

Faktorska ali clusterska analiza - katero tehniko uporabiti?

VALENTIN BUCIK

Nekontrolirane skoke v razvoju na področju metodologije družboslovnega raziskovanja, torej tudi v psihologiji, gre pripisati med drugim tudi razvoju zmogljivosti raziskovalnega orodja - velikih in malih računalniških sistemov ter pripadajoče programske opreme. Današnja stopnja razvoja analiz in tehnik multivariantnega preverjanja povezav in napovedovanja nam omogoča vpogled v globlje nivoje fenomenoloških procesov in odnosov kot v preteklih obdobjih, za katere so bili značilni na eni strani okleščene eksperimentalne situacije, ki z realnostjo niso imele veliko skupnega, ter na drugi strani bolj ali manj intuitivni opisi in ugibanja o naravi preučevanih procesov. Čase, ko so se raziskovalcem ježili lasje ob misli na ročno izračunavanje statističnih parametrov iz velikih baz zbranih podatkov, so zamenjali časi naraščajoče bojazni, da s svojim metodološkim znanjem ne bodo kos zapleteni nalogi: kako med mnogimi variantami možnih pristopov obdelave podatkov, ki so pri delu z različnimi statističnimi paketi na voljo, poiskati tisto alternativo, ki bo najbolj prikladna glede na teoretične predpostavke v zvezi z zastavljenim problemom in ki bo najbolj upravičena glede na metrični nivo podatkov, torej na količino in vrsto informacije, ki jo ti nosijo.

V šestdesetih in sedemdesetih letih so se na širšem aplikativnem raziskovalnem področju uveljavile danes že klasične multivariantne tehnike, ki jih *Fornell (1987)* imenuje kar tehnike "prve generacije", kot so multipla analiza variance, analiza kovariance, multipla diskriminantna analiza, multipla regresija, multidimenzionalno skaliranje, več različnih postopkov factorske analize, kanonična korelacijska analiza itd. Te so bile pravzaprav posledica postopnega razvoja enostavnejših bivariantnih modelov. V zadnjih letih so se, predvsem v smislu bogatejšega osvetljevanja vzročno-posledičnih odnosov, pojavile multivariantne tehnike "druge generacije" (analiza redundance, konfirmatorna factorska analiza, analiza kovariančnih struktur, konfirmatorno multidimenzionalno skaliranje itd. - *Fornell, 1987*), ki so trdno bazirane na tehnikah "prve generacije", vendar zahtevajo nov način razmišljanja in dela, predvsem pa so rigoroznejše glede postavljanja teoretičnih izhodišč, ki jih s konfirmatornimi tehnikami preverjamo (o tem glej tudi v: *Vodopivec, 1988*).

Ko govorimo o klasičnih multivariantnih tehnikah, imamo v mislih predvsem tehnike, ki sodijo v eno od obeh skupin: med one, ki so namenjene multivariantnemu

napovedovanju, in tiste, s katerim opravljamo analize kovariančnih struktur. V prvo skupino uvrščamo regresijsko analizo, analizo diskriminantnih funkcij, kanonično korelacijsko analizo, multiplo analizo variance itd.; tu skuša raziskovalec dostopne informacije z multiplih prediktorjev uporabiti v napovedovanju položaja subjekta ali opazovanega fenomena na nekem kriteriju ali na kombinaciji več kriterijev. Tehnike, kot so faktorska analiza ter različni postopki hierarhičnega ali nehierarhičnega razvrščanja v skupine - clusterske analize, pa naj bi sodile v drugo skupino.

Na tem mestu bi radi prikazali nekatere značilnosti in različnosti ter kritično analizo možnosti uporabe dveh vrst postopkov, ki jih uvrščamo v drugo od omenjenih skupin - faktorske analize¹ in clusterske analize². Pri reduciranju večjega števila enot (variabel ali objektov) v manjše število homogenejših skupin, raziskovalci radi sežejo po eni od obeh tehnik. Obe namreč služita določevanju temeljnih, latentnih struktur v multivariantnem setu podatkov. Vendar je moč opaziti, da kdaj prihaja tudi do napačnih uporab tehnik - predvsem prve na račun druge - glede na karakteristike podatkov in na raziskovalni cilj. Ravno cilj raziskovanja pa je prvi in osnovni kriterij za odločitev o tehniki, ki jo bomo aplicirali. Ne bomo se spuščali v opis tehničnih detajlov FA in CA, poskušali bomo le opozoriti na glavne razlike v vsebini obeh.

