

Priloženo
Priloženo

**UKUPNA KOLIČINA VALJANE VARIJANCE KAO OSNOV
KRITERIJA ZA ODREĐIVANJE BROJA ZNAČAJNIH GLAV-
NIH KOMPONENATA**

asist. Janez ŠTALEC i prof. dr Konstantin MOMIROVIC,
iz Odjela za kineziološki psihologiju i sociologiju

Summary:

**TOTAL VALIDITY VARIANCE AS A FUNDAMENTAL
CRITERION IN DETERMINING THE NUMBER OF
CARDINAL COMPONENTS**

The analysis of the criteria for the determination of the number of significant components defined by the smallest number of roots necessary to obtain the sum of the coefficients of the variables determination in a common space proved a reasonable procedure, very often superior to the criterion based on the magnitude of the characteristic roots. The proposed criterion may be considered worth investigating in greater detail.

(1) UVOD

Kako je dobro poznato (Harman, 1962; Cattell, 1966; Überla, 1968), problem određivanja značajnih glavnih komponenata nereducirane matrice interkorelacija nije riješen na zadovoljavajući način. Uobičajeni kriterij, po kome se značajnim smatraju glavne komponente kojih su karakteristični korjenovi veći od 1.00 ima najmanje dva nedostatka:

1) analize matrica dobivenih na uzorcima iz iste populacije ispitanika daju često različiti broj značajnih faktora

2) u većini slučajeva, osobito ako razlike između korjenova nisu velike, ovaj kriterij daje suviše veliki broj faktora, koji se često ne gube kod transformacijskih postupaka, pa se tako konačna faktorska struktura kontaminira eror faktorima.

Kako su upozorili Kaiser i Dickman, a zatim dokazao Horn (Horn, 1965), taj je kriterij, s psihometrijske točke gledišta, u stvari kriterij za određivanje gornje granice broja značajnih faktora.

Pokušaj da se problem zaobide unaprijed fiksiranim brojem faktora ima nekog smisla samo onda, ako je »pravi« rang korelacijske matrice unaprijed poznat, što je doduše moguće kod konfirmativnih, ali ne i kod eksplorativnih analiza.

Kriterij osnovan na unaprijed određenom postotku cjelokupne varijance ima također bar dva nedostatka:

1) analizom matrica interkorelacija dobivenih na uzorcima iz iste populacije ispitanika dobiva se, u pravilu, vrlo različit broj značajnih faktora

2) praktički nije moguće unaprijed znati koliki postotak ukupne varijance odgovara pravoj ukupnoj varijanci latentnih dimenzija za različite sustave manifestnih varijabli.

Nešto bolji rezultati dobiveni su primjenom kriterija koji se osniva na unaprijed određenom najmanjem postotku varijance koji mora iscrpljivati svaka glavna komponenta da bi se mogla smatrati značajnom. Nije, na žalost, utvrđen nikakav racionalni osnov za određivanje veličine toga postotka, iako su neki intuitivni kriteriji dali zanimljive rezultate.

Osnovni motiv bilo kojeg kriterija je, da glavne komponente koje se smatraju značajnim budu što manje kontaminirane eror varijancom. Opravdano je stoga smatrati značajnim samo onoliki broj glavnih komponenata koji je dovoljan za eksplikaciju noneror varijance čitavog sustava manifestnih varijabli, tretirajući, s obzirom na osnovni cilj ma koje metode faktorske analize, cijelu uniku varijancu kao eror varijancu.

Ako je k stvarni rang neke korelacijske matrice R reda n , osnovna struktura te matrice je

$$R = QDQ' + PCP'$$

gdje je Q matrica prvih k karakterističnih vektora, D dijagonalna matrica prvih k karakterističnih korjenova, a P i C matrice preostalih $(n - k)$ karakterističnih vektora, odnosno korjenova.

* Autori zahvaljuju A. Hošek koja je sudjelovala u razradi ove ideje.

