

JEDAN NOVI METOD ZA ROTACIJU U DISKRIMINACIONOJ ANALIZI

Lazar Tenjović i Aleksandar Zorić
Filozofski fakultet Univerziteta u Beogradu

Radi jednostavnije interpretacije diskriminacionih funkcija dobijenih kanoničkom diskriminacionom analizom primenjuju se različiti postupci rotacije matrice strukture ili matrice koeficijenata za funkciju. U ovom radu predložen je Centrovax postupak rotacije matrice strukture i matrice koeficijenata koji se zasniva na prethodnoj rotaciji matrice centroida grupa na značajnim diskriminacionim funkcijama. Program za implementaciju predloženog algoritma napisan je u MATRIX jeziku, a predloženi postupak rotacije prikazan je na jednom ilustrativnom numeričkom primeru.

KLJUČNE REČI: kanonička diskriminaciona analiza / rotacija centroida

„...i za metodu koja se ovde predlaže potrebne su brojne empirijske provere njenog stvarnog ponašanja da bi se mogao doneti konačan sud o njenoj vrednosti, upotrebljivosti i korisnosti, bez obzira na formalnu adekvatnost operacija na osnovu kojih je ta metoda izvedena“.

Konstantin Momirović

UVOD

Kanonička diskriminaciona analiza uobičajeno se primenjuje za analizu razlika između dveju ili više subpopulacija u pogledu skupa kvantitativnih normalno distribuiranih varijabli (Rao, 1965; Anderson, 1984; Momirović, 1997). Pri tumačenju strukture razlika između subpopulacija od prvorazrednog značaja je otkrivanje sadržaja značajnih diskriminacionih funkcija. Sadržaj diskriminacionih funkcija identifikuje se na osnovu matrice koeficijenata i matrice strukture ovih funkcija (Tatsuoka, 1971; Momirović & Zorić, 1996). Radi lakše identifikacije diskriminacionih funkcija matrice koeficijenata i strukture se često rotiraju (Rencher, 1992; Lunneborg & Lunneborg, 1978).

Kanonička diskriminaciona analiza često se koristi i za analizu razlika između grupa koje su rezultat neke prethodno izvedene analize grupisanja (Tenjović, 2002). U tom slučaju osnovni problem koji treba rešiti nije definisanje prirode razlika između grupa već teorijsko definisanje samih grupa. Definisanje grupa u takvim situacijama vršimo prvenstveno na osnovu matrice centroida na diskriminacionim funkcijama. Jedan od načina da se olakša korišćenje matrice centroida u definisanju grupa mogao bi se sastojati u rotaciji ove matrice u poziciju u kojoj su grupe predstavljene centroidima koji su što je moguće više međusobno udaljeni i, istovremeno, raspoređeni tako da njihov raspored na različitim funkcijama bude što jednostavniji.

PRELIMINARNE DEFINICIJE

Neka je E skup entiteta koji je na slučajan način izvučen iz jedne nehomogene populacije P koja se sastoji od g subpopulacija

$$E = \{e_i, i = 1, \dots, n\} \subset P = \cup P_k, \quad k = 1, \dots, g$$

i neka skup V predstavlja reprezentativan uzorak varijabli iz univerzuma varijabli U

$$V = \{v_j, j = 1, \dots, m\} \subset U$$

koje su multivarijatno normalno distribuirane u svakoj od subpopulacija P_k . Tada se opisom entiteta iz skupa E na skupu varijabli V dobija matrica

$$\mathbf{Z} = (z_{ij}) = E \otimes V$$

pri čemu element z_{ij} predstavlja meru i -tog entiteta na j -toj varijabli. Radi jednostavnosti u notaciji pretpostavimo da važe sledeće jednakosti

$$\mathbf{Z}^t \mathbf{e} = \mathbf{0}$$

$$\text{diag}(\mathbf{Z}^t \mathbf{Z}) = \mathbf{I}$$

pri čemu \mathbf{e} predstavlja sumacioni vektor reda n , a \mathbf{I} matricu identiteta reda m . Tada se matrica interkorelacija varijabli iz skupa V , u oznaci \mathbf{R} , definiše na sledeći način:

$$\mathbf{R} = \mathbf{Z}^t \mathbf{Z}$$

Neka je W jedna nominalna varijabla koja na dovoljan i potpun način definiše subpopulacije iz P . Varijabla W sadrži g modaliteta:

$$W = \{w_k, k = 1, \dots, g\}$$

U tom slučaju opisom entiteta iz skupa E na varijabli W dobijamo matricu \mathbf{S} ,

$$\mathbf{S} = (s_{ik}) = E \otimes W$$

koja predstavlja selektorsku matricu, dakle matricu u kojoj je varijabla W predstavljena u kompletnoj disjunktnoj formi. Elementi matrice \mathbf{S} , u oznaci s_{ik} određeni su na osnovu sledećeg pravila:

$$s_{ik} = 1 \quad | \quad e_i \in w_k$$

$$s_{ik} = 0 \quad | \quad e_i \notin w_k$$

Unakrsni proizvod matrice \mathbf{S} predstavlja regularnu matricu. Prema tome, nema nepraznih kategorija nominalne varijable W u uzorku entiteta E .

