XX. ANALIZA GLAVNIH KOMPONENATA

Poglavlje iz: Statistika u psihologiji, prošireno i dopunjeno izdanje, 2020.

Cilj analize glavnih komponenata (engl. Principal components analysis) je transformisanje m izmerenih i međusobno povezanih varijabli u m ortogonalnih, tj. nekorelisanih glavnih komponenata. Svaka glavna komponenta je linearna kombinacija merenih, tj. izvornih varijabli. Dakle, u analizi glavnih komponenata broj glavnih komponenata teorijski, tj. matematički jednak je broju izvornih, merenih varijabli. To nije slučaj i sa metodama faktorske analize, pa je zbog ove suštinske konceptualne razlike bitno razlikovati analizu glavnih komponenata od faktorske analize. Dok faktorska analiza pretpostavlja određeni teorijski model o direktno nemerljivim latentnim varijablama ili o faktorima kao determinantama kovariranja (korelacija) merenih ili manifestnih varijabli, analiza glavnih komponenata najčešće rezultira svođenjem većeg broja merenih varijabli na manji broj glavnih komponenata ali bez teorijskih pretpostavki o kauzalnom uticaju komponenata na merene, opažene varijable.[170]

---

[170] Metode faktorske analize nisu obuhvaćene ovom knjigom. Budući da je veoma važno razlikovati analizu glavnih komponenata od metoda faktorske analize, prenosimo deo teksta A. Zorića iz njegovog diplomskog rada u kojem su dobro objašnjeni osnovni principi faktorske analize:

„Faktorska analiza je multivarijaciona statistička metoda koja služi otkrivanju unutrašnje strukture matrica podataka. Model faktorske analize se genuino smatra psihološkim […]

Metodama faktorske analize nastoji se da se iz složene (multidimenzionalne) pojave, tj. skupa manifestnih obeležja, ekstrahuje manji broj važnih latentnih činilaca, tj. **faktora** koji su odgovorni za datu pojavu. Faktorskom analizom, dakle, pokušavamo da iza direktno izmerenih, manifestnih ili pojavnih varijabli koje smo izmerili na nekom reprezentativnom uzorku entiteta, i pomoću kojih nastojimo da opišemo izučavanu pojavu, pronađemo manji broj latentnih, skrivenih varijabli koje nazivamo **faktorima**, a koje objašnjavaju posmatranu pojavu, tj. čijom se kombinacijom mogu dobiti pojavne varijable. Kako o pojavi koju proučavamo saznajemo samo preko varijabli koje predstavljaju mere pojedinih aspekata te pojave (iz ovoga sledi i posledica da naše varijable moraju biti reprezentativan skup svih varijabli koje mere datu pojavu, tj. da naš skup varijabli mora dobro, adekvatno opisivati izučavanu pojavu), to je očigledno da termin „objasniti pojavu" znači naći odgovarajuću matematičku vezu između izmerenih varijabli, tj. objasniti dobijene veze između njih, ili pronaći uzroke pojave – objasniti dobijenu varijansu izmerenih varijabli. Na osnovu ove razlike u pristupu objašnjenju izučavane pojave, faktorske metode možemo podeliti u dve velike kategorije:

- **faktorske metode u užem smislu** – metode kojima pokušavamo da objasnimo veze, tj. korelacije između manifestnih, izmerenih varijabli;
- **komponentne metode** – postupci kojima nastojimo da objasnimo variranja u merama izučavane pojave iznalaženjem njenih uzroka.

Faktorske metode u užem smislu nastoje da objasne veze među posmatranim varijablama. One polaze od pretpostavke da su manifestne, direktno izmerene varijable međusobno povezane, dakle u korelaciji, ne zato što su one u nekom uzročno-posledičnom odnosu, već zato što ih sve određuju neki njima zajednički faktori. Na ovaj način, cilj ove grupe metoda postaje rekonstrukcija korelacione matrice, dakle veza među varijablama preko manjeg broja zajedničkih faktora. U literaturi se zbog ovoga faktorski modeli u užem smislu nazivaju i analizom zajedničkih faktora.

U ovom modelu varijansa svake direktno merene, tj. manifestne varijable, razbija se na dva dela:

- deo koji je zajednički sa drugim varijablama na osnovu kojega one koreliraju. Ovaj deo varijanse objašnjavaju zajednički faktori. Varijable koreliraju međusobno zato što koreliraju sa **zajedničkim** faktorima, te se ovaj deo varijanse varijable zove **komunalitet** i obeležava se sa h^2;
- deo varijanse varijable koji ne objašnjavaju zajednički faktori, te koji je zbog toga samo njoj svojstven, i koji objašnjava njoj odgovarajuća, takođe latentna varijabla koja se naziva **unikni** faktor. Jasno je da po ovoj podeli unikni faktor može biti povezan samo sa jednom odgovarajućom varijablom, jer bi u suprotnom i on predstavljao uzrok zajedničkog variranja varijabli, a samim tim bi po definiciji bio zajednički faktor. Ovaj deo varijanse se naziva unikvititet i obeležava se sa u^2. Unikna varijansa se dalje deli na dva dela: deo prave varijanse koji je specifičan za datu varijablu i koja se zato zove **specificitet**, a obeležava sa s^2, i na deo koji je posledica **greške** i koji se obeležava sa e^2.

U analizi glavnih komponenata ne pravi se distinkcija između zajedničke i unikne varijanse merenih varijabli (ili faktorskim jezikom rečeno – manifestnih varijabli) kao što je to slučaj u faktorskom modelu.[171] Sve glavne komponente obuhvataju celokupnu varijansu izvornih, merenih varijabli, ali tako da na svaku sledeću glavnu komponentu otpada sve manji deo ove ukupne varijanse. Ipak, praktični cilj ove analize je da se od svih m glavnih komponenti koje se dobijaju na osnovu m izvornih, tj. izmerenih varijabli, zadrži samo mali broj komponenata u kojima će pritom biti sadržan što veći deo ukupne varijanse (ili količine informacija) svih izvornih, tj. izmerenih varijabli. Budući da ova analiza ima smisla samo ako su izvorne varijable međusobno korelisane,[172] tj. bar u izvesnoj meri redundantne, analiza glavnih komponenata najčešće rezultira zadržavanjem manjeg broja komponenata koje se koriste u daljim analizama, na primer kao prediktorske varijable u regresionoj analizi.

Analiza glavnih komponenata nije skalno invarijantna: glavne komponente dobijene na osnovu matrice kovarijansi, tj. iz nestandardizovanih izvornih varijabli, ne moraju odgovarati komponentama koje se dobijaju iz matrice interkorelacija, tj. na osnovu standardizovanih izvornih varijabli.[173]

Struktura podataka

$E = \{e_i;\ i = 1,..., n\} \subset P$ je slučajni uzorak izvučen iz homogene populacije;

$V = \{v_j;\ j = 1,2,\ldots, m\} \subset U$ je skup kvantitativnih varijabli koji reprezentuje univerzum U svih mogućih varijabli koje se odnose na pojavu koja se proučava;

Pretpostavićemo da su podaci standardizovani, što je, s obzirom na osetljivost, tj. neinvarijantnost analize glavnih komponenata na metriku (tj. merne jedinice ili skalu varijable), preporučljiv postupak u psihološkim istraživanjima. Naime, s obzirom na to da

---

Komponentne metode faktorske analize nastoje da pronađu latentnu strukturu koja stoji iza direktno izmerenih varijabli i koja je odgovorna za dobijena variranja na tim varijablama. Komponentne metode dakle nastoje da pronađu uzroke pojave, koju registrujemo preko variranja rezultata u merenjima koja vršimo nad njom. Pod uzrokom se onda podrazumeva latentna varijabla, faktor ili u ovom slučaju komponenta koja je odgovorna za nastalu varijansu u rezultatima merenja. Za ovu varijansu može biti odgovorna i greška koja nastaje prilikom merenja, te drugi problem u komponentnim metodama jeste i određivanje da li izolovana latentna varijabla predstavlja realno egzistirajuću strukturu ili predstavlja samo grešku..." (A. Zorić, Faktorska analiza binarnih varijabli u generalizovanom imaž prostoru, Diplomski rad, Beograd: Filozofski fakultet, 1996. Navedeno sa dozvolom autora.)

