

**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Matea Mijatović

**ANALIZA INTERESA  
SREDNJOŠKOLACA I STUDENATA  
EKSPLOLATIVNOM FAKTORSKOM  
ANALIZOM**

Diplomski rad

Voditelj rada:  
prof. dr. sc. Anamarija Jazbec

Zagreb, Veljača 2020.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

*Želim zahvaliti svojim roditeljima, sestrama, bratu i Lovri na svojoj podršci tokom studiranja. Uz vas je sve bilo puno lakše. Posebna zahvala prof. dr. sc. Anamariji Jazbec na pomoći i savjetima pri izradi diplomskog rada.*

# Sadržaj

<b>Sadržaj</b>	<b>iv</b>
<b>Uvod</b>	<b>1</b>
<b>1 Faktorska analiza</b>	<b>3</b>
1.1 Općenito . . . . .	3
1.2 Definicije . . . . .	4
1.3 Statistički model . . . . .	5
1.4 Problemi . . . . .	6
1.4.1 Teorijski problemi . . . . .	6
1.4.2 Praktični problemi . . . . .	7
1.4.2.1 Veličina uzorka . . . . .	7
1.4.2.2 Normalnost . . . . .	7
1.4.2.3 Linearnost . . . . .	7
1.4.2.4 Faktorabilnost matrice korelacija . . . . .	7
1.5 Eksplorativna faktorska analiza . . . . .	8
1.6 Analiza glavnih komponenti . . . . .	8
1.7 Rotacije . . . . .	9
1.7.1 Ortogonalne rotacije . . . . .	9
1.7.2 Kose rotacije . . . . .	11
1.8 Komunaliteti . . . . .	11
1.9 Adekvatnost ekstrakcije i broj faktora . . . . .	12
1.10 Važnost i unutarnja konzistentnost faktora . . . . .	14
1.11 Faktorski rezultati . . . . .	14
<b>2 Primjer</b>	<b>15</b>
2.1 Analiza anketa . . . . .	15
2.2 Faktorska analiza podataka . . . . .	25
<b>3 Prilog</b>	<b>43</b>

<i>SADRŽAJ</i>	v
3.1 Anketa . . . . .	43
3.2 Kod . . . . .	49
<b>Bibliografija</b>	<b>55</b>

# Uvod

Faktorska analiza je statistička metoda koja se bavi opisivanjem varijabilnosti između promatranih, koreliranih varijabli u svrhu potencijalnog smanjivanja broja na manje neopaženih varijabli koje nazivamo faktorima. Faktorska analiza traži zajedničke varijacije kao odgovor na neopažene latentne varijable.

Eksplorativna faktorska analiza (EFA) služi kako bi se utvrdili temeljni faktori ili izvori varijacija i kovarijacija među promatranim varijablama. Ova vrsta analize daje nam faktorsku strukturu (grupaciju varijabli temeljenu na snažnoj korelaciji). EFA metoda ne koristi a priori znanje o pripadnosti predmeta određenoj strukturi, tj. ne postoji model od kojeg kreće.

Među srednjoškolcima i studentima iz Hrvatske provedena je online anketa po uzoru na anketu provedenu u SAD-u [6]. Upitnik je sadržavao pitanja vezana za interese o glazbi, filmovima, sportu, zdravlju i fobijama. Cilj ovog rada je pomoću eksplorativne faktorske analize pronaći koje se varijable (manifestne) međusobno grupiraju.



# Poglavlje 1

## Faktorska analiza

### 1.1 Općenito

Faktorska analiza (FA) je zajedničko ime za više metoda kojima je cilj kondenzacija većeg broja manifestnih varijabli, među kojima postoji povezanost (korelacija), na manji broj latentnih dimenzija (faktora) koje su izvor te povezanosti. Želimo otkriti koje varijable u skupu čine koherentne podskupove koji su relativno nezavisni jedan od drugoga. Varijable koje su korelirane jedna s drugom, a nezavisne od ostalih podskupova varijabli, kombiniraju se u **faktore**. Za faktore se smatra da odražavaju pozadinske (latentne) procese koji su stvorili korelacije među varijablama [7].

Specifičan cilj faktorske analize je sažeti uzorak korelacija između opaženih varijabli, kako bi se reducirao njihov broj na manji broj faktora, s ciljem pružanja definicije pozadinskih procesa opaženih varijabli, ili kako bi se testirala teorija o prirodi tih pozadinskih procesa.

Faktorska analiza proizvodi nekoliko linearnih kombinacija opaženih varijabli, pri čemu je svaka linearna kombinacija faktor. Faktori sažimaju uzorke korelacija u opaženoj matrici korelacija i mogu se koristiti za stvaranje opažene matrice korelacija. Kada se procjenjuju faktori za svaki predmet, oni su često pouzdaniji od rezultata za pojedine promatrane varijable.

Koraci u faktorskoj analizi su izabiranje i mjerenje skupa varijabli, pripremanje matrice korelacija, vađenje skupa faktora iz matrice korelacija, određivanje broja faktora, moguće rotiranje faktora za povećanje interpretabilnosti i na kraju interpretiranje rezultata. Iako za svaki korak postoje relevantna statistička razmatranja, važan test ove analize je njegova interpretabilnost.

Dobra faktorska analiza "ima smisla", loša nema. Interpretacija i davanje imena faktorima ovisi o značenju pojedine kombinacije opaženih varijabli koje su jako korelirane s tim faktorom. Lakše ga je interpretirati ako varijable koje su jako korelirane s njim, nisu



korelirane s drugim faktorima.

Nakon što je interpretacija odgovarajuća, posljednji i veliki korak je provjera strukture faktora utvrđivanjem konstruktne valjanosti faktora. Želimo pokazati da rezultati na latentnim varijablama (faktorima) kovariraju s rezultatima na drugim varijablama, ili da se rezultati na latentnim varijablama mijenjaju s eksperimentalnim uvjetima kako je teorija i predviđjela.

Prvi problem s faktorskom analizom je nepostojanje spremnih kriterija protiv kojih testiramo rješenje. Ne postoji neki vanjski kriterij poput članstva u grupi protiv kojeg treba testirati rješenje. Drugi problem, koji nastaje nakon izvlačenja, je postojanje beskonačnog broja mogućih rotacija gdje svaka računa na isti iznos varijance u originalnim podacima, ali s malo drukčije definiranim faktorima. Konačni izbor među alternativama ovisi o našoj procjeni interpretabilnosti i znanstvene korisnosti. U prisutnosti beskonačnog broja matematički istih rješenja, dolazi do razlikovanja po pitanju koja je najbolja.