Tehnike FA na eni strani in razvrščanja v skupine s CA na drugi lahko apliciramo tako v analizah, kjer je predmet preučevanja grupa variabel, kot tudi če je predmet preučevanja grupa ljudi ali objektov, za katere so bile izmerjene določene variable. V prvem primeru so podatki za vhod v analizo organizirani v obliki interkorelacijske matrike (pri FA) ali matrike kakšnih drugih mer podobnosti ali različnosti (pri CA) med variablami, izmerjenimi na vseh subjektih. V drugem primeru pa začnemo z matriko korelacij ali podobnosti med subjekti. V FA je analiza drugega tipa znana kot transponirana oz. invertirana FA ali Q-tehnika (*Borgen in Weiss, 1971*). Obe vrsti postopkov pa sta lahko uporabni kot eksploratorni, konfirmatorni ali simplifikirajoči metodološki pripomoček. Z obema lahko torej odkrivamo nove strukture, ki se skrivajo v velikih setih multivariantnih podatkov, nove vrste klasifikacij itd.³; lahko testiramo obstoječe klasifikacije enot v grupe, potrdimo ali ovržemo postavljene teoretične modele; in končno lahko z grupiranjem poenostavimo zapletenost in obsežnost velikih baz podatkov, kjer bi bila tolmačenja fenomenov brez uporaba FA ali CA če drugega ne vsaj neobjektivna, nejasna in na moč utrudljiva.

Najprej na kratko nekaj besed o vsaki od obeh tehnik, potem pa o razlikah med njima.

FAKTORSKA ANALIZA

FA je, kot vemo, matematični postopek za analizo korelacij med variablami⁴. Namen analize je, da za veliko število manifestnih variabel, med katerimi obstajajo različne

¹ V nadaljevanju FA.

² V nadaljevanju CA. Kljub temu, da so postopki razvrščanja v skupine mnogoštevilni in po vsebini raznovrstni (podrobneje o tem glej v: Ferligoj, 1989), v našem tekstu pod splošnejšim imenom CA pojmujeemo predvsem hierarhične aglomerativne postopke, ki so po svoji logiki najbližje postopkom faktorizacije pri FA.

³ Kar je danes v praksi gotovo najpogostejša oblika uporabe tako FA kot CA.

⁴ Glede na to, da za odgovor na vprašanje, ki smo si ga zastavili v naslovu, ni bistveno, ali gre za relacije med atributi ali med subjekti, bomo tu govorili le o t.i. R-tipu FA in o združevanju variabel v grupe pri CA.

stopnje povezanosti, najdemo manjše število latentnih, temeljnih, med seboj čim bolj neodvisnih variabel (oz. dimenzij ali faktorjev), ki bodo lahko odnose med manifestnimi variablami razložile na višjem nivoju, kot pa to lahko storimo s korelacijami. Določiti moramo število faktorjev, jih ovrednotiti oz. interpretirati in poiskati stopnjo, do katere si vsaka variabla "lasti" vsak faktor.

Vsako manifestno variabla si lahko predstavljamo kot linearno funkcijo skupnega seta latentnih variabel (skupni faktorji) in latentne variable, ki pripada samo določeni manifestni variabli (specifični faktor; *Mulaik, 1987*). Torej gre za razlago kovariantnih odnosov med mnogimi eksperimentalnimi - opazovanimi variablami v obliki linearne odvisnosti s pomočjo (številčno reduciranih) "idealnih", "intervenirajočih", abstraktnih, konceptualnih variabel (*Cattell, 1966*). Pri eksploratorni opisni FA (za razliko od konfirmatorne) po induktivni poti skušamo odkriti in identificirati nove latentne skupne faktorje, ki nam pokažejo količino skupne informacije, ki jo ti faktorji delijo z manifestnimi variablami.