[171] Ključna razlika ovog postupka u odnosu na **komponentni model faktorske analize** je u tome što se faktorska analiza izvodi na matrici interkorelacija, pa je, prema tome, metrički invarijantna, i u tome što se u faktorskoj analizi zadržavaju samo neke glavne komponente (a ne sve glavne komponente) kao faktori. Komponentni model faktorske analize je zapravo korišćenje analize glavnih komponenata za aproksimaciju modela faktorske analize u užem smislu, tj. za aproksimaciju zajedničkih faktora.

[172] Provera statističke hipoteze: **P** = **I** (matrica interkorelacija varijabli u populaciji iz koje je uzorak jednaka je matrici identiteta, tj. varijable u populaciji nisu linearno povezane) zove se **Bartletov test sferičnosti**. Za testiranje ove hipoteze koristi se statistik $B = -[n - \frac{2m+11}{6}]\sum_{p=1}^{m} \ln \lambda_p$, pri čemu je n veličina uzorka, m broj varijabli, a λ_p svojstvene vrednosti uzoračke matrice interkorelacija. Ovaj statistik, ako je hipoteza tačna, ima aproksimativno hi-kvadrat distribuciju sa m(m - 1) / 2 stepeni slobode.

[173] Budući da se analiza glavnih komponenata zasniva na nalaženju svojstvenih vrednosti matrice kovarijansi ili matrice korelacija, osetljivost metode na „metriku" varijabli posledica je činjenice da svojstvene vrednosti matrice kovarijansi i matrice korelacija dobijenih od istih varijabli ne samo da nisu iste nego nisu ni u kakvim jednostavnim relacijama. Podrobnije o ovome može se pročitati u Wilks (1962).

je metrika (nulta, početna tačka i skala) za skoro sve psihološke varijable arbitrarna, najčešće nema mnogo smisla raditi analizu glavnih komponenata na nestandardizovanim varijablama.
Opisom slučajnog uzorka, tj. skupa E u pogledu skupa kvantitativnih varijabli V, pri čemu su podaci na varijablama transformisani u standardizovani oblik, dobijamo matricu **Z**:

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

Matrica $\mathbf{R} = \mathbf{Z}^t\mathbf{Z}n^{-1}$ je matrica interkorelacija varijabli iz skupa V.

Formalno definisanje i rešenje problema glavnih komponenata:

Metodu glavnih komponenata predložio je H. Hotelling 1933. godine. Ova metoda svodi se na traženje vektora koji sadrže koeficijente za linearne transformacije varijabli iz **Z** tako da na taj način dobijene linearne kombinacije (glavne komponente) merenih varijabli imaju što veću varijansu. Rečeno geometrijskim terminima, rezultujući vektor (tj. glavna komponenta) treba da bude što duži, uz odgovarajuća ograničenja. Da bi formalna postavka i rešenje problema (koji su dati u nastavku teksta) bili lakši za praćenje, zamislimo sledeću situaciju: zadali smo uzorku ispitanika upitnik depresivnosti sa deset pitanja. Postavlja se pitanje da li za dalju analizu moramo koristiti rezultate ispitanika na svih deset pitanja ili (najčešće, standardizovane) rezultate na pojedinačnim stavkama, tj. izvornim varijablama, možemo kondenzovati (sažeti) u manji broj varijabli, glavnih komponenata, ali tako da zadržimo što veću količinu informacija koja je sadržana u odgovorima na svih deset stavki. Ukoliko sve ove stavke mere, na primer, depresivnost, mogli bismo očekivati da je to moguće. Pri tome treba odlučiti kakvu ćemo težinu dati svakoj stavki (tj. odrediti ponder za linearnu kombinaciju stavki). Matematički posmatrano, na osnovu deset stavki, tj. merenih varijabli, možemo dobiti deset glavnih komponenata. Ali neće sve ove komponente biti praktično važne. Može nam se desiti ponekad da u ovakvim slučajevima zapravo možemo da, umesto deset izvornih varijabli kojima merimo depresivnost, dođemo do **jedne** glavne komponente, tj. generalne mere depresivnosti. Ova, u idealnom slučaju **jedna**, glavna komponenta nastaje optimalnim ponderisanjem rezultata na pojedinačnim stavkama tako da sadrži praktično sve relevantne informacije o depresivnosti koje su sadržane u svim stavkama. Na taj način postižemo veliku uštedu (parsimoničnost), uz mali gubitak informacija.
Formalna postavka problema glavnih komponenata izgleda ovako:

$$\mathbf{Z}\,\mathbf{x}_p = \mathbf{k}_p \quad \left|\begin{array}{l} 1.\ \sigma_p^2 = \mathbf{k}_p^t\,\mathbf{k}_p\, n^{-1} = \mathbf{x}_p^t\,\mathbf{R}\,\mathbf{x}_p = \text{maximum.} \\ 2.\ \mathbf{x}_p^t\,\mathbf{x}_p = 1 \end{array}\right.$$

Dakle, tražimo vektor $\mathbf{x}_p$ u kojem su linearni koeficijenti pomoću kojih iz matrice podataka **Z** dobijamo vektor, tj. glavnu komponentu $\mathbf{k}_p$ koja ima maksimalnu varijansu (uslov 1.), što znači da smo iz matrice **Z** izvukli što je moguće više informacija.[174] Naravno, ta bi varijansa bila utoliko veća ukoliko su proizvoljno veći koeficijenti u vektoru $\mathbf{x}_p$. Stoga uvodimo još jedan uslov (uslov 2.), a to je da je $\mathbf{x}_p$ jedinične dužine, tj. da je suma kvadriranih pondera za pravljenje date linearne kombinacije jednaka jedinici.
Problem rešavamo metodom Lagranžovih multiplikatora, konstruišući funkciju

[174] Varijansa je, kao što je poznato, kod kvantitativnih varijabli mera količine informacija koju varijabla sadrži.

