

Sport za sve u funkciji unapređenja kvalitete života – Konferencija Zagreb

Tematska cjelina: Pedagoški, socijalni, psihološki, nutricionistički i ekološki aspekti kvalitete života

MODEL ZADRŽAVANJA MAKSIMUMA KORISNIH INFORMACIJA KOD ISTRAŽIVAČKIH PROJEKATA U KINEZILOGIJI

Pozvano predavanje

Dobromir Bonacin PhD

Fakultet sporta i tjelesnog odgoja Sarajevo – Institut sporta

Uvod

Uobičajeni pristup analizi podataka danas oslonjen je u maksimalno mogućoj mjeri na kibernetički, odnosno upravljački koncept, a ovaj opet na Huserlov fenomenološki pristup (Grlić, 1967) koji, prirodno, najveću vjerodostojnost daje tzv. sustavu zajedničkih informacija i iz njih izvedenih svih ostalih pojmova moderne spoznajne teorije (npr. Pauše, 1978, Mraković, 1992. ili Momirović, 1984). To je, naravno, jedan sasvim legitiman i prihvatljiv model, jer u najvećem broju situacija apsolutno zadovoljava, a napose kad god se radi o relativno nepoznatim pojavama u kojima treba otkrivati poznate fragmente i iz njih polako slagati nekakvu prepoznatljivu sliku (Cote i Buckley, 1987, Šošić i Rađo, 1998, Krković, Momirović i Petz, 1966).

Čak i klasične statističke definicije ovakvu situaciju naglašavaju, samom činjenicom da je po nekim viđenjima statistika disciplina koja sustavno zanemaruje pojedinosti dok ne dosegne pravilo (Jacob, 1978, Serdar i Šošić 1981). Ovome se, naizgled, nema što prigovoriti i stvari se, naravno, dalje provode jedinstveno i bez dvojbi. Na takvim temeljima, naravno narastaju svi ostali, a napose multivarijantni postupci i modeli (Rao, 1973, Cooley i Lohnes, 1971, Momirović 1984, Harman, 1970, Johnson i Wichern, 1992, Parkinson, 1980). Maksimizacija ovog koncepta vidljiva je upravo kod analiza redukcija dimenzionalnosti (Bonacin, 2004), ili kako ih neki još zovu, kod faktorskih analiza (Fulgosi, 1979), kojima se dobivaju latentne, direktno nemjerljive dimenzije, i koje nam opisuju mehanizme ili nadređene suprasustave odgovorne za manifestacije pojava koje stoje u nižim hijerarhijskim redovima ma kakvih odlučivanja ili izvršavanja.

Nažalost, situacija je u mnogočemu istinski drugačija. Naime, sve te i takve analize, zbog prirode svoje suštine okrenute su minimizaciji bitnoga (fenomenologija) i sustavnom zanemarivanju svega što ne pripada, po nekom kriteriju, zajedničkom prostoru, a što se tretira kao zanemarljivi rezidual. To, naravno, dovodi do situacije, prema kojoj je sve što ne pripada takvom zajedničkom prostoru, ili određenoj količini zajedničkog dijela spektra, s pozicije tretiranog problema - ustvari greška. Međutim, lako je dokazati, čak i kad se radi o reprezentativnom skupu parametara ili upotrebljenih varijabli, i kad skup opisanih entiteta također zadovoljava uvjet reprezentativnosti, da rezidualni varijabilitet nije rezultat greške, već skup sustavnih korisnih informacija.

Ovaj će članak na različitim primjerima, dokazati da rezidualnih i nebitnih informacija nema.

Model

Neka je iz populacije E nekim kriterijem definicije izvučen bilo kakav uzorak objekata e_i ($i=1..x_1$) opisani skupom varijabli v_j ($j=1..y$) izvučenih iz populacije V. Jednostavnim Hadamardovim pridruživanjem vrijednosti odabranim objektima dobiva se matrica brutto podataka $B = E \otimes V$ s elementima b_{ij} ($i=1..x_1, j=1..y$). Standardizirajmo podatke u matrici B tako da je $z_{ij} = (b_{ij} - m_j) / \sigma_j$, pri čemu je vektor aritmetičkih sredina sadržan u $m = \sum b_j / x$, dok su standardne devijacije definirane kao $\sigma = \sqrt{\sum (b_{ij} - m_j)^2 / x}$. U matricnom obliku iz tako standardiziranih podataka dobivaju se korelacije varijabli pod modelom najveće vjerodostojnosti kao $R = Z' Z^{-1} / x$.

