KSENIJA BOSNAR, KONSTANTIN MOMIROVIĆ I FRANJO PROT
Institut za kineziologiju Fakulteta za fizičku kulturu Sveučilišta u Zagrebu
Odjel za informatiku i statistiku

Izvorni znanstveni članak
UDK 519.254:519.237.35
Primljeno 13. 10. 1983.

ALGORITAM ZA DISKRIMINATIVNU ANALIZU U MAHALANOBISOVOM PROSTORU

/ diskriminativna analiza / mahalanobisov prostor /

Predložen je algoritam i napisan program za kanoničku diskriminativnu analizu skupa entiteta opisanih nad skupom multivarijatno normalno distribuiranih varijabli s nesingularnom matricom kovarijanci, transformiranih u Mahalanobisov oblik.

Algoritam se razlikuje od standardnih algoritama za kanoničku diskriminativnu analizu po tome što diskriminativne koeficijente i diskriminativne faktore određuje i u metrici standardiziranih i u metrici varijabli transformiranih u Mahalanobisov oblik; što rotira značajne diskriminativne faktore u položaj koji maksimizira normalizirani quartimax kriterij, pri čemu broj zadržanih faktora određuje prema informacijskom doprinosu faktora objašnjenju ukupne varijance; što određuje centroide grupa i za nerotirane i za rotirane diskriminativne varijable na koordinatnom sistemu s ishodištem u centroidu ukupnog uzorka; što statistički provjerava značajnost diskriminativnih funkcija zadržanih informacijskim kriterijem putem χ^2 testa (Rao, 1951); što provjerava rezultate χ^2 testa jednofaktorskom, univarijatnom analizom varijance za svaku zadržanu rotiranu i nerotiranu diskriminativnu varijablu.

1. UVOD

Analiza homogenosti nekog skupa entiteta opisanog nad skupom varijabli dozvoljava više strategija empirijskih istraživanja koje uglavnom ovise o preciznosti hipoteze o postojanju podskupova unutar osnovnog skupa. Kada je moguća apriorna podjela entiteta u mutualno ekskluzivne podskupove, na osnovi nekog obilježja koje ne pripada analiziranim varijablama, i kad postoji hipoteza o kvantitativnim razlikama među ako stvorenim grupama, značajnost razlika i strukturu činilaca koji utiču na te razlike najpodesnije je ispitati kanoničkom diskriminativnom analizom.

Kanonička diskriminativna analiza je, zapravo, poseban slučaj kanoničke korelacijske analize ako se nominalna varijabla koja određuje pripadanje entiteta pojedinim grupama definira kao binarna selektorska matrica. Zbog toga je ova metoda invarijantna na metriku varijabli i može se definirati na više različitih načina. Neki od načina su, u pojedinim slučajevima, pogodniji za interpretaciju i/ili izračunavanje rezultata.

Predloženi algoritam i njemu pridruženi program DIANA razlikuje se od standardnih algoritama za kanoničku diskriminativnu analizu u više svojstava:

(1) diskriminativna analiza izvodi se u Mahalanobisovom prostoru, a diskriminativni koeficijenti i diskriminativni faktori određuju se i u metrici standardiziranih i u metrici varijabli transformiranih u Mahalanobisov oblik;

(2) broj diskriminativnih funkcija zadržanih za interpretaciju određuje se na temelju informacijskih svojstava diskriminativnih varijabli. Diskriminativna varjabla se smatra informacijski značajnom ako je odgovorna za neki zadovoljavajući postotak od ukupne intergrupne varijance. Ukoliko se ne odredi drugačije, program će prihvatiti kao značajnu svaku diskriminativnu varijablu koja objašnjava 5% ili više od ukupne intergrupne varijance;

(3) značajnost zadržanih diskriminativnih funkcija provjerava se i statistički, χ^2 testom (Rao, 1951);

(4) diskriminativni faktori rotiraju se u poziciju koja maksimizira normalizirani quartimax kriterij (Ferguson, 1954) i određuju se matrice koeficijenata i strukture za rotirane faktore i u metrici standardiziranih i u metrici varijabli transformiranih u Mahalanobisov oblik;

(5) centroidi grupa i za rotirane i za nerotirane diskriminativne varijable određuju se na koordinatnom sustavu s ishodištem u centroidu ukupnog uzorka;

(6) značajnost diskriminativnih varijabli još se jednom provjerava jednofaktorskom univarijatnom analizom varijance za svaku zadržanu rotiranu i nerotiranu diskriminativnu varijablu.