$F(\mathbf{x}_p, \lambda_p) = \mathbf{x}_p^{\mathrm{t}} \mathbf{R}\, \mathbf{x}_p - \lambda_p(\mathbf{x}_p^{\mathrm{t}} \mathbf{x}_p - 1)$

Parcijalnom derivacijom ove funkcije po $\mathbf{x}_p$, izjednačavanjem sa 0 i skraćivanjem sa 2, dobija se jednačina

$$(\mathbf{R} - \lambda_p\mathbf{I})\mathbf{x}_p = \mathbf{0},$$

te se problem tako svodi na problem svojstvenih vrednosti[175] matrice **R**, tj. matrice interkorelacija varijabli iz V.[176]

Svojstvene vrednosti λ_p matrice **R** jednake su varijansama glavnih komponenata σ_p^2. Iz $(\mathbf{R} - \lambda_p\mathbf{I})\mathbf{x}_p = \mathbf{0}$ sledi da je

$\mathbf{R}\mathbf{x}_p - \lambda_p\mathbf{x}_p = \mathbf{0}.$

Ako ovu jednačinu sa leva pomnožimo sa $\mathbf{x}_p^{\mathrm{t}}$ i sredimo je, dobijamo

$\sigma_p^2 = \mathbf{x}_p^{\mathrm{t}}\mathbf{R}\mathbf{x}_p = \lambda_p\mathbf{x}_p^{\mathrm{t}}\mathbf{x}_p = \lambda_p$

jer je

$\mathbf{x}_p^{\mathrm{t}}\mathbf{x}_p=1,$

a

$\mathbf{x}_p^{\mathrm{t}}\mathbf{R}\mathbf{x}_p = \sigma_p^2.$

Rešenje matrične jednačine $(\mathbf{R} - \lambda_p\mathbf{I})\mathbf{x}_p = \mathbf{0}$ ne može se dobiti njenim množenjem sa $(\mathbf{R}-\lambda_p\mathbf{I})^{-1}$, jer kada bismo to uradili, dobili bismo

$$(\mathbf{R} - \lambda_p\mathbf{I})^{-1}(\mathbf{R} - \lambda_p\mathbf{I})\mathbf{x}_p = (\mathbf{R} - \lambda_p\mathbf{I})^{-1}\mathbf{0},$$
$$\mathbf{I}\mathbf{x}_p = \mathbf{0},$$
$$\mathbf{x}_p = \mathbf{0}$$

što je u kontradikciji sa zahtevom da je $\mathbf{x}_p^{\mathrm{t}}\mathbf{x}_p = 1$. Zato $(\mathbf{R} - \lambda_p\mathbf{I})$ ne sme imati inverz, što znači da tražimo λ_p takvo da je $|\mathbf{R}-\lambda_p\mathbf{I}| = 0$. Time se rešavanje svodi na rešavanje sistema m homogenih jednačina sa m nepoznatih. Razvojem determinante dobijamo polinomsku jednačinu po λ stepena m:

$f(\lambda) = (-\lambda)^m + b_{m-1}(-\lambda)^{m-1} + ... + b_1(-\lambda) + b_0 = 0$

Ova jednačina je **karakteristična jednačina** matrice **R**, a λ_1, λ_2,..., λ_m su karakteristični koreni ili **svojstvene vrednosti** (engl. eigenvalues) matrice **R**. Njihovim nalaženjem dobijamo i njima odgovarajuće karakteristične, svojstvene vektore $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$.

Konstruišimo sada dijagonalnu matricu

$\mathbf{\Lambda} = (\lambda_p) \quad (p = 1, 2, ..., m),$

---

[175] Realni broj (skalar) λ je **svojstvena vrednost** (engl. eigenvalue, čita se kao *ajgenvelju*) kvadratne matrice **A** ako postoji vektor **x** sa istim brojem redova koliko ima matrica **A**, takav da mu svi elementi nisu jednaki nuli (tzv. nenulti vektor) i takav da važi $\mathbf{Ax} = \lambda\mathbf{x}$. Ukoliko izraz $\lambda\mathbf{x}$ prebacimo na levu stranu jednakosti, dobijamo $\mathbf{Ax} - \lambda\mathbf{x} = \mathbf{0}$, tj. $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x}$, pri čemu je **I** matrica identiteta. Dakle, svojstvene vrednosti, kada se oduzmu od dijagonalnih elemenata matrice **A**, pretvaraju tu matricu u singularnu matricu, u matricu $\mathbf{A} - \lambda\mathbf{I}$ koja nema regularni inverz, tj. u matricu čija je determinanta $|\mathbf{A} - \lambda\mathbf{I}|$ jednaka nuli.

[176] Naravno, ukoliko je reč o nestandardizovanim varijablama, problem se svodi na rešavanje problema svojstvenih vrednosti matrice **kovarijansi**.

dakle, matricu čiji su dijagonalni elementi svojstvene vrednosti, poređane po veličini od gornjeg levog ka donjem desnom uglu matrice, i kvadratnu matricu

$$\mathbf{X} = (\mathbf{x}_p) \quad (p = 1, 2, ..., m)$$

u čijim kolonama su svojstveni vektori.

Budući da je matrica **R** pozitivno definitna (tj. $\mathbf{x}^t \mathbf{R}\, \mathbf{x} > 0$) simetrična matrica, slede sledeća tvrđenja:.

- $\lambda_p > 0 \quad (p = 1,..., m)$, tj. sve svojstvene vrednosti su veće od nule;
- $|\mathbf{\Lambda}| = |\mathbf{R}|$, ili, budući da je Λ dijagonalna: $\prod_{p=1}^{m} \lambda_p = |\mathrm{R}|$, pri čemu je $|\bullet|$ oznaka determinante;
- trag(Λ) = trag(**R**), tj. zbir svojstvenih vrednosti ili varijansi glavnih komponenata jednak je tragu matrice (zbiru vrednosti na glavnoj dijagonali matrice) iz koje su dobijene glavne komponente. Ako su glavne komponente dobijene na osnovu standardizovanih varijabli, tj. na osnovu matrice interkorelacija, onda je zbir varijansi svih glavnih komponenata, kao i zbir varijansi početnih varijabli, jednak broju tih varijabli.
- $\mathbf{X}^t\mathbf{X} = \mathbf{X}\mathbf{X}^t = \mathbf{I}$, tj. matrica svojstvenih vektora je ortogonalna matrica.
- $\mathbf{X}^t\mathbf{R}\mathbf{X} = \Lambda$, a odatle $\mathbf{X}\Lambda\mathbf{X}^t = \mathbf{R}$
- $\mathbf{X}^t\mathbf{R}^{-1}\mathbf{X} = \Lambda^{-1}$, pa je $\mathbf{X}\Lambda^{-1}\mathbf{X}^t = \mathbf{R}^{-1}$, ili, u opštem slučaju $\mathbf{X}^t\mathbf{R}^w\mathbf{X} = \Lambda^w$, tj. $\mathbf{X}\Lambda^w\mathbf{X}^t = \mathbf{R}^w$ ($\forall w \in \Re$, pri čemu je $\Re$ skup realnih brojeva). Dakle, svojstveni vektori ostaju isti za sve stepenovane matrice, pri čemu stepen pripada skupu realnih brojeva, samo se svojstvene vrednosti dižu na isti stepen. Na taj način dobijeni svojstveni vektori čine skup ortonormiranih vektora, jer je $\mathbf{x}_p{}^t\mathbf{x}_q = \delta_{pq}$, pri čemu je δ_{pq} Kronekerova delta.[177] Tako dobijeni svojstveni vektori su zapravo vektori koeficijenata za glavne komponente, pri čemu je svaka glavna komponenta linearna kombinacija merenih varijabli.[178]

Matrica nestandardizovanih glavnih komponenata je

$$\mathbf{K} = \mathbf{Z}\mathbf{X},$$

a matrica kovarijansi glavnih komponenata

$$\mathbf{K}^t\mathbf{K}\, n^{-1} = \mathbf{X}^t\mathbf{Z}^t\mathbf{Z}\mathbf{X}\, n^{-1} = \mathbf{X}^t\mathbf{R}\mathbf{X} = \Lambda.$$

