

## **KAKAOBEJZ: MAKRO ZA AJTEM ANALIZU DIHOTOMNO SKOROVANIH AJTEMA – TEORIJA AJTEMSKOG ODGOVORA<sup>1</sup>**

**Stanislav Fajgelj<sup>2</sup>,**

Fakultet za pravne i poslovne studije, Novi Sad

**Bojan Janičić**

Odsek za psihologiju, Filozofski fakultet, Novi Sad

*U radu je prikazan KakaoBejz, makro iz serije Kakao, koja je namenjena ajtem analizi mernih instrumenata iz oblasti nauka o ponašanju. Programi su izrađeni u Matrix jeziku, kao makroi u okviru statističkog paketa SPSS. To su makro programi otvorenog koda (open source), besplatni su, a za njihovo odvijanje potreban je statistički paket SPSS i vrlo su jednostavni za upotrebu.*

*Makro je namenjen ajtem analizi zasnovanoj na teoriji ajtemskog odgovora. Za ocenjivanje parametara ajtema koristi se metoda maksimalne verodostojnosti, koja je dopunjena a priori očekivanjima izvedenim iz Bayesove teoreme. Makro sadrži osnovne indikatore kontrole konvergencije i omogućuje korisniku da oceni da li je iterativni proces protekao korektno. Pomoću ulaznih opcija moguće je kontrolisati odvijanje programa, birati model ajtemskog odgovora i zadavati početne vrednosti. KakaoBejz prihvata samo dihotomno skorovane ajteme, jednakog usmerenja u odnosu na predmet merenja.*

***Ključne reči:*** ajtem analiza, teorija ajtemskog odgovora, maksimalna verodostojnost, Bayesova teorema

---

<sup>1</sup> Rad u okviru projekata broj 149027 „Integralni razvoj, fizička aktivnost i aberantno ponašanje predškolske dece“, kojeg sufinansira Ministarstvo za nauku i tehnološki razvoj i broj 149008b „Psihološke karakteristike društva u tranziciji“ kojeg finansira Ministarstvo nauke i zaštite životne sredine Republike Srbije.

<sup>2</sup> E-mail: stanef@beocity.net.

## **Uvod**

Skup programa pod zajedničkim nazivom Kakao namenjen je ajtem analizi mernih instrumenata iz oblasti nauka o ponašanju. Programi Kakao namenjeni su psiholozima i svima onima koji konstruišu ili evaluiraju takve merne instrumente. Takođe, programi su namenjeni studentima, kao pomoćno sredstvo u univerzitetskoj nastavi. Kakao se nastavlja na seriju programa „*rtt9*“, „*rtt10*“ i dr., čiji je glavni autor bio Konstantin Momirović u oblasti klasične testne teorije (KakaoKTT), ali i dodaje novu oblast teorije ajtemskog odgovora (TAO). Programi KakaoBejz i KakaoMiks, pisani kao makroi u metajeziku Matrix, ne mogu zameniti neke poznate komercijalne programe, kao što su Winsteps (Linacre, 2002), Bilog i Multilog (Du Toit, 2003) ili RUMM (Andrich, Sheridan i Luo, 2004). Oni nemaju brojne prateće pokazatelje koje ovi programi imaju, nemaju robusnost i brzinu izvođenja, nemaju grafičke prikaze itd. Međutim, pošto je softver za TAO ajtem analizu skup i slabo dostupan psiholozima generalno, a pogotovo studentima, smatramo da će naši makroi olakšati praksu konstrukcije i evaluacije mernih instrumenata u priličnoj meri<sup>3</sup>.

### *Ajtem analiza po teoriji ajtemskog odgovora*

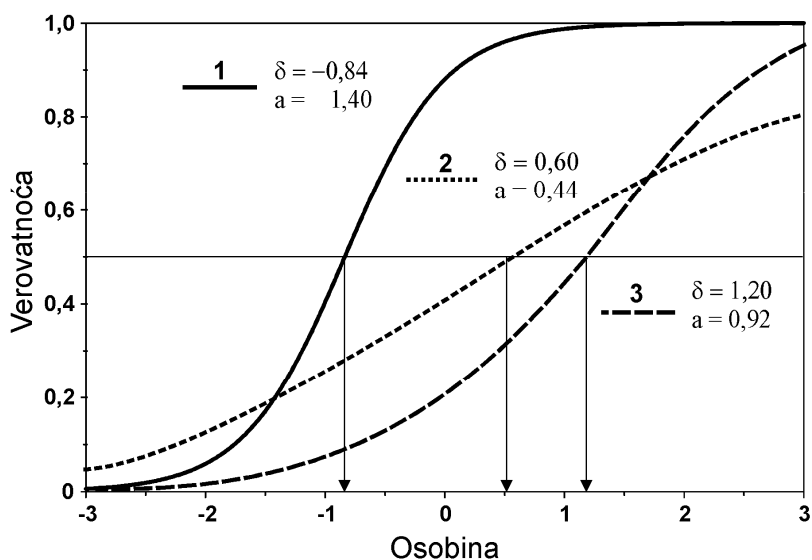
Teorija ajtemskog odgovora je usmerena na ajtem, a ne na testni skor. U TAO, zapravo, ne postoji testni skor nego mera osobine kod ispitanika. TAO je čak usmerena na pojedinačan odgovor, ona matematički egzaktno modelira verovatnoću odgovora pojedinačnog ispitanika na dati ajtem (vidi kasnije izraz 1). Dok je KTT deterministički model, bitna odlika TAO je da se doprinos kako ajtema oceni osobine ispitanika, tako i ispitanika oceni ajtema zasniva na verovatnoći.