2. ALGORITAM

2. 1 Uvodne operacije

Neka je $B=(b_{ij})$; $i=1,\dots,n$; $j=1,\dots,m$; matrica izvornih podataka entiteta skupa $E=\{e_i,\ i=1,\dots,n\}$ opisanih nad skupom multivarijatno normalno distribuiranih varijabli $V=\{v_j;\ j=1,\dots,m\}$. Neka je skup entiteta podijeljen u podskupove G_k, $k=1,\dots,g$, i neka matrica $S=(s_{ik})$ apriorno određuje pripadnost entiteta ovim podskupovima tako da je $s_{ik}=1$, ako je entitet e_i element podskupa G_k, a $s_{ik}=0$, ako e_i nije element podskupa G_k.

Odredimo matricu kovarijanci varijabli iz V na cijelom skupu entiteta E

$$C = (B^TB - B^T11^T1/nB)1/n,$$

gdje je 1 vektor jedinice S n elemenata i dijagonalnu matricu varijanci

$$V^2 = \text{diag } C.$$

Standardizacija varijabli na nultu aritmetičku sredinu i jediničnu varijancu rezultat je operacije

$$Z = (B - 11^T1/nB)V^{-1}.$$

2. 2 Interkorelacije varijabli i transformacija rezultata u Mahalanobisov oblik

Interkorelacije varijabli na cijelom skupu entiteta definirane su operacijom

$$R = V^{-1}CV^{-1} = Z^TZ1/n.$$

Eliminacija kovarijanbiliteta među varijablama izvedena je Mahalanobisovom transformacijom

$$\Psi = ZR^{-1/2}.$$

Korelcije originalnih varijabli i varijabli transformiranih u Mahalanobisov oblik su

$$Z^T\Psi 1/n = R^{1/2}.$$

Očito, varijable iz Ψ su, pod kriterijem najmanjih kvadrata, ortogonalne varijable najsličnije varijablama iz Z, jer je

$$tr(Z-\Psi)^T\,(Z-\Psi) = \text{minimum}.$$

Aritmetičke sredine rezultata u grupama entiteta su elementi vektora redaka matrice

$$M_Z = (\text{diag}(S^TS))^{-1}S^TZ$$

za standardizirane rezultate, i elementi vektora redaka matrice

$$M_M = (\text{diag}(S^TS))^{-1}S^T\Psi$$

za rezultate transformirane u Mahalanobisov oblik.

2. 3 Diskriminativna analiza u Mahalanobisovom prostoru i redukcija broja diskriminativnih funkcija

Problem diskriminativne analize u pravilu se definira na osnovi dekompozicije matrice kovarijabiliteta u ukupnom uzorku na matricu kovarijabiliteta unutar grupa i matricu kovarijabiliteta između grupa. Neka je T matrica kovarijabiliteta u cijelom uzorku, W matrica kovarijabiliteta unutar grupa, a A matrica kovarijabiliteta između grupa. Tada je

$$T = W + A.$$

Na osnovi ove definicije problem utvrđivanja diskriminativnih funkcija moguće je formulirati na više ekvivalentnih načina, kao što je npr. pakazao Romeder (1973). Jedan od njih je maksimizacija omjera

$$\frac{XAX^T}{XTX^T} = p^2$$

uz neki pogodan uvjet, najčešće $X^TTX = 1$, gdje je X vektor diskriminativnih koeficijenata koji tražimo.

Ako su varijable transformirane u Mahalanobisov oblik,

$$T = \Psi^T\Psi 1/n = I,$$

maksimizira se

$$X^TAX \,\Big|\, X^TX = 1,$$

što problem svodi na rješavanje karakteristične jednadžbe

$$(A-p^2)X = 0;\ X^TX = 1,$$

gdje je Lagrangeov množitelj p^2 traženi maksimum i kvadrat koeficijenta kanoničke diskriminacije.