---

[177] Ako su $\mathbf{x}_p$ i $\mathbf{x}_q$ različiti svojstveni vektori matrice **R**, onda važi i $\mathbf{R}\mathbf{x}_p = \lambda_p\mathbf{x}_p$ i $\mathbf{R}\mathbf{x}_q = \lambda_q\mathbf{x}_q$. Odatle je i $\mathbf{x}_p{}^t\mathbf{R}\mathbf{x}_q = \lambda_p\mathbf{x}_p{}^t\mathbf{x}_q$ i $\mathbf{x}_q{}^t\mathbf{R}\mathbf{x}_p = \lambda_q\mathbf{x}_q{}^t\mathbf{x}_p$. Budući da je $\mathbf{x}_p{}^t\mathbf{R}\mathbf{x}_q - \mathbf{x}_q{}^t\mathbf{R}\mathbf{x}_p = 0$, sledi $(\lambda_p - \lambda_q)\mathbf{x}_p{}^t\mathbf{x}_q = 0$. Prema tome, ako $\lambda_p \neq \lambda_q$, onda $\mathbf{x}_p{}^t\mathbf{x}_q$ mora biti jednako 0.

[178] Geometrijski posmatrano, glavne komponente su ose hiperelipsoida kojeg pravi raspored entiteta, tj. jedinica posmatranja u m-dimenzionalnom prostoru, pri čemu su koordinatne ose tog prostora izvorne, merene varijable.

Očigledno je, dakle, da različite glavne komponente nisu međusobno u korelaciji,[179] tj. da su vektori $\mathbf{k}_p$ ortogonalni na sve ostale vektore $\mathbf{k}_q$.

Standardizovane glavne komponente dobijamo u matrici $\mathbf{K}_s$,

$$\mathbf{K}_s = \mathbf{K}\mathbf{\Lambda}^{-1/2} = \mathbf{Z}\mathbf{X}\mathbf{\Lambda}^{-1/2}.$$

Korelacije originalnih varijabli i glavnih komponenenata dobijamo iz

$$\mathbf{F} = \mathbf{Z}^t\mathbf{K}_s\, n^{-1} = \mathbf{R}\mathbf{X}\mathbf{\Lambda}^{-1/2} = \mathbf{X}\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} = \mathbf{X}\mathbf{\Lambda}^{1/2}.$$

Dakle, iz $\mathbf{F} = \mathbf{Z}^t\mathbf{K}_s n^{-1} = \mathbf{X}\mathbf{\Lambda}^{1/2}$, vidimo da se korelacija između određene glavne komponente p i izvorne standardizovane varijable j koja ulazi u linearnu kombinaciju za pravljenje glavne komponente dobija tako što se koeficijent za varijablu j (koja služi za pravljenje glavne komponente) pomnoži kvadratnim korenom iz varijanse glavne komponente: $r_{jp} = x_{jp}\sqrt{\lambda_p}$.
Na osnovu toga što je

$$\mathbf{R} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^t = \sum_{p=1}^{m} \lambda_p x_p x_p^t$$

vidimo da smo matricu interkorelacija raslojili na m nezavisnih slojeva od kojih je svaki sledeći sve „tanji“, pošto su varijanse glavnih komponenata (tj. svojstvene vrednosti matrice) sve manje i manje, tako da svaki sledeći sloj sve manje doprinosi rekonstrukciji matrice interkorelacija originalnih varijabli.[180]

Određivanje broja važnih glavnih komponenata

Za određivanje broja značajnih komponenata (tj. važnih komponenata) koje ima smisla zadržati od svih m dobijenih komponenata, postoji mnoštvo kriterijuma, ali ćemo ovde podrobnije razmotriti samo one kriterijume koji se mogu primeniti na osnovu ispisa iz programa SPSS. To su: Katelov skri-dijagram ili dijagram osulina (engl. scree plot),[181] Gutman–Kajzerov kriterijum i procenat ukupne varijanse merenih varijabli koji je obuhvaćen zadržanim glavnim komponentama.

Katelov skri-kriterijum (kriterijum osulina) zasniva se na grafičkoj proceni broja važnih komponenata.[182] U tu svrhu konstruiše se dijagram svojstvenih vrednosti, koje su (kao što

---

[179] Glavne komponente su, dakle, i geometrijski i statistički ortogonalne. Geometrijska ortogonalnost postoji, jer su svojstveni vektori koji odgovaraju različitim svojstvenim vrednostima međusobno ortogonalni. Statistička ortogonalnost se vidi iz toga što su kovarijanse (ili korelacije) različitih glavnih komponenata jednake nuli.

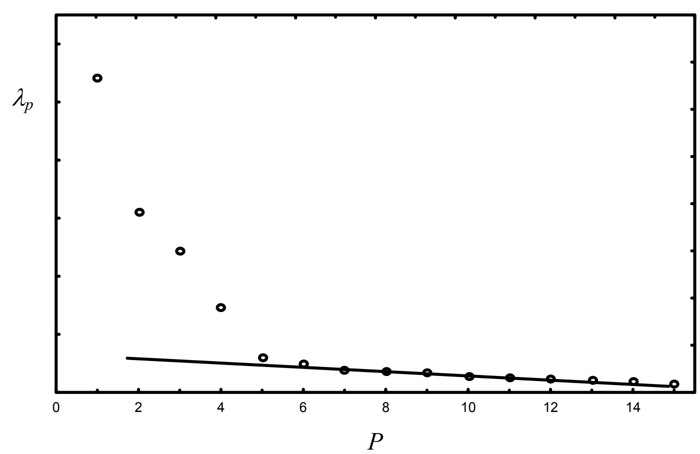
[180] Budući da je **R** simetrična matrica, ovde je reč o posebnom slučaju teoreme o „spektralnoj dekompoziciji matrice“ koja glasi: ako je **A** simetrična matrica, postoji matrica **X** takva da je $\mathbf{A} = \mathrm{X}\Lambda\mathbf{X}^t$, pri čemu su λ_p, p = 1, ..., m svojstvene vrednosti matrice **A**, a kolone matrice **X** su svojstveni vektori matrice **A**. Ime ove teoreme potiče otuda što se skup svih različitih svojstvenih vrednosti naziva spektralnim skupom.

[181] O testovima statističke značajnosti pri odabiranju broja glavnih komponenata koje treba zadržati neće biti govora u ovom tekstu, jer se ovi testovi praktično nikad ne koriste. Distribucije uzorkovanja svojstvenih vrednosti inače su veoma kompikovane. O nekima od tih testova značajnosti može se pročitati u Kovačić (1994), str. 193–195 ili Wilks (1962), str. 568–573.