Sledeći ova svojstva, TAO je reformulisala ne samo pravi skor i skor greške, nego i druge pojmove KTT, uključujući i merna svojstva. Pre svih možemo spomenuti karakterističnu krivu stavke (KKS), koja osim što predstavlja model ajtemskog odgovora, odražava i pogled na merenje ponašanja uopšte. Na slici 1 prikazane su KKS jednog lakog i visoko diskriminativnog ajtema (ajtem 1), jednog srednje teškog ali nisko diskriminativnog ajtema (ajtem 2) i jednog teškog a srednje diskriminativnog ajtema (ajtem 3). Lokacija ajtema na kontinuumu osobine definiše njegovu težinu, nagib ajtema u tački preloma definiše njegovu diskriminativnost, a verovatnoća da će ispitanici odgovoriti pozitivno

---

<sup>3</sup> Svi Kakao makroi mogu se dobiti direktno od autora putem e-maila [janicic@uns.ns.ac.yu](mailto:janicic@uns.ns.ac.yu), ili na: <http://www.cpijm.org.yu/sfstuff/index.htm>.

na sva tri ajtema dobija se kombinovanjem verovatnoća odgovora na pojedinačne ajteme. Na tu združenu verovatnoću utiče nivo osobine ispitanika, kao i težina i diskriminativnost ajtema. Uočava se još jedno važno svojstvo, a to je da se težina ajtema i osobina iskazuju na istoj skali.



**Slika 1.** Karakteristične krive tri dihotomna ajtema različite težine i diskriminativnosti

Nastojanje da se postigne nezavisnost ocenjivanja parametara ajtema od konkretnog postignuća ispitanika (i obrnuto) centralna je tačka teorije ajtemskog odgovora, a najdoslednije je izvedena u Raschovom modelu – kao specifična objektivnost. Specifična objektivnost je učinila da se danas bez TAO ne mogu zamisliti detekcija diferencijalnog funkcionisanja ajtema (DIF) i jednačenje testova (horizontalno i vertikalno). Nezamenljiva je, takođe, u računarskom adaptivnom testiranju (KAT) i u fleksibilnom skorovanju testova koji se sastoje od ajtema različitih formata.

Naravno, ništa nije idealno, pa tako i teoriji ajtemskog odgovora predstoji praktična verifikacija sa istorijskom perspektivom. Na primer, TAO ima jedan problem, koji nije teorijski, nego je tehnički, ali vrlo ozbiljan. To je ocenjivanje parametara koje je matematički i računski vrlo kompleksno, intenzivno i zavisno od metode i podataka (u ovom poslednjem se, doduše, ne razlikuje od faktorske analize i sličnih tehnika). Ovaj problem se uočava u svojoj punoj veliči-

ni kada se pristupi pisanju softvera za TAO, a bez računara je TAO neprimenljiva.

### *Ocenjivanje parametara u teoriji ajtemskog odgovora – dihotomni ajtemi*

Centralni deo svakog modela teorije ajtemskog odgovora je veličina  $Z_j = a_j(\theta - \delta_j)$ , koja u normalnim modelima predstavlja  $z$ -skor, a u logističkim modelima *logit* (od "log-odds unit"). Parametri  $a_j$  i  $\delta_j$  su parametri ajtema i odnose se na diskriminativnost i težinu, a parametar  $\theta$  je parametar ispitanika i odnosi se na nivo latentne osobine (Van der Linden i Hambleton, 1997, Baker i Kim, 2004). Program kojim se ovde bavimo namenjen je evaluaciji testova, konkretnije ajtem analizi, i u tom smislu se ne bavi ocenjivanjem  $\theta$ .

Najjednostavniji slučaj predstavljaju dihotomno skorovani ajtemi, sa dve kategorije odgovora: pozitivnom i negativnom. Uopšteni model odgovora na dihotomni ajtem  $j$  može se predstaviti funkcijom  $P_j(\theta) = P(a_j, \delta_j, \theta)$ , u kome  $P_j(\theta)$ , po konvenciji, predstavlja verovatnoću pozitivnog odgovora kod zadatog nivoa  $\theta$ . Napominjemo da ćemo u ovom tekstu koristiti samo dvoparametarski model (2P). Funkcija  $P(a_j, \delta_j, \theta)$  ima dva osnovna oblika: logistički i normalni. Iako mnogi autori smatraju da je normalni model "prirodniji", u statističkom smislu, u praksi, ali i u teoriji, vrlo je prisutan logistički model (Baker i Kim, 2004). Sledstveno tome, naši programi su takođe bazirani na logističkim modelima.

Funkcija  $P_j(\theta)$  nije ništa drugo do *karakteristična kriva stavke – KKS* sa slike 1 (engl. Item Characteristic Curve – ICC). Ona spada u kumulativne krive verovatnoće, ili kako se u žargonu kaže, u ogive. McDonald (1999) ovu ogivu naziva veznom funkcijom jer povezuje verovatnoću pozitivnog odgovora sa linearnom nezavisnom varijablom – latentnom osobinom  $\theta$  i u logističkim modelima glasi:

$$P_j(\theta) = \frac{1}{1 + e^{-Z_j}} = \frac{e^{Z_j}}{1 + e^{Z_j}}, \quad (1)$$

pri čemu smo  $Z_j$  ranije definisali kao logit:  $Z_j = a_j(\theta - \delta_j)$ . Parametar  $\delta_j$  se iskazuje u  $\theta$  vrednostima i predstavlja lokaciju ajtema na kontinuumu osobine, odnosno njegovu težinu. Parametar  $a_j$  ima višestruku interpretaciju. On predstavlja diskriminativnost ajtema, ali i nagib KKS u tački preloma. Takođe, obrnuto je proporcionalan standardnoj devijaciji logističke funkcije, pa se zato naziva faktorom skaliranja KKS. Za potrebe ocenjivanja parametara pokazala

se korisnom takva transformacija koja linearizira KKS, pretvarajući je u regresionu liniju iz koje se vrednost logita (a ne verovatnoća pozitivnog odgovora) može predvideti iz osobine. U ovoj linearnoj verziji, logit dobija oblik  $Z_j = \zeta_j + \lambda_j \theta$ , uz sledeća dva odnosa:  $\delta_j = -\zeta_j/\lambda_j$  i  $a_j = \lambda_j$ . Veličina  $\zeta_j$  se, u skladu sa regresionom terminologijom, naziva odsečkom, a  $\lambda_j$  nagibom, pri čemu on logički i numerički zadržava značenje parametra  $a_j$ .