Aritmetičke sredine varijabli transformiranih u Mahalanobisov oblik za ukupni uzorak jednake su nuli, pa je matrica A u ovom slučaju jednostavno

$$A = ((S^TS)M_M)^TM_M1/n.$$

U diskriminativnoj analizi moguće je odrediti do $k-1$ ortogonalnih diskriminativnih funkcija ako je $k<m$. Neka ih u daljnjoj analizi bude zadržano samo q, tj.

$$(A-p^2_p)X_p = 0; \qquad p = 1,\ldots,q;$$

$$q = \text{num}(p^2_p/(\sum_{p=1}^{v} p^2 \div v) > 0.05) \qquad v = \min\,((k-1),\ m)$$

samo one funkcije čiji je doprinos ukupnoj diskriminaciji jednak ili veći od 5 %.

Definirajmo matricu

$$X = (X_p),\ \text{za koju vrijedi } X^TX = I,$$

formiranu od zadržanih diskriminativnih vektora, i matrio

$$p = (p_p)$$

u čijoj su dijagonali koeficijenti kanoničke diskriminacije za zadržane diskriminativne funkcije.

Značjanost zadržanih diskriminativnih funkcija može se tesirati i statistički, χ^2 testom, Raovom aproksimacijom Wilksovog kriterija (Rao, 1951)

$$\chi^2_p = (\eta^* p^2_p)\ (1-p^2_p)^{-1};\ \eta = n-k,$$

sa stupnjevima slobode $df_p = m+k-2p$, što odgovara vrijednosti na F—distribuciji

$$f_p = \chi^2_p \div df_p,$$

sa stupnjevima slobode $df_1 = m+k-2p$, $df_2 = +\infty$.

Vrijednosti entiteta u ovom diskriminativnom prostoru su

$$D = \Psi X = ZR^{-1/2}X,$$

a centroidi grupa su određeni sa

$$G_D = (S^TS)^{-1}S^TD = (S^TS)^{-1}S^T\Psi X = (S^TS)^{-1}S^TZR^{-1/2}X = M_MX.$$

2.4 Diskriminativne funkcije u matrici standardiziranih varijabli

Diskriminativne funkcije su standardizirane

$$D^TD1/n = X^T\Psi^T\Psi X1/n = I,$$

pa su korelacije standardiziranih varijabli i diskriminativnih funkcija, tj. diskriminativni faktori standardiziranih varijabli jednaki

$$F = Z^TD1/n = Z^T\Psi X1/n = R^{1/2}X.$$

Diskriminativni ponderi za standardizirane varijable su, naravno,

$$Y = R^{1/2}X,$$

jer

$$D = ZR^{1/2}X = ZY.$$

2. 5 Rotacija diskriminativnih faktora

Diskriminativne funkcije nije lako interpretirati. Ovaj se problem može ponekad pojednostaviti ortogonalnom rotacijom diskriminativnih faktora u neku parsimoničnu poziciju.

U ovoj se analizi izvodi ortogonalna transformacija matrice X tako da rezultat zadovoljava Fergusonov quartimax kriterij (Ferguson, 1954), tj.

$$P = XQ = (p_{jp})$$

uz uvjet

$$Q^TQ = QQ^T = I,$$

tako da je

$$\sum_{j=1}^{m}\sum_{p=1}^{q} p^4_{jp} = \max.$$

Vrijednosti entiteta na rotiranim diskriminativnim faktorima, procijenjene regresijskim postupkom, su

$$\Phi = \Psi P(P^TP)^{-1},$$

no kako je

$$P^TP = Q^TX^TXQ = I,$$

to je

$$\Phi = \Psi P = \Psi XQ = ZR^{-1/2}XQ = ZYQ = DQ$$

Varijable iz Φ su i dalje ortogonalne, jer

$$\Phi^T\Phi1/n = P^T\Psi^T\Psi P1/n = P^TP = I.$$

Centroidi grupa na rotiranim diskriminativnim faktorima su

$$G\Phi = (S^TS)^{-1}S^T\Phi = (S^TS)^{-1}S^TDQ =$$
$$= (S^TS)^{-1}S^TZYQ = (S^TS)^{-1}S^TZR^{-1/2}XQ =$$
$$= M_MXQ = M_MP.$$

