XIX. KANONIČKA DISKRIMINACIONA ANALIZA I KVAZIKANONIČKA DISKRIMINACIONA ANALIZA

(Glava iz novog izdanja knjige Statistika u psihologiji...)

XIX. 1. Kanonička diskriminaciona analiza

Struktura podataka

$E = \{e_i; i = 1,\ldots, n\} \subset P = \bigcup_k^g P_k$ je slučajni uzorak izvučen iz populacije koja čini uniju g subpopulacija koje prirodno postoje ili su formirane eksperimentalno;

$Q = \{q_k; k = 1,2,\ldots, g\}$ je kategorička varijabla koja se sastoji od g kategorija;

$V = \{v_j; j = 1,2,\ldots, m\} \subset U$ skup kvantitativnih varijabli;

Merenjem jedinica uzorka E u pogledu skupa kvantitativnih varijabli V, dobijamo matricu **X**:

$E \otimes V = \mathbf{X} = (x_{ij}); i = 1,\ldots, n; j = 1,\ldots, m.$

Zasnivanje postupaka diskriminacione analize počelo je tridesetih godina dvadesetog veka u radovima P. C. Mahalanobisa („On tests and measures of group divergence“, 1930) i R. Fišera („The use of multiple measurements in taxonomic problems“, 1936) (prema Anderson, 1966). U našoj zemlji ovi postupci koriste se pedesetih godina dvadesetog veka u radovima B. Ivanovića, u kojima su rešavani problemi kontrole znanja učenika predviđenih školskim programom, analize ekonomske razvijenosti srezova, stepena izgrađenosti poljoprivrednih dobara Vojvodine, kao i u pojedinim medicinskim istraživanjima. Zanimljivo je da su metode diskriminacione analize korišćene i pri određivanju granice između Socijalističke Federativne Republike Jugoslavije i Italije (prema Ivanović, 1963, str. 2). U psihološkim istraživanjima kod nas ove metode počinju da se koriste zahvaljujući, pre svega, radovima K. Momirovića i programskim implementacijama koje su izvedene u Sveučilišnom računskom centru u Zagrebu sedamdesetih godina dvadesetog veka (cf. Momirović, Gredelj & Szirovicza, 1977). Budući da je postala sastavni deo statističkih paketa koji se standardno koriste u analizama podataka psiholoških istraživanja (Systat, SPSS, Statistica), diskriminaciona analiza je uobičajeni postupak gde god je, s obzirom na prirodu problema i varijabli, njena upotreba neophodna.

U zavisnosti od primarne namene, diskriminaciona analiza ima dva (međusobno blisko povezana) aspekta, pri čemu za svaki od aspekata postoji određeni skup različitih specifičnih postupaka (cf. Geiser, 1977; Huberty, Wisenbaker, Smith, 1987; McKay & Campbell, 1982a, 1982b).

Postupci *deskriptivne* diskriminacione analize prvenstveno su namenjeni ispitivanju postojanja razlika između grupa, tj. dveju ili više eksperimentalnih ili realnih populacija na skupu kvantitativnih obeležja, i analizi prirode strukture ili složajeva tih obeležja koji stoje u osnovi postojećih razlika. Ovi postupci se u osnovi svode na transformisanje kvantitativnih multivarijacionih podataka kako bi se što ekonomičnije i jasnije sagledale razlike između populacija definisanih kategorijama nekog kvalitativnog obeležja.

Postupci *prediktivne* ili *klasifikacione* diskriminacione analize služe za klasifikaciju eksperimentalnih ili realno postojećih entiteta (u psihologiji najčešće ljudi) na osnovu većeg broja obeležja u dve ili više dobro definisanih grupa ili kategorija. Ovi postupci diskriminacione analize namenjeni su, dakle, rešavanju problema za koje C. R. Rao koristi termin *identifikacija*, podrazumevajući pod tim „...odlučivanje o pripadnosti posmatranog entiteta jednoj iz datog skupa populacija kojima entitet može potencijalno pripadati." (Rao, 1974, str. 574). U matematičkom razvijanju postupaka klasifikacione diskriminacione analize naglasak je, dakle, na otkrivanju ili izvođenju matematičkog pravila za optimalno svrstavanje entiteta, čija grupna pripadnost nije poznata, u neku od više precizno definisanih grupa. U okviru deskriptivne diskriminacione analize pažnja se pre svega usmerava na definisanje matematičke funkcije koja, u uslovima određenih pretpostavki i ograničenja, pravi najveću moguću razliku između dveju ili više populacija ili grupa. Deskriptivni i klasifikacioni aspekti diskriminacione analize su u veoma bliskoj vezi: pod određenim uslovima moguće je korišćenjem kanoničkih diskriminacionih funkcija dobijenih postupkom deskriptivne diskriminacione analize postići klasifikaciju praktično istovetnu onoj koja se dobija u tim uslovima postupcima klasifikacione diskriminacione analize (cf. Welch, 1939; formalni dokazi ovih tvrdnji mogu se naći u Kshirsagar & Arseven, 1975 i Williams, 1982). Postupci klasifikacione diskriminacione analize minimizuju greške u klasifikaciji i zasnovani su na maksimumu posteriornih verovatnoća

Zavisno od oblasti u kojima se diskriminaciona analiza primenjuje veći naglasak stavlja se na deskriptivne ili na klasifikacione postupke. U psihologiji se, pri primeni diskriminacione analize u istraživanjima, znatno veća pažnja poklanja deskriptivnim aspektima ove analize, dok se u nekim drugim oblastima (antropologija, arheologija, biologija, medicina...) mnogo veći naglasak stavlja na njene prediktivne, tj. klasifikacione aspekte (cf. Maxwell, 1977). Kakvi „razumni" razlozi su doveli do zanemarivanja klasifikacionih aspekata diskriminacione analize u psihologiji nije nam sasvim jasno, posebno kada se uzme u obzir korist koju bi psihološka dijagnostika mogla imati u korišćenju diskriminacione analize kao statističkog sredstva za kalibrisanje psihodijagnostičkih instrumenata. I mi ćemo u ovom tekstu, ne odstupajući od tog osnovnog trenda, imati u vidu postupke deskriptivne diskriminacione analize, jer ovaj aspekt njene primene preovlađuje u psihološkim istraživanjima.

Osnovni model deskriptivne kanoničke diskriminacione analize može se formalno opisati na pojavno različite načine. Mi ćemo prikazati dva pojavno različita formalna opisa razvoja ovog modela. Prva vrsta opisa modela deskriptivne diskriminacione analize, koja je ovde prikazana u obliku koji je modifikovan za potrebe ovoga teksta, uobičajeno se sreće u statističkim časopisima i „klasičnim" knjigama iz primenjene multivarijacione statističke analize (cf. Cooley & Lohnes, 1962; Tatsuoka, 1971; Overall & Klett, 1972; Rao, 1974; Kovačić, 1994; Tacq, 1997; Stevens, 2002) i služi kao osnova za algoritam po kojem se ova metoda izvodi u programu SPSS – programu koji se najčešće koristi u analizama podataka psiholoških istraživanja. Drugi oblik formalnog opisa ovog modela dajemo kako bi se lakše razumelo izvođenje kvazikanoničke diskriminacione analize.

Formalni opis definisanja i rešavanja problema deskriptivne diskriminacione analize

„Klasični" način definisanja i rešavanja problema deskriptivne diskriminacione analize

Polazna teorijska struktura za razvoj i primenu diskriminacione analize jeste postojanje heterogene populacije sastavljene od mešavine g (sub)populacija ili grupa. Svaki član populacije može se opisati vektorom podataka sastavljenim od m + 1 komponenata, pri čemu pored m komponenti, koje sadrže informacije o kvantitativnim obeležjima, postoji dodatna komponenta koja specifikuje subpopulacijsku ili grupnu pripadnost datog člana populacije.

Cilj koji stoji u osnovi definisanja metoda deskriptivne diskriminacione analize može se formulisati na sledeći način: potrebno je pronaći linearnu transformaciju podataka na kvantitativnim obeležjima, tj. odrediti kanoničku diskriminacionu funkciju D, $D = v_1x_1 + v_2x_2 + \cdots + v_mx_m$, (linearnu kombinaciju rezultata $x_1, x_2,...,x_m$ dobijenih na skupu V od m početnih,[157] tj. merenih kvantitativnih varijabli, $V = \{v_j; j = 1,2,\ldots, m\}$) tako da aritmetičke sredine subpopulacija, tj. grupa na toj linearnoj kombinaciji, budu međusobno što je više moguće različite, u skladu sa sledećim kriterijumom:

$$\lambda_p = \frac{\mathbf{v}_p^{\mathrm{t}}\mathbf{B}\mathbf{v}_p}{\mathbf{v}_p^{\mathrm{t}}\mathbf{W}\mathbf{v}_p}.$$

U formuli kojom je definisan ciljni kriterijum maksimalnog razlikovanja grupa, $\mathbf{v}_p$ je vektor pondera (koeficijenata) ($\mathbf{v}_p = v_1, v_2,..., v_m$) kojima se množe merene varijable da bi se dobila diskriminaciona funkcija, $\mathbf{v}^{\mathrm{t}}_p$ je transpon vektora $\mathbf{v}_p$, $\mathbf{B}$ je matrica suma kvadrata i unakrsnih proizvoda između grupa, a $\mathbf{W}$ je matrica suma kvadrata i unakrsnih proizvoda unutar grupa. Matrice suma kvadrata i unakrsnih proizvoda između i unutar grupa dobijaju se na sledeći način:

$$\mathbf{W} = \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_{ik} - \mathbf{m}_k)(\mathbf{x}_{ik} - \mathbf{m}_k)^{\mathrm{t}},$$

$$\mathbf{B} = \sum_{k=1}^{g} n_k(\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^{\mathrm{t}},$$

pri čemu je $\mathbf{x}_{ik}$ vektor sa rezultatima jedinice posmatranja e_i na m početnih varijabli, k je oznaka subpopulacije ili grupe ($k = 1,..., g$), $\mathbf{m}_k$ je vektor aritmetičkih sredina na početnim varijablama za subpopulaciju ili grupu k, n_k je veličina grupe k, a $\mathbf{m}$ je vektor aritmetičkih sredina početnih varijabli za sve jedinice posmatranja zajedno. Rešenje problema sastoji se, dakle, u nalaženju vektora $\mathbf{v}_p$, tj. vektora koeficijenata za kanoničku diskriminacionu funkciju koji će dovesti do toga da međugrupna varijabilnost na diskriminacionoj funkciji ($\mathbf{v}_p^{\mathrm{t}}\mathbf{B}\mathbf{v}_p$) bude što je moguće veća u odnosu na unutargrupnu varijabilnost ove funkcije ($\mathbf{v}_p^{\mathrm{t}}\mathbf{W}\mathbf{v}_p$).

Matematički posmatrano, problem se očigledno sastoji u traženju maksimuma količnika funkcija, tj. funkcije koja predstavlja količnik dveju kvadratnih formi.

Diferenciranjem funkcije, tj. količnika λ po vektoru $\mathbf{v}$, i izjednačavanjem prvog izvoda sa vektorom $\mathbf{0}$, dobijamo:

[157] Termin *početne* ili *izvorne* varijable koristimo u ovom tekstu kao ime za varijable na osnovu kojih se grade linearne kombinacije bez obzira na njihov oblik, tj. početnu (nultu) tačku i skalu („metriku“).

$$\frac{\partial\lambda}{\partial\mathbf{v}} = \frac{2[\mathbf{B}\mathbf{v}_p)(\mathbf{v}^t_p\mathbf{W}\mathbf{v}_p) - (\mathbf{v}^t_p\mathrm{B}\mathbf{v}_p)(\mathbf{W}\mathbf{v}_p)}{(\mathbf{v}^t_p\mathbf{W}\mathbf{v}_p)^2} = \mathbf{0}\,.$$

Deljenjem celog izraza sa $\mathbf{v}_p{}^t\mathbf{W}\mathbf{v}_p$ i stavljanjem λ umesto količnika

$$\frac{\mathbf{v}^t_p\mathbf{B}\mathbf{v}_p}{\mathbf{v}^t_p\mathbf{W}\mathbf{v}_p},$$

gornja jednačina svodi se na:

$$\frac{2(\mathbf{B}\mathbf{v}_p - \lambda_p\mathbf{W}\mathbf{v}_p)}{\mathbf{v}^t_p\mathbf{W}\mathbf{v}_p} = \mathbf{0}\,.$$

Rešavanje postavljenog problema, uz uslov $\mathbf{v}_p{}^t\mathbf{W}\mathbf{v}_p = 1$, svodi se, dakle, na rešavanje problema:

$$(\mathbf{B} - \lambda_p\mathbf{W})\mathbf{v}_p = \mathbf{0}$$

koji je poznat u statističkoj literaturi kao opšti problem svojstvenih vrednosti. Rešavanje ovog problema najčešće se izvodi njegovim svođenjem na tzv. obični problem svojstvenih vrednosti. Množenjem sleva inverzom matrice **W**, prethodna jednačina dobija sledeći oblik:

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda_p\mathbf{I})\mathbf{v}_p = \mathbf{0}\,.$$

Budući da je matrica **B** pozitivno semidefinitna ($\mathbf{x}^t\mathbf{B}\mathbf{x} \geq 0$, za svako **x**), matrica **W** (ako skup početnih varijabli čini skup linearno nezavisnih vektora, tj. ako se nijedna među početnim varijablama ne može izraziti kao linearna kombinacija preostalih) pozitivno definitna ($\mathbf{x}^t\mathbf{W}\mathbf{x} > 0$, za svako **x**), a obe matrice simetrične, svojstvene vrednosti matrice $\mathbf{W}^{-1}$ **B** predstavljaju tražene ekstreme funkcije λ, tj. količnika $\lambda_p = \dfrac{\mathbf{v}^t_p\mathbf{B}\mathbf{v}_p}{\mathbf{v}^t_p\mathbf{W}\mathbf{v}_p}$. Prva, tj. najveća svojstvena vrednost dobijena rešavanjem ovog problema predstavlja traženi maksimum (precizna formulacija teoreme o ekstremima funkcije koja predstavlja količnik dveju kvadratnih formi i njen dokaz mogu se naći, na primer, u McDonald, 1968, str. 355 i Schott, 1997, str. 117). Vektor neskaliranih koeficijenata za kanoničku diskriminacionu funkciju **v**, $\mathbf{v} = (v_1, v_2,..., v_m)$, odgovara karakterističnom vektoru pridruženom datoj svojstvenoj vrednosti matrice $\mathbf{W}^{-1}\mathbf{B}$. Matrica $\mathbf{W}^{-1}\mathbf{B}$ je očigledno proizvod dveju matrica od kojih matrica **B** ima u opštem slučaju niži rang. Rang matrice **B** jednak je manjem od sledeća dva broja: broju početnih varijabli (m) ili broju grupa (g) umanjenom za jedan. Prema tome, rang ($\mathbf{W}^{-1}\mathbf{B}$) = r = min(m; g - 1). Broj nenultih svojstvenih vrednosti i njima pridruženih svojstvenih vektora te, shodno tome, broj različitih kanoničkih diskriminacionih funkcija koje je moguće definisati na osnovu skupa od m početnih varijabli za međusobno razlikovanje g subpopulacija ili grupa, identičan je rangu matrice $\mathbf{W}^{-1}\mathbf{B}$. Različite diskriminacione funkcije međusobno su statistički ortogonalne, tj. nekorelisane. Međutim, budući da matrica $\mathbf{W}^{-1}\mathbf{B}$ nije u opštem slučaju simetrična, diskriminacione funkcije nisu „geometrijski ortogonalne“, tj. svojstveni vektori koji sadrže neskalirane koeficijente za različite diskriminacione funkcije nisu međusobno ortogonalni ($\mathbf{v}_l^t\mathbf{v}_n \neq 0 \mid l \neq n$).