## 1.2 Definicije

Prije nego što nastavimo dalje, definirat ćemo neke osnovne pojmove. Matrica koja je nastala od opaženih varijabli naziva se **opažena korelacijska matrica** (eng. *observed correlation matrix*). Korelacijska matrica koja je nastala iz faktora naziva se **reproducirana korelacijska matrica** (eng. *reproduced correlation matrix*). Razlika između tih dviju matrica je **korelacijska matrica reziduala** (eng. *residual correlation matrix*). U dobro napravljenoj faktorskoj analizi korelacije u matricama reziduala su male, što ukazuje na sličnost opažene i reproducirane matrice.

Drugi skup pojmova odnosi se na matrice proizvedene i interpretirane kao dio rješenja. Rotacija faktora je proces pomoću kojega rješenje postaje jasnije bez da se mijenjaju temeljna matematička svojstva. Postoje dva osnovna tipa rotacije: ortogonalna i kosa. Ako je rotacija **ortogonalna** (tako da su svi faktori nekorelirani jedni sa drugim) tada nastaje **početna matrica** (eng. *loading matrix*). To je matrica korelacija između opaženih varijabli i faktora. Veličina matrice odražava opseg odnosa između svake opažene varijable i svakog faktora. Ortogonalna faktorska analiza se tumači iz matrice učitavanja gledanjem koja opažena varijabla korelira sa svakim faktorom.

Ako je rotacija **kosa** (tako da su faktori međusobno korelirani), nastaje nekoliko dodatnih matrica. **Faktorska korelacijska matrica** (eng. *factor correlation matrix*) sadrži korelacije između faktora. Matrica učitavanja iz ortogonalne rotacije se rastavlja na dva dijela: **strukturna matrica** (eng. *structure matrix*) korelacija između faktora i varijabli i **matrica uzoraka** (eng. *pattern matrix*) jedinstvenih veza između svakog faktora i svake opažene varijable. Nakon kose rotacije, značenje faktora utvrđuje se iz matrice uzoraka.

Na kraju, oba tipa rotacije imaju **faktorsku matricu koeficijenata** (eng. *score coefficients matrix*), matricu koja se koristi u više regresijskih jednadžbi za predviđanje rezultata

na faktorima iz rezultata na opažene varijable za svakog pojedinca.

### 1.3 Statistički model

Pretpostavimo da imamo skup od  $p$  opaženih varijabli,  $x_1, \dots, x_p$  s očekivanjima  $\mu_1, \dots, \mu_p$  ([5], [2]).

Pretpostavimo da za nepoznate konstante  $\omega_{ij}$  i  $m$  neopaženih slučajnih varijabli  $F_j$  (te varijable nazivamo 'zajednički faktori' jer utječu na sve opažene varijable), gdje je  $i \in 1, \dots, p$  i  $j \in 1, \dots, m$ ,  $m < p$ , imamo

$$X_{px1} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, F_{mx1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \mu_{px1} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

$$Cov(X) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

$$X_1 = \mu_1 + \omega_{11}F_1 + \omega_{12}F_2 + \dots + \omega_{1m}F_m + \varepsilon_1$$

$$X_2 = \mu_2 + \omega_{21}F_1 + \omega_{22}F_2 + \dots + \omega_{2m}F_m + \varepsilon_2$$

$$\vdots$$

$$X_p = \mu_p + \omega_{p1}F_1 + \omega_{p2}F_2 + \dots + \omega_{pm}F_m + \varepsilon_p.$$

Ovdje su  $\varepsilon_i$  neopaženi uvjeti stohastičkih greški s očekivanjem 0 i konačnom varijancom, koja ne mora biti jednaka za sve  $i$ . Matrični zapis ovih jednadžbi glasi:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

Iz toga slijedi:

$$X = \mu + \omega F + \varepsilon$$

$$X - \mu = \omega F + \varepsilon \longrightarrow \text{FAKTORSKI MODEL}$$

Vrijede sljedeće pretpostavke:

- $F$  i  $\varepsilon$  su nezavisne
- $\mathbb{E}(X) = \mu$ ,  $\mathbb{E}(F) = 0$  i  $\mathbb{E}(\varepsilon) = 0$
- $Cov(X) = \Sigma$ ,  $Cov(F) = Cov(FF^T) = I$

$$Cov(\varepsilon) = \varepsilon = \begin{bmatrix} \varepsilon_{11} & 0 & \dots & 0 \\ 0 & \varepsilon_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \varepsilon_{pp} \end{bmatrix}.$$

## 1.4 Problemi

### 1.4.1 Teorijski problemi

Istraživački projekt koji je posebno osmišljen za faktorsku analizu, razlikuje se od drugih u nekoliko važnih aspekta.

Prvi zadatak u takvom projektu je generirati hipoteze o faktorima za koje se vjeruje da podupiru domenu interesa. Statistički, važno je napraviti istraživanje dovoljno širokim da uključuje pet ili šest pretpostavljenih faktora kako bi istraživanje bilo stabilno. Kako bi otkrili pozadinske procese područja istraživanja, svi relevantni faktori moraju biti uključeni. Neuspjeh u mjerenju nekih važnih faktora može narušiti odnose između mjerenih faktora. Uključivanje svih relevantnih faktora predstavlja logički, ali ne i statistički problem.

Idući korak je izbor varijabli koje ćemo promatrati. Za svaki pretpostavljeni faktor uključeno je pet ili šest varijabli, i svaka se smatra relativno čistom mjerom faktora. Takve varijable su jako korelirane s jednim i samo s jednim faktorom neovisno o načinu ekstrakcije ili tehnike rotiranja. Korisne su jer jasno opisuju prirodu faktora.

Složenost varijabli je isto uzeta u obzir. Složenost je označena brojem faktora s kojima varijabla korelira. Čista varijabla (eng. *pure variable*), što je preferirano, korelira samo s jednim faktorom dok složena varijabla korelira s više njih. Ako su varijable s različitom složenošću uključene u analizu, one sa sličnom složenošću mogu biti spojene u faktor koji nema nikakve veze s pozadinskim procesima. Varijable sa sličnom složenošću mogu korelirati jedne s drugima zbog njihove složenosti, a ne zato što su povezane s istim faktorom. Procjena (ili izbjegavanje) složenosti varijable je dio generiranja hipoteza o faktorima i biranje varijabli kako bi ih izmjerili.

