возможны осциляции реальных латентных величин.

O. UVOD

Na temelju Guttmanove image teorije (Guttman, 1953) relacije između image varijabli mogu biti definirane ili kao kovarijance, ili kao korelacije. Ako je

$$Z = (z_{ij}) \quad \begin{array}{l} i = 1, \dots, N \\ j = 1, \dots, n \end{array}$$

matrica standardiziranih i normiranih, po varijablama, rezultata N entiteta S_i u n varijabli V_j matrica interkorelacija tih varijabli je

$$R = Z' Z,$$

a matrica unikviteta

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

Operacijom

$$Z(I - R^{-1} U^2) = Y$$

bit će originalne varijable V_j transformirane u image varijable W_j , pa operacija

$$Y' Y = C$$

daje matricu varijanci-kovarijanci image varijabli W . Najčešće se matrica C izračunava ekvivalentnom operacijom

$$C = R + U^2 R^{-1} U^2 - 2 U^2$$

i osnov je za grupu transformacijskih i kondenzacijskih postupaka poznatih pod imenom Image faktorska analiza (IFA).

Varijance image varijabli su

$$D_j^2 = 1 - U_j^2$$

pa su prema tome jednake koeficijentima determinacije varijabli V_j na temelju svih ostalih varijabli V_k , $k = 1, \dots, n-1$; ($j \neq k$).

Normiranjem vektora matrice C može se dobiti matrica interkorelacija image varijabli. Prema tome, u matrici dobijenoj operacijom

$$D^{-1} C D^{-1} = P$$

Ovo istraživanje provedeno je u okviru programa »Utjecaj tjelesne aktivnosti na psihosomatski status«. Dio programa posvećen istraživanju metoda za transformaciju i kondenzaciju kinezioloških informacija sufinanciralo je Sveučilište u Zagrebu.

gdje je

$$D^2 = I - U^2 = (D_j^2)$$

$$j = 1, \dots, n$$

nalaziti će se korelacije image varijabli, ako je varijanća svake od njih svedena na 1. Matrica P relativno se rijetko upotrebljava prilikom faktorskih postupaka. Jedan od vjerojatno najvažnijih razloga za to je odsustvo bilo kakvog razboritog kriterija za određivanje broja značajnih faktora (tj. za određivanje minimalnog ranga neke matrice koja dobro aproksimira matricu P , naravno vodeći računa i o dijagonalnim elementima te matrice).

Guttmanov teorijski maksimum za broj značajnih image faktora ($n/2$) vredi naravno i za matricu P . Ali, u stvari, taj je kriterij, kako je dobro poznato, sasvim neadekvatan. Horst predlaže (Horst, 1965), da se broj značajnih faktora (k , recimo) fiksira unapred, ili da se odredi na temelju nekog unapred definiranog postotka ukupne varijance koga mora »objasniti« prvih k karakterističnih korjenova matrice P . Oba su kriterija sasvim subjektivna, a njihovo je ponašanje, u praksi, vrlo loše. Neki autori smatraju značajnim svaki karakteristični korijen koji je veći ili jednak 1; no taj kriterij, kako je pokazao Browne (Browne, 1968) daje gornju granicu broja značajnih faktora već pri analizi kompletnih interkorelacijskih matrica, uslijed čega su neki faktori u znatnoj mjeri kontaminirani uniknom varijancom (Cattell, 1966). Primijenjen na matricu P , čiji su vektori namjerno kontaminirani uniknom varijancom, ovaj kriterij mora proizvesti, u općem slučaju, određeni broj faktora kojih je varijanća najvećim dijelom unikna, ili čak samo error varijanća. Zaista, kako je varijanća svake varijable iz P

$$D_i^2 + U_j^2 = 1$$

$$j = 1, \dots, n$$

racionala ovog kriterija (smislena linearna kombinacija skupa varijabli ne može emitirati manju količinu informacija, dakle imati manju varijancu, od bilo koje varijable iz tog skupa) potpuno je neprimjenljiva pri analizi image varijabli čija je varijanća normirana na 1.

U ovom radu predložen je kriterij, koji ima sličnu logiku onoj koju ima i PB kriterij, koga su Štalec i Momirović (Štalec i Momirović, 1971) predložili za određivanje broja značajnih faktora pri analizi kompletnih interkorelacijskih matrica. Osnov je toga kriterija, da broj značajnih faktora mora biti upravo toliki, koliki je potreban da bi se mogla objasniti minimalna količina valjane varijance nekog skupa varijabli.