Postoji još nekoliko načina da se rešavanje opšteg problema svojstvenih vrednosti svede na rešavanje „običnog problema svojstvenih vrednosti“ (cf. McDonald, 1968). Osim prikazanog i (sa stanovišta linearne algebre) „najočiglednijeg“ načina, prikazaćemo i onaj koji je u osnovi algoritma po kojem se ovaj problem rešava u programu SPSS:

– Matrica suma kvadrata i unakrsnih proizvoda unutar grupa faktorizuje se korišćenjem dekompozicije Čoleskog na sledeći način:[158]

$$\mathbf{W} = \mathbf{LU},$$

pri čemu je $\mathbf{L}$ donjetrougaona matrica, a $\mathbf{U} = \mathbf{L}^t$.

– Pomoću matrica $\mathbf{L}$ i $\mathbf{U}$ konstruiše se simetrična matrica $\mathbf{L}^{-1}\mathbf{B}\mathbf{U}^{-1}$. Na taj način se $(\mathbf{B} - \lambda_p\mathbf{W})\mathbf{V}$ svodi na rešavanje problema:

$$(\mathbf{L}^{-1}\mathbf{B}\mathbf{U}^{-1} - \lambda_p\mathbf{I})(\mathbf{UV}) = \mathbf{0}.$$

– Svojstvene vrednosti i matrica ortonormalnih svojstvenih vektora $\mathbf{UV}$ računski se potom dobijaju rešavanjem sistema jednačina postupkom tridijagonalizacije i QR faktorizacije. Matrica sa vektorima $\mathbf{v_p}$ u kojima su neskalirani koeficijenti za kanoničke diskriminacione funkcije dobija se na sledeći način:

$$\mathbf{V} = \mathbf{U}^{-1}(\mathbf{UV}).$$

Matrica standardizovanih koeficijenata za diskriminacione funkcije dobija se na sledeći način:

$$\mathbf{K} = \mathbf{DV},$$

pri čemu je $\mathbf{D}$ dijagonalna matrica koja u dijagonali sadrži kvadratne korenove zbirova kvadrata unutar grupa za početne varijable. Na taj način, koeficijenti za diskriminacione funkcije skaliraju se tako da bude $\mathbf{V}^t\mathbf{W}\mathbf{V} = \mathbf{I}$. Matrica nestandardizovanih koeficijenata za diskriminacione funkcije, u oznaci $\mathbf{L}$, dobija se na osnovu matrice standardizovanih koeficijenata:

$$\mathbf{L} = (n - g)^{1/2}\, \mathbf{D}^{-1}\mathbf{K},$$

pri čemu je n veličina uzorka, g broj grupa, a $\mathbf{D}^{-1}$ dijagonalna matrica čiji nenulti elementi predstavljaju recipročne vrednosti kvadratnih korenova elemenata iz glavne dijagonale matrice $\mathbf{W}$, tj. matrice suma kvadrata i unakrsnih proizvoda unutar grupa.

Matrica strukture diskriminacionih funkcija, tj. matrica korelacija početnih varijabli i kanoničkih diskriminacionih funkcija, u oznaci $\mathbf{F}_w$, dobija se na sledeći način:[159]

[158] Dekompozicija Čoleskog zasniva se na sledećoj teoremi: Ako je $\mathbf{A}$ pozitivno definitna matrica reda mxm, tada postoji jedinstvena mxm donjetrougaona matrica $\mathbf{T}$, reda mxm, koja sadrži pozitivne dijagonalne elemente takve da je $\mathbf{A} = \mathbf{TT}^t$ (dokaz opštijeg oblika ove teoreme može se naći u Schott (1997), str. 139).

[159] Ovakav postupak za dobijanje matrice strukture diskriminacionih funkcija, koji je inače implementiran u programskom paketu SPSS, veoma je osporavan, i to sa valjanim razlozima (cf. Momirović, 1997). Ovakav postupak zapravo predstavlja ocenu matrice strukture po principu maksimalne verodostojnosti, ali samo ako je

$$\mathbf{F}_\mathrm{w} = \mathbf{D}^{-1}\,\mathbf{WV}.$$

Centroidi grupa, tj. aritmetičke sredine grupa na diskriminacionoj funkciji p, u oznaci m_{pk} (p = 1,..., r; k = 1,..., g), dobijaju se kao prosta linearna kombinacija aritmetičkih sredina grupa na početnim varijablama, pri čemu se kao koeficijenti za linearnu kombinaciju koriste nestandardizovani koeficijenti za diskriminacione funkcije iz matrice **L**:

$$\mathrm{m}_{pk} = \mathrm{a}_p + \sum_{j=1}^{\mathrm{m}} \mathrm{l}_{jp}\mathrm{M}_{jk}\,,$$

pri čemu je M_{jk} aritmetička sredina grupe k na početnoj varijabli j, a a_p je konstanta koja služi za izračunavanje diskriminacionog skora na osnovu početnih nestandardizovanih varijabli. Konstanta a_p za diskriminacionu funkciju p definiše se na sledeći način:

$$\mathrm{a}_p = -\sum_{j=1}^{\mathrm{m}} \mathrm{l}_{jp}\mathrm{M}_j\,,$$

pri čemu je M_j aritmetička sredina celokupnog uzorka ispitanika na početnoj nestandardizovanoj varijabli v_j.

Kao mera udaljenosti centroida na diskriminacionoj funkciji p služi kanonička korelacija, u oznaci R_{cp}, koja se dobija iz svojstvene vrednosti koja odgovara datoj diskriminacionoj funkciji:

$$\mathrm{R}_{\mathrm{c}p} = \sqrt{\frac{\lambda_p}{1+\lambda_p}}\,.$$

Premda je, s obzirom na rang matrice $\mathbf{W}^{-1}\mathbf{B}$, moguće izvesti r diskriminacionih funkcija (r = minimum(broj početnih varijabli, broj grupa - 1)), broj diskriminacionih funkcija koje sadrže bitne informacije za razlikovanje subpopulacija obično je nešto manji. Testiranjem statističke značajnosti diskriminacionih funkcija određuje se dimenzionalnost diskriminacionog prostora i za dalju analizu (i tumačenje) zadržavaju se samo značajne kanoničke diskriminacione funkcije.

Za testiranje statističke značajnosti diskriminacione funkcije p, tj. za testiranje hipoteza:

$$\mathrm{H}_{0p}\!:\!\acute{\eta}_p = 0,$$

---

u osnovi diskriminacione analize model sa g populacija, sa identičnim matricama kovarijansi unutar grupa (cf. Huberty, 1975). Ovakva definicija strukture diskriminacionih funkcija može, međutim, dovesti do toga da je struktura diskriminacionih funkcija u određenim slučajevima definisana najmanje onim početnim varijablama na kojima se grupe najviše razlikuju! To zaista može istraživačima koji nemaju dovoljno formalnog statističkog znanja predstavljati veliki problem s obzirom na važnost koju struktura diskriminacione funkcije ima u interpretaciji ishoda diskriminacione analize. Stoga je najbolje ocenu matrice strukture u programu SPSS napraviti i kao matricu običnih linearnih korelacija između sačuvanih diskriminacionih skorova i početnih varijabli.

pri čemu su $\acute{\eta}_p$ svojstvene vrednosti matrice $\mathbf{W}^{-1}\mathbf{B}$ u populaciji, najčešće se koristi Bartletov V_s statistik (cf. Stevens, 2002). Ovaj statistik se najčešće koristi jer je implementiran u programu SPSS. Bartletov V_s-statistik definisan je kao funkcija Vilksovog statistika Λ, $\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W}+\mathbf{B}|}$, ili, ekvivalentno, uzoračkih svojstvenih vrednosti matrice $\mathbf{W}^{-1}\mathbf{B}$ (budući da je $\Lambda_s = \prod_{p=s+1}^{r} \frac{1}{1+\lambda_p}$; $p = 1,\ldots,r$; $s = 0,\ldots,r-1$):[160]

$$V_s = -(n - \frac{m+g}{2} - 1)\ln \Lambda_s = (n - \frac{m+g}{2} - 1)\sum_{s=s+1}^{r} \ln(1+\lambda_p); \quad p = 1,\ldots r; \quad s = 0,\ldots,r-1$$

U izrazu za V_s n je veličina uzorka, m broj početnih varijabli, g broj grupa, λ_p uzoračka svojstvena vrednost koja odgovara diskriminacionoj funkciji p, a $p = s + 1$. Bartletov V_s-statistik, ima, ako je nulta hipoteza tačna (tj. ako su centroidi subpopulacija na diskriminacionoj funkciji jednaki) približno hi-kvadrat raspodelu sa $\{(m - s)[g - (s + 1)]\}$ stepeni slobode.

<u>Alternativni način definisanja i rešavanja problema deskriptivne diskriminacione analize</u>

(modifikovano prema Momirović & Zorić, 1996; Momirović, 1997 i Tenjović, 2002)

Kanonička diskriminaciona analiza može se definisati i kao poseban slučaj kanoničke korelacione analize između skupa kvantitativnih varijabli i drugog skupa varijabli koji nastaje pogodnim projektovanjem kvantitativnih varijabli u prostor definisan binarnim varijablama. Binarne varijable sadrže u sebi informacije o kategorijalnoj, tj. subpopulacijskoj pripadnosti entiteta.

Neka je E ($E = \{e_i; i = 1,\ldots,n\} \subset P = \bigcup_k^g P_k$) slučajni uzorak n entiteta izvučen iz populacije koja predstavlja uniju g subpopulacija; neka je Q ($Q = \{q_k; k = 1,2,\ldots, g\}$) kategorička varijabla koja definiše subpopulacijsku pripadnost entiteta iz skupa E; neka je V ($V = \{v_j ; j = 1,2,\ldots, m\}$) skup od m kvantitativnih varijabli. Merenjem jedinica uzorka E u pogledu skupa kvantitativnih varijabli V, dobijamo matricu sirovih podataka $\mathbf{X}$, reda nxm:

$$\mathbf{X} = (x_{ij});\ i = 1,\ldots, n;\ j = 1,\ldots, m.$$

Pretpostavićemo da je izvršena standardizacija varijabli iz skupa V, tj. da su u matrici $\mathbf{Z}$ podaci koji su dobijeni iz matrice $\mathbf{X}$ na sledeći način:

$$\mathbf{Z} = (\mathbf{X} - \mathbf{\Pi X})\mathbf{D}^{-1}$$

pri čemu je $\mathbf{\Pi}$ centroidni projektor. Centroidni projektor je matrica reda nxn čiji su svi elementi isti i jednaki $\frac{1}{n}$, a dobija se na sledeći način:

$$\mathbf{\Pi} = \mathbf{e}(\mathbf{e}^t\mathbf{e})^{-1}\mathbf{e}^t,$$

[160] Dokaz jednakosti izraza za V_s na osnovu Vilksovog statistika Λ i na osnovu svojstvenih vrednosti matrice $\mathbf{W}^{-1}\mathbf{B}$ može se naći u Tatsuoka (1971).

pri čemu je **e** kolonski vektor reda nx1 čiji svi elementi su jedinice. (Pri tome skalarni proizvod $\mathbf{e}^t\mathbf{e}$ odgovara n, tj. veličini uzorka.) Proizvod $\mathbf{\Pi X}$ u izrazu za **Z** je matrica reda nxm u kojoj su kao „podaci" ispitanika upisane aritmetičke sredine svih ispitanika na datoj varijabli. Budući da je

$$\mathbf{D}^2 = \mathrm{diag}\mathbf{C},$$

pri čemu je **C**, $\mathbf{C} = (\mathbf{X}^t\mathbf{X} - \mathbf{X}^t\mathbf{\Pi X})(\mathbf{e}^t\mathbf{e})^{-1}$, matrica kovarijansi kvantitativnih varijabli iz skupa *V*; dijagonalna matrica $\mathbf{D}^{-1}$ sadrži recipročne vrednosti standardnih devijacija ovih varijabli.

Matrica **R**, pri čemu je

$$\mathbf{R} = \mathbf{Z}^t\mathbf{Z}n^{-1},$$

tada predstavlja matricu interkorelacija kvantitativnih varijabli iz skupa *V*. Pretpostavićemo da je matrica **R** nesingularna, tj. da kvantitativne varijable iz skupa *V* čine linearno nezavisni skup vektora. Prema tome, matrica **R** ima regularni inverz, koji ćemo označiti sa $\mathbf{R}^{-1}$.

Kodirajmo subpopulacijsku pripadnost entiteta u kompletni disjunktivni oblik:

$$E \otimes Q = \mathbf{S} = (s_{ik});\ i = 1,\dots, n;\ k = 1,\dots, g.$$

Matrica **S** je selektorska ili indikatorska matrica reda nxk koja sadrži g binarnih varijabli (varijabli na kojima kao podaci postoje samo cifre 0 i 1), a čiji svaki element je definisan na sledeći način:

$$s_{ik} = \begin{cases} 1 \mid e_i \in q_k \\ 0 \mid e_i \notin q_k \end{cases} \quad i = 1,\dots, n;\ k = 1,\dots, g\ .$$

Prema tome, u matrici **S** član uzorka koji pripada subpopulaciji *k* ima u koloni koja označava subpopulaciju *k* oznaku 1, a u svim drugim kolonama oznaku 0.
Definišimo pomoću matrice **S** projektorsku matricu **P** na sledeći način:

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

(Matrica $(\mathbf{S}^t\mathbf{S})^{-1}$ koja se pojavljuje u izrazu za **P** je dijagonalna matrica reda gxg čije dijagonalne vrednosti odgovaraju recipročnim vrednostima veličina grupa definisanih kategorijama varijable *Q*).
Matrica **P**, reda nxn, zapravo služi kao projektor u g-dimenzionalni hiperkub, tj. prostor koji grade binarne varijable iz **S**. Opšti element ove matrice, u oznaci p_{is}, definiše se na sledeći način:

$$p_{is} = \begin{cases} \dfrac{1}{n_k} \mid e_i \in q_k \cap e_s \in q_k & i = 1,\dots, n;\ s = 1,\dots, n \\ \\ 0 \mid e_i \notin q_k \cap e_s \notin q_k \end{cases} .$$

Dakle, element matrice **P** u redu *i* i koloni *s* jednak je recipročnoj vrednosti veličine grupe *k* ako članovi uzorka iz reda *i* i kolone *s* pripadaju istoj grupi *k*. Ukoliko to nije slučaj, element matrice **P** je nula.