Definišimo projektorsku matricu

$$\mathbf{P} = \mathbf{S}(\mathbf{S}^t \mathbf{S})^{-1} \mathbf{S}^t$$

i matricu

$$\mathbf{G} = \mathbf{PZ} = \mathbf{S}(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{Z}$$

koja predstavlja projekcije vektora iz \mathbf{Z} u prostor koji razapinju vektori iz \mathbf{S} . Element g_{ij} u matrici \mathbf{G} predstavlja aritmetičku sredinu varijable v_j za subpopulaciju kojoj element e_i pripada.

Unakrsni proizvod matrice \mathbf{G} je matrica \mathbf{A} :

$$\mathbf{A} = \mathbf{G}^t\mathbf{G} = \mathbf{Z}^t\mathbf{P}^t\mathbf{PZ} = \mathbf{Z}^t\mathbf{PZ} = \mathbf{Z}^t\mathbf{S}(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{Z}$$

koja, budući da je jednaka i proizvodu matrica \mathbf{Z} i \mathbf{G} , predstavlja matricu međugrupnog varijabiliteta. Naravno, u tom slučaju je matrica unutar-grupnog varijabiliteta \mathbf{W} jednaka

$$\mathbf{W} = \mathbf{R} - \mathbf{A}$$

KANONIČKA DISKRIMINACIONA ANALIZA

Kanonička diskriminaciona analiza je metoda koja ekstremizuje linearne veze između linearne kombinacije kvantitativnih varijabli iz skupa V i nominalne varijable W , ili, drugačije rečeno, metoda koja ekstremizuje linearne veze između vektora matrice \mathbf{Z} i matrice \mathbf{G} . Na ovaj način definisan problem može se zapisati ovako:

$$\left. \begin{array}{l} \mathbf{Zx}_p = \mathbf{k}_p \\ \mathbf{Gy}_p = \mathbf{l}_p \end{array} \right| \begin{array}{l} \rho_p = \mathbf{k}_p^t \mathbf{l}_p = \mathbf{x}_p^t \mathbf{Z}^t \mathbf{G} \mathbf{y}_p = \mathbf{x}_p^t \mathbf{A} \mathbf{y}_p = \max \\ \mathbf{k}_p^t \mathbf{k}_q = \mathbf{x}_p^t \mathbf{Z}^t \mathbf{Z} \mathbf{x}_q = \mathbf{x}_p^t \mathbf{R} \mathbf{x}_q = \delta_{pq} \\ \mathbf{l}_p^t \mathbf{l}_q = \mathbf{y}_p^t \mathbf{G}^t \mathbf{G} \mathbf{y}_q = \mathbf{y}_p^t \mathbf{A} \mathbf{y}_q = \delta_{pq} \\ \mathbf{k}_p^t \mathbf{l}_q = 0, \quad p \neq q \\ p, q = 1, \dots, b, \quad b = \min(g-1, m) \end{array}$$

Pri tome δ_{pq} predstavlja Kronekerov simbol. Postavljeni problem ekstremizacije sa konzistentnim linearnim ograničenjima može se rešiti metodom Lagranžovih multiplikatora definisanjem funkcije

$$F(\mathbf{x}_p, \mathbf{y}_p, \lambda_p, \eta_p) = \mathbf{x}_p^t \mathbf{A} \mathbf{y}_p - \lambda_p (\mathbf{x}_p^t \mathbf{R} \mathbf{x}_p - 1) - \eta_p (\mathbf{y}_p^t \mathbf{A} \mathbf{y}_p - 1),$$

te njenim parcijalnim diferenciranjem po nepoznatim argumentima,

$$\partial F / \partial \mathbf{x}_p = \mathbf{A}\mathbf{y}_p - 2\lambda_p \mathbf{R}\mathbf{x}_p$$

$$\partial F / \partial \mathbf{y}_p = \mathbf{A}\mathbf{x}_p - 2\eta_p \mathbf{A}\mathbf{y}_p$$

i izjednačavanjem dobijenih izvoda sa nulom. Sređivanjem ovih izraza i množenjem sleva prvog izraza sa \mathbf{x}_p^t , a drugog sa \mathbf{y}_p^t dobija se

$$\rho_p = \mathbf{x}_p^t \mathbf{A}\mathbf{y}_p = 2\lambda_p \mathbf{x}_p^t \mathbf{R}\mathbf{x}_p = 2\lambda_p$$

$$\rho_p = \mathbf{y}_p^t \mathbf{A}\mathbf{x}_p = 2\eta_p \mathbf{y}_p^t \mathbf{A}\mathbf{y}_p = 2\eta_p$$