## 1.4.2 Praktični problemi

### 1.4.2.1 Veličina uzorka

Koeficijenti korelacije obično su manje pouzdani ako se procjenjuju iz malih uzoraka. Stoga je važno da veličina uzorka bude dovoljno velika kako bi se korelacije pouzdano procijenile. Potrebna veličina uzorka također ovisi o veličini korelacije populacije i broju faktora: ako su korelacije jake i postoji nekoliko različitih faktora, mali uzorak je dovoljan.

### 1.4.2.2 Normalnost

Sve dok se FA koristi opisno kao prikladni način za sažimanje odnosa u velikom uzorku opaženih varijabli, pretpostavke na distribuciju varijabli nisu od važnosti. Ako su varijable normalno distribuirane, rješenje se poboljša. U slučaju da normalnost ne uspije, rješenje je degradirano, ali možda i dalje vrijedi. Međutim, multivarijatna normalnost je pretpostavljena kada se statistički zaključak koristi za određivanje broja faktora. Multivarijatna normalnost je pretpostavka da su sve varijable, i sve linearne kombinacije varijabli, normalno distribuirane.

### 1.4.2.3 Linearnost

Multivarijatna normalnost implicira da je veza između parova varijabli linearna. Analiza je degradirana u slučaju kad nema linearnosti zato što korelacija mjeri linearne veze i ne reflektira nelinearne veze. Linearnost među parovima varijabli procjenjuje se grafički.

### 1.4.2.4 Faktorabilnost matrice korelacija

Matrica koje je faktorabilna (eng. *factorable*) treba uključivati nekoliko značajnih korelacija. Očekivana veličina ovisi, donekle, o veličini uzorka, ali ako ne postoji korelacija koja je veća od 0.30 upitno je korištenje faktorske analize jer vjerojatno nemamo što faktorski analizirati.

Visoke bivarijatne korelacije nisu čvrsti dokaz da korelacijska matrica sadrži faktore. Moguće je da su korelacije samo između dvije varijable i ne odražavaju pozadinske procese koji istovremeno utječu na nekoliko varijabli. Zbog tog razloga, korisno je prvo ispitati matricu parcijalnih korelacija gdje su korelacije u parovima prilagođene efektima svih drugih varijabli.

Bartlettov test sferičnosti je osjetljiv test hipoteze kojim testiramo jesu li korelacije u korelacijskoj matrici jednake nuli. Testira hipotezu da je matrica korelacija matrica identiteta. Zbog svoje osjetljivosti i ovisnosti o veličini uzorka, test će vjerojatno biti značajan za uzorke velikih veličina, čak i ako su korelacije vrlo male.

## 1.5 Eksplorativna faktorska analiza

Postoje dva glavna tipa faktorske analize: eksplorativna i konfirmatorna. U eksplorativnoj faktorskoj analizi (EFA) se nastoji opisati i sažeti podatke tako što grupiramo varijable koje su korelirane [1]. Varijable mogu i ne moraju biti izabrane imajući na umu pozadinske procese. Koristi se u ranoj fazi istraživanja kada pruža alat za ujedinjavanje podataka i za generaciju hipoteza o pozadinskim procesima. Konfirmatorna analiza je tehnika koja se koristi u istraživanju za testiranja već unaprijed definiranih modela.

EFA služi za identifikaciju skupa latentnih konstrukcija koje leže u bazi izmjerenih varijabli. Temelji se na modelu zajedničkog faktora. U ovom modelu manifestne varijable su izražene kao funkcija zajedničkih faktora, jedinstvenih faktora i pogreške u mjerenju. Svaki jedinstveni faktor utječe na samo jednu manifestnu varijablu i ne objašnjava korelaciju između manifestnih varijabli. Zajednički faktori utječu na više od jedne manifestne varijable. U ovom radu zanima nas prepoznavanje zajedničkih faktora i povezanih manifestnih varijabli.

EFA pretpostavlja da bilo koja izmjerena varijabla može biti povezana s bilo kojim faktorom. Kod razvoja skale, prvo je potrebno provesti EFA prije nego se pređe na konfirmatornu faktorsku analizu.

## 1.6 Analiza glavnih komponenti

Cilj analize glavnih komponenti (eng. *principal component analysis-PCA*) je izvući maksimalnu varijancu iz skupa podataka za svaku komponentu. Glavne komponente su linearne kombinacije manifestnih varijabli izračunate na način da prva glavna komponenta objašnjava maksimalan moguć dio ukupne varijance manifestnih varijabli te da druga, kao i svaka sljedeća glavna komponenta, objašnjava najveći dio preostale varijance manifestnih varijabli, odnosno najveći dio varijance manifestnih varijabli koji nije objašnjen prethodnim glavnim komponentama.

Glavne komponente su poredane tako da prva izvlači najveću varijancu, a zadnja najmanju. Rješenje je matematički jedinstveno i ako su sve komponente zadržane, točno reproducira opaženu matricu korelacija. Nadalje, s obzirom da su komponente ortogonalne njihovo korištenje u drugim analizama (tipa kao ovisne varijable u multivarijatnoj analizi varijance; eng. *multivariate analysis of variance*, MANOVA) može uveliko olakšati interpretaciju rezultata.

PCA je izabrano rješenje kad smo prije svega zainteresirani za smanjenje velikog broja varijabli na manji broj komponenti. PCA je također korisna kao početni korak u FA gdje otkriva mnogo o maksimalnom broju i prirodi faktora.

## 1.7 Rotacije

Rotacija faktora je postupak interpretacije faktorskih matrica. Za svako rješenje koje ima dva ili više faktora postoji beskonačan broj orijentacija faktora koje će podatke jednako dobro objasniti. Pošto ne postoji jedinstveno rješenje, izabire se jedna od beskonačno mogućnosti. Cilj rotacije je rotirati faktore u višedimenzionalnom prostoru kako bi došli do rješenja s najboljom jednostavnom strukturom. Postoje dvije vrste rotacija: ortogonalna i kosa.

Rezultati ekstrakcije faktora bez pratnje rotacije vjerojatno je teško protumačiti bez obzira koju metodu ekstrakcije koristimo. Nakon ekstrakcije, rotacija se koristi kako bi poboljšali interpretabilnost i znanstvenu korisnost rješenja. Ne koristi se kako bi se poboljšalo matematičko podudaranje (eng. *fit*) između opaženih i reproduciranih korelacijskih matrica zato što sva ortogonalno rotirana rješenja su matematički ekvivalentna jedna drugome kao i rješenju prije rotacije. Kao i što su drukčije metode ekstrakcije sklone davanju sličnih rezultata u dobrom skupu podataka, tako su i drukčije metode rotacija sklone davanju sličnih rješenja ako je uzorak korelacije u podacima prilično jasan. Drugim riječima, stabilno rješenje se javlja neovisno o metodi rotacije koja se koristi.