Projektovanjem vektora standardizovanih varijabli iz matrice **Z** u prostor („g-dimenzionalni hiperkub“) koji razapinju binarni vektori selektorske matrice **S**, operacijom

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

dobija se matrica **G** reda nxm, čiji elementi su definisani na sledeći način:

$$g_{ij} = (\mathbf{s}_k{}^t\mathbf{s}_k)^{-1}\mathbf{s}_k{}^t\mathbf{z}_j \mid e_i \in q_k$$

$$i = 1,..., n; j = 1,..., m.$$

Dakle, u matrici **G** se kao „podatak“ na varijabli v_j za entitet e_i koji pripada populaciji k ili grupi k nalazi aritmetička sredina za varijablu v_j grupe kojoj entitet pripada.
Matrica **B** u izrazu za **G**, pri čemu je

$$\mathbf{B} = (\mathbf{S}^t\mathbf{S})^{-1}\,\mathbf{S}^t\mathbf{Z},$$

predstavlja zapravo regresionu matricu reda gxm koja sadrži aritmetičke sredine varijabli po pojedinim grupama. Ova matrica, kao što se iz izraza za **G** može videti, omogućuje – kada se pomnoži sleva matricom **S** – predviđanje rezultata ispitanika na nekoj varijabli samo na osnovu njegove grupne pripadnosti, i to tako da zbir kvadriranih grešaka predviđanja bude minimalan (Tenjović, 2002). Dakle, matrica **B** definisana je na sledeći način:

$$\mathbf{SB} = \mathbf{Z} - \mathbf{E} \mid \varepsilon^2 = \text{trag}\,(\mathbf{E}^t\mathbf{E}) = \text{minimum},$$

pri čemu je **E** matrica sa greškama predviđanja standardizovanih rezultata iz matrice **Z** na osnovu grupne pripadnosti, tj. informacija sadržanih u matrici **S** (dokaz o tome da je matrica **B** regresiona matrica sa zadatim svojstvima može se naći u Momirović, 1997)
Matrica kovarijansi varijabli iz matrice **G** dobija se na sledeći način:

$$\mathbf{A} = \mathbf{G}^t\mathbf{G}n^{-1} = \mathbf{Z}^t\mathbf{PZ}n^{-1}.$$

.
Budući da je **G** = **PZ** matrica, **A** je istovremeno i matrica kovarijansi varijabli iz matrica **Z** i **G**.

Kanonička diskriminaciona analiza se, na osnovu matrica **Z** i **G**, može definisati kao poseban slučaj kanoničke korelacione analize, tj. kao rešenje sledećeg problema:

$$\boxed{\begin{array}{c}\mathbf{Zx}_k = \mathbf{k}_k;\ \mathbf{Gy}_k = \mathbf{l}_k \mid \rho_k = \mathbf{k}_k{}^t\mathbf{l}_k\,n^{-1} = \text{maximum},\ \mathbf{k}_k{}^t\mathbf{k}_q\,n^{-1} = \mathbf{l}_k{}^t\mathbf{l}_q\,n^{-1} = \delta_{kq} \\ \\ \mathbf{x}_k = ?;\ \mathbf{y}_k = ?\end{array}}$$

Pri tom je δ_{kq} Kronekerova delta (ako je $k = q$, tada je $\delta_{kq} = 1$, a ako je $k \neq q$, onda je $\delta_{kq} = 0$). Dakle, problem se sastoji u određivanju vektora $\mathbf{x}_k$ i $\mathbf{y}_k$ (vektora sa koeficijentima za kanoničke varijable $\mathbf{k}_k$ i $\mathbf{l}_k$). Kanoničke varijable $\mathbf{k}_k$ predstavljaju linearne kombinacije

skupova početnih standardizovanih varijabli (varijabli iz matrice $\mathbf{Z}$). Kanoničke varijable $\mathbf{l}_k$ su linearne kombinacije varijabli koje nastaju projektovanjem varijabli iz matrice $\mathbf{Z}$ u prostor što ga grade binarni vektori koji definišu subpopulacijsku pripadnost ispitanika (varijable iz matrice $\mathbf{G}$). Vektore $\mathbf{x}_k$ i $\mathbf{y}_k$ treba odrediti tako da: a) kanoničke korelacije linearnih kombinacija sa istim indeksom budu maksimalne ($\rho_k = \mathbf{k}_k^t\mathbf{l}_k\, n^{-1}$ = maximum); b) različito indeksirane kanoničke varijable unutar svakog skupa i različito indeksirane kanoničke varijable iz različitih skupova budu međusobno ortogonalne ($\delta_{kq} = 0$, $k \neq q$); c) kanoničke varijable da budu standardizovane ($\delta_{kq} = 1$, $k = q$).
Za određivanje funkcije koju treba maksimizovati polazimo od sledećih jednakosti:
a) $\rho_k = \mathbf{k}_k^t\mathbf{l}_k n^{-1} = \mathbf{x}_k^t\mathbf{Z}^t\mathbf{G}\mathbf{y}_k n^{-1} = \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k$
b) $\mathbf{k}_k^t\mathbf{k}_k n^{-1} = \mathbf{x}_k^t\mathbf{R}\mathbf{x}_k = \mathbf{x}_k^t\mathbf{R}\mathbf{x}_k$
c) $\mathbf{l}_k^t\mathbf{l}_k n^{-1} = \mathbf{y}_k^t\mathbf{Z}^t\mathbf{G}\mathbf{y}_k n^{-1} = \mathbf{y}_k^t\mathbf{A}\mathbf{y}_k$.

Za $k = 1$, funkcija f čije uslovne ekstreme tražimo je, prema tome:

$$f(\mathbf{x}_k, \mathbf{y}_k, \lambda_k, \eta_k) = \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k - (1/2)\lambda_k(\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k - 1) - (1/2)\eta_k(\mathbf{y}_k^t\mathbf{A}\mathbf{y}_k - 1).$$

Diferenciranjem funkcije f po vektoru $\mathbf{x}_k$, dobijamo:

$$\partial f/\partial \mathbf{x}_k = \mathbf{A}\mathbf{y}_k - \lambda_k \mathbf{R}\mathbf{x}_k.$$

Diferenciranjem iste funkcije po vektoru $\mathbf{y}_k$, dobijamo:

$$\partial f/\partial \mathbf{y}_k = \mathbf{A}\mathbf{x}_k - \eta_k\mathbf{A}\mathbf{y}_k.$$

Izjednačavanjem dobijenih izvoda sa nulom i njihovim sređivanjem, dobijamo sledeće dve jednakosti:

$$\mathbf{A}\mathbf{y}_k = \lambda_k \mathbf{R}\mathbf{x}_k$$

i

$$\mathbf{A}\mathbf{x}_k = \eta_k\mathbf{A}\mathbf{y}_k.$$

Množenjem sleva prve jednakosti sa $\mathbf{x}_k^t$, a druge jednakosti sa $\mathbf{y}_k^t$, dobijaju se sledeće jednakosti:

$$\mathbf{x}_k^t\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k$$

i

$$\mathbf{y}_k^t\mathbf{A}\mathbf{x}_k = \eta_k\mathbf{y}_k^t\mathbf{A}\mathbf{y}_k.$$

Na osnovu uslova $\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k = 1$ i $\mathbf{y}_k^t\mathbf{A}\mathbf{y}_k = 1$, očigledno sledi da je $\lambda_k = \eta_k$.

Množenjem, pak, jednakosti $\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{R}\mathbf{x}_k$ sleva matricom $\mathbf{R}^{-1}$, dobija se:

$$\mathbf{R}^{-1}\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{R}^{-1}\mathbf{R}\mathbf{x}_k = \lambda_k\mathbf{x}_k,$$

odakle sledi da je

$$\mathbf{x}_k = \mathbf{R}^{-1}\mathbf{A}\mathbf{y}_k\lambda_k^{-1}.$$

Na osnovu jednakosti $\mathbf{A}\mathbf{x}_k = \eta_k\mathbf{A}\mathbf{y}_k$ i $\lambda_k = \eta_k$ sledi da je:

$$\mathbf{A}\mathbf{x}_k\lambda_k^{-1} = \mathbf{A}\mathbf{y}_k,$$

pa je, prema tome, $\mathbf{x}_k\lambda_k^{-1} = \mathbf{y}_k$.

Na osnovu $\mathbf{R}^{-1}\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{x}_k$ dobija se, zamenom $\mathbf{y}_k$ sa $\mathbf{x}_k\lambda_k^{-1}$, sledeća jednakost:

$$\mathbf{R}^{-1}\mathbf{A}\mathbf{x}_k\lambda_k^{-1} = \lambda_k\mathbf{x}_k.$$

Množenjem ove jednakosti sa λ_k, dobijamo:

$$\mathbf{R}^{-1}\mathbf{A}\mathbf{x}_k = \lambda_k^2\mathbf{x}_k.$$

Na taj način rešenje problema kanoničke diskriminacione analize kao posebnog slučaja kanoničke korelacione analize svodi se na rešavanje problema svojstvenih vrednosti matrice $\mathbf{R}^{-1}\mathbf{A}$:

$$(\mathbf{R}^{-1}\mathbf{A} - \lambda_k^2)\mathbf{x}_k = \mathbf{0}.$$

Budući da je

$$\rho_k = \mathbf{k}_k^t\mathbf{l}_k n^{-1} = \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k \text{ i } \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k = \lambda_k,$$

to je, na osnovu $\mathbf{y}_k = \mathbf{x}_k\lambda_k^{-1}$,

$$\rho_k = \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k = \mathbf{x}_k^t\mathbf{A}\mathbf{x}_k\lambda_k^{-1},$$

a, što je još važnije, na osnovu $\mathbf{x}_k^t\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k$ i $\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k = 1$

$$\rho_k = \mathbf{x}_k^t\mathbf{A}\mathbf{y}_k = \lambda_k\mathbf{x}_k^t\mathbf{R}\mathbf{x}_k = \lambda_k.$$

Svojstvene vrednosti matrice $\mathbf{R}^{-1}\mathbf{A}$ su, prema tome, kvadrati kanoničkih korelacija između linearnih kombinacija skupova početnih standardizovanih varijabli (varijabli iz matrice **Z**) i varijabli iz matrice **G** (varijabli koje nastaju projektovanjem varijabli iz matrice **Z** u prostor koji grade binarni vektori koji definišu subpopulacijsku pripadnost ispitanika). Svojstveni vektori matrice $\mathbf{R}^{-1}\mathbf{A}$ predstavljaju tražene vektore $\mathbf{x}_k$, koji treba da posluže za pravljenje kanoničkih varijabli iz **Z**, tj. linearnih kombinacija varijabli iz matrice **Z**. Vektori $\mathbf{y}_k$, koji služe za izvođenje linearnih kombinacija varijabli iz matrice **G**, dobijaju se na osnovu sledeće jednakosti:

$$\mathbf{y}_k = \mathbf{x}_k\lambda_k^{-1}.$$

Kanoničke korelacije, tj. korelacije između odgovarajućih linearnih kombinacija varijabli iz **Z** i **G**, proporcionalne su diferencijaciji centroida u prostoru varijabli iz matrice **Z**. (Formalni dokaz ekvivalentnosti maksimizacije koeficijenata kanoničke korelacije između kanoničkih diskriminacionih varijabli iz **Z** i **G** i maksimizacije udaljenosti centroida grupa na kanoničkim diskriminacionim funkcijama može se naći u Momirović (1997).)
Rang matrice $\mathbf{R}^{-1}\mathbf{A}$, budući da je matrica **R** nesingularna, jednak je rangu matrice **A**. Prema tome, na osnovu nenultih svojstvenih vrednosti matrice $\mathbf{R}^{-1}\mathbf{A}$ moguće je konstruisati ukupno s kanoničkih diskriminacionih varijabli, pri čemu je s = minimum(m, g - 1).

Matrica kanoničkih diskriminacionih varijabli iz matrice **Z**, reda nxs, u oznaci **K**, dobija se iz početnih standardizovanih varijabli na sledeći način:

$$\mathbf{K} = \mathbf{ZX},$$

pri čemu je $\mathbf{X} = (\mathbf{x}_k)$, $k = 1,..., s$ matrica svojstvenih vektora $\mathbf{R}^{-1}\mathbf{A}$.
Matrica kanoničkih diskriminacionih varijabli za varijable iz matrice **G**, reda nxs, u oznaci **L**, dobija se na sledeći način:

$$\mathbf{L} = \mathbf{GY} = \mathbf{GX}\Lambda^{-1},$$

pri čemu je Λ^{-1} inverz dijagonalne matrice čiji nenulti elementi odgovaraju kvadratnim korenovima svojstvenih vrednosti matrice $\mathbf{R}^{-1}\mathbf{A}$.
Dijagonalna matrica kanoničkih korelacija, u oznaci ρ, reda s, definisana je na sledeći način:

$$\rho = \mathbf{K}^t\mathbf{L}n^{-1} = \mathbf{X}^t\mathbf{A}\,\mathbf{X}\Lambda^{-1} = \mathbf{X}^t\mathbf{AY}.$$

Matrica korelacija skupa početnih varijabli iz matrice **Z** i kanoničkih diskriminacionih varijabli tog skupa, tj. matrica strukture kanoničkih diskriminacionih varijabli, dobija se na sledeći način:

$$\mathbf{F} = \mathbf{Z}^t\mathbf{K}n^{-1} = \mathbf{RX}.$$

Matrica **K** sadrži vektore rezultata na kanoničkim diskriminacionim funkcijama (diskriminacione skorove), a matrica **X** vektore koeficijenata za diskriminacione funkcije. Na osnovu vektora u matrici strukture koji odgovara datoj diskriminacionoj funkciji moguće je, na osnovu sadržaja izvornih varijabli koje sa tom funkcijom najviše koreliraju, odrediti sadržaj diskriminacione funkcije.

Matrica koja sadrži centroide grupa E_k, k = 1,..., g iz uzorka E na kanoničkim diskriminacionim varijablama iz **Z**, u oznaci **M**, može se izraziti na različite načine:

$$\mathbf{M} = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{K} = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{ZX} = \mathbf{BX}.$$

Za testiranje statističke značajnosti kanoničkih korelacija, tj. nultih hipoteza:

$$H_{0k} : \varphi_k = 0, (k = 1,..., r),$$

pri čemu su φ_k kanoničke korelacije u populaciji, najčešće se koristi Bartletov V_k-statistik

$$V_k = -(n - \frac{m+g}{2} - 1)\ln\Lambda_k = -(n - \frac{m+g}{2} - 1)\sum_{k=s+1}^{r}\ln(1-\rho_{s+1}^2);\; k = 1,\dots r;\; s = 0,\dots, r-1.$$

U izrazu za V_k-statistik n je veličina uzorka, m broj početnih varijabli, g broj grupa, a ρ_s koeficijenti kanoničkih korelacija dobijeni na uzorku. Nulta distribucija uzorkovanja statistika V_k približno odgovara hi-kvadrat distribuciji sa (m - k + 1)(g - k) stepeni slobode. Dakle, testiranje statističke značajnosti kanoničkih korelacija praktično je identično sa testiranjem statističke značajnosti diskriminacionih funkcija, što smo objasnili u odeljku o „klasičnom“ načinu definisanja i rešavanja problema deskriptivne diskriminacione analize. Pored toga, testiranje statističke značajnosti diskriminacione funkcije identično je F-testu za testiranje statističke značajnosti razlika aritmetičkih sredina grupa na kanoničkoj diskriminacionoj funkciji.

Bez obzira na to na da li se problem deskriptivne diskriminacione analize definiše i rešava na „klasični“ ili „alternativni” način, testiranje statističke značajnosti diskriminacionih funkcija zaustavlja se kada se naiđe na prvu kanoničku korelaciju (tj. diskriminacionu funkciju) za koju se ne može odbaciti nulta hipoteza. Sve diskriminacione funkcije koje slede posle prve diskriminacione funkcije koja nije statistički značajna smatraju se statistički neznačajnim. Pri tumačenju rezultata uzimaju se u obzir samo one kanoničke diskriminacione funkcije koje daju značajne kanoničke korelacije.