Ako je V noneror varijanca matrice R , vrijedi

$$V = \text{tr}(QDQ') \\ = \text{tr} D$$

pa se k može odrediti ako je poznata noneror varijanca V .

(2) ALGORITAM ZA ODREĐIVANJE BROJA ZNAČAJNIH GLAVNIH KOMPONENATA

Kako su pokazala Guttmanova istraživanja, najmanja veličina valjane varijance ma koje varijable iz nekog sustava varijabli koji pokriva područje sa zajedničkim obilježjima jednaka je koeficijentu determinacije te varijable na temelju svih ostalih varijabli iz toga sustava (Darroch, 1965; 1969; Browne, 1968, 1).

Neka je R matrica interkorelacija n manifestnih varijabli dobivena na uzorku od N subjekta. Dijagonalna matrica koeficijenata determinacije analiziranih varijabli bit će

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

Donja granica ukupne količine valjane varijance matrice R je, očito,

$$V = \text{tr} W^2$$

Razložito je definirati najmanji broj značajnih glavnih komponenata tako, da iscrpljuju upravo toliku količinu ukupne varijance.

Neka je

$$(R - D)X_i = 0$$

$$i = 1, \dots, n$$

karakteristična jednadžba matrice R , gdje su D_i karakteristični korjenovi a X_i karakteristični vektori.

Broj k značajnih glavnih komponenata može se odrediti tako da budu zadovoljena oba slijedeća uvjeta

$$\sum_{i=1}^k D_i \geq V$$

$$k = \min$$

Na taj način sigurno se dobiva donja granica broja značajnih glavnih komponenata. Taj će broj u pravilu biti manji, ili jednak broju dobivenom na temelju uobičajenog kriterija $D_i \geq 1.00$. No kako su pokazala nekoja istraživanja (Linn, 1968; Tucker, Koopman i Linn, 1969), realni broj faktora obično je manji od onoga, koji se dobiva na temelju uobičajenih statističkih kriterija, jer eror varijanca povezana sa greškama uzorka ili greškama mjerenja tendira produkciji suvišnih faktora (Cartwright, 1965; Cliff i Pennell, 1967; Browne, 1968, 2).

Treba upozoriti, da predloženi kriterij nije jednak Guttmanovu postupku koji se temelji na određivanju karakterističnih vrijednosti matrice

$$r = R - I + W^2$$

jer se u tom slučaju smatraju značajnim sve komponente povezane sa pozitivnim korjenima matrice r koja očito nije pozitivno definitna. Iskustvo je pokazalo da je tako određen broj faktora pretjeran i da se približava teorijskoj gornjoj granici koja iznosi $n/2$. Osim toga, kod matrica koje nisu pozitivno definitne numerička greška kod određivanja korjenova bliskih nuli može biti znatna, pa nije uvijek moguće ocijeniti koji su korjenovi stvarno pozitivni.

(3) NEKI REZULTATI DOBIVENI PRIMJENOM PREDLOŽENOG KRITERIJA

Ponašanje predloženog kriterija ispitano je na većem broju korelacijskih matrica. U svrhu usporedbe prikazani su i rezultati dobiveni primjenom uobičajenog kriterija, prema kojemu se značajnim smatraju glavne komponente čiji su karakteristični korjenovi veći ili jednaki 1.00. U svim tabelama koje slijede simboli imaju slijedeće značenje:

- N = broj ispitanika
 n = broj varijabli
 H = broj značajnih glavnih komponenata po kriteriju $D_i \geq 1.00$
 $\%_0H$ = postotak ukupne varijance koji iscrpljuje H komponenta
 $\%_0V$ = donja granica postotka valjane varijance
 F = broj komponenata određen na temelju veličine valjane varijance V.

U tabeli 1. prikazani su rezultati dobiveni analizom 56 antropometrijskih, funkcionalnih i motoričkih varijabli, primijenjenih na uzorcima muškaraca (M) i žena (Ž) starih 15 i 17 godina. S obzirom na primijenjene instrumente i fazu razvoja, mogao se očekivati jednaki broj latentnih dimenzija u svim uzorcima.