Transformacijska matrica Q je istovremeno i matrica korelacija rotiranih i nerotiranih faktora

$$D^T\Phi1/n = X^T\Psi^T\Psi P1/n = X^TP = X^TXQ = Q.$$

Testovi značajnosti učinjeni za nerotirane diskriminativnefunkcije, naravno, ne vrijede za rotirane faktore pa se značajnost pojedinog rotiranog diskriminativnog faktora može provjeriti jednofaktorskom, univarijatnom analizom varijance.

2. 6 Rotirani faktori u metrici standardiziranih varijabli

Korelacije standardiziranih varijabli i rotiranih diskriminativnih varijabli, tj. rotirani diskriminativni faktori standardiziranih varijabli su

$$H = Z^T\Phi1/n = Z^TZR^{-1/2}XQ1/n = R^{1/2}XQ = FQ.$$

Rotirani diskriminativni ponderi za standardizirane varijable su

$$U = YQ,$$

jer

$$\Phi = \Psi XQ = ZR^{-1/2}XQ = ZYQ = ZU.$$

2. 7 Diskriminativne funkcije originalnih varijabli

Na kraju, algoritmom se određuju i diskriminativni ponderi za originalne, nestandardizirane varijable.

Ako je

$$Z = (B - 11^T1/nB)V^{-1},$$

iz toga slijedi da je

$$ZV = (B - 11^T1/nB)$$

pa, jer je

$$D = ZR^{-1/2}X = ZVV^{-1}R^{-1/2}X,$$

ponderi za nestandardizirane varijable su

$$K = V^{-1}R^{-1/2}X.$$

3. PROGRAM

Program DIANA napisan je u meta jeziku SS (Zakrajšek, Štalec i Momirović, 1974), verzija 5.2/M, i implementiran je na računalu UNIVAC Sveučilišnog računskog centra u Zagrebu gdje je pohranjen u programskoj biblioteci SRCE * SS—MAKRO.

U ovoj verziji program omogućuje analizu do 10.000 entiteta svrstanih u do 20 grupa i opisanih s najviše 250 varijabli. Broj grupa je moguće povećati na 250, no u tom slučaju neće biti izvedene univarijatne analize varijance na originalnim varijablama i nerotiranim i rotiranim diskriminativnim varijablama.

Korisnik programa dužan je pripremiti matricu originalnih podataka tako da u njoj nema entiteta s nepotpunim podacima, a također je potrebno da prethodno kreira selektorsku matricu, najbolje uz pomoć jednog od programa za binarizaciju kvalitativnih varijabli, npr. BINAR ili VESNA, također dostupnih na Sveučilišnom računskom centru. Detaljnije upute o korištenju programa nalaze se unutar samog programa.

4. NUMERIČKI PRIMJER

Efikasnost algoritma i programa DIANA provjerena je na rezultatima 653 klinički zdrava muškarca, stara 19-27 godina. Ispitanici su podijeljeni u četiri grupe prema slijedećim obilježjima mjesta u kojem žive:

(1) Republičko središte (MJES 4)
(2) Mjesto u kom je sjedište okružnog suda (MJES 3)
(3) Mjesto u kom je sjedište općine (MJES 2)
(4) Mjesto koje nema ni jednu od navedenih karakteristika (MJES 1).

Ispitanicima je izmjereno deset morfoloških mjera: (1) visina tijela (VISINA), (2) dužina ruke (DUZIRU), (3) dužina noge (DUZINO), (4) težina (TEZINA), (5) opseg nadlaktice (OPNAD), (6) opseg podlaktice (OPPODL), (7) opseg potkoljenice (OPPOTK), (8) nabor na leđima (NANALE), (9) nabor na trbuhu (NATRBU) i (10) nabor na nadlaktici (NANADL) na način kako propisuje Internacionalni Biološki Program.