[182] Skri (engl. scree) je osulina, a u geologiji se ovaj termin koristi za ruševine koje se skupljaju na donjem delu stenovitog nagiba. U ovom slučaju „ruševine“ koje se skupljaju na dnu strmine su svojstvene vrednosti koje se nalaze ispod „prelomne“ tačke.

smo rekli prilikom izvođenja) sortirane u nerastućem nizu. Na apscisu se nanosi redni broj svojstvene vrednosti, a na ordinatu veličina svojstvene vrednosti. Broj komponenata koje treba zadržati određuje se tako što se iz poslednje, najmanje svojstvene vrednosti povuče linija tako da prati rast nekoliko zadnjih svojstvenih vrednosti (Grafik 6). Redni broj svojstvene vrednosti koja prva bude uočljivo iznad ove prave predstavlja i broj značajnih komponenata (na Grafiku 6 to je broj 5). U osnovi ovog postupka leži pretpostavka da su nevažne komponente posledica greške merenja i da je njihov doprinos objašnjenju varijanse skupa analiziranih varijabli zato slučajan, te se zbog toga i njima odgovarajuće svojstvene vrednosti ne razlikuju bitno jedna od druge (dakle fluktuiraju oko neke srednje vrednosti). Prva komponenta čija svojstvena vrednost odstupa od ovog trenda smatra se značajnom, tj. smatra se da ona objašnjava varijansu koja nije posledica samo greške, te iz tog razloga i odstupa od narednih svojstvenih vrednosti. Veliki problem sa ovim kriterijumom je što očitavanje dijagrama zavisi od iskustva istraživača i što ponekad nimalo nije jednostavno povući pravu liniju i odrediti koja najniža svojstvena vrednost je uočljivo iznad nje. Stoga ne postoji uvek dovoljno slaganja između različitih istraživača u očitavanju istog dijagrama. Ovaj kriterijum je opravdano primenjivati onda kada postoji sasvim očigledan nagli pad u veličini svojstvenih vrednosti.

Grafik 6. Demonstracija Katelovog skri-kriterijuma za određivanje broja značajnih komponenata (grafik preuzet uz dozvolu autora iz A. Zorić, (1996), Faktorska analiza binarnih varijabli u generalizovanom imaž prostoru, Diplomski rad, Beograd: Filozofski fakultet.)



Gutman–Kajzerov kriterijum zasniva se na meri pouzdanosti za glavnu komponentu p, u oznaci α_p, koju su ovi autori definisali na sledeći način:[183]

$$\alpha_p = \frac{m}{m-1}\left(1 - \frac{1}{\lambda_p}\right).$$

Kako pouzdanost može da se kreće u segmentu [0, 1], to se neznačajnim komponentama proglašavaju sve one komponente čija je pouzdanost, po Gutman–Kajzerovom određenju mere pouzdanosti, manja od nule. Ova anomalija mere pouzdanosti objašnjava se nerealnošću date komponente, tj. ona se proglašava artificijelnom, te zato i nevažnom. Prema ovom kriterijumu treba zadržati sve komponente čija je pouzdanost veća od nule:

[183] Pouzdanost pokazuje stepen tačnosti kojom se meri ono što je predmet merenja.

$$k = \mathrm{num}(\alpha_p > 0).$$

Očigledno je da je to ekvivalentno kriterijumu zadržavanja svih komponenata čija je varijansa, tj. svojstvena vrednost, veća od 1 ako su izvorne varijable standardizovane:

$$k = \mathrm{num}(\lambda_p > 1).$$

Istovremeno, glavne komponente koje imaju varijansu manju od 1 sadrže u sebi manje informacija no jedna standardizovana izvorna varijabla.
Ako se glavne komponente dobijaju na osnovu matrice kovarijansi, tj. iz nestandardizovanih izvornih varijabli, tada se zadržavaju glavne komponente sa svojstvenim vrednostima većim od proseka svih svojstvenih vrednosti:[184]

$$k = \mathrm{num}\,(\lambda_p > \frac{\sum_{1}^{m} \lambda_p}{\mathrm{m}}).$$

Gutman−Kajzerov kriterijum se za sada najčešće koristi za određivanje broja komponenata koje treba zadržati, premda mnoga statistička istraživanja ukazuju na njegovo slabo funkcionisanje (cf. Zorić & Opačić, 2013). Najverovatniji razlog za to je što je taj kriterijum implementiran kao automatska (difolt) opcija u programu SPSS.

Kriterijum koji počiva na procentu ukupne varijanse izvornih varijabli koja je obuhvaćena zadržanim komponentama svodi se na to da se zadrži k glavnih komponenata tako da

$$\frac{\sum_{1}^{k} \lambda_p}{\sum_{1}^{m} \lambda_p} > \mathrm{p}\,,$$

pri čemu p predstavlja proporciju ukupne varijanse merenih varijabli koja je obuhvaćena sa prvih k komponenata i određuje se najčešće kao 0.6, 0.7 ili 0.8. Dakle, treba zadržati onoliko glavnih komponenata dok proporcija ukupne varijanse izvornih varijabli koja je sadržana u tim komponentama ne pređe 60%, 70% ili 80%. Očigledno, kriterijum za vrednost p u ovom slučaju određuje sam istraživač.

Pored navedenih kriterijuma, za određivanje broja važnih glavnih komponenti relativno često se koriste i **Bartletov hi-kvadrat test** (Bartlet's Chi-square test), **kriterijum zasnovan na mimimumu prosečne parcijalne korelacije** (Minimum Average Partial Method − MAP) i **Hornova paralelna analiza** (Horn's Parallel analysis) i (CHall method). O detaljima ovih i još nekih postupaka koje nismo pomenuli u ovom tekstu može se pročitati u Zwick & Velicer (1986), Peres-Neto, Jackson, & Somers (2005) i Zorić & Opačić (2013).

Komponentni skorovi

Svojstveni vektori matrice **R** sadrže pondere kojima se množe rezultati na pojedinim izvornim varijablama da bi se dobile glavne komponente.[185] Prema tome, matrica sa

[184] Ponekada se, umesto aritmetičke sredine, u ove svrhe koristi geometrijska sredina svojstvenih vrednosti.

[185] Rezultat ispitanika e_i na glavnoj komponenti je linearna kombinacija njegovih rezultata u standardizovanom obliku na m izvornih, merenih varijabli:

$$k_{1i} = x_{11}z_{1i} + x_{12}z_{2i} + \cdots + x_{1m}z_{mi} = \mathbf{Zx_1}$$
$$k_{2i} = x_{21}z_{1i} + x_{22}z_{2i} + \cdots + x_{2m}z_{mi} = \mathbf{Zx_2}$$

komponentnim skorovima **K** dobija se množenjem matrice podataka, u kojoj su standardizovani rezultati na izvornim varijablama, matricom čije kolone sadrže svojstvene vektore matrice **R**:

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

Tako dobijene nekorelisane glavne komponente mogu se koristiti u daljim analizama, na primer kao prediktorske varijable u multiploj regresionoj analizi.
Glavne komponente dobijene na osnovu standardizovanih varijabli nužno imaju aritmetičke sredine jednake nuli, ali im varijanse nisu jednake jedinici.[186] Ukoliko želimo da ih standardizujemo, potrebno je da svaku glavnu komponentu podelimo korenom iz njene varijanse. Matrica standardizovanih glavnih komponenata $\mathbf{K}_s$ se iz matrice **K** nestandardizovanih komponenata (koje su nastale linearnom kombinacijom standardizovanih merenih varijabli) dobija množenjem sa dijagonalnom matricom $\Lambda^{-1/2}$ u čijoj dijagonali su recipročne vrednosti kvadratnih korenova varijansi glavnih komponenata:

$$\mathbf{K}_s = \mathbf{K}\Lambda^{-1/2}.$$

Uslovi za korišćenje analize glavnih komponenata:

Da bi se moglo imati poverenje u rezultate dobijene analizom glavnih komponenata, poželjno je da uzorak ispitanika bude ne samo slučajan već i dovoljno veliki (preko sto ispitanika ili bar pet puta više ispitanika od broja merenih varijabli). Osim toga, budući da je u ovom tekstu prikazan postupak koji se zasniva na matrici kovarijansi ili linearnih korelacija, potrebno je da budu ispunjeni uslovi za korišćenje koeficijenta linearne korelacije da bi se primenjivao prikazani postupak analize glavnih komponenata: podaci moraju poticati bar sa intervalne skale, distribucije varijabli treba da budu normalne (ili bar simetrične), a povezanost među varijablama linearna.