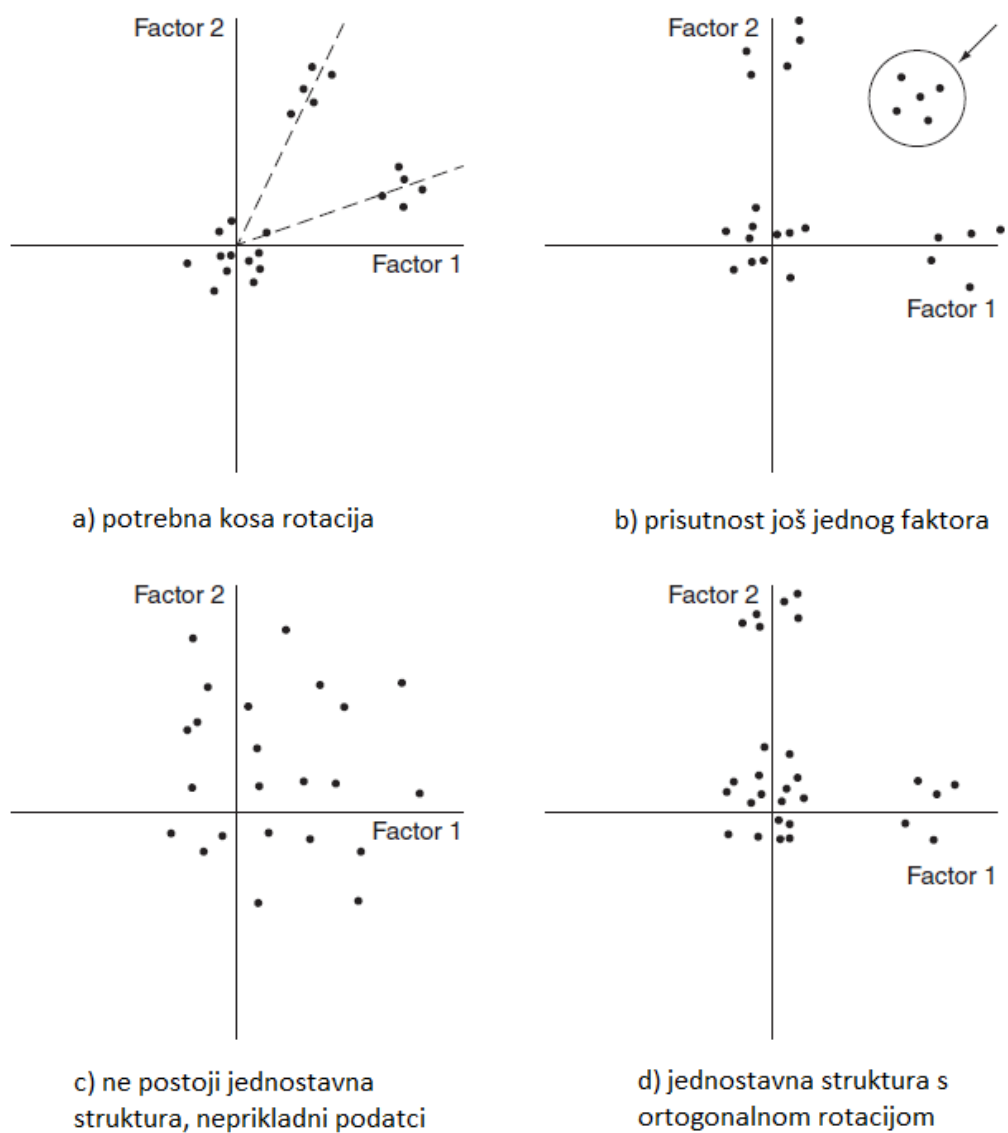
Potrebno je donijeti odluku između ortogonalne i kose rotacije. U ortogonalnoj rotaciji rješenja su nekorelirana. Ortogonalno rješenje nudi lako tumačenje, opis i izvještavanje o rezultatima. Ipak ona naprežu "stvarnost", osim ako nismo uvjereni da su pozadinski procesi gotovo neovisni. Ako vjerujemo da su pozadinski procesi korelirani koristit ćemo kosu rotaciju. U kosoj rotaciji faktori mogu biti korelirani s konceptualnim prednostima, ali s praktičkim nedostacima u tumačenju, opisivanju i izvještavanju o rezultatima.

Na Slici 1.1 možemo vidjeti nekoliko različitih načina raspoređenosti varijabli i različite mogućnosti za rotacije u takvim slučajevima.

### 1.7.1 Ortogonalne rotacije

Ortogonalna rotacija ograničava faktore da budu nekorelirani. Varimax je najkorištenija od svih rotacija. Kao što ekstrakcijske procedure imaju različiti statistički cilj, tako i sve rotacije maksimiziraju ili minimiziraju drukčije statistike. Cilj varimax-a je pojednostaviti faktore maksimizirajući varijancu unutar faktora, preko varijabli. Tumačenje faktora postaje jednostavnije jer je očito koje su varijable korelirane s njim. Varimax također ima tendenciju ponovne razmjene varijance među faktorima tako da oni postaju relativno jednaki po važnosti, varijanca se uzme iz prvog izvučenog faktora i rasporedi se među kasnijima.

Quartimax čini s varijablama ono što varimax čini s faktorima. Varimax radi na stupcima početne matrice, dok quartimax radi na redcima. Quartimax se ne koristi toliko često koliko varimax jer nas više zanimaju pojednostavljeni faktori od pojednostavljenih varijabli.



Slika 1.1: Četiri slučaja raspoređenosti varijabli i mogućnosti rotacija u tim slučajevima

Equamax je spoj varimax-a i quartimax-a i pokušava istovremeno pojednostaviti faktore i varijable. Equamax se ponaša neizvjesno osim u slučaju kada možemo pouzdano odrediti broj faktora.

### 1.7.2 Kose rotacije

Kosa rotacija nudi kontinuirani raspon korelacija između faktora. Faktori ne moraju nužno korelirati kad se koristi kosa rotacija. U stvari, često ne koreliraju i rezultat je jednostavnija ortogonalna rotacija. Kosu rotaciju koristimo ako nakon ortogonalne rotacije i dalje nemamo čistu sliku o faktorima. Komunaliteti ostaju isti prije i poslije rotacije, dok se varijanca koju objašnjava svaki od faktora mijenja.

Nekoliko kosih rotacija su često korištene, kao direktna oblimin rotacija, direktna quartimin rotacija, promax rotacija i Harris-Kaiser orthoblique rotacija. Prednost je u tome što dopušta korelaciju između faktora pa proizvodi rješenja s jednostavnijim strukturama i daje procjene korelacija među faktorima.