Uzmimo da imamo test od  $m$  dihotomno skorovanih ajtema, administriran na uzorku od  $n$  ispitanika. Pretpostavimo da su ajtemi skorovani sa 0 i 1, što znači da neki ispitanik  $i$  može da ima sklop odgovora na  $m$  ajtema: 01011...0. Sklop odgovora  $i$ -tog ispitanika možemo predstaviti vektorom  $\mathbf{u}_i$ . Sklopovi odgovora ispitanika su osnov za ocenjivanje parametara kako ajtema, tako i ispitanika.

### *Metoda maksimalne verodostojnosti*

U kolokvijalnom engleskom jeziku izraz "likelihood" je jedna od varijanti pojma "verovatnoće". Embretson i Rajs (Embretson i Reise, 2000) kažu da kada se relativna frekvencija na velikom broju događaja računa pre nego što se oni dese, onda je to verovatnoća. Međutim, kad se relativna frekvencija računa tek pošto se oni dogode, onda je to verodostojnost (likelihood). Tačnija definicija je da *funkcija verodostojnosti*  $L(p | X)$  određuje verovatnoće dobijenih podataka  $X$  u zavisnosti od vrednosti parametra  $p$ . Uporedimo to sa definicijom verovatnoće:  $P(X | p)$ . Parafrazirajući Edvardsa (Edwards, 1972), u verovatnoći  $X$  je varijabla a  $p$  konstanta, dok je u verodostojnosti  $p$  varijabla a  $X$  konstanta. Takođe, verodostojnost nije isto što i verovatnoća, a to se najbolje dokumentuje time što integral funkcije verodostojnosti nije 1 (Edwards, 1972).

Ono što nas ovde interesuje, to je vrh funkcije verodostojnosti, odnosno mod, odnosno *maksimum*. Mesto maksimuma na apscisi određuje vrednost parametra  $p$  kod koje je verodostojnost dobijenih podataka najveća. U našem slučaju to su oni parametri ajtema pod kojima dobijeni podaci imaju najveću verovatnoću (verodostojnost) pojavljivanja. Ovakav način ocene parametara naziva se *ocenjivanjem putem maksimalne verodostojnosti* (engl. maximum likelihood estimation – MLE).

Metoda maksimalne verodostojnosti je danas veoma popularna jer je fleksibilna i prilagodljiva različitim namenama, a vrednosti parametara koje su dobijene ovom metodom imaju mnoga poželjna svojstva. Jedno od neželjenih svojstava je to što se funkcija verodostojnosti kao i način određivanja njenog maksimuma moraju izvesti za svaku aplikaciju posebno. Drugo, za maksimiziranje funkcije potrebna je "teška" matematika. Treće, vrednosti maksimuma

ne mogu se izračunati u zatvorenoj formi, nego se koriste složene i računski veoma intenzivne numeričke metode.

Ako se sada vratimo na naš zadatak, opšti oblik funkcije verodostojnosti sklopa odgovora  $\mathbf{u}_i$  za ispitanika  $i$  možemo napisati kao:

$$L(\boldsymbol{\xi} | \mathbf{u}_i, \theta_i) = \prod_{j=1}^m P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}}, \quad (2)$$

pri čemu je  $Q_j(\theta_i) = 1 - P_j(\theta_i)$ , a  $P_j(\theta_i)$  se razlikuje od izraza 1 samo po tome što se odnosi na konkretnog ispitanika, sa nivoom osobine  $\theta_i$ . Vektor  $\boldsymbol{\xi}$  sadrži sve parametre svih ajtema. Kao što je rečeno u uvodu, za ocenjivanje parametara u  $\boldsymbol{\xi}$ , a na osnovu dobijenih podataka iz  $\mathbf{u}_i$ , koji opet zavise od nivoa osobine kod ispitanika, na raspolaganju nam je metoda maksimalne verodostojnosti. U osnovi, maksimum funkcije verodostojnosti se dobija iz prvog izvoda te funkcije.

Za svaki TAO model se mora izvesti konkretna funkcija verodostojnosti i prilagoditi konkretnom algoritmu. Kao prvo, funkcija  $P_j(\theta)$  se izvodi iz konkretnog modela. Zatim, umesto za jednog ispitanika, u stvarnosti se radi sa matricom sklopova odgovora  $\mathbf{U}$ . Takođe,  $\theta$  se klasifikuje u neki krajnje svedeni broj razreda i svaki sklop odgovora dobija mesto u nekom od tih razreda. Za taj slučaj se koristi diskretan oblik funkcije verodostojnosti i to njen negativni logaritam  $-\ln L$  (Hambleton, Swaminathan i Rogers, 1991, Baker i Kim, 2004).

Sledeća varijacija u primeni ML metode za ocenu parametara se sastoji u uvođenju Bayesove teoreme, koja u načelu pomaže određivanju parametara ukoliko je poznata njihova a priori distribucija. A priori distribucija parametara se u našim programima svodi na momente distribucije parametra diskriminativnosti  $\mu_\alpha$  i  $\sigma_\alpha$ . Ovi momenti, koji se nazivaju i hiperparametrima (parametri parametara), zadaju se na početku i kasnije se "dodaju" na standardne formule za prvi i drugi izvod funkcije verodostojnosti.

*Expectation-maximization* algoritam (EM) se definiše kao iterativna tehnika maksimalne verodostojnosti kada su prisutne nedostupne slučajne varijable. U našem slučaju nedostupna varijabla je latentna osobina. Ono čime raspoložemo je opažena matrica  $\mathbf{U}$ . Znamo da ona zavisi od skupa parametara ajtema  $\boldsymbol{\xi}$ , koje trebamo da odredimo, i od latentne osobine  $\theta$  koju ne možemo da izmerimo, bar ne direktno.