****

U SPSS programu analiza glavnih komponenata može se izvesti pomoću menija **Analyze/Dimension Reduction/Factor**. U prozoru za dijalog **Factor Analysis** merene varijable na kojima će se primeniti analiza treba prebaciti iz levog u desni okvir. Klikom na dugme **Descriptives,** može se ući u poddijalog **Factor Analysis: Descriptives** u kojem je moguće, ukoliko je to potrebno, zatražiti aritmetičke sredine, standardne devijacije i broj rezultata za svaku varijablu (opcija **Univariate Descriptives**), matricu interkorelacija izvornih varijabli (opcija **Coefficients** u okviru **Correlation Matrix**) ili test sferičnosti matrice interkorelacija (opcija **KMO and Bartlett's test of Sphericity**). Bartletovim testom sferičnosti testira se nulta hipoteza koja glasi: populaciona matrica interkorelacija **P** jednaka je matrici identiteta (H_0: $\mathbf{P} = \mathbf{I}$). Na taj način proverava se da li uopšte ima smisla primenjivati analizu glavnih komponenata na dobijenim podacima. Naime, ako se navedena nulta hipoteza ne odbaci, analiza glavnih komponenata nema smisla.

....

$$k_{mi} = x_{m1}z_{1i} + x_{m2}z_{2i} + \cdots + x_{mm}z_{mi} = \mathbf{Zx_m}$$

ili u matričnom obliku: $\mathbf{K} = \mathbf{ZX}$. U matrici **K** (koja je reda nxm) su komponentni skorovi za n ispitanika na m komponenata, u matrici **Z** (reda nxm) su standardizovani rezultati n ispitanika na m izvornih varijabli, a kolona *p* matrice **X** (reda mxm) sadrži koeficijente ili pondere za glavnu komponentu *p*.

[186] O aritmetičkoj sredini i varijansi linearne kombinacije može se pogledati u odeljku **XIV.1**.

Klikom na dugme **Extraction** u polju **Method**, treba samo proveriti da li je aktivna opcija **Principal Components**. Ova opcija se podrazumeva i ne treba je menjati kada se želi izvođenje analize glavnih komponenata. Promena polja **Method** ima smisla samo u faktorskoj analizi. U novijim verzijama programa (SPSS 7.5 ili 8.0) moguće je (u okviru **Analyse**) definisati da li želimo analizu glavnih komponenata na osnovu standardizovanih varijabli (opcija **Correlation Matrix**) ili na osnovu nestandardizovanih izvornih varijabli (opcija **Covariance Matrix**). Analiza glavnih komponenata, kao što smo već ranije istakli, nije metrički invarijantna: ako se izvodi na osnovu matrice kovarijansi, analiza glavnih komponenata uzima više u obzir varijable koje imaju veću varijansu, a potcenjuje one sa manjom varijansom. To je posledica osnovnog cilja kojem služi analiza glavnih komponenata: napraviti ortogonalne komponente sa što većom varijansom. Iako ima situacija kada je analiza glavnih komponenata na osnovu matrice kovarijansi smisaona i opravdana, s obzirom na prirodu psiholoških varijabli, to najčešće nije smisaono. Stoga je, ako ne postoji jasan razlog protiv toga, najbolje ovu analizu uraditi na osnovu korelacione matrice izvornih varijabli (dakle, sa uključenom opcijom **Correlation Matrix** u novijim verzijama programa). Starije verzije programa (na primer SPSS 6.1) podrazumevaju da se analiza glavnih komponenata radi na osnovu korelacione matrice.
U okviru **Extract** treba uključiti opciju **Number of factors** i u ulazno polje treba upisati broj koji odgovara broju početnih, merenih varijabli.
U okviru **Display**, pored uključene opcije **Unrotated factor solution**, može se uključiti opcija **Scree Plot** na osnovu koje je moguće primeniti Katelov postupak za određivanje broja važnih komponenata.
Klikom na dugme **Scores** u glavnom prozoru faktorske analize, ulazimo u poddijalog **Factor Analysis: Factor Scores** u kojem se može, uključivanjem opcije **Save as variables**, zatražiti da se rezultati na glavnim komponentama snime kao poslednje varijable u radni prozor sa podacima. Uključenu opciju **Regression** u okviru **Method** ne treba menjati. Posle klika na dugme **Options** u glavnom prozoru, možemo definisati postupak u slučaju podatka koji nedostaje: uključivanjem **Exclude cases listwise** – izbacivanje ispitanika iz svih analiza ako samo na jednoj varijabli nema podatka, i **Exclude cases pairwise** – izbacivanje ispitanika samo iz računanja korelacija za one varijable na kojima nema podatka. Opcija **listwise** je u ovoj analizi smislena, dok opcija **pairwise** nije uopšte smislena, jer se tako ne dobija prava matrica korelacija, dakle Gramova matrica.[187] Uključivanjem opcije **Replace with mean**, moguće je definisati da se podatak koji nedostaje u nekoj varijabli zameni aritmetičkom sredinom za datu varijablu. Postoje i dve opcije koje kontrolišu izgled matrice korelacija izvornih varijabli i glavnih komponenata: opcija **Sorted by size** omogućuje da se koeficijenti korelacije izvornih varijabli i komponenata prikažu u ispisu poređani po veličini, a opcija **Suppress absolute values less then** *r* dovodi do toga da se koeficijenti korelacija izvornih varijabli i komponenata koji su po apsolutnoj vrednosti manji od *r* ne pojavljuju u ispisu.

<u>Primer:</u>

Na uzorku izbeglica sa ratnih područja BiH i Hrvatske koje su pribežište našle u Srbiji (n = 198), a doživele su neki od ozbiljnijih traumatskih događaja tokom rata, dobijeni su odgovori na pojedinačne stavke upitnika IES (Impact of event scale – Skale uticaja događaja). Autori upitnika su Horovic, Vilner i Alvarez, a namenjen je merenju postraumatskih stresnih reakcija. Skala je bila prvobitno namenjena utvrđivanju

[187] Gramova ili Gramijanska matrica je simetrična i pozitivno semidefinitna ($\mathbf{x}^t\mathbf{R}\mathbf{x} \geq 0$). Ova matrica nastaje kao proizvod određene matrice i njenog transpona ($\mathbf{R} = \mathbf{Z}^t\mathbf{Z}$).