Iz matrice R se operacijom inverzije, uz uvjete $I = R * R^{-1}$, kao i $I = R^{-1} * R$, izračunavaju dijagonalni elementi kao $D_{jj} = \text{dg}(R)^{-1}$, što predstavlja upravo procjenu dijela neobjašnjene varijabilnosti svake pojedine varijable iz R procjenjene na temelju preostalih, odnosno unikatnosti varijabli (u_j^2). Naravno, komunaliteti varijabli iz R su jednostavno $h_j^2 = 1 - u_j^2$. Suma ovih komunaliteta može se definirati kao kriterij za zaustavljanje ekstrakcije latentnih dimenzija (faktora), baš kao što se mogu upotrijebiti i drugi poznati kriteriji ($GK > 1.00$, $\text{Sum}(\%SMC > 80 \text{ ili } 90)$, Scree tehnike,...). No, ono što jest činjenica, bez obzira na upotrebljeni kriterij, jest da je dio varijabilnosti jednostavno odbačen. Tu je vrhunac tzv. faktorske analize, neovisno o tome provode li se u daljnjim postupcima određeni postupci rotacije kao izrazi parsimonijskih ciljeva zbog postizanja jednostavne strukture latentnih dimenzija.

Međutim, prema teoriji harmonika i dekompoziciji totalnih podataka i totalnog varijabilnosti (Bonacin, 2004) u analizi procesa, na opisani način se favoriziraju takozvani harmonici višeg reda, tj. oni s većom amplitudom a manjom frekvencijom, jer su i primjetni baš zato što im je amplituda izražena. Kombinacijom nekoliko upravo takvih viših harmonika nastaju svojstvene amplitudne pojave, koje nisu ništa drugo nego faktori, tj. latentne dimenzije. Međutim, u takvim modelima, izrazito se zanemaruju niži visokofrekventni harmonici malih amplituda, ali koji i dalje postoje. U slučaju da imamo veći broj varijabli (npr. $m=50$) i ako po nekom čak jako objektivnom kriteriju zadržavamo manji broj «bitnih faktora», recimo 5, tada je sasvim sigurno da najmanje 45 frekventnih područja našeg problema uopće nismo adekvatno obradili niti analizirali. Možda zaista nije isključeno da su zanemarljivi, ali nije isključeno i da su možda zaista važni.

Takvu potvrdu ili negiranje značaja nam nipošto ne može dati količina informacija koju dakle, ti harmonici emitiraju, jer ono što je «sitno» ne mora biti i nevažno. Najbolji primjer za to je gotovo trivijalna situacija prema kojoj treba «iskočiti» izvan standardnih klišeja promišljanja: ukoliko promatramo nekog čovjeka u prostoru mase, registrirat ćemo npr. 75 kg tijela (prvi faktor), daljnjih 3-4 kg razne opreme poput odjeće, obuće, putne torbe, kišobrana i slično (drugi faktor), te eventualno do 1 kg nekih tehničkih fragmenata, poput novčanika, ukrasnog lančića, ručnog sata, mobitela (treći faktor). U opisanoj situaciji potpuno ćemo zanemariti komarca koji mu je sletio na ruku, jer zaista ne predstavlja neki posebni teret ni masu. To očito vrijedi sve dok stvari promatramo u prostoru fizike tj. mase, no ako uključimo neki drugi prostor poput kemije, tada više onaj komarac nipošto ne mora biti nevažan, posebno ako prenosi virus ce-ce bolesti ili gripe, a budući je sve povezano, u slučaju da obolimo, sva ona pusta masa neće nam biti od koristi. Kako se vidi, zanemarivanje informacija može dovesti do toga da se zanemare i one bitne, iako s pozicije početnog modela ne moraju tako izgledati. Međutim, u znanstvenom smislu, mi ne želimo u tolikoj mjeri biti ovisni o početnim uvjetima ili subjektivnom sagledavanju početne pozicije. Stoga ćemo u naš model uključiti i analizu reziduala, što se može izvesti na više načina, a za ovaj prikaz odabran je najjednostavniji.