Iz prethodnih jednakosti sledi da je

$$\lambda_p = \eta_p$$

Na osnovu izvedenih relacija dobijaju se sledeće jednakosti:

$$\mathbf{A}\mathbf{y}_p = \rho_p \mathbf{R}\mathbf{x}_p$$

$$\mathbf{A}\mathbf{x}_p = \rho_p \mathbf{A}\mathbf{y}_p$$

Sređivanjem drugog izraza dobija se relacija između vektora linearnih koeficijenata za varijable iz matrice \mathbf{Z} i koeficijenata za varijable iz matrice \mathbf{G}

$$\mathbf{y}_p = \rho_p^{-1} \mathbf{x}_p$$

Na osnovu ove relacije i izraza $\mathbf{A}\mathbf{y}_p = \rho_p \mathbf{R}\mathbf{x}_p$ sledi

$$\rho_p^{-1} \mathbf{A}\mathbf{x}_p = \rho_p \mathbf{R}\mathbf{x}_p$$

Daljim sređivanjem dobijamo

$$(\mathbf{A} - \rho_p^2 \mathbf{R})\mathbf{x}_p = \mathbf{0}$$

čime se problem svodi na opšti problem svojstvenih vrednosti. Svojtvene vrednosti dobijene rešavanjem ovog problema predstavljaju kvadrate kanoničkih korelacija, a svojstveni vektori koji odgovaraju dobijenim svojstvenim vrednostima služe za konstruisanje kanoničkih diskriminacionih funkcija.

Neka je

$$\mathbf{X} = (\mathbf{x}_p), \quad p = 1, \dots, b$$

matrica čije su kolone netrivialni svojstveni vektori matrice $\mathbf{R}^{-1}\mathbf{A}$, a

$$\mathbf{P} = (\rho_{pp}), \quad p = 1, \dots, b$$

dijagonalna matrica sa koeficijentima kanoničkih korelacija na dijagonali. Tada je

$$\mathbf{K} = \mathbf{Z}\mathbf{X}$$

matrica kanoničkih diskriminacionih funkcija. Na osnovu uslova pod kojima se rešava problem ekstremizacije važe i sledeće jednakosti:

$$\mathbf{X}^t\mathbf{R}\mathbf{X} = \mathbf{I}$$

$$\mathbf{X}^t\mathbf{A}\mathbf{X} = \mathbf{P}^2$$

Prema tome linearne kombinacije početnih varijabli iz \mathbf{Z} i početnih varijabli iz \mathbf{G} čine biortogonalni sistem varijabli, ali koeficijenti u matricama \mathbf{X} i \mathbf{Y} nisu ortogonalni.

Interpretacija kanoničkih diskriminacionih funkcija uobičajeno se zasniva na matrici koeficijenata za diskriminacione funkcije (matrici \mathbf{X}) i matrici strukture. Matrica strukture diskriminacionih funkcija, tj. matrica korelacija početnih varijabli iz matrice \mathbf{Z} i kanoničkih diskriminacionih funkcija iz matrice \mathbf{K} , u oznaci \mathbf{F} , definisana je na sledeći način:

$$\mathbf{F} = \mathbf{Z}^t\mathbf{K} = \mathbf{Z}^t\mathbf{Z}\mathbf{X} = \mathbf{R}\mathbf{X}$$

Identifikacija sadržaja kanoničkih diskriminacionih funkcija vrši se i na osnovu položaja ispitivanih grupa na dobijenim diskriminacionim dimenzijama. U tu svrhu koristi se matrica centroida, u oznaci \mathbf{C} , koja je definisana na sledeći način:

$$\mathbf{C} = (\mathbf{c}_{kj}) = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{K}, \quad k = 1, \dots, g \quad j = 1, \dots, b.$$

Pri tome, element \mathbf{c}_{kj} matrice \mathbf{C} predstavlja aritmetičku sredinu grupe k na diskriminacionoj funkciji j .

DEFINICIJA CENTROVAX POSTUPKA ROTACIJE

Radi lakše interpretacije prirode diskriminacionih funkcija primenjuju se različiti postupci rotacije matrica koeficijenata i strukture. Na taj način pojednostavljuje se i struktura razlika između ispitivanih subpopulacija. Kada je primarni cilj kanoničke diskriminacione analize identifikacija grupa, a ne identifikacija strukture razlika između grupa, što je vrlo čest slučaj kada su grupe definisane kao rezultat neke taksonomske tehnike, nameće se problem transformacije matrice centroida u interpretativno pogodnu poziciju. Takvu transformaciju matrice centroida moguće je izvesti nekom od ortogonalnih tehnika rotacije, koje zadržavaju ortogonalnost diskriminacionih funkcija.