Tumačenje ishoda deskriptivne diskriminacione analize

Budući da je otkrivanje psihološke prirode razlika između postojećih populacija jedan od najčešćih ciljeva primene deskriptivne diskriminacione analize u psihologiji, veoma važan segment u primeni ove metode predstavlja tumačenje statistički značajnih kanoničkih diskriminacionih funkcija. Potrebno je, naime, u kanoničkoj diskriminacionoj varijabli prepoznati smisaonu psihološku strukturu, složaj koji može da se podvede pod neki od teorijskih konstrukata psihološke nauke, ili, pak, prepoznati složaj koji sugeriše smisaono definisanje novog konstrukta. Ovaj segment primene kanoničke diskriminacione analize često je skopčan s teškoćama i nekonzistentnošću. Poreklo nekonzistentnosti potiče jednim delom iz različitih „preporuka“ i „recepata“ koje konstruktori najčešće korišćenih statističkih paketa i udžbenika namenjenih primeni multivarijacionih metoda daju u vezi sa interpretacijom dobijenih rezultata. Pored toga, izvor nekonzistentnosti može biti i to što, osim u slučaju sa dve grupe (i jednom diskriminacionom funkcijom), nema eksplicitnog izraza za vektor koeficijenata za diskriminacionu funkciju u terminima početnih struktura podataka.[161] Ponekad se, naime, za tumačenje diskriminacionih funkcija u psihološkim istraživanjima analiziraju složajevi (standardizovanih) koeficijenata za diskriminacione funkcije (cf. Campbell, 1982; Huberty, 1984; Rencher, 1992; Stevens, 2002). Budući da koeficijenti za kanoničku diskriminacionu funkciju spadaju u tzv. „parcijalne“ koeficijente, tj. pokazatelje koji odražavaju povezanost svake početne varijable i diskriminacione funkcije u uslovima u kojima je iz ove povezanosti isključen uticaj preostalih početnih varijabli, analiza složajeva ovih koeficijenata u tumačenju prirode i strukture kanoničkih diskriminacionih varijabli može da navede na pogrešne zaključke (cf. Klecka, 1980). Koeficijenti za kanoničku diskriminacionu funkciju pokazuju specifičan „matematički“ doprinos svake početne varijable, *u kontekstu preostalih varijabli u skupu*, građenju diskriminacione funkcije. Za tumačenje same strukture dobijenih diskriminacionih funkcija (tj. njihove prirode) smisaonije je u uobičajenom kontekstu međusobno korelisanih početnih varijabli analizirati matricu strukture, tj. složajeve korelacija između početnih varijabli i kanoničkih diskriminacionih funkcija (mada ima i sasvim drugačijih mišljenja,

[161] U slučaju kada postoje dve grupe ili subpopulacije, vektor koeficijenata za diskriminacionu funkciju, u oznaci **v**, može se, u klasičnoj formulaciji diskriminacione analize, izraziti eksplicitno na sledeći način: $\mathbf{v} = m\,\mathbf{T}^{-1}\mathbf{d}$, pri čemu je $\mathbf{T}$ ($\mathbf{T} = \mathbf{B} + \mathbf{W}$) matrica ukupnih suma kvadrata i unakrsnih proizvoda, $\mathbf{d}$ ($\mathbf{d} = \mathbf{m}_1 - \mathbf{m}_2$) vektor razlika između vektora aritmetičkih sredina dveju grupa, a m nepoznati skalar. Vektor **v** je na taj način eksplicitno definisan do na arbitrarnu konstantu proporcionalnosti, a ova konstanta se određuje najčešće skaliranjem vektora **v** na jediničnu dužinu (cf. Tatsuoka, 1971, str. 170–173).

cf. Rencher, 1992). Odnos između koefijenata korelacije iz matrice strukture i koeficijenata za kanoničke diskriminacione funkcije može se uočiti iz prethodno prikazanog izraza za matricu strukture diskriminacionih funkcija, u oznaci **F**:

$$\mathbf{F} = \mathbf{RX},$$

pri čemu je **R** matrica interkorelacija početnih varijabli, a **X** matrica neskaliranih koeficijenata za diskriminacione funkcije.
Množenjem datog izraza sleva matricom $\mathbf{R}^{-1}$, dobijamo:

$$\mathbf{X} = \mathbf{R}^{-1}\mathbf{F}.$$

Prema tome, koeficijenti za diskriminacione funkcije pored „uticaja“ početnih varijabli na strukturu diskriminacione funkcije (matrica **F**) uključuju istovremeno i „uticaj“ drugih varijabli preko strukture interkorelacija među njima (matrica **R**). Koeficijenti za diskriminacione funkcije mogu svakako biti od koristi u analizi rezultata kanoničke diskriminacione analize: složajevi ovih koeficijenata, kada se analiziraju naporedo sa odgovarajućim složajevima matrice strukture, mogu ukazivati na početne varijable koje su redundantne u datom kontekstu početnih varijabli.

Za određivanje „relativne važnosti“ početnih kvantitativnih varijabli u razlikovanju subpopulacija predložena su dva pokazatelja (cf. Thomas & Zumbo, 1996):

1. Paralelni koeficijent diskriminacionog količnika, tj. proizvod standardizovanog koeficijenta za diskriminacionu funkciju i koeficijenta linearne korelacije date početne varijable i diskriminacione funkcije, u oznaci PDRC (v_j) (engl. Parallel Discriminant Ratio Coefficient), koji je definisan sledećom formulom:

$$\mathrm{PDRC}(\mathrm{v}_j) = \mathrm{k}_{jp} * \mathrm{r}_{jp}$$

Pri tome, k_{jp} je standardizovani koeficijent varijable v_j za diskriminacionu funkciju p, a r_{jp} je koeficijent linearne korelacije varijable v_j i diskriminacione funkcije p. Što je paralelni koeficijent diskriminacionog količnika za neku početnu varijablu viši, to je uloga te varijable u definisanju diskriminacione funkcije, tj. razlikovanju grupa u kontekstu preostalih varijabli, važnija. Paralelni koeficijent diskriminacionog količnika može biti negativan, što znatno može otežati njegovo tumačenje. Ukoliko je taj koeficijent negativan ali jako nizak, može se tretirati kao da je praktično jednak nuli. Ovaj koeficijent predstavlja analogon Pratovog indeksa u multiploj regresionoj analizi (v. glavu **XIV**).

2. Ukupni koeficijent diskriminacionog količnika za varijablu v_j, u oznaci TDRC (v_j) (engl. Total Discriminant Ratio Coefficient), koji je definisan sledećom formulom:

$$\mathrm{TDRC}(\mathrm{v}_j) = \left|\mathrm{k}_{jp}\right| \sqrt{\frac{1 + \dfrac{\mathrm{df_b} * \mathrm{F}_j}{\mathrm{df_w}}}{1 + \lambda_p}}\,.$$

Pri tome, k_{jp} je standardizovani koeficijent varijable v_j za diskriminacionu funkciju p, λ_p je svojstvena vrednost za diskriminacionu funkciju p, a F_j, df_b i df_w su F-količnik, broj stepeni slobode između i unutar grupa (tim redom) iz jednofaktorske analize varijanse sa grupom kao faktorom a varijablom v_j kao zavisnom varijablom. Ako je ukupni koeficijent diskriminacionog količnika za neku početnu varijablu veoma blizu nuli ili jednak nuli za sve diskriminacione funkcije, takva varijabla je redundantna u modelu i može se ukloniti iz modela. Odnos paralelnog i ukupnog koeficijenta diskriminacionog količnika može biti dragocen za otkrivanje tzv. supresorskog efekta određene početne varijable: ukoliko je paralelni koeficijent diskriminacionog količnika za određenu početnu varijablu relativno visok i negativan a ukupni koeficijent diskriminacionog količnika za tu varijablu relativno visok i pozitivan, po svoj prilici ta početna varijabla ima ulogu supresora u diskriminacionom modelu.[162]Supresorska varijabla u diskriminacionom modelu vrši „supresiju“ onog dela varijanse u nekoj od preostalih početnih varijabli koji je nevažan za razlikovanje subpopulacija i na taj način bitno doprinosi razlikovanju grupa.

Jedan od važnih aspekata tumačenja rezultata kanoničke diskriminacione analize kojem se obično poklanja malo pažnje u njenoj primeni jeste „relativna važnost“ ili „reprezentativnost“ kanoničkih diskriminacionih varijabli, tj. stepen u kojem je ukupna varijabilnost skupa početnih kvantitativnih varijabli predstavljena diskriminacionom funkcijom. Osnovni razlog za zanemarivanje ovog aspekta tumačenja rezultata diskriminacione analize, po našem mišljenju, posledica je načina na koji je ispis iz diskriminacione analize organizovan u najčešće korišćenim statističkim paketima u psihologiji, pre svega u programu SPSS. Naime, po analogiji sa analizom glavnih komponenata, u ispisu iz diskriminacione analize osnovna informacija o „važnosti“ kanoničke diskriminacione funkcije u navedenom smislu data je pomoću količnika:

$$\text{Procenat objašnjene varijanse} = 100 * \frac{\lambda_p}{\sum_{p=1}^{r} \lambda_p},$$

pri čemu je λ_p svojstvena vrednost matrice $\mathbf{W}^{-1}\mathbf{B}$, tj. (po algoritmu programa SPSS) matrice $\mathbf{L}^{-1}\mathbf{B}\mathbf{U}^{-1}$. Kolona u ispisu iz programa SPSS koja je naslovljena „*% of Variance*“ neposredno uz kolonu koja sadrži svojstvene vrednosti sugeriše na taj način, analogno analizi glavnih komponenata, da je ovaj procenat pokazatelj relativne važnosti diskriminacione funkcije. Međutim, ukoliko se na osnovu odnosa kanoničke korelacije i svojstvene vrednosti (koji smo prikazali u prethodnom delu teksta) svojstvena vrednost iskaže preko odgovarajućeg koeficijenta kanoničke korelacije:

$$\lambda_p = \frac{R_{cp}^2}{1 - R_{cp}^2},$$

postaje odmah jasno pravo značenje svojstvene vrednosti koja odgovara datoj kanoničkoj diskriminacionoj funkciji. Svojstvena vrednost, naime, ne ukazuje na relativnu važnost diskriminacione funkcije, već samo na to koliko puta je zajednička varijansa para kanoničkih varijabli (jedne kanoničke varijable iz skupa početnih kvantitativnih varijabli – kanoničke diskriminacione varijable – i druge kanoničke varijable iz skupa binarnih varijabli koje definišu grupnu pripadnost) veća od varijanse tog para kanoničkih varijabli koja im nije zajednička. U uslovima kada u analizi postoje više od dve grupe, svojstvena vrednost nije analogna onome što predstavlja svojstvena vrednost u analizi glavnih

[162] Supresorski efekat u kontekstu multiple regresione analize objašnjen je u odeljku **XIV.4**.

komponenata, te, prema tome, ne može biti pokazatelj relativne važnosti ili „reprezentativnosti“ diskriminacione funkcije.

Jedan od pokazatelja „reprezentativnosti“ koji bi na adekvatan način ukazivao na stepen u kojem je varijabilnost skupa početnih kvantitativnih varijabli predstavljena određenom diskriminacionom funkcijom mogla bi biti **p**roporcija **u**kupne **v**arijanse **p**očetnih **v**arijabli koja je obuhvaćena diskriminacionom funkcijom *p* ($PUVPV_p$):

$$\mathrm{PUVPV}p = \frac{\mathrm{f}_{jp}^{\mathrm{t}}\mathrm{f}_{jp}}{\mathrm{tr(R)}} = \frac{\sum_{j=1}^{\mathrm{m}} \mathrm{f}_{jp}}{\mathrm{m}}.$$

U izrazu za $PUVPV_p$ oznakom $\mathbf{f}_{jp}$ označen je vektor korelacija iz matrice strukture diskriminacione funkcije **F** ($\mathbf{F} = \mathbf{Z}^t\mathbf{K}\ n^{-1}$) koji odgovara funkciji *p*, f_{jp} predstavlja koeficijent linearne korelacije između diskriminacione funkcije *p* i početne varijable v_j, m je broj početnih varijabli, a tr(**R**) je trag matrice interkorelacija početnih varijabli.[163]

Uslovi za primenu kanoničke diskriminacione analize

1. multivarijaciona normalna raspodela početnih varijabli u svakoj od g subpopulacija; kanonička diskriminaciona analiza je relativno robustna na neispunjenost ovog uslova ako su grupe dovoljno velike; ono o čemu svakako treba voditi računa jeste eventualno postojanje multivarijacionih autlajera, tj. iznimaka u podacima;
2. homogenost matrica kovarijansi svih populacija; ovaj uslov je posebno važan kada su grupe male i/ili nejednake po veličini, jer tada ne možemo imati poverenja u ishod testova statističke značajnosti;
3. prisutnost barem dva ispitanika u svakoj od grupa;
5. linearna nezavisnost vektora početnih varijabli;
6. nezavisnost opservacija: ne sme se rezultat jednog istog ispitanika pojaviti više puta u jednoj ili u različitim grupama.

Testiranje homogenosti matrica kovarijansi subpopulacija – Boksov test

Jedna od pretpostavki za korišćenje diskriminacione analize (osim multivarijacione normalne raspodele kvantitativnih varijabli iz skupa *V* u svakoj od subpopulacija) jeste homogenost matrica kovarijansi varijabli iz skupa *V* za sve subpopulacije. Za testiranje zadovoljenosti ove pretpostavke Boksovim testom testiramo nultu hipotezu:

$$H_0: \mathbf{\Sigma}_k = \mathbf{\Sigma},\ \forall k.$$

Za testiranje H_0 o jednakosti matrica kovarijansi subpopulacija računa se, preko determinanti, tj. generalizovanih varijansi, M-statistik:

[163] Matrica strukture o kojoj je ovde reč ne odgovara, kao što smo već objasnili, matrici strukture koja se pod imenom **Structure matrix** pojavljuje u ispisu iz diskriminacione analize programa SPSS. Kako bi se u programu SPSS napravila ova matrica strukture, potrebno je sačuvati diskriminacione skorove i potom izračunati koeficijente linearne korelacije između početnih varijabli i sačuvanih diskriminacionih skorova.

$$M = \frac{\prod_{k=1}^{g} |S_k|^{(n_k-1)/2}}{|\bar{S}|^{(n-g)/2}}$$

Matrica **S** (sa crticom iznad) u obrascu za M-statistik predstavlja količnik **W** / (n - g), tj. ocenu opšte matrice kovarijansi, dok su matrice $\mathbf{S}_k$ ocene matrica kovarijansi u subpopulacijama. Na osnovu M-statistika može se izračunati približni F-statistik i testirati H_0 na osnovu približnog F-statistika.

Ako definišemo b na sledeći način: $b = [1 - c_1 - v_1 / v_2] / v_1$, pri čemu je

$$c_1 = \frac{(2m^2+3m-1)}{6(m+1)(g-1)}\left(\sum_{k=1}^{g}\frac{1}{n_k-1} - \frac{1}{n-g}\right)$$

$$c_2 = \frac{(m-1)(m+2)}{6(g-1)}\left(\sum_{k=1}^{g}\frac{1}{(n_k-1)^2} - \frac{1}{(n-g)^2}\right)$$

dok je

$v_1 = (1/2)m(m + 1)(g - 1)$ a $v_2 = (v_1 + 2) \,|\, c_2 - c_1^2|$, tada približno F možemo odrediti na osnovu sledećeg izraza:

$$-2b^*\ln M \approx F.$$

Približno F, koje je dobijeno na ovaj način, distribuira se po Snidikorovoj F-distribuciji sa $df_1 = v_1$ i $df_2 = v_2$. U slučaju da je $c_2 - c_1^2$ manje od 0, količnik $(-2b_1v_2\ln M)/(v_1 + 2b_1v_1\ln M)$, pri čemu je $b_1 = [1- c_1 - (2/ v_2)]/ v_2$), distribuira se približno kao F-statistik sa $df_1 = v_1$ i $df_2 = v_2$. Ukoliko je p-vrednost za F-statistik manja od kriterijuma za odbacivanje H_0 (npr. 0.05), F-statistik je statistički značajan i nulta hipoteza o homogenosti matrica kovarijansi može se odbaciti. Boksov test je veoma osetljiv i ne „reaguje" samo na nehomogenost matrica kovarijansi već i na odstupanje multivarijacione raspodele od normalne.