## 1.8 Komunaliteti

Komunaliteti su zbroj kvadratnog faktorskog opterećenja za sve faktore. Za danu varijablu, varijanca u toj varijabli je objašnjena svim faktorima, i zove se komunalitet. Komunalitet mjeri postotak varijance u danoj varijabli koji je objašnjen svim faktorima zajedno i može se tumačiti kao pouzdanost pokazatelja u korist faktora koji se postavljaju. Komunalitet nam govori u kojoj mjeri se predmet podudara sa svim ostalim stavkama. Visoki komunaliteti su bolji.

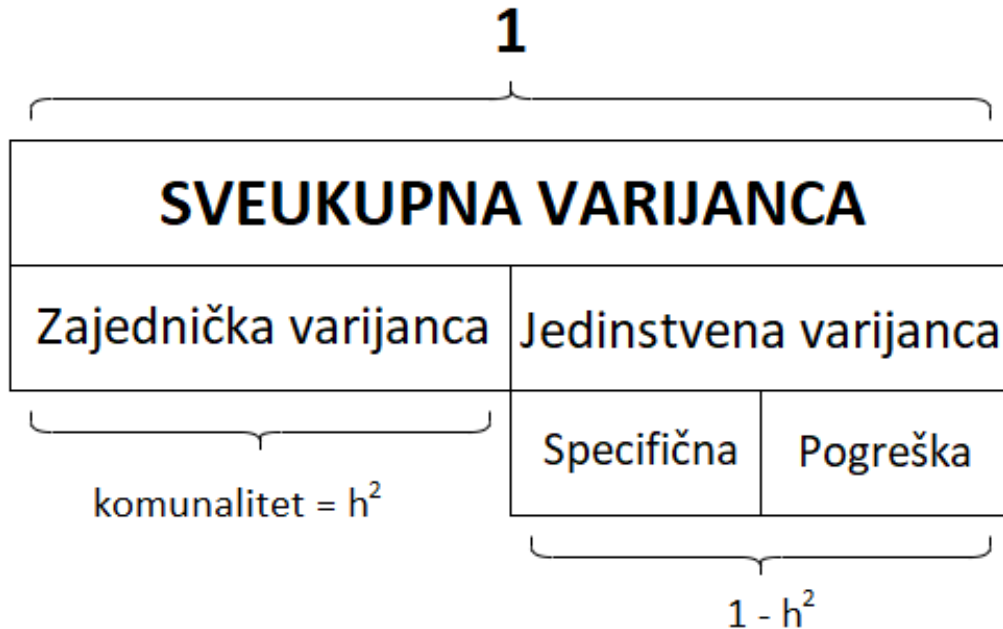
FA se razlikuje od PCA u tome što su vrijednosti komunaliteta zamijenjene s onima na pozitivnoj dijagonali korelacijske matrice prije nego se napravi ekstrakcija faktora. Komunaliteti se koriste kako bi uklonili jedinstvenu varijancu i varijancu pogreške svake opažene varijable. Za rješenje se jedino koristi varijanca koju varijabla dijeli s faktorom.

Kako se rješenje razvija, procjene komunaliteta se prilagođavaju iterativnim postupkom kako bi uklopili reproducirane i opažene korelacijske matrice s najmanjim brojem faktora. Iteracija staje kad su uzastopne procjene komunaliteta vrlo slične.

Krajnje procjene komunaliteta su također višestruke kvadrirane korelacije (eng. *squared multiple correlation, SMC*), ali sada između svake varijable kao zavisne varijable i faktora kao nezavisne varijable. Krajnja vrijednost komunaliteta predstavlja omjer varijance u varijabli koja je predvidljiva iz faktora koji su joj podloga.

Ako su vrijednosti komunaliteta 1 ili veće, to nam ukazuje na probleme s rješenjem, ili imamo premalo podataka, ili početni komunaliteti nisu točni ili je broj dobivenih faktora kriv. Dodavanje ili brisanje faktora može smanjiti vrijednost na ispod 1. Vrlo niski komunaliteti ukazuju da varijable s takvim komunalitetima nisu povezane s drugim varijablama u skupu.





Slika 1.2: Objašnjenje varijance pomoću komunaliteta

## 1.9 Adekvatnost ekstrakcije i broj faktora

Kod odabira broja faktora koje želimo uključiti u model, potrebno je balansirati škrtoš (model sa relativno malo varijabli) i vjerodostojnost (postojanje dovoljno faktora kako bi se adekvatno objasnila korelacija između izmjerenih varijabli). Bolje je uključiti previše nego premalo faktora.

Budući da uključivanje više faktora u rješenje poboljšava podudaranje između opažene i reproducirane matrice korelacija, adekvatnost ekstrakcija je povezana s brojem faktora. Što više faktora izvadimo, to je bolje podudaranje i veći postotak varijance u podacima je "objašnjen" faktorskim rješenjem. Ali što više faktora dobijemo, rješenje postaje manje škrto/štedljivo. Želimo zadržati dovoljno faktora za adekvatno uklapanje, ali ne previše jer time gubimo glavni cilj faktorske analize da smanjimo dimenzionalnost prostora.

Izbor broja faktora je vjerojatno kritičniji od izbora ekstrakcije i rotacije ili vrijednosti komunaliteta. Postoji nekoliko načina za procjenu adekvatnosti ekstrakcije i broja faktora.

Prva brza procjena broja faktora dobiva se iz veličine svojstvenih vrijednosti dobivenih kao dio početnog pokretanja s ekstrakcijom glavnih komponenti. Svojstvene vrijednosti predstavljaju varijancu. Zato što je varijanca, koju standardizirana varijabla doprinosi ek-

strakciji glavne komponente, jednaka 1, komponenta sa svojstvenom vrijednošću manjom od jedan nije važna, iz perspektive varijance, kao opažena varijabla. Ovo se zove Kaiserovo pravilo. Ono izračunava svojstvene vrijednosti za korelacijsku matricu i određuje koliko je od svih tih svojstvenih vrijednosti veće od 1. Taj je broj upravo broj faktora koji se uključuju u model. Nedostatak ove procedure je u tome što je dosta proizvoljna (uključujemo svojstvenu vrijednost koja je 1.01, ali ne i onu koja je 0.99). Napravljena je varijacija Kaiserovog pravila kako bi se smanjila ozbiljnost problema kod izračuna intervala pouzdanosti za svaku svojstvenu vrijednost i zadržava samo one faktore čiji je cijeli interval pouzdanosti veći od 1.0.