Problem pogodne transformacije matrice centroida \mathbf{C} može se definisati na sledeći način:

$$\mathbf{CT} = \mathbf{D} \mid f(\mathbf{D}) = \text{ext}, \quad \mathbf{T}^t\mathbf{T} = \mathbf{TT}^t = \mathbf{I}$$

pri čemu matrica \mathbf{T} predstavlja transformacionu matricu, a funkcija f neku parsimonijsku funkciju kakva je, na primer, Varimaks H.Kajzera. Varimaks funkcija, kao što je poznato, maksimizuje varijansu kolona matrice čiji su elementi kvadrirani. U skalarnom obliku ova funkcija može se zapisati na sledeći način (Mulaik 1972, str. 259):

$$v = \sum_{j=1}^b \frac{g \sum_{k=1}^g d_{kj}^4 - \left(\sum_{k=1}^g d_{kj}^2 \right)^2}{g^2}$$

Pri tome, d_{kj} predstavlja element u redu k i koloni j , a g broj elemenata u koloni matrice koju treba rotirati.

Korišćenje transformacione matrice \mathbf{T} za rotaciju matrice centroida u neku parsimonijsku poziciju može se protumačiti i kao rotacija diskriminacionih funkcija u položaj u kojem centriidi grupa na tim funkcijama grade jednostavnu strukturu. Naime, budući da je

$$\mathbf{CT} = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{KT} = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{ZXT}$$

očigledno je da se ortogonalna rotacija matrice centroida istovremeno može posmatrati i kao rotacija diskriminacionih funkcija. Prema tome transformacionu matricu \mathbf{T} moguće je upotrebiti za ortogonalnu rotaciju matrice koeficijenata i matrice strukture.

Na osnovu ovoga sledi da je rotirana matrica koeficijenata za diskriminacione funkcije u stvari matrica \mathbf{X}_r

$$\mathbf{X}_r = \mathbf{X}\mathbf{T}$$

tj. da su rotirane diskriminacione funkcije u matrici \mathbf{K}_r

$$\mathbf{K}_r = \mathbf{Z}\mathbf{X}_r = \mathbf{Z}\mathbf{X}\mathbf{T} = \mathbf{K}\mathbf{T}$$

a matrica strukture rotiranih funkcija \mathbf{F}_r

$$\mathbf{F}_r = \mathbf{Z}^t\mathbf{K}_r = \mathbf{Z}^t\mathbf{K}\mathbf{T} = \mathbf{F}\mathbf{T}$$

predstavlja rotiranu originalnu matricu strukture. Naravno, diskriminacione funkcije posle rotacije ostaju ortogonalne jer je matrica kovarijansi rotiranih diskriminacionih funkcija:

$$\mathbf{K}_r^t\mathbf{K}_r = \mathbf{T}^t\mathbf{X}^t\mathbf{Z}^t\mathbf{Z}\mathbf{X}\mathbf{T} = \mathbf{T}^t\mathbf{X}^t\mathbf{R}\mathbf{X}\mathbf{T} = \mathbf{T}^t\mathbf{T} = \mathbf{I}$$

Koeficijenti kanoničkih korelacija koji odgovaraju rotiranim diskriminacionim funkcijama jednaki su dijagonalnim elementima matrice \mathbf{P}_r koja je definisana na sledeći način:

$$\mathbf{P}_r = \mathbf{K}_r^t\mathbf{L} = \mathbf{T}^t\mathbf{X}^t\mathbf{Z}^t\mathbf{G}\mathbf{Y} = \mathbf{T}^t\mathbf{X}^t\mathbf{A}\mathbf{Y} = \mathbf{T}^t\mathbf{X}^t\mathbf{A}\mathbf{X}\mathbf{P}^{-1}$$

PROGRAMSKA IMPLEMENTACIJA PREDLOŽENOG METODA

Makro program CENTVAX za Centrovax postupak rotacije implementiran je u MATRIX jeziku statističkog paketa SPSS.¹ Način njegovog aktiviranja objašnjen je na početku priloženog simboličkog koda.

¹ Makro program nije mogao biti nazvan po nazivu postupka jer SPSS (barem do verzije 11.5) ne dozvoljava da ime makro programa ima više od osam karaktera!

Numerički primer

Slučajnom uzorku 307 srednjoškolaca muškog pola iz učeničkih internata u Beogradu, uzrasta od 15 do 18 godina, zadat je Ahenbahov "Upitnik samoprocene za mlade". Na osnovu ukupnih rezultata na osam skala problema u ponašanju (TWS = skala povlačenja, TSCS = skala somatskih teškoća, TADS = skala anksioznosti/depresivnosti, TSPS = skala socijalnih problema, TTPS = skala problema u mišljenju, TAPS = skala problema pažnje, TDBS = skala delinkventnog ponašanja, TABS = skala agresivnog ponašanja) dečaci su grupisani u pet grupa primenom postupka "pokretnih sredina" (K-means algoritam). Radi identifikacije dobijenih grupa primenjena je kanonička diskriminaciona analiza u kojoj su korišćene iste varijable na osnovu kojih su učenici grupisani. U Tabeli 1 prikazane su veličine grupa koje su dobijene postupkom grupisanja, a u tabelama 2, 3, 4 i 5 rezultati kanoničke diskriminacione analize pre primene postupka rotacije.