Osnovni model FA zahteva matriko indeksov, koeficientov ali drugih ustreznih mer podobnosti ali različnosti med variablami (največkrat Pearsonov koeficient produkt-moment korelacije⁵). Izražen je v faktorski enačbi

$$Z_j = a_{j1} * F_1 + a_{j2} * F_2 + \dots + a_{ji} * F_i + \dots + a_{jm} * F_m + d_j * U_j$$

kjer Z_j pomeni standardizirano vrednost (celotna varianca) za variabla j , a_{ji} standardiziran koeficient multiple regresije variable j s faktorjem i , kar pomeni faktorsko obtežitev ali saturacijo variable j (to pa ni nič drugega kot koeficient korelacije faktorja i z manifestno variabla j - višji ko je a_{ji} , močnejše F_i pojasnjuje variabla j); F_i je skupno faktor - ti so razporejeni hierarhično. Prvi faktor pojasni največ skupne variance in je zato navadno najmočnejši, naslednji ima na voljo pojasnjevati le še varianco, ki je ostala po ekstrahiranju prvega faktorja itd.; d_j pomeni obtežitev na specifičnem faktorju U pri variabli j (kar pomeni nepojasnen del variance oz. reziduum).

V osnovni faktorski matriki lahko razberemo najvažnejše parametre faktorske strukture. Komunaliteta variable j pomeni propore celotne variance, ki je vsebovana v kombinaciji vseh skupnih faktorjev pri variabli j , torej koliko celotne variance je za neko variabla sposobna pojasniti kombinacija vseh skupnih faktorjev in koliko ostane nepojasnjene. Lastna vrednost, "eigenvalue" skupnega faktorja i preko vseh n variabel pomeni delež celotne variance, ki jo je preko vseh variabel sposoben pojasniti skupni faktor. Odstotek celotne variance faktorja i pa pove, kolikšen delež te variance je faktor v povprečju glede na število variabel zmožen pojasniti. Večja, ko je lastna vrednost faktorja, več celotne variance pojasnjuje - to imenujemo tudi "moč faktorja"; ta s številom ekstrahiranih faktorjev upada. Celotna varianca preučevanega pojava na neki variabli je torej razdeljena med vse faktorje, ki to variabla pač bolj ali manj pojasnjujejo.

Od količine celotne variance, ki jo je še sposoben pojasniti posamezni faktor, oz. od moči faktorja, bo odvisen tudi kriterij za določevanje števila faktorjev, ki jih bomo ekstrahirali.

Ekstrahirane faktorje lahko zaradi jasnejše faktorske strukture in s tem lažje interpretacije le-teh pravokotno ali poševnokotno rotiramo. Zatem iz matrike rotiranih

⁵ Oziroma Točkovno-biseralni ali Fi-koeficient, če je ena oz. sta obe variabli naravno dihotomizirani.

faktorjev izračunamo matriko koeficientov faktorskih skorov (oz. točk) na takle način (Nie in dr., 1975)

$$F = (R^T * R)^{-1} * R^T$$

kjer je R^T transponirana rotirana faktorska matrika R in $(R^T * R)^{-1}$ inverz produkta R^T in R .

V matriki F torej nastopajo regresijski koeficienti, ki jih potrebujemo za ocenjevanje faktorjev z vidika opazovanih variabel. Iz matrike koeficientov faktorskih skorov pa lahko za vsakega respondenta izračunamo faktorski skor za vsakega od ekstrahiranih skupnih faktorjev posebej.

Predznak in višina faktorskega skora na nekem faktorju torej pomeni stopnjo pripadnosti posameznika "lastnostim in imenu" tega faktorja preko vseh analiziranih variabel (npr. koliko skupne, pojasnjene variance, v obnašanju respondenta na celotni lestvici trditev, je sposoben ta faktor razložiti).

CLUSTERSKA ANALIZA

Skupine posameznih enot bi po aplikaciji CA morale izražati izoliranost navzven in kohezivnost navznoter (Milligan, 1980) - tj. podobnost znotraj grup in različnost med grupami. S spajanjem enot v grupe po podobnosti zmanjšujemo varianco znotraj grupe in s tem ustvarjamo več manjših skupin, ki pa so v odnosu na osnovno, začetno grupo navznoter bolj homogene. Tu nas torej ne zanima posebej, kolikšen delež celotne variance pri posamezni enoti lahko pripišemo kateremu od evaluiranih clustrov, temveč se trudimo, da bi enoto kot celoto, z vso njeno variabilnostjo uvrstili v tisto skupino, ki ji po svoji podobnosti z drugimi enotami v tej grupi, pripada.