psiholoških posledica civilne traume, a potom se (pa i u ovom istraživanju) pokazala korisnom u ispitivanju psiholoških posledica ratnih trauma.
Na pojedine stavke IES upitnika (na primer: „Mislio sam o tome i kada nisam nameravao“; „Pokušao sam da ne razgovaram o tome“; „Sve što me podseća na to ponovo me uznemiri“) odgovara se zaokruživanjem jednog od ponuđenih odgovora:
Uopšte ne (0 pena); Veoma retko (1 poen); Ponekad (3 poena); Često (5 poena).
Pitanje na koje želimo da odgovorimo glasi: da li se skup ovih petnaest merenih varijabli može svesti na manji broj dimenzija koje bi u sebi sadržavale bitne informacije sadržane u izvornim varijablama? Drugim rečima: da li se ovaj skup manifestnih postraumatskih reakcija može svesti na manji broj dimenzija?
Da bismo napravili analizu glavnih komponenata ovog skupa merenih varijabli, odabrali smo **Statistics/Data Reduction/Factor**, a zatim u okvir **Variables** ubacili imena varijabli od IES1 do IES15. Posle klika na dugme **Descriptives**, u okviru **Correlation Matrix** uključili smo opciju **KMO and Bartlett's test of Sphericity**. U poddijalogu **Extraction** ostavili smo difolt opciju **Principal Components**, u okviru **Extract** uključili opciju **Number of factors** i u ulazno polje upisali **15**. U okviru **Display** ostavili smo uključenu opciju **Unrotated factor solution**. Po kliku na dugme **Scores**, uključena je opcija **Save as variables**, a u okviru **Method** ostavljena je (automatski uključena) opcija **Regression**. Na taj način zatraženo je da program u aktivnom fajlu sa podacima upiše u novim varijablama (čija imena program prikazuje na kraju ispisa) komponentne skorove ispitanika. U okviru **Coefficient display format** poddijaloga **Options**, uključivanjem **Sorted by size**, definisano je da se u ispisu koeficijenti korelacija izvornih varijabli i glavnih komponenata prikazuju sortirani po veličini i da se ne prikazuju korelacije koje su manje po apsolutnoj vrednosti od 0.20 (uključena je opcija **Suppress absolute values less than** i u ulazno polje upisano **.20**).

Ispis iz analize glavnih komponenata iz SPSS 6.1:[188]

(objašnjenja pojedinih delova ispisa uokvirena su crticama)

- - - - - - - - - - -   F A C T O R   A N A L Y S I S   - - - - - - - - - - -

Analysis number 1   Listwise deletion of cases with missing values

Kaiser–Meyer–Olkin Measure of Sampling Adequacy = .92892

Kajzer-Mejer-Olkinova mera adekvatnosti uzorkovanja je mera reprezentativnosti merenih varijabli za univerzum varijabli koje se odnose na dato područje. Budući da je viša od 0.80, to pokazuje zadovoljavajuću reprezentativnost.[189]

Bartlett Test of Sphericity = 1715.2642, Significance =   .00000

----

[188] Budući da je ispis iz ove metode praktično isti i u novijim verzijama programa, ovde će biti dat ispis samo iz verzije 6.1, dok je prikazani grafik osulina (Grafik 6) napravljen u verziji 8.0 ovog programa.
[189] O ovoj meri može se pročitati u Momirović, Wolf i Popović (1999).

Budući da je **Significance** manje od .05, odbacujemo nultu hipotezu da je matrica interkorelacija merenih varijabli u populaciji jednaka matrici identiteta, što znači da ima smisla raditi analizu glavnih komponenata na datim varijablama.

Extraction 1 for analysis 1, Principal Components Analysis (PC)
Initial Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
IES1	1.00000	*	1	7.31400	48.8	48.8
IES2	1.00000	*	2	1.53233	10.2	59.0
IES3	1.00000	*	3	1.04987	7.0	66.0
IES4	1.00000	*	4	.79555	5.3	71.3
IES5	1.00000	*	5	.70775	4.7	76.0
IES6	1.00000	*	6	.66197	4.4	80.4
IES7	1.00000	*	7	.49282	3.3	83.7
IES8	1.00000	*	8	.41794	2.8	86.5
IES9	1.00000	*	9	.37669	2.5	89.0
IES10	1.00000	*	10	.36448	2.4	91.4
IES11	1.00000	*	11	.32272	2.2	93.6
IES12	1.00000	*	12	.30965	2.1	95.6
IES13	1.00000	*	13	.24115	1.6	97.2
IES14	1.00000	*	14	.22127	1.5	98.7
IES15	1.00000	*	15	.19179	1.3	100.0

PC extracted 15 factors.

U koloni **Eigenvalue** su svojstvene vrednosti matrice interkorelacija, a u koloni **Pct of Var** procenat ukupne varijanse svih merenih varijabli koji je obuhvaćen datom glavnom komponentom. U koloni **Cum Pct** je procenat ukupne varijanse merenih varijabli koji je obuhvaćen datom komponentom i svim komponentama pre nje. Treba uočiti da je zbir svojstvenih vrednosti jednak 15, tj. broju varijabli. To je uvek slučaj kada glavne komponente ekstrahujemo iz matrice interkorelacija, jer je tada varijansa svake izvorne varijable jednaka 1.

Factor Matrix:[190]

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
IES5	.84694	-.24407			
IES14	.81931	-.21023			
IES10	.81714	-.30879			
IES7	.75377	.40631			
IES13	.74361	.44439			
IES1	.73602	-.31559	-.21449		
IES4	.73298				-.37912
IES12	.72901	-.32217			
IES6	.72549				-.33474
IES11	.68684	-.31617			.43260
IES2	.68383		-.32265	.22764	.35778
IES3	.64801	.49851	-.24949		
IES9	.62206	.46391			
IES8	.30588	.38983	.66545	-.31362	
IES15	.40001		.53152	.72762	

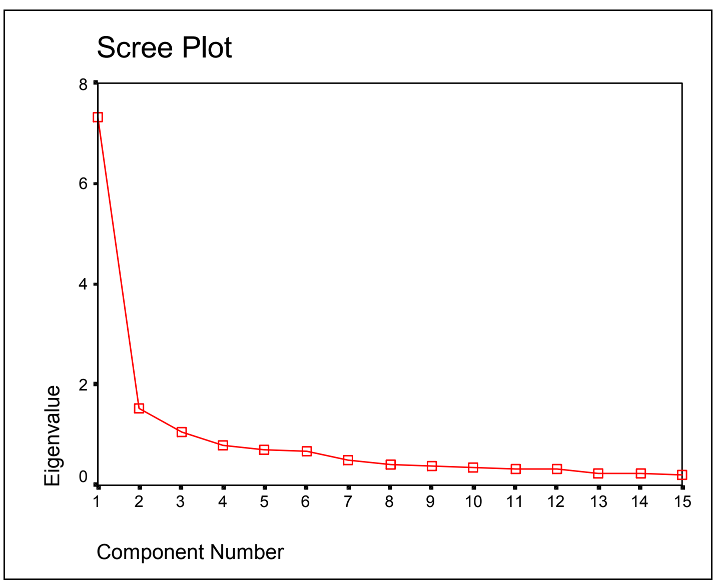
	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
IES5					
IES14			-.24975		
IES10					
IES7			-.25945		
IES13					

[190] U novijim verzijama programa SPSS ova matrica zove se **Component Matrix**.