U faktorskom modelu se uvijek vrši redukcija dimenzionalnosti bilo to sukcesivnim koracima, bilo to nekom transformacijom cijelog raspoloživog seta ((tri)dijagonalizacijom,...) do eigen vrijednosti i svojstvenih vektora bilo kako definirane operacionalne matrice. Nakon određivanja broja zadržanih dimenzija (k), u svim tim situacijama lako se određuje dio varijabilnosti podataka za koji su odgovorni «zadržani» faktori tj. latentne dimenzije. Ono što preostaje jest rezidual. Prema klasičnim, ali ne i adekvatno teorijski utemeljenim postavkama, za pojavu rezidualnih relacija među podacima odgovorni su «slučajni» faktori, koje dakle ne možemo promatrati u kontekstu istih pravila. Međutim, već Guttman (Bonacin, 2004, Šoš i Rado 1999, Bonacin 2006 a i b) je dokazao da su moguće nenulte i neslučajne relacije među tzv. nezajedničkim faktorima, što direktno vodi u elemente teorije reprezentativnosti, ali također i u situaciju po kojoj analiza reziduala postaje obavezni dio svake ozbiljne faktorske analize, s direktnim reperkusijama i u diskriminativne, i regresijske, i kanoničke, i taksonomske analize, a time očito i u analize promjena. U ovom članku obradit će se elementarne reperkusije u faktorskom modelu, zbog očite jednostavnosti i značenja u svakodnevnom radu. Neka je dakle, potpuna matrica asocijacija upotrebljenih varijabli zadana kao R , i neka je dio te matrice koji reproduciraju zadržani faktori jednak baš F , tada je očito da se u matrici $S = R - F$, nalaze dijelovi asocijacija među varijablama za koje nije odgovorno k zadržanih faktora tj. latentnih dimenzija.

Ova matrica S nije potpunog ranga, ali također ima određena svojstva koja dopuštaju operacije nad njom. Najjednostavniji način je analiza koherentnosti reziduala, tj. generalizacija multigrupne metode, pri čemu je selektorska matrica definirana samim jedinicama. Predmultiplikacija matrice S jediničnim vektorom $\mathbf{1}$ proizvodi vektor projekcija $\mathbf{v}$ rezidualne dimenzionalnosti na početne varijable i u suštini predstavlja generalizirani faktor reziduala nakon ekstrakcije tzv. bitnih informacija početne metrice R . Ova operacija još uvijek zadovoljava kriterije entropijske funkcije $f(e) \rightarrow R_0$ (gdje R_0 jest nula matrica dimenzija $m \times m$) pod pretpostavkom da među rezidualnim faktorima nisu asocijacije zadane samim nulama, u kojem slučaju bi već početna matrica R izgubila gramianska svojstva i puni rang, pa, prirodno, o rezidualima ne bi bilo ni govora.

Numerički primjeri

Primjer 1. Edukacija

Uzorak od 249 učenika muškog spola uzrasta 7 godina (+/- 2 mjeseca) opisan je s 26 varijabli na samom ulasku u prvi razred osnovne škole. Primjenjene varijable bile su: visina tijela (AVIT), duljina noge (ADUN), duljina ruke (ADUR), dijаметar ručnog zgloba (ADRZ), dijаметar koljena (ADIK), širina ramena (ASIR), širina zdjelice (ASIK), tjelesna težina (ATEZ), opseg podlaktice (AOPL), opseg podkoljenice (AOPK), opseg grudnog koša (AOGK), kožni nabor nadlaktice (AKNN), kožni nabor leđa (AKNL), kožni nabor trbuha (AKNT), koraci u stranu (MKUS), poligon natraške (MPOL), ravnoteža otvorenim očima (MP2O), pretklon u sjedu raznožno (MPRR), taping rukom (MTAP), taping nogom (MTAN), skok u dalj s mjesta (MSDM), bacanje loptice u daljinu (MBLD), trčanje 20 m s visokim startom (M20V), podizanje trupa iz ležanja (MDTS), izdržaj u visu zgibom (MVIS) i istrajno trčanje 3 minute (MT3M).