TABELA 1.

grupa	N	n	H	$\%_0H$	$\%_0V$	F
M15	430	56	13	67.41	55.71	8
M17	436	56	14	65.82	51.57	8
F15	426	56	15	65.98	50.32	8
F17	423	56	13	62.32	49.88	8

Kako se vidi, predloženi kriterij na konzistentan način određuje broj značajnih faktora. Bitnije je međutim to, što je nakon transformacije H glavnih komponenata u varimax i oblimin poziciju bilo moguće nedvosmisleno interpretirati samo 8 ili 9 faktora, dok su ostali bili dual ili single faktori, ili čak virtualno nul vektori.

Ukupno 68 kognitivnih i konativnih varijabli primijenjeno je skupini od 1342 maloljetna delinkventa, koja je predstavljala reprezentativan uzorak jugoslavenske populacije maloljetnih delinkvenata. Iz te skupine izvučeni su subuzorci onih, kojima je istekla sankcija (A), onih, kojima je sankcija upravo izrečena (B) te onih, koji žive na teritoriju SR Hrvatske, SR Slovenije i SR Bosne i Hercegovine (C). Rezultati su prikazani u tabeli 2.

TABELA 2.

grupa	N	n	H	$\%_0H$	$\%_0V$	F
SFRJ	1342	68	12	62.42	51.10	6
A	624	68	12	63.00	53.08	7
B	595	68	13	64.40	54.38	8
C	728	68	12	62.91	43.24	7

U ovom slučaju predloženi kriterij bio je inkonzistentniji od standardnog kriterija. No uzorci nisu bili slučajni, već su se sistematski razlikovali bilo zbog toga, što je penalni tretman mogao djelovati na strukturu, pa i broj latentnih dimenzija, bilo uslijed toga što su maloljetnici iz skupine C poticali iz sredine koja se po ekonomskim, sociološkim, etničkim i kulturnim obilježjima razlikovala od sredine iz koje su poticali ostali maloljetnici. Nakon transformacija u varimax i oblimin poziciju samo je 9 faktora bilo kongruentno (od H izoliranih) i samo su se oni mogli smisljeno interpretirati. Ostali su bili većinom dual faktori.

Iako su se svi faktori, nakon transformacija koji su bili dobiveni po predloženom kriteriju mogli smisljeno interpretirati, ipak je očito da je u ovom slučaju predloženi kriterij dao manji broj latentnih dimenzija od onoga, kojemu se mogao pridati psihologijski realitet.

Baterija za procjenu patoloških konativnih faktora 18 PF bila je primijenjena na uzorcima učenika i učenica završnih razreda srednjih škola. Analize su izvršene posebno na uzorku učenika i uzorku učenica, kao i na cijeloj skupini ispitanika. Rezultati su prikazani u tabeli 3.

TABELA 3.

grupa	N	n	H	$\%_0H$	$\%_0V$	F
MUŠKI	120	18	3	67.39	62.93	3
ŽENE	356	18	3	62.98	55.53	2
UKUPNO	476	18	2	59.71	57.04	2

Premda se na prvi pogled ne vide bitne razlike između oba kriterija, treći faktor u skupini žena bio je, nakon varimax i oblimin transformacija virtualno nul vektor, pa je prema tome tri dimenzije u toj skupini bilo previše.

Zanimljiva i prilično ubjedljiva informacija o vrijednosti predloženog kriterija dobivena je faktorskom analizom zadataka u testovima znanja. Konstruirano je po dve paralelne forme testova znanja hrvatskog jezika, matematike, fizike i općeg znanja, procijenjenog na temelju zadataka iz svih školskih predmeta. Odgovori na zadatke nisu bili binarni, pa se svaki zadatak mogao tretirati kao kontinuirana varijabla. Rezultati su prikazani u tabeli 4.