****

Kanonička diskriminaciona analiza se u programu SPSS izvodi odabirom menija **Analyze/Classify/Discriminant**, tako što se u dijalogu **Discriminant Analysis** u polje **Independents** unesu imena kvantitativnih varijabli iz skupa *V*, a u polje **Grouping Variable** ime kategoričke varijable *Q* i popune polja **Define Range** za tu varijablu. Korisno je pri tome u dijalogu **Discriminant Analysis: Statistics** unutar okvira **Descriptives** uključiti opcije **Means**, **Box's M** i **Univariate ANOVAs**.

Pod naslovom **Wilks' Lambda (U-statistic) and univariate F-ratio** dati su (samo ako je u okviru **Discriminant Analysis: Statistics** uključena opcija **Univariate ANOVAs**) Vilksove lambde (koje su ovde jednake $1 - \eta^2$ za svaku varijablu) i F-statistici identični onima iz procedure **One-Way ANOVA**.

U odeljku ispisa **Summary of Canonical Discriminant Functions** date su svojstvene vrednosti (kolona **Eigenvalue**), procenat od ukupne međugrupne varijanse koji otpada na datu svojstvenu vrednost (**% of variance)**, tj. procenat ukupne diskriminacione moći skupa varijabli koji otpada na datu diskriminacionu funkciju (količnik date svojstvene vrednosti i zbira svih svojstvenih vrednosti pomnožen sa 100), kanoničke korelacije (**Canonical Correlation**) i rezultati testiranja značajnosti diskriminacionih funkcija (kanoničkih korelacija) pomoću Bartletovog hi-kvadrat testa (kolone **Wilks' Lamda, Chi-square, df** i **Sig.**). Testiranje značajnosti kanoničkih korelacija u suštini je isto što i testiranje

$$H_0: \mu_{kp} = \mu_p \ \forall k,$$

(centroidi subpopulacija na diskriminacionoj funkciji *p* su jednaki)

korišćenjem F-statistika u proceduri **One-Way ANOVA**, u kojoj je diskriminaciona funkcija *p* zavisna varijabla, a *Q* nezavisna varijabla, tj. faktor.
U nastavku ispisa dati su standardizovani (**Standardized Canonical Discriminant function coefficients**), nestandardizovani koeficijenti (samo ukoliko je u dijalogu **Discriminant Analysis:Statistics** izabrana opcija **Unstandardized** u okviru **Function Coefficients**) za pravljenje linearnih kombinacija, tj. diskriminacionih funkcija (**Unstandardized Canonical Discriminant Function Coefficients**), matrica korelacija izvornih varijabli i diskriminacionih funkcija (**Structure matrix**) i centroidi grupa na pojedinim diskriminacionim funkcijama /**Canonical discriminant functions evaluated at group means (Functions at Group Centroids)**/. Standardizovani koeficijenti za diskriminacionu funkciju pokazuju parcijalni doprinos izvornih varijabli definisanju date diskriminacione funkcije. U delu ispisa koji nosi naslov **Test of Equality of Group Covariance Matrices Using Box's M** (**Box's Test of Equality of Covariance Matrices**) prikazani su rezultati testiranja homogenosti matrica kovarijansi subpopulacija.

****

Napomenimo ovde da je jednofaktorijalnu analizu varijanse sa jednim neponovljenim faktorom moguće napraviti i diskriminacionom analizom, odabirom menija **Analyze/Classify/Discriminant**, tj. napraviti je kao specijalni slučaj diskriminacione analize sa jednom kvantitativnom varijablom (polje **Independents!**) i jednom kategoričkom varijablom (polje **Grouping Variable**). Korisno je pri tome u dijalogu **Discriminant Analysis:Statistics** unutar okvira **Descriptives** uključiti opcije **Means** i **Univariate ANOVAs**. Ispis iz diskriminacione analize u kojoj je kategorička varijabla **Grouping Variable** a jedna zavisna kvantitativna varijabla predstavlja skup **Independent** (!) – ima sve standardne komponente ispisa iz diskriminacione analize:
– Na početku ispisa date su veličine grupa (**Number of cases by group**) i (ukoliko je u okviru **Descriptives** uključena opcija **Means**) aritmetičke sredine i standardne devijacije pojedinih grupa.
– Naravno, ovde je koeficijent za kanoničku diskriminacionu funkciju (**Standardized canonical discriminant function coefficients**) jednak jedinici (jer samo jedna varijabla čini skup varijabli čijom linearnom kombinacijom treba dobiti diskriminacionu funkciju), a jedinici je jednaka i korelacija izvorne varijable sa samom sobom (tj. sa 'diskriminacionom funkcijom', jedinim elementom matrice strukture – **Structure matrix**). Dakle, diskriminaciona funkcija je u ovom slučaju identična određenoj transformaciji izvorne kvantitativne varijable (zavisne varijable iz jednosmerne analize varijanse). Matrice

kovarijansi su zapravo ukupne ili grupne ocene varijanse između i unutar grupa na toj varijabli.
Pod naslovom **Wilks' Lambda (U-statistic) and univariate F-ratio** dati su (samo ako je u okviru **Discriminant Analysis:Statistics** uključena opcija **Univariate ANOVAs**) Vilksova lambda (koja je ovde jednaka 1 - η^2) i F-statistik identičan onom iz procedure **One-Way ANOVA**.
U odeljku ispisa **Canonical Discriminant Functions** pod kanoničkom korelacijom (**Canonical Corr**) krije se zapravo η**-**koeficijent, a testiranje njegove značajnosti pomoću hi-kvadrat statistika (kolone **Wilks' Lamda, Chi-square, df** i **Sig.**) isto je što i testiranje H_0 korišćenjem F-statistika u proceduri **One-Way ANOVA**.
S druge strane, ukoliko bismo jednofaktorijalnu analizu varijanse napravili sa diskriminacionim funkcijama, dobijenim u diskriminacionoj analizi, kao zavisnim varijablama, onda bi eta-koeficijenti bili jednaki kanoničkim korelacijama iz diskriminacione analize.

Primer 1: Da li se u pogledu profila na poznatom instrumentu Minesota multifazni inventar ličnosti astmatičari razlikuju od ostalih vrsta bolesnika? Na uzorku od 256 pacijenata, od kojih su polovina bili astmatičari (u varijabli grupa označeni jedinicom), a druga polovina psihosomatski bolesnici koji nemaju astmu (u varijabli grupa označeni dvojkom), pomoću upitnika MMPI izmerene su sledeće osobine ličnosti:

MMPIHs	HIPOHONDRIČNOST
MMPID	DEPRESIJA
MMPIHy	HISTERIJA
MMPIPd	PSIHOPATSKA DEVIJANTNOST
MMPIPa	PARANOIDNOST
MMPIPt	PSIHASTENIČNOST
MMPISc	SHIZOFRENIČNOST
MMPIMa	HIPOMANIČNOST
MMPIL	SKLONOST LAGANJU NA UPITNIKU
MMPIF	SKALA FALSIFIKOVANJA
MMPIK	KONTROLNA SKALA

Drugačije iskazano pitanje iz ovog primera moglo bi glasiti: postoji li složaj ovih osobina ličnosti (diskriminaciona funkcija) koji dobro razlikuje astmatičare od ostalih psihosomatskih bolesnika?

Ispis ove analize urađen je u verziji 6.1 programa SPSS.

```
Following variables will be created upon successful completion of the procedure:

  Name         Label
--------     ----------------------------------------
DIS1_1   --- Function 1 for analysis 1



 - - - - - - - -   D I S C R I M I N A N T   A N A L Y S I S   - - - - - - - -

On groups defined by GRUPA

           256 (Unweighted) cases were processed.
             0 of these were excluded from the analysis.
           256 (Unweighted) cases will be used in the analysis.
```

Number of cases by group

GRUPA	Number of cases Unweighted	Weighted	Label
1	128	128.0	
2	128	128.0	
Total	256	256.0	

Group means

GRUPA	MMPIL	MMPIF	MMPIK	MMPIHS
1	3.95313	4.28125	11.89063	22.37500
2	3.71094	3.90625	12.03125	16.40625
Total	3.83203	4.09375	11.96094	19.39063

GRUPA	MMPID	MMPIHY	MMPIPD	MMPIPA
1	22.18750	21.03125	15.56250	12.25000
2	14.70313	16.79688	13.76563	8.41406
Total	18.44531	18.91406	14.66406	10.33203

GRUPA	MMPIPT	MMPISC	MMPIMA
1	25.45313	25.82813	13.18750
2	21.78125	24.69531	14.85938
Total	23.61719	25.26172	14.02344

Group standard deviations

GRUPA	MMPIL	MMPIF	MMPIK	MMPIHS
1	1.67830	2.17696	2.99668	4.63638
2	1.69368	1.85471	1.97997	2.28117
Total	1.68706	2.02702	2.53571	4.71582

GRUPA	MMPID	MMPIHY	MMPIPD	MMPIPA
1	5.25829	4.17329	4.05037	3.51823
2	3.26605	2.73893	3.06744	2.39266
Total	5.75692	4.11221	3.69690	3.56496

GRUPA	MMPIPT	MMPISC	MMPIMA
1	4.14284	4.42454	1.99112
2	2.84767	3.16616	2.60748
Total	3.99631	3.88131	2.46215

Wilks' Lambda (U-statistic) and univariate F-ratio
with 1 and 254 degrees of freedom

Variable	Wilks' Lambda	F	Significance
MMPIL	.99483	1.3206	.2516
MMPIF	.99141	2.2007	.1392

```
  MMPIK          .99923             .1962       .6582
  MMPIHS         .59794           170.7925       .0000
  MMPID          .57580           187.1256       .0000
  MMPIHY         .73389            92.1031       .0000
  MMPIPD         .94071            16.0095       .0001
  MMPIPA         .70941           104.0417       .0000
  MMPIPT         .78812            68.2872       .0000
  MMPISC         .97862             5.5490       .0193
  MMPIMA         .88428            33.2401       .0000



 - - - - - - - -   D I S C R I M I N A N T   A N A L Y S I S   - - - - - -


On groups defined by GRUPA

Analysis number       1
Direct method:  all variables passing the tolerance test are entered.

     Minimum tolerance level.................  .00100

Canonical Discriminant Functions
     Maximum number of functions.............       1
     Minimum cumulative percent of variance...  100.00
     Maximum significance of Wilks' Lambda....  1.0000
Prior probability for each group is  .50000



                       Canonical Discriminant Functions

                 Pct of   Cum  Canonical  After  Wilks'
 Fcn Eigenvalue Variance  Pct     Corr      Fcn  Lambda Chi-sq. df  Sig

                                         :    0 .375092  243.675 11 .0000
  1*    1.6660  100.00  100.00    .7905 :

   * Marks the 1 canonical discriminant functions remaining in the analysis.
Standardized canonical discriminant function coefficients

             Func  1

MMPIL         .30705
MMPIF        -.48041
MMPIK        -.04250
MMPIHS        .54815
MMPID         .50660
MMPIHY        .13860
MMPIPD       -.02220
MMPIPA        .42202
MMPIPT        .04194
MMPISC       -.30082
MMPIMA       -.28347

Structure matrix:

Pooled within-groups correlations between discriminating variables
                                   and canonical discriminant functions
(Variables ordered by size of correlation within function)

             Func  1

MMPID         .66498
MMPIHS        .63530
```

```
MMPIPA         .49585
MMPIHY         .46653
MMPIPT         .40171
MMPIMA        -.28027
MMPIPD         .19451
MMPISC         .11451
MMPIF          .07212
MMPIL          .05586
MMPIK         -.02153


Unstandardized canonical discriminant function coefficients

                  Func  1

MMPIL             .1821147
MMPIF            -.2375579
MMPIK            -.0167355
MMPIHS            .1500240
MMPID             .1157417
MMPIHY            .0392669
MMPIPD    -6.17839512E-03
MMPIPA            .1402749
MMPIPT            .0117990
MMPISC           -.0781924
MMPIMA           -.1221926
(Constant)      -3.2603894

Canonical discriminant functions evaluated at group means (group centroids)

   Group      Func   1

       1       1.28569
       2      -1.28569

Test of Equality of Group Covariance Matrices Using Box's M

   The ranks and natural logarithms of determinants printed are those
   of the group covariance matrices.
     Group Label                Rank   Log Determinant
         1                        11       23.849526
         2                        11       16.063686
      Pooled within-groups
      covariance matrix           11       21.685263

       Box's M  Approximate F  Degrees of freedom  Significance
     439.07879      6.35155     66,    205711.7         .0000
```

Na početku ispisa program nas obaveštava o tome da je u aktivnom fajlu sa podacima kreirana varijabla **DIS1_1** u kojoj su pohranjeni skorovi ispitanika na diskriminacionoj funkciji. Zatim su pod naslovima **Group means** i **Group standard deviations** dati proseci i standardne devijacije na pojedinačnim zavisnim varijablama za pojedine grupe i uzorak u celini. Boksov M-statistik od 439.08 (prikazan u ispisu pod naslovom **Test of Equality of Group Covariance Matrices Using Box's M**) statistički je značajan i pokazuje da su matrice kovarijansi između grupa heterogene. Ipak, budući da matrice kovarijansi po grupama nisu drastično različite, smatraćemo da je Boksov test u ovom slučaju bio preosetljiv. Pored toga, kanonička diskriminaciona analiza je u izvesnom stepenu robustna na heterogenost matrica kovarijansi ako su grupe dovoljno velike i jednake veličine, što je u ovom primeru slučaj.

Kao što vidimo, u delu ispisa pod naslovom **Canonical Discriminant Functions**, subpopulacija astmatičara se po svom profilu ličnosti razlikuje od neastmatičara. Na to ukazuje statistički značajan hi-kvadrat koji iznosi 243.675. To znači da između subpopulacijske pripadnosti (astmatičari – neastmatičari), sa jedne strane, i skupa osobina ličnosti uzetih zajedno, sa druge strane, postoji povezanost u populaciji. Pod naslovima **Standardized canonical discriminant function coefficients** i **Unstandardized canonical discriminant function coefficients** možemo videti kako nastaje diskriminaciona funkcija na osnovu standardizovanih, odnosno nestandardizovanih rezultata na izvornim osobinama. Pored imena svake varijable dat je koeficijent kojim treba pomnožiti rezultat na datoj varijabli prilikom pravljenja linearne kombinacije, tj. diskriminacione funkcije.