U E-koraku se izračunavaju verovatnoće  $P_j(\theta_i)$  za svaku od kvadraturnih tačka. Na osnovu toga se izračunavaju i očekivane verovatnoće tačnih odgovora

za svaki ajtem i svaku kvadraturnu tačku, a iz toga i funkcija verodostojnosti iz 2 za svaku tačku – razred. Za izračunavanje verovatnoća koristi se model iz izraza 1. Težine kvadrature tačaka određujemo na osnovu normalne distribucije kao prirodne distribucije za latentnu osobinu, a broj intervala (tačaka ili čvorova) zadaje korisnik.

Nakon toga se prelazi u M-korak, u kome se određuje maksimum funkcije verodostojnosti. Pošto se maksimum ne može naći u zatvorenoj formi, koristi se iterativna tehnika Fišerovog skorovanja (varijanta Newton-Raphsonove metode). U osnovi te tehnike je formiranje matrice prvih i drugih izvoda. Matrica drugih izvoda je kvadratna matrica koja se u matematici naziva Heseovom i posebno je značajna zbog toga što se množenjem sa  $-1$  iz nje dobijaju Fišerove informativnosti svih parametara  $I(\theta)$  (u dijagonali), a iz informativnosti se mogu izračunati standardne greške kao  $1/\sqrt{I(\theta)}$ . Ova matrica se transformiše i dodaje na parametre iz prethodnog EM ciklusa. Time se jedan EM ciklus (iteracija) završava i prelazi se na sledeći.

### *Logistički modeli sa normalnom metrikom*

Kao što je rečeno, logistički modeli dominiraju u TAO softveru, pa je to i naš slučaj. Međutim, još je 1952. dokazano da se normalna i logistička ogiva razlikuju manje od 0,01 u celom opsegu  $\theta$ , ukoliko se logistički model reskalira faktorom  $D=1,702$  (Baker i Kim). To se postiže jednostavnim uvođenjem faktora  $D$  u logit kao:  $Da_j(\theta - \delta_j)$ . Time, naravno, model ne postaje normalni, ali njegova metrika postaje, odnosno sada se parametri  $a$  i  $\delta$  iskazuju u  $z$ -vrednostima, a ne u logitima. Varijansa logističke distribucije je veća od varijanse normalne distribucije, pa je za isti parametar  $a$ , nagib logističke krive manji. Drugim rečima, da bi se dobio isti nagib krive, odnosno ista diskriminativnost ajtema, u normalnom modelu dovoljan je manji parametar  $a$ . Zato su parametri nagiba u normalnoj metrici niži od onih u logističkoj, dok je efekat na parametre težine minimalan.

### **Opšti opis programa**

Program je napisan u matričnom metajeziku Matrix u okviru statističkog paketa SPSS, a realizovan je kao makro u okviru istog paketa (SPSS, 2005). Makro izvedba omogućuje da se program jednostavno poziva (kao nova komanda SPSS-a) i da mu se jednostavno prenose parametri. Korisnik u redovnoj upot-

rebi ne mora da manipuliše izvornim kodom programa, niti da ima bilo kakvog dodira sa njime.

Nastojali smo da oprema svih Kakao programa bude što ujednačenija. To je dobro i za korisnika i za programere. Korisnik se susreće sa usaglašenom terminologijom i ispisom rezultata analize (output), a programerima je olakšano održavanje i razvoj programa. Osim KakaoBejza u seriji Kakao postoji i Kaka-oMiKS, makro za ajtem analizu bilo kakve smese dihotomnih i politomnih ordinalnih ajtema (po TAO). Takođe u fazi razvoja je i procedura za ajtem analizu polihotomnih nominalnih ajtema (Thissen i Steinberg, 1984).

### *Program KakaoBejz*

Program KakaoBejz, kao i ostali naši programi iz oblasti teorije ajtemskog odgovora, vode poreklo od Basic programa iz knjige Bejkera i Kima (2004). Da bi migracija u potpuno drugačije programsko okruženje bila efikasna, izmenjen je "skalarni" pristup iz Basica u "matrični" pristup, koji je neophodan i prirodan za Matrix okruženje. Programi su prerađeni za direktno učitavanje podataka i formiranje sklopova odgovora (matrica  $\mathbf{U}$ ) i njihovih frekvencija. Osim toga, programi su dopunjeni dodatnim ajtemskim pokazateljima, ispis rezultata generalno je obogaćen i ugrađena je kontrola grešaka u odvijanju.

KakaoBejz je namenjen ajtem analizi testa sastavljenog od *dihotomno ocenjivanih* ajtema. U psihologiji to mogu biti izvorno dihotomni ajtemi (tačno-netačno, da-ne), ili polihotomni ajtemi koji se ocenjuju kao tačno-netačno (po pravilu se radi o kognitivnim pitanjima sa višestrukim izborom). Kada je reč o kognitivnim ajtemima, polaritet ajtema je jasan, međutim, kod konativnih ajtema pretpostavlja se da je jedna kategorija negativna, a druga pozitivna, sa stanovišta iskazivanja prisustva osobine koja se meri. Takođe, pretpostavlja se da je ista kategorija (u numeričkom smislu) negativna, odnosno pozitivna kod svih ajtema u testu (npr. 0 i 1, ili 1 i 2).

Za ocenjivanje parametara koristi se tehnika marginalne maksimalne verodostojnosti, opisana u prethodnim poglavljima, realizovana primenom EM algoritma i dopunjena Bejzovim očekivanjima (MMLE/BME/EM). Program ima jednu osobitost. U ocenjivanju parametara ne učestvuje zapravo parametar  $a_j$ , nego  $\alpha_j = \ln a_j$ . To je posledica toga što su Bejker i Kim usvojili program PC-Bilog kao osnovu za svoj algoritam i izvorni Basic program. Unatoč tome što postoje neki matematički zdravi razlozi za korišćenje  $\alpha_j$  umesto  $a_j$ , posledica je ta da **KakaoBejz ne može da ima negativne diskriminativnosti**. Zbog toga će pogrešno skorovani ili nereflektovani ajtemi imati dramatične posledice na



tok iteracija. Sledeća posledica koju korisnik može primetiti je ta da se opcije *malfa* i *salfa*, istina, zadaju aritmetički, ali se koriste logaritamski.