Drugi kriterij je grafički test svojstvenih vrijednosti prikazanih prema faktorima, te ga još nazivamo pravilo lakta (eng. *scree plot*). Faktori su silaznim redoslijedom poslagani na os apscisu sa svojstvenim vrijednostima na osi ordinati. Graf se na početku koristi s glavnim komponentama ili analizom faktora, a kasnije kako bi se pronašao broj faktora. Obično se graf negativno smanjuje – svojstvena vrijednost je najveća za prvi faktor i umjerena, ali opada za idućih nekoliko prije nego što dosegne male vrijednosti za posljednjih nekoliko faktora. Tražimo točku u kojoj linija povučena kroz točke mijenja nagib. S obzirom da je ova metoda subjektivna (ne postoji jasna objektivna definicija onoga što predstavlja značajan pad), ne preporuča se za korištenje.

Kaiser-Meyer-Olkinov test je mjera koliko su naši podatci prilagođeni faktorskoj analizi [4]. Test mjeri adekvatnost uzorkovanja za svaku varijablu u modelu i za model u cijelosti. Statistika je mjera udjela varijance među varijablama koja može biti zajednička varijanca. Kaiser-Meyer-Olkinova mjera adekvatnosti uzorkovanja je omjer zbroja kvadriranih korelacija i zbroja kvadriranih korelacija plus zbroja kvadriranih parcijalnih korelacija. Vrijednost se približava jedinici ako su parcijalne korelacije niske. Vrijednosti od 0.6 i više su potrebne za dobru faktorsku analizu.

Formula glasi:

$$KMO = \frac{\sum_{k \neq j} r_{jk}^2}{\sum_{j \neq k} r_{jk}^2 + \sum_{j \neq k} p_{jk}^2},$$

gdje je  $r_{jk}^2$  kvadratni koeficijent korelacije između j-te i k-te varijable, a  $p_{jk}^2$  kvadratni koeficijent parcijalne korelacije između j-te i k-te varijable.

Jednom kad smo odredili broj faktora važno je pogledati rotiranu početnu matricu kako bi odredili broj varijabli koje se spajaju na svaki faktor. Ako se samo jedna varijabla visoko spaja na faktor, faktor je loše definiran. Ako se dvije varijable visoko spajaju na faktor onda njegova pouzdanost ovisi o uzorku korelacija između tih dviju varijabli i korelacija s ostalim varijablama iz matrice korelacija. Ako su dvije varijable jako korelirane jedna s drugom ( $r > 0.7$ ), a relativno nekorelirane s drugim varijablama, faktor može biti pouzdan. Interpretacija faktora opisanih samo s jednom ili dvije varijable je opasna, čak i u većini eksplorativnih faktorskih analiza.

## 1.10 Važnost i unutarnja konzistentnost faktora

Važnost faktora (ili skupa faktora) je procijenjena udjelom varijance ili kovarijance koja je objašnjena faktorom nakon rotacije. Udio varijance koji se može pripisati pojedinom faktoru razlikuje se prije i poslije rotacije jer rotacija teži preraspodjeli varijance među faktorima. Jednostavnost utvrđivanja udjela varijance za faktore ovisi o tome je li rotacija bila ortogonalna ili kosa.

Nakon ortogonalne rotacije važnost faktora je povezana s veličinom sume kvadratnih opterećenja (eng. *sum of squared loadings, SSLs*). SSLs se pretvaraju u proporciju varijance za faktor dijeljenjem s brojem varijabli. SSLs se pretvaraju u proporciju kovarijanci za faktor dijeljenjem njegovih SSL s zbrojem SSLs ili ekvivalentno zbrojem komunaliteta.

Udio varijance objašnjen faktorom je količina varijance u početnim varijablama koja se kondenzirala u faktor. Udio varijance je varijanca faktora srodna varijanci u varijabli. Udio kovarijance objašnjene faktorom ukazuje na relativnu važnost faktora u ukupnoj kovarijanci koja je objašnjena svim faktorima. Udio kovarijance je varijanca faktora srodna varijanci u rješenju. Varijancu u rješenju vjerojatno objašnjava samo djelić varijance iz izvornih varijabli.

## 1.11 Faktorski rezultati

Faktorski rezultati (eng. *factor score*) su procjene koje bi subjekt dobio na svakom od faktora da su oni bili mjereni direktno. Svaki faktor je linearna kombinacija analiziranih (manifestnih) varijabli.

Jednom kada je broj originalnih varijabli smanjen na određeni broj faktora, moguće je izračunati faktorski rezultat. Oni zapravo predstavljaju rezultate za svaku jedinicu promatranja po pojedinim faktorima. Visoka vrijednost varijable u pogledu faktorskog opterećenja će rezultirati i visokim faktorskim rezultatom.