TABELA 4.

test	N	n	H	$\%_0H$	$\%_0V$	F
FIZ 1	381	28	9	53.63	22.39	2
FIZ 2	193	28	10	59.41	26.06	3
MAT 1	381	23	6	50.92	26.38	2
MAT 2	198	23	7	58.18	32.21	2
HRV 1	377	21	7	53.33	22.80	2
HRV 2	195	21	4	50.44	37.50	2
KOM 1	383	32	8	51.14	29.16	3
KOM 2	197	32	9	60.55	40.33	4

Standardni kriterij je inkonzistentniji od predloženog. Osim toga, po tome je kriteriju izoliran sasvim pretjeran broj faktora, jer su testovi tako konstruirani, a i faktor-ska analiza tih testova u prostoru kognitivnih dimenzija, školskih ocjena i konativnih dimenzija je to pokazala, da mjere samo znanje i kognitivni generalni faktor. Samo su neki testovi imali (negativne) korelacije sa generalnim faktorom neurotizma. Otuda su dva ili tri faktora dovoljna da objasne strukturu svakog testa, pa je nerazborito pretpostavljati, da je veći broj faktora potreban da bi se objasnila struktura zadataka, pogotovo što je normalizacijom eliminirana mogućnost pojave faktora koji zavise od težine zadataka. Razlike u broju značajnih faktora kod paralelnih testova vjerojatno su posljedica pre-malog i nejednakog broja ispitanika, jer je predloženi kriterij dao 3 faktora u FIZ 2 odnosno 1 u KOM 2 samo zbog razlika reda 10^{-2} .

Prema tome, predloženi je kriterij u većini slučajeva davao bolje rezultate od uobičajenog kriterija, osnovanog na veličini karakterističnih korjenova.

(4) ZAKLJUČAK

Analiza kriterija za određivanje broja značajnih glavnih komponenata definiranog najmanjim brojem korjenova potrebnih da bi se dostigao zbroj koeficijenata determinacije varijabli u zajedničkom prostoru, pokazala je da se radi o razboritom postupku, često superiornom kriteriju osnovanom na veličini karakterističnih korjenova. Može se smatrati da predloženi kriterij zaslužuje da ga is-trraživanja.

W. D. NO

LITERATURA

- 1) Browne, M. W.
A note on lower bounds for the number of common factors. Psychometrika, 33, No 2, June 1968, pp 233-236.
- 2) Browne, M. W.
A comparison of factor analytic techniques. Psychometrika, 33, No 3, September 1968, pp 267-334.
- 3) Cartwright, D. S.
A note on some modifications of latent roots and vectors. Psychometrika, 30, No 3, September 1965, pp 319-321.
- 4) Cattell, R. B.
Handbook of multivariate experimental psychology. Rand McNally, Chicago, 1966.
- 5) Cliff, N. and R. Pennell.
The influence of communalities, factor strength and loading size on the sampling characteristics of factor loadings. Psychometrika, 32, No 3, September 1967, pp 309-326.
- 6) Darroch, J. N.
A set of inequalities in factor analysis. Psychometrika, 30, No 4, December 1965, pp 449-453.
- 7) Darroch, J. N.
Some further inequalities and an identity in factor analysis. Psychometrika, 34, No 1, March 1969, pp 45-49.
- 8) Harman, H. H.
Modern factor analysis. The University of Chicago Press, Chicago, 1962.
- 9) Horn, J. L.
A rationale and test for the number of factors in factor analysis. Psychometrika, 30, No 2, June 1965, pp 179-185.
- 10) Linn, R. L.
A Monte Carlo approach to the number of factor problem. Psychometrika, 33, No 1, March 1968, pp 37-71.
- 11) Tucker, L. R., Koopman, R. F. and R. L. Linn.
Evaluation of factor analytic research procedures by means of simulated correlation matrices. Psychometrika, 34, No 4, December 1969, pp 421-459.
- 12) Überla, K.
Faktorenanalyse. Springer, Berlin, 1968.