Ipak, budući da su koeficijenti za diskriminacionu funkciju osetljivi na međusobne korelacije među sastavnim varijablama, za tumačenje sadržaja diskriminacione funkcije (tj. za definisanje složaja osobina koji najbolje diferencira astmatičare od ostalih bolesnika) pogledaćemo i koeficijente korelacije izvornih varijabli i diksriminacione funkcije koji su dati pod naslovom **Structure matrix**. Vidimo da je reč o složaju koji grade pre svega depresivnost, hipohondričnost, paranoidnost, histeričnost i psihasteničnost. Ovu strukturu bismo možda najbolje opisali kao generalni neurotizam depresivno-anksioznog tipa. Budući da je, kao što vidimo pod naslovom **Canonical discriminant functions evaluated at group means (group centroids)**, centroid (tj. prosek astmatičara na dobijenoj diskriminatnoj funkciji) pozitivan (tj. veći), možemo zaključiti i da su astmatičari generalno neurotičniji od ostalih psihosomatskih bolesnika. Ukoliko bismo sa dobijenom diskriminacionom funkcijom kao zavisnom varijablom i grupnom pripadnošću kao nezavisnom varijablom napravili jednofaktorijalnu univarijacionu analizu varijanse, dobili bismo sledeći ispis:

```
 - - Analysis of Variance - -

Dependent Variable   DIS1_1     Function 1 for analysis 1
      By levels of   GRUPA

     Value  Label                     Mean    Std Dev  Sum of Sq    Cases

         1                       1.2856889  1.2229464 189.940927      128
         2                      -1.2856889   .7102128 64.0590734      128
                                ----------------------------------------
Within Groups Total             -4.441E-16  1.0000000 254.000000      256

                           Sum of                   Mean
Source                    Squares        d.f.      Square          F       Sig.

Between Groups            423.1670          1      423.1670   423.1670    .0000

Within Groups             254.0000        254        1.0000

                       Eta =  .7905    Eta Squared =  .6249
```

Uočimo da je eta-koeficijent iz ove analize jednak kanoničkoj korelaciji za diskriminacionu funkciju.

.

<u>Primer 2</u>: Da li različiti aspekti subjektivnog doživljaja vremena (sukcesija, integracija, vremenske distinkcije, usmerenost na ciljeve i usredsređenost na sadašnjost) razlikuju osobe sa dijagnozom PTSP-a (posttraumatskog stresnog poremećaja), osobe sa drugim psihijatrijskim poremećajima i tzv. normalne osobe (osobe bez psihijatrijske dijagnoze)? I

ako ih razlikuju, kakva je priroda tih razlika? Drugačije rečeno: da li osobe sa dijagnozom PTSP-a, generalno gledano, imaju drugačiji subjektivni doživljaj vremena od osoba koje imaju druge psihijatrijske poremećaje i od tzv. normalnih osoba?
U kanoničku diskriminacionu analizu u ovom slučaju uključeni su rezultati dobijeni na slučajnom uzorku od 227 odraslih osoba (pacijenata psihijatrijskih ustanova i iz opšte populacije) sa upitnika koji ispituju pet pomenutih aspekata subjektivnog doživljaja vremena. Početne kvantitativne varijable nose sledeća imena: **SUCCESSION, INTEGRATION, Temporal distinction iz TII_veci skor integrisanost, Goal Directedness iz TII_veci skor veca inetgrisanost** i **Present concentration.** Kategorička varijabla koja definiše grupu (**pts_dr_z**) ima tri kategorije: **PTSP**, **drugi poremećaji** i **zdrav**.
U ovom primeru biće razmatrani samo delovi ispisa koji su specifični za situaciju kada ima više diskriminacionih funkcija, tj. više od dve grupe. Delovi ispisa iz programa SPSS (verzija 23) izgledaju ovako:

Discriminant

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		227	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		227	100.0

Group Statistics

Grupa: ptsp, drugi porem. zdrav		Mean	Std. Deviation	Valid N (listwise) Unweighted	Valid N (listwise) Weighted
zdrav	SUCCESSION	3.5852	.63392	135	135.000
	INTEGRATION	3.6263	.43052	135	135.000
	Temporal distinction iz TII_veci skor integrisanost	4.6546	.64523	135	135.000
	Goal Directedness iz TII_veci skor veca inetgrisanost	4.2531	.74085	135	135.000
	Present concentration	2.6252	1.10560	135	135.000

drugi porem.	SUCCESSION	3.0722	.66636	34	34.000
	INTEGRATION	3.5327	.47222	34	34.000
	Temporal distinction iz TII_veci skor integrisanost	4.1728	.81708	34	34.000
	Goal Directedness iz TII_veci skor veca inetgrisanost	3.9951	.92704	34	34.000
	Present concentration	2.8412	1.11928	34	34.000
PTSP	SUCCESSION	2.5674	.57730	58	58.000
	INTEGRATION	3.3551	.44579	58	58.000
	Temporal distinction iz TII_veci skor integrisanost	3.7241	.75646	58	58.000
	Goal Directedness iz TII_veci skor veca inetgrisanost	3.1839	.78341	58	58.000
	Present concentration	3.7828	1.12844	58	58.000
Total	SUCCESSION	3.2483	.76066	227	227.000
	INTEGRATION	3.5430	.45366	227	227.000
	Temporal distinction iz TII_veci skor integrisanost	4.3447	.80580	227	227.000
	Goal Directedness iz TII_veci skor veca inetgrisanost	3.9413	.90079	227	227.000
	Present concentration	2.9533	1.21310	227	227.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
SUCCESSION	.669	55.381	2	224	.000
INTEGRATION	.936	7.693	2	224	.001
Temporal distinction iz TII_veci skor integrisanost	.753	36.822	2	224	.000
Goal Directedness iz TII_veci skor veca inetgrisanost	.746	38.039	2	224	.000
Present concentration	.835	22.126	2	224	.000

Box's Test of Equality of Covariance Matrices

Log Determinants

Grupa: ptsp, drugi porem.		
zdrav	Rank	Log Determinant
zdrav	5	-5.415
drugi porem.	5	-5.197
PTSP	5	-4.725
Pooled within-groups	5	-5.074

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

Box's M		29.930
F	Approx.	.952
	df1	30
	df2	35492.816
	Sig.	.541

Tests null hypothesis of equal population covariance matrices.

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.587[a]	93.7	93.7	.608
2	.040[a]	6.3	100.0	.195

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.606	111.137	10	.000
2	.962	8.620	4	.071

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
SUCCESSION	.631	.718
INTEGRATION	-.146	.073
Temporal distinction iz TII_veci skor integrisanost	.331	.542
Goal Directedness iz TII_veci skor veca inetgrisanost	.266	-.985
Present concentration	-.044	.620

Structure Matrix

	Function	
	1	2
SUCCESSION	.917*	.146
Goal Directedness iz TII_veci skor veca inetgrisanost	.749*	-.520
Temporal distinction iz TII_veci skor integrisanost	.747*	.152
Present concentration	-.566*	.497
INTEGRATION	.341*	-.120

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Functions at Group Centroids

Grupa: ptsp, drugi porem. zdrav	Function	
	1	2
zdrav	.577	.064
drugi porem.	-.233	-.467
PTSP	-1.208	.124

Unstandardized canonical discriminant functions evaluated at group means

Correlations

Correlations

		Discriminant Scores from Function 1 for Analysis 1
SUCCESSION	Pearson Correlation	.945**
	Sig. (2-tailed)	.000
	N	227
INTEGRATION	Pearson Correlation	.415**
	Sig. (2-tailed)	.000
	N	227
Temporal distinction iz TII_veci skor integrisanost	Pearson Correlation	.817**
	Sig. (2-tailed)	.000
	N	227
Goal Directedness iz TII_veci skor veca inetgrisanost	Pearson Correlation	.815**
	Sig. (2-tailed)	.000
	N	227
Present concentration	Pearson Correlation	-.651**
	Sig. (2-tailed)	.000
	N	227

**. Correlation is significant at the 0.01 level (2-tailed).

Ovoga puta Boksov test ne ukazuje na heterogenost matrica kovarijansi u subpopulacijama ($p = 0.541$). Budući da ima pet početnih kvantitativnih varijabli a tri grupe, bilo je moguće definisati dve diskriminacione funkcije. Prva kanonička korelacija iznosi 0.608 i statistički je značajna (p u testu značajnosti prve diskriminacione funkcije manje je od 0.001). Druga kanonička korelacija iznosi 0.195 i nije statistički značajna ($p = 0.071$), te drugu diskriminacionu funkciju ne uzimamo u obzir u tumačenju.
Na osnovu standardizovanih koeficijenata za prvu diskriminacionu funkciju vidimo da u definisanju diskriminacione funkcije najviše učestvuju aspekti sukcesije, vremenske distinkcije i usmerenosti na ciljeve. S obzirom na znatne razlike između standardizovanih koeficijenata i koeficijenata strukture za pojedine početne varijable, radi jednostavnijeg tumačenja, izračunali smo paralelne i ukupne koeficijente diskriminacionog količnika za prvu funkciju:

	Koeficijenti diskriminacionog količnika	
	Paralelni	Ukupni

SUCCESSION	.58	.55
INTEGRATION	-.05	.13
Temporal distinction iz TII_veci skor integrisanost	.25	.30
Goal Directedness iz TII_veci skor veca inetgrisanost	.20	.22
Present concentration	.02	.04

Dakle, ključni aspekti subjektivnog doživljaja vremena koji razlikuju ove tri populacije su sukcesija, vremenska distinkcija i usmerenost na ciljeve. Da li se iza ovih aspekata subjektivnog doživljaja vremena krije neka prepoznatljiva latentna psihološka struktura, može proceniti istraživač koji se bavi ovom oblašću. Uočimo da aspekt integracije ima negativan paralelni koeficijent diskriminacionog količnika a pozitivan ukupni koeficijent diskriminacionog količnika, ali su njihove vrednosti jako niske i zbog toga nećemo ulaziti u razmatranje moguće uloge integracije kao supresora. Pre bi se moglo reći da je nizak negativan paralelni koeficijent posledica redundantnosti te varijable u odnosu na ostale početne varijable. I zaista, koeficijent multiple korelacije varijable integracije sa preostale četiri varijable iznosi 0.59. Ukoliko se varijabla integracija izostavi iz diskriminacionog modela, prva kanonička korelacija ostaje praktično nepromenjena i iznosi 0.605. Dakle, ova varijabla može se izostaviti iz modela.
Pregledom tabele **Functions at Group Centroids**, možemo uočiti da prva funkcija zapravo dobro razlikuje grupu sa PTSP-om od preostalih dveju grupa: subjektivni doživljaj vremena najviše je poremećen (centroid = -1.208) kod osoba sa PTSP-om, znatno manje kod osoba sa drugim psihijatrijskim dijagnozama (centroid = -0.233), a najmanje kod tzv. zdravih osoba. Dakle, budući da prva funkcija dobro razlikuje grupu sa PTSP-om ne samo od zdravih osoba nego i od osoba sa drugim psihijatrijskim poremećajima, moglo bi se reći da su poremećeni aspekti sukcesije, vremenske distinkcije i usmerenosti na ciljeve specifična karakteristika PTSP-a.
****

XIX. 2. Kvazikanonička diskriminaciona analiza

Kvazikanoničku diskriminacionu analizu formalno su definisali Štalec i Momirović (1984), a algoritam i program za računarsku implementaciju ove metode napravljen je ubrzo po njenom formalnom definisanju (Dobrić & Momirović, 1984). Kvazikanonička diskriminaciona analiza formalno je definisana kao poseban slučaj opštijeg modela kvazikanoničke korelacione analize ili kanoničke analize kovarijansi koji su prethodno predložili Momirović, Dobrić i Karaman (1983). Kvazikanonička korelaciona analiza dva skupa izvornih kvantitativnih varijabli zasniva se na određivanju linearnih kombinacija varijabli iz jednog i drugog skupa, tzv. kvazikanoničkih varijabli, koje će, ako su isto indeksirane, imati maksimalnu kovarijansu.[164] Pri tome, vektori, u kojima su koeficijenti kojima se grade ove linearne kombinacije, jesu normirani, tj. definisani su tako da zbir

[164]Termine *početne* ili *izvorne* varijable koristimo ovde kao ime za varijable bez obzira na njihov oblik, tj. kao početnu tačku i skalu. Prema tome, početnim ili izvornim varijablama nazivaćemo i nestandardizovane i standardizovane varijable.

kvadrata koeficijenata za određenu linearnu kombinaciju bude jednak jedinici. Dakle, za razliku od „klasičnog“ kanoničkog korelacionog modela u kojem se linearne kombinacije početnih varijabli (kanoničke varijable) dva skupa grade tako da isto indeksirane kanoničke varijable budu maksimalno *korelisane*, u kvazikanoničkoj korelacionoj analizi nastoji se na tome da isto indeksirane linearne kombinacije varijabli iz dva skupa imaju maksimalnu *kovarijansu*.

Postupak izvođenja osnovnog oblika kvazikanoničke diskriminacione analize, kada se ova metoda definiše kao poseban slučaj opštijeg modela kvazikanoničke korelacione analize, može se formalno predstaviti na sledeći način:

$$\mathbf{Z}\mathbf{v}_k = \mathbf{t}_k;\ \mathbf{G}\mathbf{w}_k = \mathbf{h}_k \,\big|\, \phi_k = \mathbf{t}_k^t\mathbf{h}_k n^{-1} = \text{maximum},\ \mathbf{v}_k^t\mathbf{v}_q = \mathbf{w}_k^t\mathbf{w}_q = \delta_{kq}$$
$$\mathbf{v}_k = ?;\ \mathbf{w}_k = ?$$

Pri tome je:
$\mathbf{Z} = (\mathbf{X}\text{-}\mathbf{\Pi X})\mathbf{D}^{-1}$ matrica početnih kvantitativnih varijabli u standardizovanom obliku;[165]
$\mathbf{G} = \mathbf{S}(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\ \mathbf{Z} = \mathbf{PZ} = \mathbf{SB}$ je matrica varijabli iz $\mathbf{Z}$ projektovanih u prostor definisan binarnim vektorima koji definišu subpopulacijsku pripadnost entiteta; drugim rečima, u matrici $\mathbf{G}$ kao „podatak“ na datoj varijabli služi aritmetička sredina te varijable za onu grupu kojoj pripada dati entitet;[166]
$\phi_k = \mathbf{t}_k^t\mathbf{h}_k n^{-1}$ je kovarijansa varijabli koje će nastati linearnim kombinovanjem početnih varijabli iz matrice $\mathbf{Z}$ ($\mathbf{t}_k$) i varijabli iz matrice $\mathbf{G}$ ($\mathbf{h}_k$), a δ_{kq} Kronekerova delta.
Problem koji treba rešiti jeste, dakle, određivanje vektora $\mathbf{v}_k$ i $\mathbf{w}_k$ u kojima su koeficijenti za kvazikanoničke varijable $\mathbf{k}_k$ i $\mathbf{l}_k$, tj. linearne kombinacije skupova početnih standardizovanih varijabli (varijabli iz matrice $\mathbf{Z}$), i linearne kombinacije varijabli iz matrice $\mathbf{G}$, tako da kovarijanse linearnih kombinacija sa istim indeksom budu maksimalne ($\phi_k = \mathbf{t}_k^t\mathbf{h}_k n^{-1}$= maximum), vektori $\mathbf{v}_k$ i $\mathbf{w}_k$ normirani ($\delta_{kq} = 1$, $k = q$) i međusobno ortogonalni ($\delta_{kq} = 0$, $k \neq q$).

2. Postupak izvođenja:

Za određivanje funkcije koju treba maksimizovati polazimo od sledećih jednakosti:
a) $\phi_k = \mathbf{t}_k^t\mathbf{h}_k n^{-1} = \mathbf{v}_k^t\mathbf{Z}^t\mathbf{G}\mathbf{w}_k n^{-1} = \mathbf{v}_k^t\mathbf{A}\mathbf{w}_k$;
b) $\mathbf{A} = \mathbf{Z}^t\mathbf{PZ}n^{-1} = \mathbf{Z}^t\mathbf{G}n^{-1}$;
c) $\mathbf{A} = \mathbf{Z}^t\mathbf{P}^t\mathbf{PZ}n^{-1} = \mathbf{G}^t\mathbf{G}\, n^{-1} = \mathbf{A}$.
Za k = 1, funkcija f čije uslovne ekstreme tražimo je, prema tome:

$$f(\mathbf{v}_k, \mathbf{w}_k, \lambda_k, \eta_k) = \mathbf{v}_k^t\mathbf{A}\mathbf{w}_k - (1/2)\lambda_k(\mathbf{v}_k^t\mathbf{v}_k - 1) - (1/2)\eta_k(\mathbf{w}_k^t\mathbf{w}_k - 1),$$

pri čemu su λ_k, η_k Lagranžovi multiplikatori.