	T	F	h2
AVIT	0.11	0.14	0.02
ADUN	0.26	0.34	0.11
ADUR	0.17	0.22	0.05
ADRZ	0.04	0.06	0.00
ADIK	0.10	0.12	0.02
ASIR	0.00	0.01	0.00
ASIK	-0.23	-0.30	0.09
ATEZ	-0.04	-0.05	0.00
AOPL	-0.08	-0.10	0.01
AOPK	-0.16	-0.20	0.04
AOGK	-0.08	-0.10	0.01
AKNN	-0.01	-0.01	0.00
AKNL	0.08	0.11	0.01
AKNT	-0.03	-0.04	0.00
MKUS	0.77	0.98	0.97
MPOL	0.78	1.00	1.00
MP2O	0.54	0.70	0.48
MPRR	-0.06	-0.08	0.01
MTAP	0.11	0.14	0.02
MTAN	0.44	0.57	0.32
MSDM	0.40	0.52	0.27
MBLD	0.03	0.04	0.00
M20V	0.35	0.45	0.20
MDTS	0.10	0.13	0.02
MVIS	0.58	0.74	0.55
MT3M	-0.30	-0.38	0.15
P	1.28	Var	4.35
		%	16.72

Kako se jasno vidi u tabeli 1. rezidual, nakon ekstrakcije 6 zadržanih latentnih dimenzija po GK kriteriju, zadržava još uvijek 16.72 % informacija koje se mogu podvesti pod model različit od slučajnog. U tom modelu primjećuju se saturacije F koje pokazuju da postoji bimodalni karakter dijela informacija. Na pozitivnom polu su sve redom varijable koje predstavljaju strukturne elemente **mreže kapacitivnih temelja** bioloških funkcija (visina, duljina noge i ruke, dij. koljena, ravnoteža, frekvencija kanala, eksplozivnost, izdržaj). Na negativnom polu su sve redom varijable koje opisuju **nadgradnju** i učenje (opsezi, koordinacija, trčanje).

Vidljivo je da se na kapacitivne strukturne elemente teže može utjecati, dok se na segmente koji definiraju nadgradnju dugotrajnim transformacijskim postupcima može izvršiti adekvatan utjecaj. Ovakva interpretacija očito oslikava rezidualnu analizu kao jako bitnu i pokazuje da olako proglašavanje rezidualnih relacija nevažnima nema dobro utemeljenje, te se predlaže obvezno uključivanje analize reziduala u svaki ozbiljan projekt.

Tabela 1. Rezidualni faktor

(T = transformacijski vektor, F = rezidualni faktor, h2 = komunalitet)
(P = ponder za normalizaciju, V i % = varijabilitet i postotak faktora)