IES1	.26293			.32462	
IES4		.28399	.38716		
IES12		.32431			-.41093
IES6		-.43350			-.27843
IES11				-.36122	
IES2	.22952	-.25921			
IES3	.25291				
IES9	-.43047		.27344	.20220	
IES8	.42351				
IES15					

	Factor 11	Factor 12	Factor 13	Factor 14	Factor 15
IES5			-.27677	.20024	-.20638
IES14				.20656	
IES10				-.35157	
IES7	.20630				.21815
IES13	-.21175	-.30899			-.20684
IES1	.24642				
IES4					
IES12					
IES6					
IES11					
IES2					
IES3		.28813			
IES9					
IES8					
IES15					

Koeficijenti u „faktorskoj" matrici (**Factor Matrix**), kako je SPSS naslovljava, jesu koeficijenti linearnih korelacija svake merene varijable i glavnih komponenata. Ovi koeficijenti sortirani su po veličini, jer smo to pri definisanju procedure tražili. Prazna mesta su umesto koeficijenata koji su manji od 0.20, što smo takođe zahtevali pri definisanju procedure (uključivanjem **Sorted by size** i **Suppress absolute values less than 0.20** u okviru **Coefficient display format** u poddijalogu **Options**).



Grafik pod naslovom **Scree Plot** pokazuje promene u veličinama svojstvenih vrednosti idući od prve do poslednje glavne komponente

----

Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
		*				
IES1	1.00000	*	1	7.31400	48.8	48.8
IES2	1.00000	*	2	1.53233	10.2	59.0
IES3	1.00000	*	3	1.04987	7.0	66.0
IES4	1.00000	*	4	.79555	5.3	71.3
IES5	1.00000	*	5	.70775	4.7	76.0
IES6	1.00000	*	6	.66197	4.4	80.4
IES7	1.00000	*	7	.49282	3.3	83.7
IES8	1.00000	*	8	.41794	2.8	86.5
IES9	1.00000	*	9	.37669	2.5	89.0
IES10	1.00000	*	10	.36448	2.4	91.4
IES11	1.00000	*	11	.32272	2.2	93.6
IES12	1.00000	*	12	.30965	2.1	95.6
IES13	1.00000	*	13	.24115	1.6	97.2
IES14	1.00000	*	14	.22127	1.5	98.7
IES15	1.00000	*	15	.19179	1.3	100.0

Skipping rotation 1 for extraction 1 in analysis 1

15 PC EXACT factor scores will be saved.

```
Following factor scores will be added to the working file:

  Name       Label

FAC1_1       REGR factor score   1 for analysis    1
FAC2_1       REGR factor score   2 for analysis    1
FAC3_1       REGR factor score   3 for analysis    1
FAC4_1       REGR factor score   4 for analysis    1
FAC5_1       REGR factor score   5 for analysis    1
FAC6_1       REGR factor score   6 for analysis    1
FAC7_1       REGR factor score   7 for analysis    1
FAC8_1       REGR factor score   8 for analysis    1
FAC9_1       REGR factor score   9 for analysis    1
FAC10_1      REGR factor score  10 for analysis    1
FAC11_1      REGR factor score  11 for analysis    1
FAC12_1      REGR factor score  12 for analysis    1
FAC13_1      REGR factor score  13 for analysis    1
FAC14_1      REGR factor score  14 for analysis    1
FAC15_1      REGR factor score  15 for analysis    1
```

Ispis pod naslovom **Final Statistics** potpuno je istovetan onom pod naslovom **Initial Statistics**. Razlike u ova dva dela ispisa pojavljuju se tokom rotacije faktora u faktorskoj analizi. Kolona **Communality** u oba slučaja sadrži jedinice za svaku merenu varijablu. Komunaliteti pojedinih merenih varijabli (ili u terminima faktorskog modela – manifestnih varijabli) takođe nisu od interesa u analizi glavnih komponenata. Pojam komunaliteta veoma je važan u faktorskoj analizi u kojoj predstavlja proporciju ukupne varijanse svake manifestne varijable koja je objašnjena zajedničkim faktorima.

Deo ispisa na kraju ispod naslova **`Following factor scores will be added to the working file`** služi kao obaveštenje o imenima varijabli (kolona **Name**) u kojima su, u aktivnom fajlu sa podacima, zapisani skorovi ispitanika na glavnim komponentama.

Pre odgovora na pitanje koje smo postavili pre izvođenja analize glavnih komponenata treba uočiti da bi prema Gutman–Kajzerovom kriterijumu postojale tri važne glavne komponente. Sličnu sugestiju daje i Katelov dijagram osulina. Ukoliko bismo zadržali prve tri glavne komponente, onda bismo njima obuhvatili 66% varijanse svih merenih varijabli, tj. svih stavki iz upitnika. Međutim, pre konačne odluke dobro je pokušati sa interpretacijom glavnih komponenata. To, naravno, podrazumeva dobro poznavanje sadržaja merenih varijabli. Na osnovu matrice korelacija pojedinih stavki i glavnih komponenata (**Factor Matrix**) vidimo da sa prvom komponentom najvišu korelaciju pokazuju stavke 5 („Navirala su mi snažna osećanja u vezi sa tim“), 14 („Sve što me podseća na taj događaj ponovo mi vraća i osećanja vezana za njega“) i 10 („Te slike se same javljaju i kad mislim o nečem drugom“). Očigledno je reč o komponenti koja u sebi sadrži elemente nametanja, tj. nevoljnog ponovnog preživljavanja traumatskog događaja. Druga komponenta ima najviše korelacije sa stavkama 3 („Pokušao sam da to izbacim iz sećanja“), 9 („Pokušao sam da ne pričam o tome“), 7 („Izbegavao sam sve što me na taj događaj podseća“) i 13 („Pokušao sam da ne mislim o tome“) i predstavlja reakcije izbegavanja stimulusa u vezi sa traumom, ali te reakcije nisu u vezi sa nametanjem. Reakcije izbegavanja koje su u vezi sa nametanjem sadržane su u prvoj glavnoj komponenti. Treća komponenta korelira najviše sa stavkama 8 („Imao sam osećaj kao da se to nije desilo ili nije bilo stvarno“) i 15 („Moja osećanja vezana uz taj događaj kao da su zaleđena“) i podseća na reakciju poricanja (denial) i emocionalne otupljenosti.

Dakle, skup od 15 merenih varijabli (stavki) mogao bi se u ovom slučaju transformisati, tj. svesti na tri glavne komponente koje bi u sebi sadržale sve bitne informacije (tj. veliki deo

varijanse 15 merenih varijabli). Naravno, treća zadržana komponenta je u izvesnoj meri problematična, jer sadrži uglavnom informacije iz dveju stavki. Detaljnija analiza upitnika mogla bi dovesti do toga da ove stavke eliminišemo iz upitnika i da se, na taj način, celokupna relevantna informacija koja je sadržana u odgovorima ispitanika na preostalih 13 stavki svede na dve glavne komponente.
****

Reference na koje se upućuje u ovom tekstu:

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

Momirović, K., Wolf, B. i Popović, D. A. (1999). *Uvod u teoriju merenja I: Interne metrijske karakteristike kompozitnih mernih instrumenata*. Priština: Univerzitet u Prištini, Fakultet za fizičku kulturu.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*(4), 974–997. doi:10.1016/j.csda.2004.06.015

Zorić, A. (1996). *Faktorska analiza binarnih varijabli u generalizovanom imaž prostoru*, Diplomski rad. Filozofski fakultet, Beograd.

Zorić, A., & Opačić, G. (2013). Impact of different conditions on accuracy of five rules for principal components retention. *Psihologija*, *46*(3), 331–347. DOI:10.2298/PSI130801008Z

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442.

Wilks, S. S. (1962). *Mathematical Statistics*. New York: John Wiley & Sons, Inc.