S obzirom da je svaki programski paket razvojni projekt, treba očekivati dalji razvoj programa, pa prema tome i nove opcije, nove pokazatelje i nove mogućnosti. Sve što je bitno za najnoviju verziju programa i za njeno izvršavanje biće priloženo u obliku komentara na početku makroa.<sup>4</sup>

### *Pozivanje i parametri programa KakaoBejz*

KakaoBejz se aktivira na sledeći način:

```
include 'kakaobejz.sps'.
kakaobejz file={'datoteka'|*} ajtemi={spisak ajtema}/
metrika={log|norm} GKvadrat=n EMciklus=n float={0|1}
fishiter=n missing={omit|accept} malfa=n salfa=n
konv=n.
```

Sve što je rečeno za komandu *include* i za opcije *file* i *ajtemi* za makro KakaoKTT (v. rad u ovom časopisu) važi i ovde. Nijedna od opcija osim *ajtemi* nije obavezna. Ako se izostave, uzima se automatska vrednost.

*Metrika.* Metrika je "norm" ili "log" (automatski je "log"). Opcijom "norm" se u logit uvodi faktor *D*, na način kako je ranije opisano.

*Missing.* Dozvoljene opcije su "omit" ili "accept". Automatska vrednost je "omit", koja tretira ispuštene vrednosti na "listwise" način po terminologiji SPSS-a. Dakle, odbacuje se svaki slučaj koji na bar jednom ajtemu ima ispuštenu vrednost.

Opcijom "accept" se sve ispuštene vrednosti pretvaraju u 0, tj u negativan odgovor i svi slučajevi se prihvataju. U psihometrijskom pogledu to izjednačavanje nije opravdano ako je test ubrzan, a u statističkom pogledu je problematično ako ispuštenih vrednosti ima mnogo.

---

<sup>4</sup> Svi Kakao makroi mogu se dobiti direktno od autora putem e-maila janicic@uns.ns.ac.yu, ili na: <http://www.cpijm.org.yu/sfindex/sfindex.htm>.

*Malfa i salfa.* Ovim opcijama se zadaju aritmetička sredina i standardna devijacija distribucije hiperparametra, tj. a priori očekivanja parametra nagiba ( $\mu_\alpha$  i  $\sigma_\alpha$ ). Automatska vrednost je *malfa*=1 i *salfa*≈1,649, što odgovara tačnim vrednostima  $\mu_\alpha=0$  i  $\sigma_\alpha=0,5$ . Skrećemo pažnju na ranije opisanu parametrizaciju  $\alpha_j = \ln a_j$ , što znači da program interno radi sa logaritmom parametra nagiba. Smatrali smo da je korisniku komfornije da ovu  $M$  i  $\sigma$  zadaje aritmetički (u  $a$  jedinicama), a ne logaritamski, ali se u rezultatima analize vrednosti prikazuju kao  $\exp(\textit{malfa})$  i  $\exp(\textit{salfa})^2$ . Delovanje ovih parametara je opisano u poglavlju o kvalitetu parametara.

*Float.* Ova opcija utiče na proces modifikacije a posteriori distribucije parametra nagiba na osnovu a priori očekivanja. Naime, ako se stavi *float*=1, prilikom ocenjivanja parametra  $a$  koristiće se prosek parametara  $a$  svih ostalih ajtema kao Bejzovo očekivanje, a ako je *float*=0, koristiće se a priori hiperparametar za aritmetičku sredinu koju program na početku postavlja automatski. Drugim rečima, ako je opcija uključena, a priori očekivanja prosečne diskriminativnosti će postupno biti rekonstruisana iz samih podataka, a neće biti nametana inicijalno postavljena vrednost. Efekat ove opcije na rezultate je obično minimalan, ali će o tome kasnije biti još reči.

*GKvadrat.* Broj Gausovih kvadrturnih tačaka se u programu automatski postavlja na 10, a program sam izračunava tačke ne apscisi normalne distribucije (počev od  $z=-4,5$ ) i njihove težine (koje približno odgovaraju površinama – verovatnoćama). Težine se izračunavaju tako da njihova suma daje jedinicu. Podsećamo da kvadrturne tačke predstavljaju intervale razreda za pretpostavljenu distribuciju osobine  $\theta$  (u našem slučaju, normalnu).

*EMciklus.* Maksimalni broj EM iteracija koje će program izvršiti (ako se konvergencija ne postigne) automatski se postavlja na 25. Obično je to sasvim dovoljno. Broj EM ciklusa *presudno utiče* na dužinu rada programa. Povećanje broja ciklusa ne vodi nikuda ukoliko rešenje stalno osciluje, a smanjenje može dovesti do neoptimalnog konačnog rešenja, posebno parametra nagiba.

*Fishiter.* Broj Newton-Raphson iteracija automatski se postavlja na 6. Obično je to sasvim dovoljno, više od 10 nije produktivno, a i 4 može biti zadovoljavajuće.

*Konv.* Opcija *konv* (od "konvergencija") upravlja iterativnim delovima programa: EM ciklusima i Newton-Raphsonovima iteracijama (NR). To je izvedeno na sledeći način. Ako priraštaj bilo kog parametra za neki ajtem nije veći od "konv", NR iteracije se prekidaju, tj. završava se maksimizacija u  $M$ -koraku za taj ajtem u tom EM ciklusu i prelazi se na sledeći ajtem. Dalje, ako se takav slučaj desi za sve ajteme u jednom EM ciklusu, smatra se da je konvergencija

parametara svih ajtema dostignuta i EM iteracije se prekidaju. Automatska vrednost je 0,01.