Dobre lastnosti hierarhičnega razvrščanja v grupe so predvsem enostavnost metode, jasni rezultati (v obliki drevesne strukture), ki so močni, veljavni in uporabni; predvsem pa ni pretirano strogih pogojev za metrični nivo vhodnih podatkov. Uporabimo pravzaprav lahko kakršenkoli tip podatkov, vendar je skladno s tem izbira pravilnega algoritma razvrščanja v grupe za veljavnost izida grupiranja bistvenega pomena. Končna mera različnosti pa je vedno v celoti intervalna (Jackson, 1983).

Ko smo se odločili za aplikacijo CA, se moramo držati nekaterih osnovnih pravil:

1) *Pravilna organiziranost vhodnih podatkov*: ali imamo namen grupirati attribute preko vzorca subjektov, subjekte preko vzorca atributov, ali v istem združevati subjekte in različne attribute (Brenk in Kline, 1988) itd.

2) *Izbira mere za merjenje podobnosti, bližine, različnosti itd.*, med variablami, ki je odvisna predvsem od organiziranosti in še bolj od narave (tipa) vhodnih podatkov (ali gre za številske, binarne, nominalne, mešane ... podatke - več o tem v: Ferligoj, 1989, str. 31-48). Za številske podatke najpogosteje uporabljana Evklidska distanca odraža vse tri bistvene elemente profila večdimenzionalnih podatkov, to je povprečno vrednost, oziroma relativno naraščanje skozi profil seta variabel, razpršenost, ki je definirana s standardno deviacijo in obliko profila variabel (Blashfield, 1980; Borgen in Weiss, 1971). Zato je uporabnejša kot Pearsnova r , saj le-ta s standardizacijo podatkov v z-

vrednosti izgubi informacijo o aritmetični sredini in standardni deviaciji, torej upošteva le obliko profila (*Borgen in Barnett, 1987*).

3) *Izbira ustreznega algoritma*, po katerem bomo združevali enote v skupine. Primerjalne raziskave (npr. *Ferligoj in Batagelj, 1980; Milligan, 1980*) so pokazale, da je tako na merljivih (ordinalne in intervalne), kot na kategorialnih variabliah najboljši algoritem za hierarhično razvrščanje Wardova metoda minimiziranja variance znotraj grup in maksimiziranja variance med grupami na vsaki zaporedni stopnji združevanja. Algoritem išče optimalno stopnjo v hierarhično dobljenih clustrih, kar je adekvaten kompromis med minimiziranjem števila clustrov in maksimiziranjem znotraj grupne homogenosti (*Berven, 1985*). Izbira pravega algoritma razvrščanja je po mnenju strokovnjakov veliko pomembnejša kot izbira prave mere podobnosti oz. različnosti.

4) *Evaluacija pravega števila skupin*; če smo se pri grupiranju enot v clustre ustavili prezgodaj in imamo še večje število clustrov, potem verjetno še nismo izčrpali celotne clusterske strukture vzorca, če pa smo mero prekoračili, smo naravno distinktivne clustre nasilno združevali. Bistven kriterij za določevanje števila clustrov je vsebinska analiza clustrov na posameznih nivojih združevanja. Velja pa tudi pravilo, da so na dendrogramu (prikaz združevanja enot v clustre v drevesni obliki) že na pogled vidne razlike med t.i. naravnimi skupinami. Te so določene z večjimi skoki višin v drevesu (*Ferligoj in Batagelj, 1980; Batagelj, 1982*). Če so te naravne skupine še nasilno združevane, se večja napaka združevanja, večja se varianca znotraj grupe, manjša pa se varianca med grupami. To pa pomeni rušenje osnovnih principov hierarhičnega razvrščanja v grupe.

5) *Vsebinska interpretacija dobljene clusterske strukture s pomočjo enot*, ki so v posameznem clustru (zelo podobno kot interpretiramo faktorje v faktorjski analizi s pomočjo manifestnih variabel).

6) *Izvedba adckvatne validacije clusterske strukture*. Vsako končno rešitev razvrščanja v grupe moramo gledati s skepto in vedno s ponavljanjem rešitve (preko vzporednih baz podatkov, preko različnih setov variabel - ali subjektov, preko različnih združevalnih algoritmov, ki so nam na voljo in ki so relevantni (*Blashfield, 1980*)), preverjati na začetku dobljeno strukturo.

RAZLIKA MED FA IN CA

Kakšne so težave, ki nastopijo pri grupiranju s FA?