Primjer 2. Rekreacija

Za svrhe jednog pilot istraživanja izrađena je posebno konstruirana anketa s 24 pitanja koja su se odnosila na stavove entiteta o kineziološkoj rekreaciji i turizmu. Skala je bila Likhertovog tipa sa 5 ponuđenih tvrdnji u rasponu od maksimalnog slaganja do potpunog neslaganja, te s nekoliko izraženih zahtjeva kojom se ovakva skala interno validira i provjerava. Bila su uključena slijedeća pitanja/varijable: SRRI – Rijeke i jezera nemaju većih mogućnosti za zadovoljenje turističkih interesa, SRPP – U turističku ponudu treba puno više uključiti pješačke ture i lakše planinarenje, SRDU – Duhovna svetišta i rekreacija potpuno se međusobno nespojivi pojmovi, SRSA – Nemoguće je dobro pripremiti odgovarajuće sadržaje zbog prevelikih razlika turista u željama, SRPO – Prevelike su razlike među pojedincima u turizmu da bi se moglo išta ozbiljno napraviti, SRTU – O kompletnoj ponudi u turizmu isključivo moraju brinuti turoperatori, tj. oni koji ih i dovode, SRRE – Treba javno istaknuti ponudu sadržaja, a turisti ako to žele neka se izvole uključiti, SRKT – Kineziološka rekreacija je iznimno važna za turizam, SRTN – Nije moguće ozbiljno djelovati sportom u turizmu zbog opće situacije, SRNV – Sredstva koja treba izdvojiti za abavljenje sportom u turizmu su previsoka, SRTU – Turisti dolaze samo kako bi se malo odmorili i relaksirali, a ne naporno vježbali, SREN – Angažman rekreacijom u turizmu je prevelik a efekti praktično nikakvi, SRMZ – Turistima je samo do mora, kupanja i zabave, SRES – Rekreacija u turizmu je besmislena jer se oni poslije vrate kući i na posao a efekta nema, SRND – Turisti niti ne znaju što hoće, treba ih naučiti da se odrede što hoće, pa im to dati, SRTK – Turizam je iznimno važan za kineziološku rekreaciju, SRKA – Stručna razina kadra u rekreaciji je preniska za ikakve ozbiljne efekte, SRPZ – Posjete povijesnim znamenitostima nemaju u suštini nikakve veze s rekreacijom, SRPL – Potrebno je isplanirati unaprijed 5-10 godina što se može pripremiti i onda to stalno provoditi, SRST – Nema se što pitati turiste, jer su oni nekompetentni pa o tome može odlučivati samo struka, SRTP – Turistička ponuda treba sadržavati obvezno sportsku i sličnu ponudu, SRME – Turistima treba samo medicinski tretman, a ne rekreativni, SRPN – Planine su izuzetno zahvalne sredine za primejnu rekreativnih sadržaja, SRUN – Sve je to zamišljeno samo kako bi se turistima izvukla sredstva iz novčanika.

	T	F	h ²
SRRI	0.11	0.08	0.01
SRPP	0.13	0.10	0.01
SRDU	-0.20	-0.16	0.02
SRSA	-0.11	-0.08	0.01
SRPO	0.10	0.07	0.01
SRTU	1.28	1.00	1.00
SRRE	0.00	0.00	0.00
SRKT	-0.44	-0.34	0.12
SRTN	-0.19	-0.15	0.02
SRNV	0.31	0.24	0.06
SRTU	0.73	0.57	0.33
SREN	-0.11	-0.08	0.01
SRMZ	0.15	0.11	0.01
SRES	-0.12	-0.09	0.01
SRND	0.16	0.13	0.02
SRTK	0.36	0.28	0.08
SRKA	0.29	0.23	0.05
SRPZ	0.60	0.47	0.22
SRPL	0.08	0.06	0.00
SRST	-0.04	-0.03	0.00
SRTP	-0.04	-0.03	0.00
SRME	0.06	0.05	0.00
SRPN	0.39	0.30	0.09
SRUN	-0.26	-0.20	0.04
P	0.78	V	2.11
		%	8.78

Tabela 2. Rezidualni faktor

(T = transformacijski vektor, F = rezidualni faktor, h² = komunalitet)
(P = ponder za normalizaciju, V i % = varijabilitet i postotak faktora)

Anketa je u prvom koraku aplicirana na studentima fizičke kulture u Zapadnom Mostaru. U konačni efektiv ušlo je 48, po brojnim parametrima, pozitivno selektiranih entiteta. Vjerodostojnost predložene ankete, naravno, bit će višestruko provjerena, na različitim uzorcima, stratificiranim po ciljanim obilježjima, koja su potencijalno bitna za varijabilitet rezultata. Inicijalni podaci su prikazani histogramima, a zatim je primjenjena uobičajena i neophodna normalizacija, kako bi se podaci pripremili za multivarijantne obrade. Prije provođenja faktorske analize s orthoblique rotacijom, još je izvršena redirekcija svih varijabli tako da veći rezultat znači i bolji rezultat. Prema podacima u tabeli 2. vidljivo je da je rezidual nakon ekstrakcije 4 najvažnija faktora, još uvijek sadrži dio informacija koje je nemoguće eliminirati na temelju same elementarne entropije matrice asocijacija. Kao i u prethodnom primjeru iz edukacije, mogu se zapaziti pozitivne i negativne vrijednosti rezidualnih saturacija na rezidualni faktor. U pozitivnom dijelu vidimo stav prema turoperatorima, prema odmoru i relaksaciji turista, prema znamenitostima, planinama, stručnosti kadra, dok su na negativnom polu stavovi prema važnosti rekreacije za turizam, «sredstvima iz novčanika», općoj situaciji koja se reflektira na sportsko djelovanje u turizmu i poziciji duhovnih središta.