### *Ispis rezultata programa KakaoBejz*

U uvodnom delu listinga korisniku se daju osnovne informacije o obradi, kao što su odabrani model, način tretiranja ispuštenih vrednosti, broj ispitanika, broj sklopova odgovora i sl. Zatim se prikazuju inicijalne vrednosti: težina kvadraturnih tačaka, parametara i hiperparametara. Zatim se daje informacija o broju različitih sklopova odgovora i broju ispitanika. Iza toga sledi osnovna deskripcija stavki: broj i proporcija negativnih i pozitivnih odgovora za svaki ajtem.

```
*****
* KAKAOBEJZ - Ajtem analiza           *
* Fajgelj - Janicic                   *
* Verzija 1.1, 28.11.2008.           *
*****
```

```
* SVI AJTEMI MORAJU BITI SKOROVANI JEDNAKO: 0=negativno,
1=pozitivno *
```

```
* LOGISTICKI MODEL, PARAMETRI U LOGISTICKOJ METRICI *
```

```
* Svi slucajevi sa ispustenim vrednostima su ODBACENI *
```

Kvadraturne tacke-apscise i tezine

```
apscise  tezine
-4,500000 ,000016
-3,500000 ,000873
-2,500000 ,017528
-1,500000 ,129518
-,500000 ,352065
,500000 ,352065
1,500000 ,129518
2,500000 ,017528
3,500000 ,000873
4,500000 ,000016
```

Inicijalne vrednosti parametara

```
          tezina  nagib
ravb1    ,0000  1,0000
ravb2    ,0000  1,0000
...      ...     ...
ravb11   ,0000  1,0000
ravb12   ,0000  1,0000
```

Parametri a priori ocekivanja za nagib

	M	Var
ravb1	,0000	,2500
ravb2	,0000	,2500
...	...	...
ravb11	,0000	,2500
ravb12	,0000	,2500

Broj EM iteracija, broj Fisher/Newton-Raphson iteracija,  
float opcija  
50 10 0

Broj sklopova odgovora i broj ispitanika  
sklopovi ispitan.  
N 426 2312

	Frekvencije i proporcije kategorija			
	Neg.	Poz.	Neg.	Poz.
ravb1	80,000	2232,000	,035	,965
ravb2	282,000	2030,000	,122	,878
ravb3	410,000	1902,000	,177	,823
ravb4	462,000	1850,000	,200	,800
ravb5	814,000	1498,000	,352	,648
ravb6	1037,000	1275,000	,449	,551
ravb7	1299,000	1013,000	,562	,438
ravb8	1571,000	741,000	,679	,321
ravb9	1487,000	825,000	,643	,357
ravb10	1302,000	1010,000	,563	,437
ravb11	1567,000	745,000	,678	,322
ravb12	1828,000	484,000	,791	,209

Postignuta konvergencija na osnovu stabilizacije parameta-  
ra.

Istorija EM iteracija		
iterac.	LRG**2	Max.prom
1,0000	12678,8179	,6122
2,0000	2353,3847	,1944
3,0000	2152,0302	,0896
4,0000	2105,7475	,0521
5,0000	2082,7572	,0301
6,0000	2066,5169	,0181
7,0000	2053,6579	,0116
8,0000	2043,0390	,0106
9,0000	2034,1020	,0104
10,0000	2026,5063	,0104
11,0000	2020,0231	,0104
12,0000	2014,4814	,0102
13,0000	2009,7510	,0100

Iz tabele "Istorija EM iteracija" korisnik može saznati kako je protekla konvergencija. U uvodnoj rečenici biva obavješten o tome da li je konvergencija postignuta i na koji način (na osnovu pune stabilizacije parametara ili neznatnih promena u parametrima). Ako je bilo problema, to će takođe pisati u uvodnoj rečenici i biće navedeni razlozi i ajtemi kod kojih su se javili problemi konvergencije.

U drugoj koloni ("LR  $G^2$ ") se nalazi  $G^2$  statistik koji se izračunava na osnovu funkcije verodostojnosti, ima približno  $\chi^2$  distribuciju i izračunava se, u principu, kao odnos dobijenih i očekivanih frekvencija, što znači da je namenjen da bude statistik tipa GOF (goodness-of-fit). Stvarne mogućnosti ovog i svih sličnih statistika nisu ispitane i prilično su sumnjive, tako da ga ovde preporučujemo prvenstveno kao pokazatelj konvergencije, jer se u regularnim okolnostima njegove vrednosti strmo približavaju nekoj asimptotskoj vrednosti. Jedan od razloga što upotreba ovog "likelihood-ratio"  $\chi^2$  statistika obično nije adekvatna je to što se sumira preko sklopova odgovora, od kojih mnogi imaju veoma niske očekivane verovatnoće.

Treća kolona ove tabele ("Max.prom") sadrži maksimalne promene parametara bilo kog ajtema u svakoj EM iteraciji. Maksimalna promena parametara je osnovni kriterijum konvergencije i ona u regularnim slučajevima brzo opada do neke vrlo male vrednosti. Već je opisano da se iz ove kolone može videti da li su iteracije oscilirale. Male oscilacije, na primer na drugom decimalnom mestu, su česte i na njih ne treba obraćati pažnju. Ako rešenje oscilira, to se obično vidi i iz kolone sa  $G^2$  statistikom.

Param. tezine i diskrim., s. greske, opterećenja na 1. faktoru i p-vrednost ajtema

	Tezina	S.G.	Diskrim	S.G.	Faktor	p-vred
ravb1	-2,5397	,1327	1,8630	,0895	,8811	,9874
ravb2	-1,5977	,0538	1,9674	,0574	,8915	,9228
ravb3	-1,2361	,0384	2,1700	,0505	,9082	,8692
ravb4	-1,1878	,0408	1,9140	,0508	,8863	,8538
ravb5	-,5812	,0297	1,8960	,0456	,8845	,6964
ravb6	-,2711	,0394	1,2265	,0487	,7750	,5832
ravb7	,2634	,0485	,9918	,0530	,7042	,4264
ravb8	,5176	,0252	2,7265	,0424	,9388	,3135
ravb9	,3991	,0258	2,5000	,0408	,9285	,3555
ravb10	,1437	,0277	2,0635	,0407	,8999	,4485
ravb11	,5115	,0252	2,7225	,0424	,9387	,3156
ravb12	1,1558	,0448	1,6880	,0469	,8604	,1600

Faktorsko opterećenje se može interpretirati i kao indeks relijabilnosti

ajtema, odnosno kao biserijalna korelacija ajtema sa osobinom.