Počiva na računu korelacije, oz. kovariance in zahteva kot vhod v analizo matriko povezanosti med merjenimi variablami. Najčešče uporabljani Pearsonov koeficient zahteva podatke na intervalni skali, zatem kolikor toliko normalno porazdelitev podatkov vsaj v osnovni populaciji. Predpostavlja linearno zvezo in je neobčutljiv za krivuljčne korelacije med variablami, njihovo pomembnost podcenjuje in rezultat so slabše definirani, nejasni faktorji (*Tinsley in Tinsley, 1987*). Koeficienti zvez med naravno dihotomnimi ali naknadno dihotomizirani podatki niso priporočljivi, ker zaradi morebitne asimetričnosti dihotomnih podatkov ne dajejo prave slike povezanosti, ali pa celo tvorijo matrike, ki izgubijo določene lastnosti, katere so pogoj za vključitev v FA (t.i. gramianske lastnosti, pri katerih morajo biti reducirane korelacijske matrike pozitivno definirane in simetrične; torej vsi njihovi kvadratni koreni (lastne vrednosti)

morajo biti večji od 0. V diagonali reducirane korelacijske matrike morajo ostati pozitivne vrednosti, sicer so neadekvatne - ravno to pa je nevarnost, do katere lahko pride pri računanju korelacijskih matrik na podlagi dihotomnih podatkov (*Fulgosi, 1979*). Za CA pa dihotomne spremenljivke ob izbiri pravega algoritma ne predstavljajo nobenih težav.

Problem aplikacije FA lahko nastopi tudi zaradi priporočila, naj v analizo ne vstopamo z majhnimi vzorci: Vzorec velikosti 100 je (kot trdijo strogi metodologi) ubog, 200 ravno še zadovoljiv, 300 dober, 500 zelo dober in 1000 odličen (*Tinsley in Tinsley, 1987*).

Zato je torej v nekaterih primerih⁶ CA bolj priporočljiva tehnika kot FA. Pri slednji se je namreč v metodoloških raziskavah pokazalo, da ne daje najboljših rezultatov v klasifikacijskih, "tipoloških" problemih in da se bolje obnaša kot pripomoček za dimenzionalne analize. CA je bolj robustna tehnika za grupiranja. Ne zanima je notranja struktura variance, ki se pri FA razpodeli med več dimenzij in nam s tem zabriše morebitno jasno klasifikacijsko strukturo. *Chatterji in Mukerjee (1986)* sta s CA vzorcev subjektov dobila manjše število jasneje določenih grup, s FA (Q-analiza glavnih komponent) pa subjektov ni bilo moč tako jasno ločiti in tudi struktura skupin (faktorjev) je bila manj homogena in konsistentna.

Obe, FA in CA sta metodi za prepoznavanje latentnih pojasnjevalnih struktur v multivariantnih podatkih. Glavna razlika med njima je le v obravnavanju variance znotraj posamezne variable. FA razdeli variiranje odgovorov posameznikov na določeno postavko preko več virov pojasnjevanja (faktorjev), medtem ko v CA celotno varianco pripišemo enemu pojasnjevalnemu viru (tistemu, h kateremu je variabla po principu bližine uvrščena). Torej CA omogoča, da se variable razvrščajo v jasno določene (diskretne) skupine oz. clustre, medtem ko so rezultati FA manj "jasni", z deli skupne variance posamezne variable, razdeljenimi med več dimenzij, faktorjev (*Borgen in Weiss, 1971; Borgen in Barnett, 1987*).

Priporoča se, naj bo FA uporabljena predvsem takrat, ko raziskovalec proučuje notranjo strukturo skupine variabel (npr. ko hočemo zmanjšati skupno varianco v testu z manjšanjem števila postavk) ali ko si hoče pojasniti, kako je vsaka posamezna enota - test ali item - strukturirana.

Odločitev - ali FA ali CA - je torej odvisna predvsem od namena raziskave: če cilj ni iskanje notranje strukture variable (tj. kakšen delček njene variance je pojasnjen s katero od latentnih dimenzij, ki jo razlagajo), pač pa postavljanje kategorične odločitve o razvrstitvi spremenljivk ali respondentov kot celote v različne, a navznoter homogene grupe (kar je sicer robustnejše delo), tedaj je CA prikladnejša in bolj upravičena tehnika (*Jackson, 1983*).