# Poglavlje 2

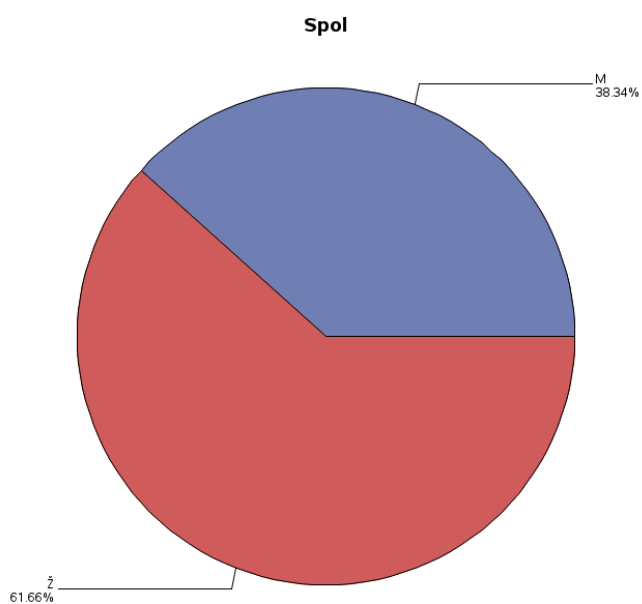
## Primjer

### 2.1 Analiza anketa

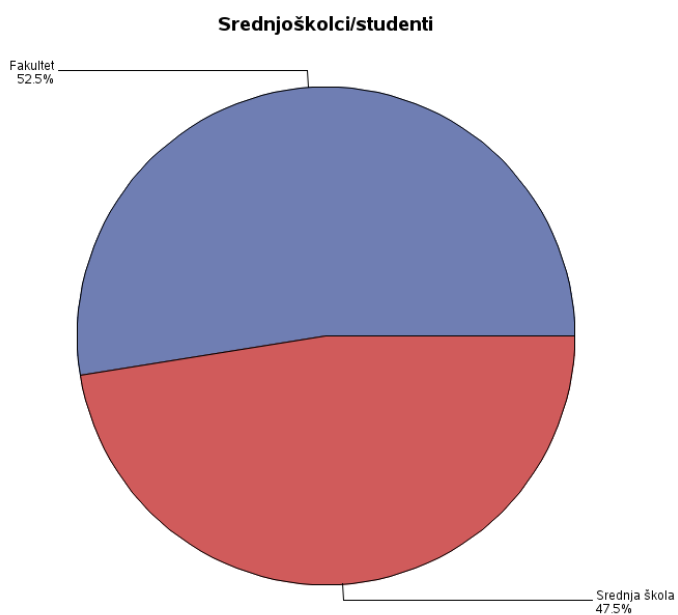
Anketa (Prilog 3.1) koja je korištena u ovom diplomskom radu napravljena je u online obliku i distribuirana je putem društvenih mreža. Bila je dostupna u razdoblju od četiri mjeseca i u tom vremenskom periodu preko 700 ispitanika diljem Hrvatske ju je ispunilo. Nakon uklanjanja svih neispravnih ispunjavanja, preostalo je 699 ispitanika na kojima je odrađena daljnja analiza. Cilj je bio postići ravnotežu između broja studenata i broja srednjoškolaca, kako bi se dobili statistički značajni rezultati. S obzirom da je anketa online i anonimna, nije bilo razloga za sumnju u istinitost odgovora.

U prvih 12 pitanja tražili su se osobni podatci o ispitaniku, kao što su spol, obrazovanje, zaposlenost, mjesto rođenja i slično. Poslije toga slijedila su pitanja podijeljena u četiri odlomka: Glazba, Filmovi, Strahovi te Hobiji i interesi. U tim odlomcima, odgovori su rangirani brojevima od 1 do 7, pri čemu 1 označava najslabiji interes, a 7 jako izraženi interes, tj. kod strahova 1 označava nemanje straha, a 7 fobiju od navedenog. Posljednji odlomak je bio Zdravstvene navike, gdje se ispitivala učestalost konzumacije alkohola, duhana te droga među mladim ljudima.

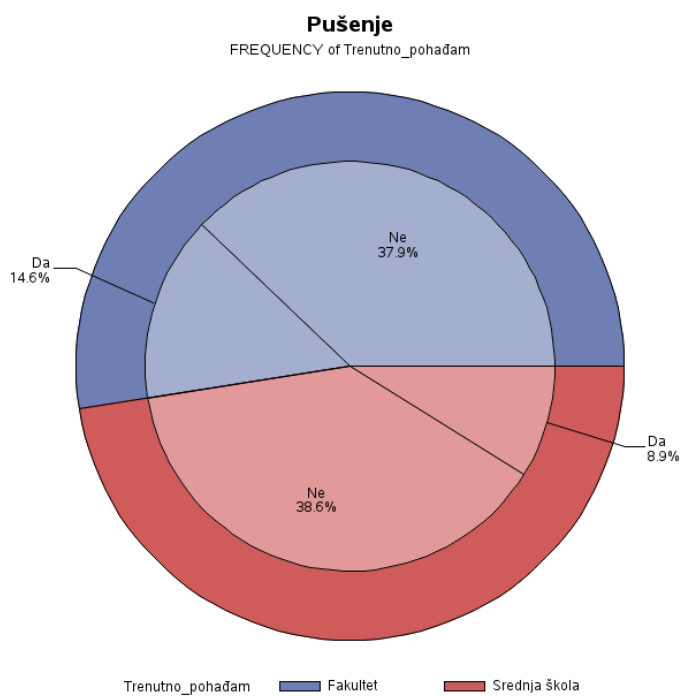
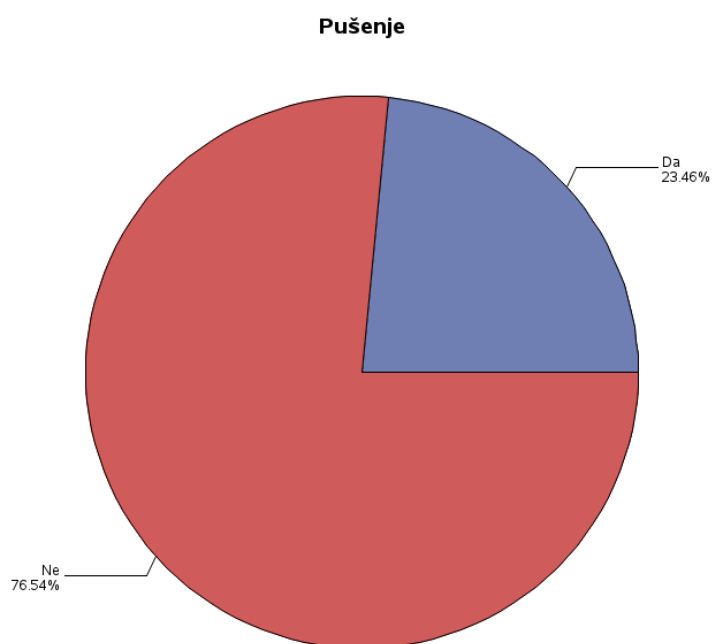
Pitanja iz ankete, tj. varijable su podijeljene na kategorijske i diskretne. Za kategorijske varijable napravljeni su strukturni krugovi (eng. *pie chart*) (Slike 2.1, 2.2, 2.3, 2.4, 2.5) i frekvencijske tablice 2.2, 2.3, 2.4, 2.5 i 2.6, 2.7, 2.8, 2.9, 2.10.