Tabela 1.
Veličine grupa dobijenih postupkom "pokretnih sredina"

Grupa	n	%
g1	79.0	25.7
g2	39.0	12.7
g3	98.0	31.9
g4	13.0	4.2
g5	78.0	25.4

Tabela 2.
Kanoničke korelacije (kolona rho) i ishodi asimptotskih testova značajnosti (kolone f i sig) kanoničkih diskriminacionih funkcija

Funkcija	rho	f	sig
f1	.913	1530.511	.000
f2	.737	361.668	.000
f3	.468	85.414	.000
f4	.056	.973	.325

Kao što se iz Tabele 2 može videti prve tri diskriminacione funkcije su statistički značajne. U kolonama Tabele 3 prikazani su koeficijenti za statistički značajne diskriminacione funkcije, a u Tabeli 4 prikazana je struktura značajnih diskriminacionih funkcija.

Tabela 3.
 Nerotirana matrica koeficijenata statistički značajnih
 kanoničkih diskriminacionih funkcija

Početne varijable	f1	f2	f3
TWS	-.092	.247	-.192
TSCS	-.063	.135	1.070
TADS	-.286	.696	-.482
TSPS	.015	.045	.045
TTPS	-.051	.082	.220
TAPS	-.181	-.101	-.360
TDBS	-.066	.022	.239
TABS	-.580	-.808	-.209

Tabela 4.
 Nerotirana matrica strukture
 kanoničkih diskriminacionih funkcija

Početne varijable	f1	f2	f3
TWS	-.498	.557	-.059
TSCS	-.559	.292	.716
TADS	-.726	.619	-.155
TSPS	-.433	.354	-.088
TTPS	-.559	.262	.205
TAPS	-.754	.112	-.150
TDBS	-.666	-.151	.226
TABS	-.876	-.457	.008

Tabela 5.
 Nerotirana matrica centroida grupa na kanoničkim
 diskriminacionim funkcijama

Grupe	f1	f2	f3
g1	-.281	.992	-.426
g2	-1.467	-.875	-.353
g3	.135	-.668	.083
g4	-2.039	1.057	1.845
g5	1.188	.095	.196

Radi pojednostavljenja identifikacije grupa dobijenih postupkom grupisanja početne matrice centroida, matrica strukture i matrica koeficijenata za značajne diskriminacione funkcije rotirane su Centrovax postupkom, tj. korišćenjem transformacione matrice koja je prikazana u Tabeli 6. Matrice dobijene rotacijom početnih matrica prikazane su u tabelama 7, 8 i 9.

Tabela 6.
 Transformaciona matrica

	f1	f2	f3
f1	.651	.730	.209
f2	-.373	.548	-.749
f3	-.661	.409	.629

Tabela 7.
 Rotirana matrica koeficijenata kanoničkih diskriminacionih funkcija

Početne varijable	f1	f2	f3
TWS	-.026	-.010	-.325
TSCS	-.799	.465	.559
TADS	-.127	-.025	-.884
TSPS	-.037	.055	-.002
TTPS	-.209	.097	.067
TAPS	.158	-.335	-.189
TDBS	-.209	.061	.120
TABS	.062	-.951	.352

Tabela 8.
 Rotirana matrica strukture kanoničkih diskriminacionih funkcija

Početne varijable	f1	f2	f3
TWS	-.493	-.083	-.558
TSCS	-.947	.045	.115
TADS	-.601	-.255	-.713
TSPS	-.356	-.158	-.411
TTPS	-.597	-.181	-.184
TAPS	-.434	-.551	-.336
TDBS	-.526	-.477	.116
TABS	-.405	-.887	.164

Tabela 9.
 Rotirana matrica centroida grupa na kanoničkim diskriminacionim funkcijama

Grupe	f1	f2	f3
g1	-.271	.164	-1.070
g2	-.394	-1.694	.126
g3	.282	-.233	.581
g4	-2.942	-.155	-.058
g5	.608	.999	.301

Rezultati dobijeni Centrovax postupkom rotacije pokazuju da grupu 4 (g4) čine mladići iz internata koji imaju veoma izražene probleme ponašanja, a pre svega sklonost somatizaciji problema, anksioznost, depresivnost i probleme mišljenja. Grupu 5 čine mladići sa najmanje izraženim problemima ponašanja, a, pre svega, smanjenom agresivnošću, delinkventnošću i problemima pažnje, a grupu 2 oni mladići kod kojih su ovi problemi veoma izraženi. Grupu 1 odlikuje sklonost povlačenju, anksioznost i depresivnost ali i nesklonost somatizacijama problema. Na kraju, grupa 3 odlikuje se društvenošću koja je udružena sa agresivnošću i odsustvom somatskih teškoća.