Primijenjenim informacijskim kriterijem značajnosti sve tri diskriminativne funkcije smatrale bi se značajnim (tabela 1) .Raov χ^2 test je međutim, znatno stroži i po njemu je značajna samo prva diskriminativna funkcija. Sudeći prema rezultatima χ^2 testa, a isto tako i prema slabo razmaknutim centroidima grupa na trećoj diskriminativnoj funkciji (tabela 3), treća funkcija je vjerojatno suvišna i trebalo bi, bar u ovoj analizi, pooštriti kriterij od 5% doprinosa ukupnoj diskriminaciji.

Morfološke varijable bile su odabrane tako da »pokriju« tri poznata morfološka faktora: longitudinalnu dimenzionalnost skeleta, volumen i masu tijela i faktor masnog tkiva. Kao što je vidljivo u tabeli 2 , samo diskriminativni faktori standardiziranih varijabli donekle slijede taj model, što je potpuno razumljivo, jer transformacija u Mahalanobisov prostor uništava postojeću zajedničku varijancu varijabli, pa treba napustiti interpretaciju u terminima zajedničkih činilaca i prihvatiti interpretaciju u terminima nezavisnih varijabli.

Prvi diskriminativni faktor u Mahalanobisovom prostoru defniran je najviše visinom, zatim naborom na leđima (koji značajno ovisi o broju masnih stanica) te opsegom nadlaktice i opsegom podlaktice s negativnim predznakom, tj. slabijom i više podložnom specifičnim utjecajima mjerom mišićne mase i boljom mjerom ukupne mišićne mase sa suprotnim predznakom. Iz vrijednosti centroida grupa na prvoj funkciji (tabela 3) vidljiv je razmak između stanovnika malih mjesta i republičkog središta, a po položaju na funkciji može se zaključiti kako je vjerojatnije da će veću visinu, više masnog tkiva dijelom endogenog porijekla i više mišićne mase koja je pretežno uvjetovana egzogenim činiocima imali stanovnici republičkog središta nego stanovnici malog mjesta ili sela.

Zanimljiv je podatak negativna vrijednost za MJES 3 na drugoj diskriminativnoj funkciji. Stanovnici mjesta koje je sjedište okružnog suda, što su obično srednje veliki gradovi, razlikuju se od ostalih ispitanika po tome što su manje visoki, kraćih udova, manje težine i manjeg opsega podlaktice, tj. manjih vrijednosti na varijablama koje definiraju drugu funkciju. Ovo može biti posljedica negativne selekcije uslijed migracionih kretanja i/ili lošijih uvjeta za rast i razvoj u srednje velikim gradovima i od malih mjesta i sela i od velikih gradova.

Rotirani diskriminativni faktori pružaju nešto drugačiju sliku (vidi tabelu 6). Donekle je reproducirana druga nerotirana diskriminativna funkcija u Mahalanobisovom prostoru (tabela 4) na kojoj centroid grupe MJES 3 ima ponovno negativnu vrijednost (tabela 5).

Prva rotirana diskriminativna funkcija u Mahalanobisovom prostoru definirana je pretežno Mahalanobisovom reprezentacijom visine i, znatno manje, težine, i ponovno dobro diferencira stanovnike malih mjesta i sela na negativnom polu od stanovnika republičkih središta na pozitivnom polu.

Treći rotirani diskriminativni faktor u Mahalanobisovom prostoru malo je neobičan. Definiraju ga uglavnom nabor na leđima i opseg nadlaktice i zanimljiv je po tome što ne razlikuje stanovnike MJES 3 i MJES 4, tj. stanovnike uglavnom srednjih i velikih gradova.

U prostoru standardiziranih varijabli prvi rotirani diskriminativni faktor nije interpretabilan, dok se preostala dva lako prepoznaju kao faktor općeg rasta bez učešća masnog tkiva (Q_2) i kao faktor mekih tkiva (Q_3).