Nije teško prepoznati da se taj faktor raslojava na bimodalni kompozit upravo sukladno stvarnim vrijednostima rekreacije u turizmu (pozitivni dio) i sukladno stavovima o općim društvenim kretanjima koja se drže odgovornim za manifestacije u turizmu (negativni dio). Na taj način je za primjetiti bitno razlikovanje dijela informacija koje opisuju **aktivni stav prema rekreaciji i turizmu**, te one koje opisuju **pasivnu društveno uvjetovanu konzumaciju**. Očito je da je predloženi model analize reziduala donio neke jako zanimljive dodatne informacije, koje bi inače bile zanemarene, iako je tih informacija čak 8.78 %.

Iako algoritam producira još niz tabela, i iako je moguće raspraviti još mnoge detalje modela, za ovu priliku čini se dovoljno ilustrativnim najjednostavniji prikaz u tabelama 1. i 2.

Zaključak

Predložen je i realiziran model analize reziduala koji na najjednostavniji mogući način pokazuje da u «odbačenom» varijabilitetu, nakon ekstrakcije «značajnih» latentnih dimenzija na temelju elementarne entropije matrice asocijacija, još uvijek ostaje značajan dio informacija. Taj dio informacija analiziran je u nekoliko desetaka slučajeva, a za potrebe ovog rada prikazani su primjeri dva slučaja, jedan iz edukacije i drugi iz rekreacije. U oba slučaja pokazalo se da producirani rezidualni faktor sadrži bitne, zanimljive i interpetabilne informacije, što je sasvim dovoljan razlog da se opisani postupak predloži za obveznu proceduru u analizi bilo kojih relacija, a posebno u analizi faktorskih reziduala.

Literatura

1. Bonacin, D. (2004). Uvod u kvantitativne metode. Vlastito izdanje, Kaštela.
2. Bonacin, D. (2006 a). ITERIM - Solution for detection and optimisation of variance in applied variable system. *Anthropological status and physical activity of children and youth - Interdisciplinary scientific conference*, NoviSad. Proceedings: 289-294.
3. Bonacin, D., Blažević, S., Bilić, Ž. (2006 b). SLIKOM – Algoritam za identifikaciju sličnosti uzoraka. *HomoSporticus*, 9, 2:6-10.
4. Cote, J.A., Buckley, R.M. (1987). Estimating Trait, Method, and Error Variance: Generalizing across 70 Construct Validation Studies. *Journal of Marketing Research*, Vol. 24, No. 3: 315-318.
5. Cooley, W.W., Lohnes, P.R. (1971). Multivariate data analysis. John Wiley and sons.inc. New York.
6. Fulgosi, A. (1979). Faktorska analiza. Školska knjiga, Zagreb.
7. Grlić, D. (1967). Filozofija (Školski leksikon). Panorama, Zagreb.
8. Harman, H.H. (1970). Modern factor analysis. The university of Chicago.
9. Jacob, F. (1978). Logika živog. Nolit, Beograd.
10. Johnson, R. A., Wichern, D. W. (1992). Applied multivariate statistical analysis. Prentice-Hall.
11. Krković, A., Momirović, K., Petz, B. (1966). Odabrana poglavlja iz psihometrije i neparametrijske statistike. Društvo psihologa i RZZZ SRH, Zagreb.
12. Momirović, K. (1984). Kvantitativne metode za programiranje i kontrolu treninga. Statističke metode 1. FFK, Zagreb.
13. Mraković, M. (1992). Uvod u sistematsku kineziologiju. FFK, Zagreb.
14. Pauše, Ž. (1978). Vjerojatnost, informacija, stohastički procesi. Školska knjiga, Zagreb.
15. Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business*, Vol. 53, No. 1: 61-65.
16. Rao, C. R. (1973). Linear Statistical Inference and its Applications. John Wiley & Sons.
17. Serdar, V., Šošić, I. (1981). Uvod u statistiku. Školska knjiga, Zagreb.
18. Šošić, H., Rađo, I. (1998). Mjerenje u kineziologiji. Fakultet za fizičku kulturu Sarajevo.