Prosečna težina i nagib  
Mtez. Mnag.  
-,3685 1,9774

Sa „S.G.“ su označene standardne greške težina i diskriminativnosti ajtema. Dve krajnje kolone su u KakaoBejzu posvećene povezivanju stare i nove ajtem analize, tj. klasične testne teorije i teorije ajtemskog odgovora. To su kolone sa faktorskim opterećenjima i očekivanim  $p$ -vrednostima.

Kolona u ispisu rezultata, označena kao faktorsko opterećenje, bazira se na jednom starom zapažanju, još iz 1936. godine, da je veličina  $\rho = a/\sqrt{1 + a^2}$  jednaka biserijalnoj korelaciji tog ajtema sa osobinom  $\theta$ . Ovo je danas široko prihvaćeno, između ostalog i u programu Bilog (Du Toit, 2003), s time da je uočena istovetnost takve interpretacije sa interpretacijom faktorskog opterećenja.

Kolona sa očekivanom  $p$ -vrednošću ajtema sadrži staru Takerovu formulaciju klasične težine ajtema preko parametara  $a$  i  $\delta$ :  $p = \Phi(-\delta\rho)$ , gde je  $\Phi$  kumulativna funkcija normalne distribucije, a  $\rho$  faktorsko opterećenje. Zanimljivo je uporediti ovu očekivanu  $p$ -vrednost sa stvarnom  $p$ -vrednošću pozitivnog odgovora. Na kraju, prikazani su prosečna težina i prosečna diskriminativnost ajtema.

### *Problemi konvergencije*

Centralnu tačku u ocenjivanju parametara predstavlja maksimizacija funkcije verodostojnosti. Traženje maksimuma je iterativni proces, čija ključna pretpostavka je da će dovesti do konačnog rešenja – da će konvergirati. U KakaoBejz (i KakaoMiksu) je ugrađena kontrola iterativnog procesa koja, između ostalog, obaveštava korisnika da maksimizacija nije tekla uredno. Korisniku se savetuje da pažljivo pogleda sve poruke, kao i da uoči ajteme sa ekstremnim parametrima. Ako funkcija verodostojnosti za konkretne podatke nije unimodalna, ili je zaravnjena, mogu se očekivati sledeće posledice: rešenje ne konvergira, postoje oscilacije, rezultati su osetljivi na izmenu ulaznih parametara ili podataka itd.

Kad god su rezultati sumnjivi, a pogotovo kada odvijanje programa bude prekinuto usled neke nekontrolisane greške, korisniku preporučujemo sistematičnost i strpljenje. Može pokušati da promeni parametre u makrou, recimo da pokuša sa drugim modelom ("log" ili "norm"). Može da poveća (ili ev. smanji, ali to je manje preporučljivo) broj kvadrata tačaka. Sve opcije koje se odnose na iterativni tok, takođe mogu biti od značaja: EMciklus, Fishiter i konv.

Ove tri opcije imaju simultano dejstvo na iterativni tok i zasada ne možemo dati opštu preporuku za njihovo korišćenje. Isto važi i za opcije *malfa* i *salfa*.

Nema konačnih preporuka o tome koliko je kvadraturnih tačaka potrebno za dobro ocenjivanje. Povećanje *GKvadrat* može delovati blagotvorno na iterativni tok, verovatno kada je distribucija osobine u uzorku stvarno normalna i ako su distribucije parametara normalne i sl. U nekim slučajevima kvalitet rešenja je bio bolji kada smo broj kvadraturnih tačaka povećali sa 10 na 20. S druge strane, ukoliko su distribucije "nemirne", do konvergencije će doći lakše ukoliko se *gkvadrat* smanji ispod 10. Zbog uticaja na brzinu odvijanja programa ne isplati se broj kvadraturnih tačaka povećavati iznad 20.

Ako je test kratak, onda se može povećati *emciklus* (i/ili *gkvadrat*) sve dok se ne stabilizuju parametri. Kod dugačkih testova, na velikim uzorcima to će biti penalizovano dugim trajanjem.

Naravno, korisnik je ovlašćen da učini suštinsku modifikaciju, a to je da pokuša da izbacij aйтеme koji deluju nepovoljno. Vrlo verovatno se radi o "lošim" aјtemima, kojima i inače nije mesto u tom testu, ali se može raditi i o specifičnostima uzorka ispitanika. Bilo kako bilo, uočavanje da neki aјtemi imaju loš uticaj na iterativni postupak može biti korisno za psihološku evaluaciju testa jer može pomoći korisniku da uoči slabosti: aјtema, testa, uzorka.

Normalna metrika u načelu teže konvergira.

Oscilacije tokom konvergencije je teško programski otkriti jer oscilacije mogu da postoje, ali da parametri ipak idu ka stabilizaciji. Zasada se samo štampa upozorenje da je bilo porasta maksimalne promene parametara, a korisniku se prepušta da reši problem.

### *Kvalitet ocenjenih parametara*

Na kvalitet konačnog rešenja utiče mnogo faktora. Pre svega to su jednodimenzionalnost i lokalna nezavisnost aјtemskih odgovora (Fajgelj, 2005). Zatim, sve što je prethodno rečeno o iterativnom postupku i problemima konvergencije od velikog je značaja za to da li ćemo imati poverenja u dobijene parametre ili ne.

Istraživanjima i simulacijama ustanovljeno je da su dobijeni parametri to bolji što ima **više aјtema**. Međutim, da bi broj aјtema bio blagotvoran on mora biti praćen **dovoljnim brojem ispitanika**. U suprotnom, ako je uzorak mali, veliki broj aјtema će samo ometati kvalitet. Na primer, ako imamo 30 aјtema i 500 ispitanika moći ćemo biti prilično sigurni u dobijene parametre i standardne

greške. Iz pregleda literature se uglavnom može zaključiti da za kvalitetno ocenjivanje parametara u TAO trebaju veći uzorci nego što je uobičajeno za KTT (uzgred, to posebno važi za 3P modele kod kojih se komforna granica postavlja na 60 ajtema i 1000 ispitanika). Nismo našli sistematska istraživanja o tome koji je poželjan odnos broja ispitanika prema broju ajtema, ali sve ukazuje na to da komforan odnos počinje oko 10:1.