Kako se vidi, algoritam i program DIANA su upotrebljivo oruđe za diskriminativnu analizu, koje može dati zanimljive informacije i onada kada se analizira dobro poznat problem kao što je to problem morfoloških karakteristika ispitanika različitog rezidencijalnog statusa.

Tabela 1

SVOJSTVENE VRIJEDNOSTI DISKRIMINATIVNE JEDNADŽBE (ρ^2), KOEFICIJENTI KANONIČKE DISKRIMINACIJE (ρ), POSTOTAK DOPRINOSA POJEDINE DISKRIMINATIVNE FUNKCIJE UKUPNOJ DISKRIMINACIJI (ρ%), REZULTATA RAOVOG χ^2 TESTA (χ^2), STUPNJEVI SLOBODE ZA χ^2 TEST (df) I ZNAČAJNOST χ^2 VRIJEDNOSTI (p). SA RB JE OZNAČEN INDEKS DISKRIMINATIVNE FUNKCIJE

RB	ρ^2	ρ	ρ%	X^2	df	p
1	.039	.197	69.8	26.08	12	.0106
2	.012	.110	21.9	7.98	10	.6316
3	.005	.68	8.3	2.99	8	.9345

Tabela 3

CENTROIDI GRUPA NA NEROTIRANIM DISKRIMINATIVNIM VARIJABLAMA

	X_1	X_2	X_3
MJES 1	—.220	.018	.059
MJES 2	—.013	.022	—.086
MJES 3	.161	—.276	.024
MJES 4	.340	.120	.059

Tabela 2

NEROTIRANI DISKRIMINATIVNI FAKTORI U MAHALANOBISOVOM PROSTORU (x), DISKRIMINATIVNI KOEFICIJENTI (y) I DISKRIMINATIVNI FAKTORI (F) STANDARDIZIRANIH VARIJABLI

	x_1	x_2	x_3	y_1	y_2	y_3	F_1	F_2	F_3
VISINA	(.67)	(49)	—.25	—.29	(—.93)	—.50	—.07	(.56)	.18
DUZIRU	—.05	(.30)	(.32)	.49	—.30	.61	.10	(.57)	.35
DUZINO	—.13	(.50)	.11	—.10	(1.84)	—.48	—.03	(.82)	.15
TEZINA	—.01	(.42)	—.20	—.24	.21	(1.23)	.11	.32	(.52)
OPNADL	(.34)	—.23	.26	(—1.08)	—.20	.45	—.01	—.03	(.62)
OPPODL	(—.32)	(.34)	.16	(.81)	.17	.26	.21	.09	(.63)
OPPOTK	21	.00	—.19	—.06	—.15	(—1.04)	.05	.05	.10
NANALE	(.48)	—.08	(.62)	(1.08)	—.42	—.14	(.56)	—.13	.24
NATRBU	.12	—.26	(—.47)	.50	.37	—.33	(.55)	.13	.04
NANADL	.18	.05	—.22	—.50	.00	.08	.21	—.02	.19

Tabela 4

ROTIRANI DISKRIMINATIVNI FAKTORI U MAHALANOBISOVOM PROSTORU (P), ROTIRANI DISKRIMINATIVNI KOEFICIJENTI (K) I DISKRIMINATIVNI FAKTORI (Q) STANDARDIZIRANIH VARIJABLI

	P_1	P_2	P_3	K_1	K_2	K_3	Q_1	Q_2	Q_3
VISINA	(.85)	.11	.13	—.46	(—.95)	—.27	.16	(.57)	—.09
DUZIRU	—.03	(.43)	.10	—.10	—.02	(.84)	.21	(.63)	.15
DUZINO	.12	(49)	—.15	(1.11)	(1.24)	(—.92)	(.34)	(.75)	—.15
TEZINA	(.31)	.23	—.26	—.65	(.90)	.62	.00	(.52)	(.34)
OPNADL	.00	—.12	(.47)	(—1.08)	.32	—.35	(—.32)	(.32)	(.42)
OPPODL	—.13	(.43)	—.20	.54	.10	.67	—.10	(.37)	(.54)
OPPOTK	.24	—.15	.01	.37	—.68	—.71	.01	.08	.09
NANALE	.00	.18	(.77)	.61	—.66	(.75)	.22	—.10	(.57)
NATRBU	.17	(—.49)	—.17	.70	.00	.00	(.44)	.00	(.36)
NANADL	.26	—.12	—.04	—.39	.16	—.28	.05	.05	(.28)