U Tabeli 10 prikazani su koeficijenti kanoničke korelacije (Rho_rot) koji odgovaraju rotiranim diskriminacionim funkcijama. Očigledno, postignuta

jednostavnost u strukturi centroida grupa ima "cenu" koja se iskazuje smanjenjem koeficijenta kanoničke korelacije. Ovo ne čudi, budući da primenjeni postupak rotacije dovodi do toga da se na svakoj diskriminacionoj funkciji izdvoji jedna grupa koja je udaljena od presotalih grupa, dok se razlike među ostalim grupama umanjuju.

Tabela 10.
Koeficijenti kanoničke korelacije posle rotacije

Funkcija	Rho_rot
f1	.594
f2	.403
f3	.294

ZAKLJUČAK

Postupak rotacije matrice centroida koji je prikazan u ovom radu može biti koristan pre svega u situacijama kada je potrebno definisati karakter grupa koje su dobijene nekom prethodnom analizom grupisanja ili eksperimentalnom selekcionom procedurom. Interpretacija ili definisanje karaktera grupa olakšano je ovim postupkom rotacije budući da postupak dovodi do toga da se na svakoj rotiranoj diskriminacionoj funkciji najčešće izdvaja po jedna grupa koja je bitno različita od ostalih. Naravno, postupak nije moguće primeniti u situacijama kada postoji samo jedna značajna diskriminaciona funkcija. Potrebne su dodatne empirijske provere ponašanja ovog postupka kako bi se utvrdila stvarna korist od njegove primene.

LITERATURA

- (1) ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.
- (2) LUNNEBORG, C.E., & LUNNEBORG, P.W. (1978). Improved counseling information through rotation of discriminant functions, *Educational and Psychological Measurement*, **38**, 737–754.
- (3) MULAİK, S.A. (1972). *The foundations of factor analysis*, New York: McGraw-Hill Inc.

Zbornik IKSI, 1-2/2005 – L. Tenjović, A. Zorić
„Jedan novi metod za rotaciju u diskriminacionoj analizi“, (str. 213-232)

- (4) MOMIROVIĆ, K. (1997). O diskriminativnim funkcijama, diskriminativnim faktorima i nekim očiglednim glupostima, *Statistička revija*, **46**, 79–100.
- (5) MOMIROVIĆ, K. & ZORIĆ, A. (1996). On the variance, reliability, significance and importance of canonical discriminant functions, U S. Bogosavljević i M. Kovačević, *Analiza grupisanja*, **2** (str. 79-91), Beograd: Savezni zavod za statistiku.
- (6) RAO, C.R. (1974). *Linear statistical inference and its application*, Second edition, New Delhi: Wiley Eastern Private Limited.
- (7) RENCHER, A. (1992). Interpretation of Canonical Discriminant Functions, Canonical Variates, and Principal Components, *American Statistician*, **46**, 217–225.
- (8) TATSUOKA, M.M. (1971). *Multivariate analysis*. New York: John Wiley & Sons, Inc.
- (9) TENJOVIĆ, L. (2002). *Statistika u psihologiji – priručnik*, Beograd: Centar za primenjenu psihologiju.