---

[165] Matrica $\mathbf{\Pi}$ je centroidni projektor, $\mathbf{\Pi} = \mathbf{e}(\mathbf{e}^t\mathbf{e})^{-1}\mathbf{e}^t$, pri čemu je e kolonski vektor reda nx1 čiji svi elementi su jedinice.
[166] Kao što smo u izlaganju modela diskriminacione analize već objasnili, $\mathbf{S}$ je selektorska ili indikatorska matrica koja sadrži binarne varijable koje definišu subpopulacijsku pripadnost entiteta; $\mathbf{P} = \mathbf{S}(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t$ je projektorska matrica za prostor koji definišu varijable iz $\mathbf{S}$, a $\mathbf{B} = (\mathbf{S}^t\mathbf{S})^{-1}\ \mathbf{S}^t\mathbf{Z}$ je regresiona matrica koja sadrži aritmetičke sredine grupa na varijablama iz $\mathbf{Z}$.

Diferenciranjem funkcije f po vektoru $\mathbf{x}_k$:

$$\partial f/\partial \mathbf{v}_k = \mathbf{A}\mathbf{w}_k - \lambda_k \mathbf{v}_k,$$

a zatim po vektoru $\mathbf{w}_k$:

$$\partial f/\partial \mathbf{w}_k = \mathbf{A}\mathbf{v}_k - \eta_k \mathbf{w}_k,$$

i izjednačavanjem dobijenih izvoda sa nulom i njihovim sređivanjem, dobijamo sledeće dve jednakosti:

$$\mathbf{A}\mathbf{w}_k = \lambda_k \mathbf{v}_k$$

i

$$\mathbf{A}\mathbf{v}_k = \eta_k \mathbf{w}_k.$$

Budući da je $\mathbf{A} = \mathbf{A}^t$, množenjem prve jednakosti sa $\mathbf{v}_k^t$, a druge sa $\mathbf{w}_k^t$ iz uslova $\mathbf{v}_k^t\mathbf{v}_k = 1$ i $\mathbf{w}_k^t\mathbf{w}_k = 1$, sledi da je $\lambda_k = \eta_k$ i da je $\mathbf{v}_k = \mathbf{w}_k$. Odatle je $\phi_k = \mathbf{v}_k^t\mathbf{A}\mathbf{w}_k = \lambda_k = \mathbf{w}_k^t\mathbf{A}\mathbf{v}_k = \eta_k$. Prema tome, rešavanje postavljenog problema svodi se na dekompoziciju singularnim vrednostima matrice **A**. Matrica **A**, kao matrica kovarijansi, u opštem slučaju je pozitivno semidefinitna ($\mathbf{x}^t\mathbf{A}\mathbf{x} \geq 0,\ \forall \mathbf{x}$). Stoga su singularne vrednosti ove matrice jednake njenim pozitivnim svojstvenim vrednostima, te je dekompozicija singularnim vrednostima ove matrice isto što i njena spektralna dekompozicija (cf. Schott, 1997, str. 135). Prema tome, rešavanje problema singularnih vrednosti svodi se u ovom slučaju na rešavanje problema svojstvenih vrednosti:

$$(\mathbf{A} - \phi_k)\mathbf{v}_k = \mathbf{0}.$$

Matrica **A**, $\mathbf{A} = \mathbf{Z}^t\mathbf{P}\mathbf{Z}n^{-1} = \mathbf{Z}^t\mathbf{G}\ n^{-1}$, predstavlja matricu kovarijansi početnih varijabli iz matrice **Z** i varijabli iz matrice **G**. Ali, budući da je $\mathbf{P}^t\mathbf{P} = \mathbf{P}$, to je $\mathbf{Z}^t\mathbf{P}^t\mathbf{P}\mathbf{Z}n^{-1} = \mathbf{G}^t\mathbf{G}\ n^{-1} = \mathbf{Z}^t\mathbf{P}\mathbf{Z}n^{-1} = \mathbf{A}$, a matrica **A** je istovremeno i matrica kovarijansi varijabli iz **G**, tj. varijabli koje kao „podatke“ za određeni entitet sadrže aritmetičke sredine varijabli u grupi kojoj entitet pripada.

Budući da je rang matrice **A** jednak s, [s = minimum(m, g - 1), pri čemu je m broj početnih kvantitativnih varijabli, a g broj grupa], broj nenultih svojstvenih vrednosti ove matrice i, prema tome, broj parova linearnih kombinacija (kvazikanoničkih diskriminacionih varijabli) za varijable iz matrica **Z** i **G** jednak je broju s.

Matrica kvazikanoničkih diskriminacionih funkcija (ili kvazikanoničkih diskriminacionih varijabli) koje nastaju kao linearne kombinacije varijabli iz matrice **Z**, u oznaci **T**, reda nxs, dobija se na sledeći način:

$$\mathbf{T} = \mathbf{Z}\mathbf{V}.$$

Pri tome je **V**, $[\mathbf{V} = (\mathbf{v}_k),\ k = 1,\ldots, s]$, matrica svojstvenih vektora koji odgovaraju nenultim svojstvenim vrednostima matrice **A**. Matrica kvazikanoničkih diskriminacionih varijabli za varijable iz matrice **G**, u oznaci **H**, reda nxs, može se dobiti na sledeći način:

$$\mathbf{H} = \mathbf{B}\mathbf{V} = \mathbf{P}\mathbf{Z}\mathbf{V}.$$

Kovarijanse i varijanse kvazikanoničkih diskriminacionih varijabli iz **T**, tj. linearnih kombinacija početnih varijabli iz matrice **Z**, nalaze se u matrici **Ω**:

$$\mathbf{\Omega} = \mathbf{T}^t\mathbf{T}n^{-1} = \mathbf{V}^t\mathbf{R}\mathbf{V}.$$

Matrica kovarijansi kvazikanoničkih diskriminacionih varijabli iz **H**, tj. linearnih kombinacija varijabli iz **G**, u oznaci **Φ**, jednaka je dijagonalnoj matrici svojstvenih vrednosti matrice **A**:

$$\mathbf{\Phi} = \mathbf{H}^t\mathbf{H}n^{-1} = \mathbf{V}^t\mathbf{A}\mathbf{V}$$

Prema tome, različite kvazikanoničke diskriminacione varijable iz **H** međusobno su nekorelisane. Kvazikanoničke diskriminacione varijable koje se nalaze u kolonama matrice **T**, nisu nužno nekorelisane, jer matrica njihovih kovarijansi nije, u opštem slučaju, dijagonalna matrica. Naime, vektori iz matrice **V** nisu svojstveni vektori matrice **R**, pa ni matrica $\mathbf{V}^t\mathbf{R}\mathbf{V}$ nije nužno dijagonalna.
Matrica unakrsnih kovarijansi kvazikanoničkih diskriminacionih varijabli iz **T** i kvazikanoničkih diskriminacionih varijabli iz **H** jednaka je matrici kovarijansi kvazikanoničkih diskriminacionih varijabli iz **H**, tj. dijagonalnoj matrici čiji dijagonalni elementi predstavljaju svojstvene vrednosti matrice **A**:

$$\mathbf{T}^t\mathbf{H}n^{-1} = \mathbf{V}^t\mathbf{Z}^t\mathbf{G}\mathbf{V}n^{-1} = \mathbf{V}^t\mathbf{A}\mathbf{V} = \mathbf{\Phi}$$

Kvazikanoničke diskriminacione varijable iz **T** i **H**, budući da nastaju kao linearne kombinacije varijabli čije su aritmetičke sredine jednake nuli, imaju takođe aritmetičke sredine jednake nuli. Da bismo ih standardizovali, potrebno je, dakle, samo pomnožiti matrice **T** i **H** odgovarajućim dijagonalnim matricama u kojima se nalaze standardne devijacije kvazikanoničkih varijabli, tj. matricama $\mathbf{\Sigma}^{-1}$ (pri čemu je $\mathbf{\Sigma}^2 = \mathrm{diag}\,\mathbf{\Omega}$) i $\mathbf{\Phi}^{-1/2}$, tim redom. Prema tome, u kolonama matrice **Ψ**:

$$\mathbf{\Psi} = \mathbf{T}\mathbf{\Sigma}^{-1} = \mathbf{Z}\mathbf{V}\mathbf{\Sigma}^{-1}$$

i matrice **Θ**:

$$\mathbf{\Theta} = \mathbf{H}\mathbf{\Phi}^{-1/2} = \mathbf{P}\mathbf{Z}\mathbf{V}\mathbf{\Phi}^{-1/2} = \mathbf{G}\mathbf{V}\mathbf{\Phi}^{-1/2}$$

nalaze se standardizovane kvazikanoničke funkcije varijabli iz **Z** i **G**, tim redom.
Kvazikanonički koeficijenti korelacija između standardizovanih kvazikanoničkih varijabli iz **Z** i **G**, u oznaci R_k, jednaki su dijagonalnim elementima dijagonalne matrice **Γ**:

$$\mathbf{\Gamma} = \mathbf{\Psi}^t\mathbf{\Theta}n^{-1} = \mathbf{\Sigma}^{-1}\mathbf{V}^t\mathbf{Z}^t\mathbf{G}\mathbf{V}\mathbf{\Phi}^{-1/2}\,n^{-1} = \mathbf{\Sigma}^{-1}\mathbf{\Phi}\mathbf{\Phi}^{-1/2} = \mathbf{\Sigma}^{-1}\mathbf{\Phi}^{1/2}.$$

Aproksimativni test statističke značajnosti koeficijenata kvazikanoničke korelacije predložio je Momirović (1989; 1997). Prema Momiroviću, asimptotske ocene varijansi kvazikanoničkih korelacija, u oznaci S^2_{Rk}, mogu se definisati na sledeći način:

$$S^2_{R_k} = \frac{(1-R_k^2)^2}{n},$$

pri čemu je R_k^2 kvadrat koeficijenta kvazikanoničke korelacije dobijenog na uzorku. Prema tome, testiranje hipoteza:

$$H_{0p}: \rho_k = 0,\ k = 1,..., s$$

(pri čemu je ρ_k koeficijent kvazikanoničke korelacije u populaciji) može se izvesti standardnim F-testom koji služi za testiranje statističke značajnosti koeficijenata linearne korelacije. F-statistik za testiranje nulte hipoteze o koeficijentu k kvazikanoničke korelacije:

$$F_k = R_k^2 \frac{n-2}{1-R_k^2},$$

ima, ako je nulta hipoteza tačna, Snidikorovu distribuciju uzorkovanja sa 1 i n - 2 stepeni slobode.

Za interpretaciju rezultata kvazikanoničke diskriminacione analize neophodna je identifikacija sadržaja kvazikanoničkih diskriminacionih funkcija početnih varijabli, tj. varijabli iz **Z**. U te svrhe se, osim analize sklopa koeficijenata kojima se grade ove funkcije (tj. koeficijenata koji se nalaze u vektorima matrice **V**), mogu upotrebiti matrica sklopa i matrica strukture kvazikanoničkih diskriminacionih funkcija varijabli iz **Z**.

Matrica strukture kvazikanoničkih diskriminacionih funkcija za varijable iz **Z,** u oznaci **Q**, može se dobiti na sledeći način:

$$\mathbf{Q} = \mathbf{Z}^t\Psi n^{-1} = \mathbf{RV\Sigma}^{-1}.$$

Matrica sklopa kvazikanoničkih diskriminacionih funkcija za varijable iz **Z**, u oznaci **J**, data je sledećim izrazom:

$$\mathbf{J} = \mathbf{QM}^{-1},$$

pri čemu je **M** ($\mathbf{M} = \Psi^t\Psi n^{-1} = \mathbf{\Sigma}^{-1}\mathbf{V}^t\mathbf{RV\Sigma}^{-1}$) matrica inerkorelacija kvazikanoničkih diskriminacionih funkcija varijabli iz **Z**.

Matrice strukture i sklopa kvazikanoničkih diskriminacionih funkcija početnih varijabli iz **Z** predstavljaju faktorske matrice za matricu **R**, tj. za matricu interkorelacija varijabli iz **Z** (dokaz o tome može se naći u Momirović (1997)).

****

Kvazikanoničku diskriminacionu analizu u programu SPSS moguće je izvesti primenom makroa Konstantina Momirovića QDISC.SPS. Pored toga, ovu analizu moguće je izvesti na internetskoj adresi www.kal.rs u statističkom paketu KAL Aleksandra Zorića (Zorić, 2013). Ukoliko se kvazikanonička diskriminaciona analiza izvodi makroom QDISC.SPS, početne varijable je neophodno standardizovati pre analize. Za program KAL to nije neophodno.

Primer: Podaci iz Primera 2 za diskriminacionu analizu analizirani su kvazikanoničkom diskriminacionom analizom u programu KAL. Dobijen je sledeći ispis:

Quasi Canonical Discriminant Analysis

Quasi discriminant canonical correlations and significance tests

	rho	F	df1	df2	sig
1.	.581	115.077	1.000	225.000	.000
2.	.189	8.335	1.000	225.000	.004

Standardized Quasi Canonical Discriminant Function Coefficients

	1.	2.
zsucc	-.917	.251
zinteg	-.594	-.072
ztd	-.886	.336
zgd	-.840	-.229
zpc	.659	.442

Structure matrix

	1.	2.
zsucc	-.873	.090
zinteg	-.606	-.177
ztd	-.827	.180
zgd	-.880	-.376
zpc	.737	.558

Functions at Group Centroids

	1.	2.
1	-.429	.069
2	.098	-.448
3	.941	.101

Dobijene su dve statistički značajne kvazikanoničke korelacije (kolone **rho** i **sig.**). Prvu kvazikanoničku funkciju grade svi uključeni aspekti subjektivnog doživljaja vremena, s tim što je koeficijent za usredsređenost na sadašnjost suprotnog predznaka (predznak koeficijenata je inače arbitraran i bitno je samo da li su svi istog predznaka). Iz centroida se vidi da prva kvazidiskriminaciona funkcija pre svega razlikuje grupu sa PTSP-om (grupa 3) od ostalih dveju grupa. Pri tome su svi aspekti subjektivnog doživljaja vremena (osim usredsređenosti na sadašnjost) kod osoba sa PTSP-om poremećeni više no kod ostalih dveju grupa. (Pri tumačenju centroida treba uzeti u obzir predznake kvazidiskriminacionih koeficijenata.) Druga kvazidiskriminaciona funkcija definisana je pre svega usredsređenošću na sadašnjost udruženu sa neusmerenošću na ciljeve i ova funkcija pre svega razlikuje grupu sa drugim psihijatrijskim poremećajima od ostalih dveju grupa. Takva vrsta usredsređenosti na sadašnjost najmanje je, generalno posmatrano, izražena kod osoba sa drugim psihijatrijskim poremećajima, a najviše kod osoba sa PTSP-om.
****

XIX. 3. Kanonička diskriminaciona i kvazikanonička diskriminaciona analiza sa varijablama u standardizovanom obliku: jedno uopštenje modela

Odnosu kanoničke i osnovnog oblika kvazikanoničke diskriminacione analize moguće je pristupiti sa stanovišta formalnih rešenja na kojima se ove metode zasnivaju. Naime, i kanonička i kvazikanonička diskriminaciona analiza svode se u osnovi na rešavanje problema svojstvenih vrednosti. Ukoliko su početne varijable standardizovane, u kanoničkoj diskriminacionoj analizi (ako se uzme u obzir alternativni način njenog definisanja kao specijalnog slučaja kanoničke korelacione analize) do rešenja – pod zadatim uslovima – dolazi se nalaženjem svojstvenih vrednosti i odgovarajućih svojstvenih vektora matrice $\mathbf{R}^{-1}\mathbf{A}$. U kvazikanoničkoj diskriminacionoj analizi rešenje se dobija pronalaženjem svojstvenih vrednosti i pripadajućih svojstvenih vektora matrice **A**. Kvazikanonička diskriminaciona analiza se može, stoga, posmatrati kao varijanta diskriminacione analize u kojoj se matrica **R**, tj. matrica interkorelacija, „skuplja“ (engl. shrink) u matricu identiteta.