Tabela 5

CENTROIDI GRUPA NA ROTIRANIM DISKRIMINATIVNIM VARIJABLAMA

	P_1	P_2	P_3
MJES 1	—.175	.096	—.111
MJES 2	.043	—.027	—.074
MJES 3	—.042	—.245	.203
MJES 4	.276	.053	.234

Tabela 6

KORELACIJE NEROTIRANIH (X) I ROTIRANIH DISKRIMINATIVNIH FAKTORA (P)

	P_1	P_2	P_3
X_1	.71	—.22	.67
X_2	.52	.80	—.28
X_3	—.47	.55	.69

LITERATURA

1. Ferguson, G. A.: The concept of parsimony in factor analisis. Psyhometrika, 19, (1954), 281-290.
2. Rao, C. R.: An asymptotic expansion of the distrition of Wilks' criterion. Bulletin of international statistics institute, 33, 2, (1951), 177-180.
3. Romeder, J.: Methodes et programmes d'analyse discriminate. DUNOD, Paris, 1973.
4. Zakrajšek, E., J. Štalec i K. Momirović: SS-programski sistem za multivarijatnu analizu podataka. Zbornik Simpozija »Kompjuter na Sveučilištu«, Sveučilišni računski centar, Zagreb, 1974, C8-1 —C8-16.

Bosnar, K., Momirović, K., Prot, F.:

THE ALGORITHM FOR DISCRIMINATIVE ANALYSIS IN MAHALANOBIS' SPACE

/ discriminative analysis / mahalanobis, space /

The algorithm was proposed and the program was written for the canonic discriminative analysis of a group of variables with non-singular matrix of co-variances, transformed into the Mahalanobis' form.

The algorithm differs from standard algorithms for discriminative analysis in that it determines the discriminative coefficients and discriminative factors in both the metrics of standardized and in the metrics of variables transformed into the Mahalanobis' form: in that it rotates the significant discriminative factors into a position which maximizes the normalized quartimax criterion in which the number of retained factors is determined according to the informational cotribution of factors to the explanation of the entire inter-group variance; in that it determines the group centroids for both the unrotated and the rotated discriminative variables in the co-ordinate system with the starting point in the centroid of the entire sample; in that it statistically cheks the significance of discriminative functions retained by the informational criterion by means of the X^2 test (Rao, 1951); in that it cheks the results of the X^2 test by means of the single-factor, univariant analysis of variance for each retained rotated and unrotaded discriminative variable.

Ксения Боснар, Константин Момирович и Франье Прот

АЛГОРИТМ ДЛЯ ДИСКРИМИНАТИВНОГО АНАЛИЗА В ПРОСТРАНСТВЕ МАХАЛАНОБИСА

Предложен алгоритм и написана программа для канонического анализа дискриминативного множества единиц, описанных над группой мультивариантно и нормально распределенных переменных с несингулярной матрицей ковариансы, которые трансформированы в форму Махаланобиса.

Алгоритм отличается от нормальных алгоритмов для канонического дискриминативного анализа тем, что дискриминативные коэффициенты и дискриминативные факторы определяются в метрике стандартизованных переменных и переменных трансформированных в форму Махаланобиса. Достоверные дискриминативные факторы ротируются в позицию, в которой максимизируется кватримакс критерий, при чем задержанное числе факторов, в зависимости от их информационной нагрузки, определяет общую интергрупповую варианцу. Определяются центроиды групп для ротированных и неротированных дискриминативных переменных в координатной системе, исходная точка которой находится в центроиде всей выборки. Статистически проверяются достоверности дискриминативных функций задержанных на основе информационной нагрузки при помощи x^2 теста (Рао, 1951). Результаты x^2 теста проверяются при помощи однофакторного, унивариантного анализа варианцы для каждой задержанной ротированной и неротировнной дискриминативной переменной.