ZAKLJUČEK

Tehniki FA in CA imata mnogo skupnega. Obe sta v funkciji zmanjševanja kompleksnosti podatkov na neke bistvene, centralne, pojasnjevalne lastnosti. Obe zahtevata od raziskovalca precej lastne iniciative, znanja in odgovornosti v odločanju -

⁶ Posebej tedaj, ko želimo klasificirati, formirati določene taksonomije, tipologije.

končni rezultati so namreč vedno odvisni tudi od tega, kako je bila analiza izvedena. V obeh primerih ostaja ne do konca razrešeno nasprotje o tem, katera je najučinkovitejša tehnika v grupah FA in CA - različni raziskovalci imajo različne pristope.

Bazična razlika v uporabi FA in CA tudi ni v naravi podatkov, ki jih ena ali druga zahtevata (saj sta lahko obe aplicirani na istih podatkih). Najvažnejše je, da FA in CA služita različnim analitičnim in raziskovalnim ciljem. S FA raziskovalec lahko razišče oz. pregleda notranjo strukturo enote (če ga to zanima), s CA pa grupira enote, ki so kot celota podobne med seboj in ko mu združevanje pomaga pri konstruiranju diskretnih, jasno določenih grup.

Glede na to, da je danes pri uporabi CA na voljo dovolj zanesljivih algoritmov za združevanje variabel, ki so z metrijskega vidika na nižjem nivoju (tudi za dihotomne variable (Cliff in dr., 1986; SPSS inc., 1988; Ferligoj, 1989; Jug, 1989)), ni več nobene resnega opravičila za uporabo FA za vsako ceno pri razvrščanju tako variabel kot tudi subjektov v grupe, kjer bi svoje mesto bolj upravičeno našla katera od tehnik CA.

REFERENCE

- Batagelj, V. (1982). CLUSE/CYBER. Priročnik, Ljubljana: FNT.
- Berven, N.L. (1985). Reliability and validity of standardized case management simulations. *J. of counseling psychology*, 32, 397-409.
- Blashfield, R.K. (1980). Propositions regarding the use of cluster analysis in clinical research, *J. of consulting and clinical psychology*, 48, 456-459.
- Borgen, F.H. & Barnett, D.C. (1987). Applying cluster analysis in counseling psychology research. *J. of counseling psychology*, 34, 456-468.
- Borgen, F.H. & Weiss, D.J. (1971). Cluster analysis and counseling research. *J. of counseling psychology*, 18, 583-591.
- Brenk, K. & Kline, M. (1988). Analysing free response data with cluster analysis. EIASM. Workshop on new developments in the study of human behavior. Brussels.
- Cattell, R.B. (1966). The measuring and strategic use of factor analysis. V R.B. Cattell (ed). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally & Company.
- Chatterji, S. & Mukerjee, M. (1986). Isolating of profile pattern of success individuals using cluster analysis and Q-technique. *Japanese psychological research*, 28.4.161-167.
- Cliff, N., McCormick, D.J., Zarkin, J.L., Cudeck, R.A. & Collins, L.M. (1986). BINCLUS: Nonhierarchical clustering of binary data. *Multivariate behavioral research*, 21, 201-227.
- Ferligoj, A. (1989). Razvrščanje v skupine. *Metodološki zvezki*, 4.
- Ferligoj, A. & Batagelj, V. (1980). Taksonomske metode v družboslovnem raziskovanju. Ljubljana: RI FSPN.
- Fornell, C. (1987). A second generation of multivariate analysis: classification of methods and implications for marketing research. *The Review of Marketing*, 87. Section 6. Am. Marketing Assoc.
- Fulgosi, A. (1984). Faktorska analiza. Zagreb: Školska knjiga.
- Jackson, B. (1983). *Multivariate data analysis: an introduction*, Richard D. Irwin inc.
- Jug, J. (1989). Razvrščanje binarnih enot s podprogramom CLUSTER v SPSS/PC+. *Metodološki zvezki*, 5.
- Milligan, G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45, 325-342.
- Mulaik, S.A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate behavioural research*, 22, 267-305.
- Nie, N.H. in drugi (1975). *SPSS*. New York: McGraw-Hill.
- SPSS Inc. (1988). *SPSS-X™ User's Guide*. 3rd Edition. Chicago: SPSS Inc.
- Tinsley, H.E.A. & Tinsley, D.J. (1987). Use of factor analysis in counseling psychology research. *J. of counseling psychology*, 34, 414-424.
- Vodopivec, B. (1988). Epistemološki vidiki analize kovariančnih struktur. *Anthropos*, 18, 46-57.