Ukoliko bismo matricu čije svojstvene vrednosti i svojstvene vektore tražimo, definisali u opštem slučaju pomoću matrica **R** i **A** na sledeći način:

$$[k\mathrm{I} + (1 - \mathrm{k})\,\mathbf{R}]^{-1}\,\mathbf{A},$$

pri čemu je k realni broj takav da je $0 \leq k \leq 1$, onda bismo ove dve metode mogli opisati kao ekstreme kontinuuma mogućih metoda diskriminacije koje se mogu definisati u zavisnosti od vrednosti koje u segmentu od 0 do 1 uzima skalar k. Naime, kvazikanonička diskriminaciona analiza bi odgovarala metodi za koju je $k = 1$, a kanonička diskriminaciona analiza postupku koji je definisan za $k = 0$.

Iz ovako definisanog odnosa dveju metoda diskriminacione analize vidi se jasno i njihova identičnost u jednom specijalnom graničnom slučaju – kada se analiza razlika među grupama izvodi samo na *jednoj standardizovanoj* varijabli. U tom se slučaju, kada je $\mathrm{k} = 0$, izraz $(k\mathrm{I} + (1 - \mathrm{k})\,\mathbf{R})^{-1}\,\mathbf{A}$, budući da je $\mathbf{R}^{-1}$ tada očigledno jednako 1, svodi na **A**, a kada je $\mathrm{k} = 1$, na **IA**. Naravno, **A** = **IA**. Pored toga, ovo uopštavanje pruža mogućnost da se kalibrisanjem, odnosno podešavanjem vrednosti skalara k (analogno „ridge“ varijantama regresije u multiploj regresionoj analizi, koje se koriste kada postoji izražena multikolinearnost među prediktorima), definišu „prelazni“ modeli između kanoničke i kvazikanoničke diskriminacione analize: oni bi, za razliku od kvazikanoničkog modela, uzeli u obzir interkorelacije između početnih kvantitativnih varijabli i bili bi, za razliku od kanoničkog modela, primenljivi i u situacijama izražene multikolinearnosti između početnih varijabli.

XIX.4. Klasifikacioni postupci diskriminacione analize

U prethodnom delu teksta u ovoj glavi prikazali smo postupke deskriptivne diskriminacione analize, jer se oni najčešće koriste u analizama podataka psiholoških istraživanja. Postupke prediktivne ili klasifikacione diskriminacione analize nećemo podrobnije prikazivati ovde, već ćemo dati samo nekoliko osnovnih informacija o ovim postupcima.[167]

Postoji mnoštvo postupaka klasifikacione diskriminacione analize razvijenih sa ciljem da se maksimizuje uspešnost klasifikovanja novih entiteta (entiteta čija grupna pripadnost nije poznata) u postojeće kategorije. Ovde ćemo navesti samo tri najpoznatija postupka:

1. Klasifikacione funkcije

Klasifikacione funkcije predstavljaju linearne kombinacije početnih varijabli, pri čemu se ponderi kojima se množe rezultati na početnim varijablama biraju tako da se maksimizuje klasifikaciona uspešnost. Klasifikacionih funkcija ima onoliko koliko ima grupa, tj. kategorija, a jedna ista početna varijabla može imati različite pondere u različitim klasifikacionim funkcijama. Vektor od m koeficijenata za klasifikacionu funkciju za grupu k, u oznaci $\mathbf{c}_k$, dobija se na sledeći način:

$$\mathbf{c}_k = \mathbf{C}_\mathrm{w}^{-1}\mathbf{m}_k \,,$$

pri čemu je $\mathbf{C}_\mathrm{w}^{-1}$ inverz kombinovane matrice kovarijansi za m kvantitativnih varijabli, a $\mathbf{m}_k$ vektor aritmetičkih sredina na m početnih varijabli za grupu k. Matrica $\mathbf{C}_\mathrm{w}$ dobija se uprosečavanjem, tj. kombinovanjem posebnih (unutargrupnih) matrica kovarijansi ($\mathbf{C}_1$, $\mathbf{C}_2$,..., $\mathbf{C}_\mathrm{g}$) svih grupa:

$$\mathbf{C}_\mathrm{w} = \frac{(\mathrm{n}_1 - 1)\mathbf{C}_1 + (\mathrm{n}_2 - 1)\mathbf{C}_2 + \cdots + (\mathrm{n}_\mathrm{g} - 1)\mathbf{C}_\mathrm{g}}{\mathrm{n}_1 + \mathrm{n}_2 + \cdots + \mathrm{n}_\mathrm{g} - \mathrm{g}} .$$

Konstanta za klasifikacionu funkciju za grupu k, u oznaci c_{0k}, određuje se na sledeći način:

$$c_{0k} = -0.5\mathbf{c}_k^\mathrm{t}\mathbf{m}_k,$$

pri čemu je $\mathbf{c}_k^\mathrm{t}$ transpon vektora koeficijenata za klasifikacionu funkciju za grupu k.

Za entitet koji treba klasifikovati u neku od postojećih grupa računaju se skorovi na svim klasifikacionim funkcijama i potom se on klasifikuje u onu grupu na čijoj klasifikacionoj funkciji ima najveći skor. Uočimo, dakle, da ove klasifikacione funkcije nisu isto što i

[167] O postupcima klasifikacione diskriminacione analize može se podrobnije pročitati u Tatsuoka (1971) ili u Johnson & Wichern (2007).

kanoničke diskriminacione funkcije iz deskriptivne diskriminacione analize (o matematičkim relacijama između ovih funkcija može se pročitati u Green (1979)). S obzirom na to kako se dobija matrica $\mathbf{C}_w$, ovaj postupak klasifikacije podrazumeva da su matrice kovarijansi svih subpopulacija homogene. Ispunjenost ovog uslova može se proveriti Boksovim testom, koji smo objasnili u prethodnom delu teksta u ovoj glavi.

2. Kvadrirana Mahalanobisova distanca

Kvadrirana Mahalanobisova distanca za entitet e_i u odnosu na grupu k, u oznaci MD^2_{ik}, definisana je na sledeći način:

$$MD^2_{ik} = (\mathbf{x}_i - \mathbf{m}_k)^t \mathbf{C}_w^{-1} (\mathbf{x}_i - \mathbf{m}_k),$$

pri čemu je $\mathbf{x}_i$ vektor sa rezultatima entiteta e_i na m kvantitativnih varijabli, $\mathbf{m}_k$ vektor aritmetičkih sredina na m varijabli za grupu k, a $\mathbf{C}_w^{-1}$ inverz kombinovane matrice kovarijansi za m kvantitativnih varijabli. Očigledno, kvadrirana Mahalanobisova distanca predstavlja kvadratnu formu, te se kao rezultat njenog računanja dobija skalar.[168] Pri primeni ovog postupka entitet se klasifikuje u onu grupu u odnosu na koju ima najmanju kvadriranu Mahalanobisovu distancu. Budući da važnu ulogu u računanju distance u ovom postupku ima matrica kovarijansi $\mathbf{C}_w$, vrlo važan uslov za primenu ovog postupka jeste homogenost matrica kovarijansi svih subpopulacija definisanih kategoričkom varijablom. Postupak klasifikacije zasnovan na Mahalanobisovoj distanci može se izvesti u programu KAL na internetskoj adresi www.kal.rs odabirom opcije **Analyze/Mahalanobis classification**.

Ukoliko su apriorne verovatnoće pripadanja pojedinim subpopulacijama (p_k) različite, tada se za potrebe klasifikacije koristi sledeća formula:

$$-\frac{1}{2}(\mathbf{x}_i - \mathbf{m}_k)^t \mathbf{C}_w^{-1}(\mathbf{x}_i - \mathbf{m}_k) + \ln(p_k).$$

3. Kvadratni klasifikacioni postupak

Kvadratni klasifikacioni postupak može se primeniti i onda kada matrice kovarijansi različitih subpopulacija nisu homogene. Postupak se zasniva na računanju vrednosti D^2_{ik} za entitet e_i u odnosu na grupu k:

$$D^2_{ik} = -\frac{1}{2}\log(|\mathbf{C}_k|) - \frac{1}{2}(\mathbf{x}_i - \mathbf{m}_k)^t \mathbf{C}_k^{-1}(\mathbf{x}_i - \mathbf{m}_k) + \ln(p_k),$$

pri čemu je $\mathbf{x}_i$ vektor sa rezultatima entiteta e_i na m kvantitativnih varijabli, $\mathbf{m}_k$ vektor aritmetičkih sredina na m varijabli za grupu k, $\mathbf{C}^{-1}_k$ inverz matrice kovarijansi za m kvantitativnih varijabli za grupu k, $|\mathbf{C}_k|$ determinanta matrice kovarijansi za m kvantitativnih varijabli za grupu k, a p_k apriorna verovatnoća pripadanja subpopulaciji k. Entitet se klasifikuje u grupu za koju ima najmanju vrednost D^2_{ik}.

[168] Pojam kvadratne forme objašnjen je u **Dodatku 2**.

Svi prikazani postupci klasifikovanja pretpostavljaju da je raspodela m kvantitativnih varijabli u svakoj od subpopulacija multivarijaciona normalna raspodela.
Za klasifikovanje novih entiteta u postojeće grupe, pored postupaka diskriminacione analize, mogu se koristiti i postupci binarne i multinomne logističke regresije (cf. Press & Wilson, 1978).

Reference na koje se poziva u ovom tekstu:

Anderson, H.E. (1966). Regression, discriminant analysis, and a standard notation for basic statistics, In R. B. Cattell (ed.). *Handbook of multivariate experimental psychology* (pp. 153–173), Chicago: Rand McNally & Company.

Cooley, W.W., & Lohnes, P.R. (1962). *Multivariate procedures for the behavioral Sciences*, New York: John Wiley & Sons, Inc.

Dobrić, V., & Momirović, K. (1984). An algorithm and program for stupid discriminant analysis, *Proceedings of "Jahorina 84",* Sarajevo, 213, 1–5.

Geisser, S. (1977). Discrimination, Allocatory and Separatory, Linear Aspects, In J. Van Ryzin (Ed.) *Classification and Clustering* (pp. 301–330), New York: Academic Press, Inc.

Green, B. F. (1979). The Two Kinds of Linear Discriminant Functions and Their Relationship. *Journal of Educational Statistics*, *4*(3), 247–263.

Huberty, C. J. (1975).The stability of three indices of relative variable contribution in discriminant analysis. *Journal of Experimental Education*, *44*(2), 59–64.

Huberty, C. (1984). Issues in the use and interpretation of discriminant analysis, *Psychological Bulletin*, *95*, 156–171.

Huberty, C., Wisenbaker, J. M., & Smith, J.C. (1987). Assessing Predictive Accuracy in Discriminant Analysis, *Multivariate Behavioural Research*, *22*, 307–329.

Ivanović, B. (1963). *Diskriminaciona analiza sa primenom u ekonomskim istraživanjima*, Beograd: Naučna knjiga.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis, Sixth edition*. Upper Saddle River NJ: Pearson Eduaction, Inc.

Klecka, W.R. (1980). *Discriminant analysis*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-019, Beverly Hills: Sage.

Kovačić, Z. (1994). *Multivarijaciona analiza*, Beograd: Ekonomski fakultet.

Kshirsagar, A. M., & Arseven, E. (1975). A note on the equivalency of two discrimination procedures, *The American statistician*, *29*, 38–39.

Maxwell, A.E. (1977). *Multivariate Analysis in behavioral research*, London: Chapman and Hall.

McDonald, R.P. (1968). A unified treatment of the weighting problem, *Psychometrika*, *33*, 351–381.

McKay, R.J., & Campbell, N.A. (1982a). Variable selection techiques in discriminant analysis I. Description, *British Journal of Mathematical and Statistical Psychology*, *35*, 1-29.

McKay, R.J. & Campbell, N.A. (1982b). Variable selection techiques in discriminant analysis II. Allocation, *British Journal of Mathematical and Statistical Psychology*, *35*, 30–41.

Momirović, K. (1989). Kvazikanonička diskriminativna analiza u Pearsonovoj diskriminativnoj metrici. *Zbornik radova 3 Sekcije za klasifkacije Saveza statističkih društava Jugoslavije*, 190–199, Beograd: Savezni zavod za statistiku.

Momirović, K. (1997). O diskriminativnim funkcijama, diskriminativnim faktorima i nekim očiglednim glupostima, *Statistička revija*, *46*, 79–100.

Momirović, K., Dobrić, V., & Karaman, Ž. (1983). Canonical covariance analysis. *Proc. 5th International Symposium "Computer at the University"*, 463-473.

Momirović, K., Gredelj, M., & Szirovicza. L. (1977). *Metode multivarijatne analize.*Zagreb: Zavod za produktivnost.

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.

Overall, J.E., & Klett, C.J. (1972). *Applied Multivariate Analysis*, New York: McGraw-Hill Book Company.

Press, S. J., & Wilson, S. (1978). Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association*, *73*(364), 699–705.

Rao, C. R. (1974). *Linear statistical inference and its application, Second edition*, New Delhi: Wiley Eastern Private Limited.

Rencher, A. C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components, *American statisticians*, *46*, 217–225.

Schott, J.R. (1997). *Matrix analysis for statistics*, New York: John Wiley & Sons, Inc.

Stevens, J.P. (2002). *Applied multivariate statistics for the social sciences*, Mahwah: Lawrence Erlbaum Associates, Publishers.

Štalec, J., & Momirović, K. (1984). On a very simple method of robust discriminant analysis, *Proceedings of 6th Intenational Symposium 'Computer at the University',* 515: 1-16.

Tacq, J. (1997). *Multivariate analysis techiques in social science research*, London: Sage Publications

Tatsuoka, M. M. (1971). *Multivariate analysis*. New York: John Wiley & Sons, Inc.

Thomas, D. R., & Zumbo, B. D. (1996). Using a Measure of Variable Importance to Investigate the Standardization of Discriminant Coefficients. *Journal of Educational and Behavioral Statistics*, *21*(2), 110−130.

Tenjović, L. (2002). *Statistika u psihologiji – priručnik*, Beograd: Centar za primenjenu psihologiju.

Welch, B.L. (1939). Note on discriminant functions, *Biometrika*, *31*, 218–220.

Williams, B. K. (1982). A simple demonstration of the relationship between classification and canonical variates analysis, *American Statistician*, *36*, 363–365.

Zorić, A. (2013). *Novi pristup statističkoj analizi podataka u psihologiji − od rigidnih statističkih paketa ka fleksibilnom sistemu*. Doktorska disertacija, Filozofski fakultet, Beograd.