Kod KakaoBejza, Bayesova očekivanja, predstavljena opcijama *malfa* i *salfa*, mogu imati značajan uticaj na parametre. Pokazano je da postoji tendencija da sa povećanjem uzorka (iznad 250), a priori očekivanja imaju sve manje dejstvo na konačne rezultate. Na manjim uzorcima ( $n < 150$ ), pogrešno zadata očekivana varijansa parametra  $a$  (*salfa*<sup>2</sup>) može značajno negativno uticati na konačno rešenje. Pri tome, male vrednosti očekivane varijanse će dati bolje rezultate, a velike lošije. Negativan uticaj pogrešno zadatih hiperparametara za nagib (posebno varijanse) se, u tom slučaju, može donekle smanjiti uključivanjem opcije *float*. Dakle, ako je uzorak mali, pa još i test ima malo ajtema i rezultati su sumnjivi (pogotovo ako postoji upozoravajuća poruka da su iteracije bile neregularne), može se pokušati sa prilagođavanjem vrednosti *malfa*, smanjivanjem *salfa*, ili sa uključivanjem opcije *float*.

### Vreme izvršavanja programa

Generalno, KakaoBejz je višestruko sporiji od KTT ajtem analize, ali znatno brži od pomenutog KakaoMiksa (najmanje dvostruko brži).

U KakaoBejzu trajanje jednog EM ciklusa zavisi (po redosledu prioriteta) od broja sklopova odgovora, broja ajtema, broja kvadrurnih tačaka i broja NR iteracija. Broj sklopova se može odrediti kao  $l \leq \min(2^m, n)$ . Broj ajtema deluje na trajanje programa dvojako: preko broja sklopova i samostalno. Pošto navedeni činioci deluju multiplikativno, ukupan broj operacija, a time i vreme izvršavanja, rastu dramatično sa povećanjem problema.

**Tabela 1.** Vremena izvršavanja makroa u zavisnosti od broja stavki i veličine uzorka

M	N	KakaoBejz
10	100	0'02"
10	500	0'04"
10	1000	0'14"
50	500	1'23"
50	1000	4'11"



Naravno, vreme izvršavanja zavisi konfiguracije računara<sup>5</sup>. Treba imati u vidu da za tako velike probleme valja obaviti oko 16 miliona operacija po EM ciklusu.

O vremenu potrebnom za izvršenje programa potrebno je voditi računa prilikom analiza sa velikim brojem stavki i velikim uzorkom. Naime, dok izvršava program, SPSS u statusnoj liniji ispisuje samo "RUNNING: MATRIX..." i nema nikakvih detaljnijih informacija o samom toku izvršavanja. Potencijalnom korisniku se zato može učiniti da program više ne funkcioniše. Preporučujemo da imate u vidu navedene informacije i mogućnosti svoga računara i strpljivo sačekate kraj analize.

### **Reference**

- Andrich, D., Sheridan, B., Luo, G. (2004). *Interpreting RUMM2020: Part I, II i III*, RUMM Laboratory Pty. Dostupno na: <http://www.rummlab.com.au>.
- Baker, F. B., Kim, S. (2004). *Item response theory: Parameter estimation techniques* (2<sup>nd</sup> edition). New York: Marcel Dekker.
- Du Toit, M. (ur.) (2003). *IRT From SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: Scientific Software International. Dostupno na: <http://www.ssicentral.com>.
- Edwards, A. W. F. (1972). *Likelihood, An account of the statistical concept of likelihood and its application to scientific inference*. London: Cambridge University Press.
- Embretson, S. E., Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum.
- Fajgelj, S. (2005). *Psihometrija: Metod i teorija psihološkog merenja* (2. izdanje). Beograd: Centar za primenjenu psihologiju.
- Hambleton, R.K., Swaminathan, H., Rogers, J.H. (1991). *Fundamentals of item response theory*. Newbury Park – CA: SAGE Publications.
- Linacre, J. M. (2002). *A User's Guide to WINSTEPS & MINISTEP, Rasch-Model computer programs*. Chicago: autor. Dostupno na: <http://www.rasch.org>.

---

<sup>5</sup> Vremena izvršavanja makroa merena su na računaru sledeće konfiguracije: AMD Athlon™ XP 2000+, 1,67 GHz, sa 768 RAM, ali izgleda da i verzija SPSS-a igra ulogu. U novijim verzijama makro se izvršava brže.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Assoc..

SPSS Inc. (2005). *SPSS 14.0 Command syntax reference*. Chicago-IL: Author.

Van der Linden, W.J., Hambleton, R.K. (eds.) (1997). *Handbook of modern item response theory*. New York: Springer Verlag.

## ABSTRACT

**KAKAOBEJZ:MACRO FOR ITEM ANALYSIS OF DICHOTOMOUSLY SCORED ITEMS – ITEM RESPONSE THEORY**

Stanislav Fajgelj and Bojan Janičić

The paper presents KakaoBejz, one of the macros from the set called "Kakao." This set of programs is intended for the item analysis of instruments in the field of behavioral sciences. It is developed in Matrix program language as macros for statistical package SPSS for Windows. These macros are open source programs, which are free and very simple to use. They can run only within SPSS statistical package.

KakaoBejz is intended for the item analysis based on Item Response Theory (IRT). Item parameters estimation is based on the maximum likelihood method with added *a priory expectations* derived from the Bayes Theorem. The macro contains basic indicators of convergence control, enabling a user to assess if the iterative process is carried out correctly. By using the macro's entry options, it is possible to control program execution, choose item response model and set starting values. KakaoBejz accepts only dichotomously scored items. Items must be scored in the same direction.

**Key words:** item analysis, Item response theory, maximum likelihood estimation, Bayes theorem