PRILOG: KOD PROGRAMA ZA CENTROVAX ROTACIJU

```
preserve
set printback=no mprint=no mxloops=999
*-----
*
*          CENTVAX MACRO
*      Program za Centrovax rotaciju matrice centroida
*      u kanonickoj diskriminacionoj analizi
*      Napisali A. Zoric & L. Tenjovic
*      Verzija 1.0.
*      10.11.2005.
*
* Program CENTVAX se aktivira na sledeci nacin:
*
* INCLUDE 'CENTROVAX.SPS'.
* CENTVAX VARS = imena kvantitativnih varijabli/
*   GROUPS = ime nominalne varijable/.
*-----

define centvax (vars=!charend('/')
                /groups=!charend('/'))

COUNT majmun = !vars (SYSMIS) .
EXECUTE .
FILTER OFF.
USE ALL.
SELECT IF(NOT(majmun)).
```

```
EXECUTE .  
*-----  
* Cuvanje originalnog zapisa.  
*-----  
save outfile='cntvx_tmp.sav'  
*-----  
* Aktiviranje Matrix jezika i preliminarne operacije  
*-----  
matrix  
print /title '*** C E N T R O V A X ***' /space 1  
print /title '  VARIMAX ROTACIJA CENTROIDA' /space 0  
print /title 'U KANONICKOJ DISKRIMINACIONOJ ANALIZI' /space 0  
get x/file=* /variables=!vars/names=nx  
get g/file=* /variable=!groups/names=ng  
compute s=design(g)  
compute kont=t(s)*s  
compute num=trace(kont)  
compute numg={"g1","g2","g3","g4","g5","g6","g7","g8","g9","g10",  
  "g11","g12","g13","g14","g15","g16","g17","g18","g19","g20",  
  "g21","g22","g23","g24","g25","g26","g27","g28","g29","g30"}  
compute ngr=ncol(kont)  
compute nvr=ncol(x)  
print num/format "f8.0"/title 'Broj entiteta'  
print ngr/format "f8.0"/title 'Broj grupa'  
print nvr/format "f8.0"/title 'Broj varijabli'  
compute numg=numg(1:ng)  
compute dkont=diag(kont)  
compute pkont=(dkont&/num)&*100  
compute majmun1={dkont,pkont}  
compute majmun2={"n","%"}  
print majmun1 /format "f8.1"/title 'Velicine grupa'  
  /mame=numg/cname=majmun2  
compute en=make(num,1,1)  
compute mmm=(csum(x))&/num  
compute mmmm=en*mmm  
compute x=x-mmmm  
compute c=(t(x)*x)&/num  
compute dc=sqrt(inv(mdiag(diag(c))))  
compute x=x*dc  
compute rmat=dc*c*dc  
release mmm,mmmm,dc,c  
call eigen(rmat,xxx,lambda)
```

```
do if lambda(nvr)=0
print /title 'GRESKA: '/space 2
print /title 'Matrica interkorelacija je singularna' /space 0
else

print rmat /format "f8.3" /title 'Interkorelacije varijabli'
      /rname=nx/cname=nx

compute amat=t(x)*s*inv(kont)*t(s)*x
compute amat=amat&/num

print amat /format "f8.3" /title 'Intergrupne kovarijanse varijabli'
      /rname=nx/cname=nx

compute wmat=rmat-amat
print wmat /format "f8.3" /title 'Intragrupne kovarijanse varijabli'
      /rname=nx/cname=nx

compute mmat=t(x)*s*inv(kont)

*-----
* Kanonicka diskriminativna analiza
*-----

do if nvr<(ngr-1)
.compute nf=nvr
else
.compute nf=ngr-1
end if

compute lambda=mdiaq(lambda)
compute lambda=sqrt(inv(lambda))
compute qmat=xxx*lambda*t(xxx)

compute mahmat=inv(qmat)
compute omega=qmat*amat*qmat

compute numf={"f1","f2","f3","f4","f5","f6","f7","f8","f9","f10",
"f11","f12","f13","f14","f15","f16","f17","f18","f19","f20",
"f21","f22","f23","f24","f25","f26","f27","f28","f29","f30",
"f31","f32","f33","f34","f35","f36","f37","f38","f39","f40",
"f41","f42","f43","f44","f45","f46","f47","f48","f49","f50",
"f51","f52","f53","f54","f55","f56","f57","f58","f59","f60",
"f61","f62","f63","f64","f65","f66","f67","f68","f69","f70",
"f71","f72","f73","f74","f75","f76","f77","f78","f79","f80",
"f81","f82","f83","f84","f85","f86","f87","f88","f89","f90",
"f91","f92","f93","f94","f95","f96","f97","f98","f99","f100"}
compute numf=numf(1:nf)

call svd(omega,y1,lambda,y2)

compute y1=y1(:,1:nf)
compute y2=y2(:,1:nf)
compute dtr=t(y1)*omega*y2
compute dtr=diag(dtr)
```

```
compute rho=sqrt(dtr)

compute df1=1
compute df2=num-2
compute ftest1=make(nf,1,1)
compute ftest2=ftest1
compute sig1=ftest1
loop s=1 to nf
. compute ftest1(s,1)=dtr(s,1)*((num-2)/(1-dtr(s,1)))
. compute ftest2=ftest1(s,1)
. compute sig1(s,1)=1-fcdf(ftest2,df1,df2)
end loop
compute srez={rho,dtr,ftest1,sig1}

print srez/format "f8.3"
. /title 'Kanonicke korelacije i asimptotski testovi znacajnosti'
. /space=2
. /rnames=numf/cnames={"rho","dtr","ftest","sig"}

compute suma=make(nvr,1,0)
loop k=1 to nf
do if sig1(k,1) < 0.01
. compute suma(k,1)=1
end if
end loop

do if k<ncol(s)
compute k=ncol(s)
else
compute k=csum(suma)
end if
print k /title 'Broj znacajnih kanonickih korelacija' /space=2

do if k>1

compute numf=numf(1:k)
compute y=y1(:,1:k)
compute rhosq=t(y)*omega*y
compute rhosq=mdiag(diag(rhosq))
compute rho=sqrt(rhosq)
compute invrho=inv(rho)

compute fmat=mahmat*y
compute rinu=inv(rmat)
compute beta=rinu*fmat
compute cent=t(mmat)*beta

print beta/format "f8.3"
. /title 'Nerotirana matrica koeficijenata kanonickih diskriminacionih funkcija'
. /rnames=nx /cnames=numf
print fmat /format "f8.3"
. /title 'Nerotirana matrica strukture kanonickih diskriminacionih funkcija'
. /rname=nx /cname=numf
print cent/format "f8.3"
. /title 'Nerotirana matrica centroida grupa na kanonickim diskriminacionim funkcijama'
```

```
/rnames=numg /cnames=numf
*-----
* Varimax rotacija
*-----
compute nfak=k
compute nkat2=ngr
compute f=cent

compute tv=0
compute nc=0
compute stabil=1

compute trans=ident(nfak,nfak)

loop if (stabil = 1 and nc le 50)
+ compute sv=0
+ loop j=1 to nfak
+ compute sa=0
+ compute sb=0
+ loop i=1 to nkat2
+ compute v2=f(i,j)**2
+ compute sa=sa + v2
+ compute sb=sb + v2 * v2
+ end loop
+ compute sv=sv + ((nkat2 * sb - sa * sa) / (nkat2 * nkat2))
+ end loop

+ compute nc=nc + 1
+ do if (abs(sv-tv) le 1e-7)
+ compute stabil=stabil+1
+ else
+ compute stabil=1
+ end if
+ compute tv=sv

+ loop j=1 to nfak-1
+ loop k=j+1 to nfak
+ compute as=0
+ compute bs=0
+ compute cs=0
+ compute ds=0
+ loop i=1 to nkat2
+ compute xs=(f(i,j) **2 - f(i,k) **2)
+ compute ys=f(i,j) * f(i,k) * 2
+ compute as=as+xs
+ compute bs=bs+ys
+ compute cs=cs + (xs * xs - ys * ys)
+ compute ds=ds + xs * ys
+ end loop
+ compute ds=ds * 2

+ compute xs=ds - ((2 * as * bs) / nkat2)
+ compute ys=cs - ((as * as - bs * bs) / nkat2)

+ do if xs>0
+ do if ys>0
```

```
+ compute p=(artan(xs/ys))
+ else if ys<0
+ compute p=(artan(xs/ys)+ 314159265359e-11)
+ end if
+ else if xs<0
+ do if ys>0
+ compute p=(artan(xs/ys))
+ else if ys<0
+ compute p=(artan(xs/ys)- 314159265359e-11)
+ end if
+ else if xs=0
+ do if ys>0
+ compute p=0
+ else if ys<0
+ compute p= 314159265359e-11
+ end if
+ else
+ compute p= 314159265359e-11 / 2
+ end if
+ compute p = p / 4
+ compute sinp=sin(p)
+ compute cosp=cos(p)
+ do if abs(sinp) gt 1e-30
+ loop i=1 to nkat2
+ compute xs=f(i,j) * cosp + f(i,k) * sinp
+ compute ys=f(i,k) * cosp - f(i,j) * sinp
+ compute f(i,j)=xs
+ compute f(i,k)=ys
+ end loop
+ loop i=1 to nfak
+ compute xs=trans(i,j) * cosp + trans(i,k) * sinp
+ compute ys=trans(i,k) * cosp - trans(i,j) * sinp
+ compute trans(i,j)=xs
+ compute trans(i,k)=ys
+ end loop
+ end if
+ end loop
+ end loop
end loop
*-----
* Rotirane strukture
*-----
compute centr=cent*trans
compute betar=beta*trans
compute fmatr=fmat*trans
compute rhor= diag(t(trans)*rho)

print trans /format "f8.3"
      /title 'Transformaciona matrica'/space=2
      /names=numf/cnames=numf
print betar /format "f8.3"
```