Spomenuti kriterij izvrsno se ponaša u praksi (Štalec i Momirović, 1971; Momirović i Štalec, 1972). Lako se može pokazati da je logika na kojoj je taj kriterij osnovan još adekvatnija pri

određivanju broja značajnih faktora matrica interkorelacija image varijabli, budući da je valjana varijanca sistema image varijabli, u stvari, tačno poznata.

1. LC KRITERIJ I ALGORITAM ZA ODREĐIVANJE BROJA ZNAČAJNIH GLAVNIH OSOVINA MATRICA INTERKORELACIJA IMAGE VARIJABLI

Neka su L_j , $j = 1, \dots, n-1$ karakteristični korjenovi, dobijeni rješavanjem karakterističnih jednadžbi

$$(P - L_j I) X_j = 0$$

$$j = 1, \dots, n-1$$

pri čemu je, naravno, i ovdje kao i kod matrice C n -ti korijen sigurno 0 ., dakako u granicama numeričke greške. Za matricu P vrijedi, kao i za svaku pozitivno semidefinitnu interkorelacijsku matricu

$$\text{tr } P = \sum_{j=1}^n L_j = n$$

Ali, istovremeno, vrijedi i

$$\sum_{j=1}^n D_j^2 + \sum_{j=1}^n U_j^2 = n$$

Međutim, valjana varijanca sistema image varijabli i dalje je

$$w = \sum_{j=1}^n D_j^2$$

pa se može smatrati razboritim zadržati samo onoliko glavnih osovina matrice P , koliko je upravo dovoljno da se dostigne w , ako je matrica minimalnog ranga, koja na zadovoljavajući način aproksimira matricu P (u smislu Eckert-Youngovog tretiranja ovog problema),

$$\hat{P} = X L X'$$

gdje su X matrica prvih k karakterističnih vektora, a L matrica prvih k karakterističnih korjenova od P . Taj broj korjenova i vektora definira se tako da vrijedi

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

ako je

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

pa je otuda

$$\text{tr } \hat{P} \geq w$$

varijanca sistema varijabli dobijenih operacijom

$$Y D^{-1} X = K$$

jer je, očito, L_j varijanca neke varijable K_j .

Prostor od $(n-1)$ dimenzija definiran varijablama iz $Y D^{-1}$ može se shvatiti kao prostor u kome su moguće slučajne oscilacije k -dimenzionalnog sistema varijabli iz K . Veličina tih oscilacija proporcionalna je veličini ukupne unikne varijance, dakle $(n-w)$.

Čini se prema tome, da je LC^* kriterij pristojna zamjena za, za sada nepostojeći, egzaktni statistički test značajnosti korjenova matrice P , i da se s valjanim razlozima može primijeniti pri određivanju broja značajnih faktora matrice interkorelacija image varijabli.

2. ZAKLJUČAK

Predloženo je, da se pri analizi matrice korelacija image varijabli smatra značajnim upravo onoliko broj faktora, koji je tek dovoljan da se objasni onaj dio traga matrice, koji je ravan tragu matrice kovarijanci image varijabli.

Preostale glavne osovine u tom slučaju omeđuju prostor, u kome su moguće slučajne oscilacije realnih latentnih dimenzija.

LITERATURA

1. Browne, M. W.
A Note on Lower Bounds for the Number of Common Factors.
Psychometrika, 33, pp. 233—236 (1968)
2. Cattell, R. B.
Handbook of Multivariate Experimental Psychology. Rand McNally, Chicago, 1966

* L (acrimae) C (hristi). Ime kriterija nije potpuno ortogonalno na dimenzije koje su definirale situaciju u kojoj je LC kriterij nastao.

3. Guttman, L.
Image Theory for the Structure of Quantitative Variates.
Psychometrika, 18, pp. 277—296 (1953)
4. Horst, P.
Factor Analysis of Data Matrices.
Holt, Rinehart and Winston, New York (1965)
5. Momirović, K. i J. Štalec
DAMEAN i DMAX kriteriji za određivanje bro-
ja značajnih image faktora pri analizi zadataka
u psihologijskim testovima.
Zbornik saopćenja referiranih na »Danima Ra-
mira Bujasa«, Zagreb, 1973 (u štampi)
6. Štalec, J. i K. Momirović
Ukupna količina valjane varijance kao osnov
kriterija za određivanje broja značajnih glav-
nih komponenata.
Kineziologija, 1. Br. 1, str. 77—81 (1971)