Zbornik IKSI, 1-2/2005 – L. Tenjović, A. Zorić
„Jedan novi metod za rotaciju u diskriminacionoj analizi“, (str. 213-232)

```
/title 'Rotirana matrica koeficijenata kanonickih diskriminacionih funkcija'  
/rnames=nx /cnames=numf  
print fmatr /format "f8.3"  
/title 'Rotirana matrica strukture kanonickih diskriminacionih funkcija'  
/rname=nx/cname=numf  
print centr /format "f8.3"  
/title 'Rotirana matrica centroida grupa na kanonickim diskriminacionim funkcijama'  
/rnames=numg /cnames=numf  
print rhor /format "f8.3"  
/title 'Kanonicke korelacije posle rotacije'  
/rnames=numf  
end if  
end if  
*-----  
* Kraj programa i završne operacije  
*-----  
end matrix  
get file='cntvx_tmp.sav'  
restore  
!enddefine
```

ONE NEW METHOD FOR THE ROTATION IN DISCRIMINANT ANALYSIS

Different methods of discriminant weights and structure matrices rotation has been applied in canonical discriminant analysis to facilitate the interpretation of the discriminant functions. A new ("Centrovax") method of discriminant structure and discriminant weights matrices rotation has been proposed in this paper. The proposed method is based on the rotation of group centroid matrix based on the significant discriminant functions. A program implementation of proposed algorithm has been written in SPSS MATRIX language. An illustrative numerical example of use of the method has been presented.

KEY WORDS: *canonical discriminant analysis / centroid rotation*