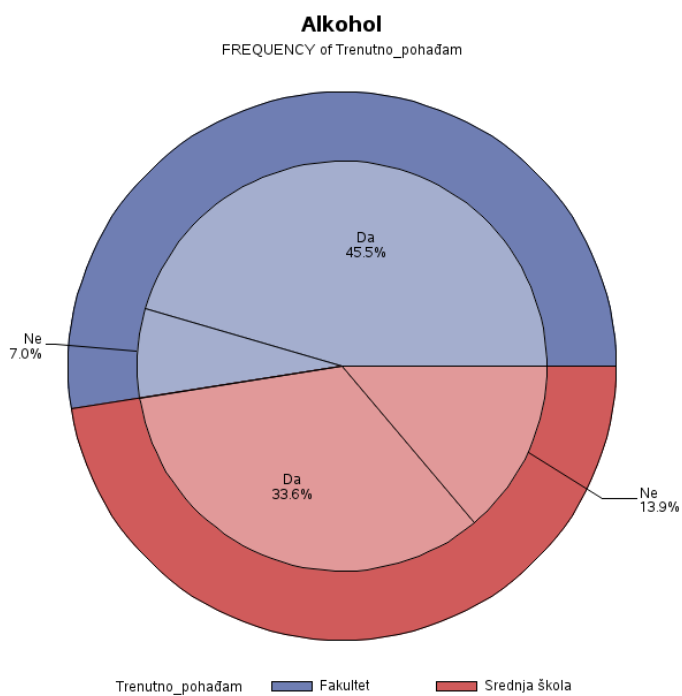
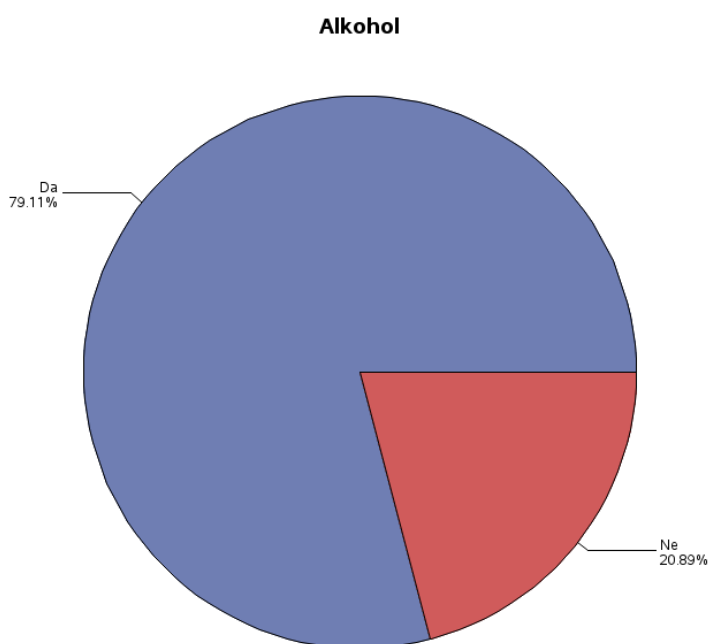
Slika 2.1: Grafički prikaz relativnih frekvencija ženskih i muških ispitanika



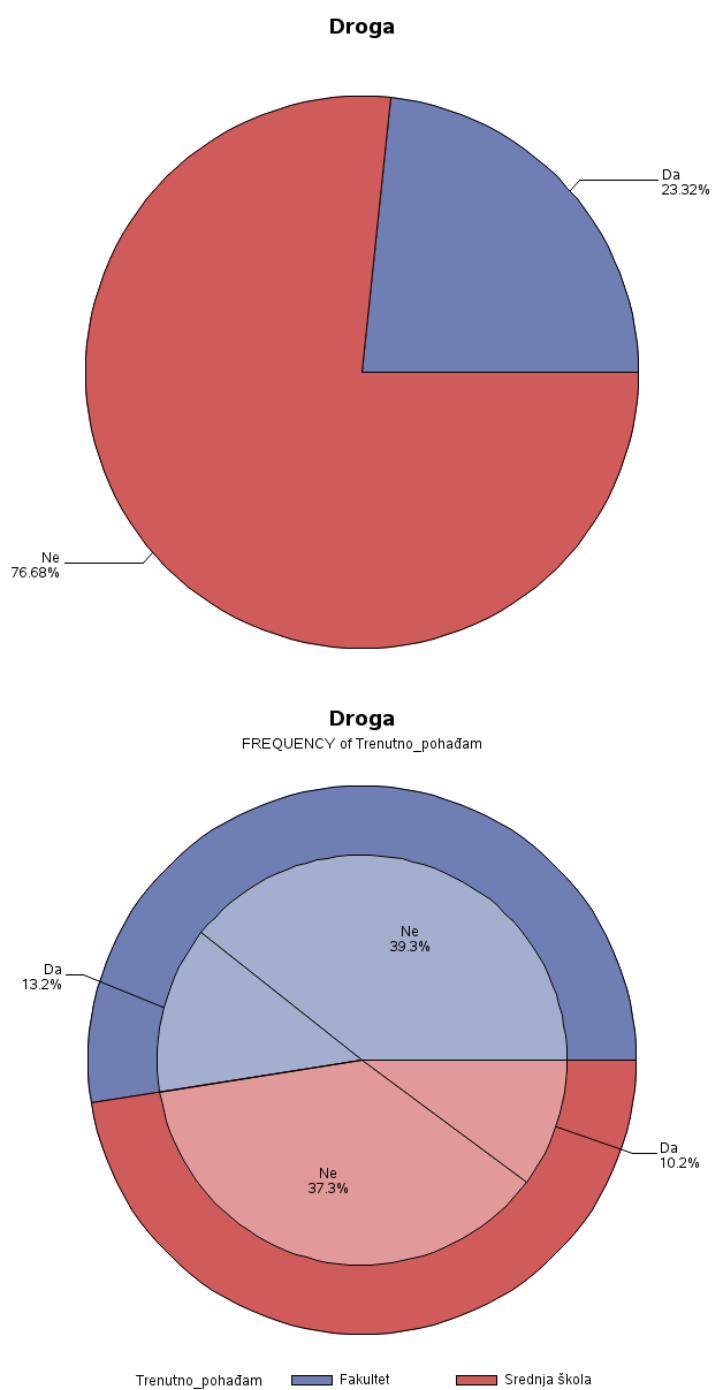
Slika 2.2: Grafički prikaz relativnih frekvencija studenata i srednjoškolaca



Slika 2.3: Grafički prikaz relativnih frekvencija pušača i nepušača u sveukupnom broju ispitanika (gornja slika) i pušača i nepušača s obzirom na podjelu ispitanika na srednjoškolce i studente (donja slika)



Slika 2.4: Grafički prikaz relativnih frekvencija ispitanika koji su probali alkohol i onih koji nisu u sveukupnom broju (gornja slika) i s obzirom na podjelu ispitanika na srednjoškolce i studente (donja slika)



Slika 2.5: Grafički prikaz relativnih frekvencija ispitanika koji su probali drogu i onih koji nisu u sveukupnom broju (gornja slika) i s obzirom na podjelu ispitanika na srednjoškolce i studente (donja slika)



Više od polovice ispitanika je ženskog spola i više je studenata nego srednjoškolaca. Što se tiče zdravstvenih navika ispitanika, primjećujemo da 76.% ispitanika ne puši cigarete, od toga ih je 37.9% studenata, a 38.6% srednjoškolaca. 79.11% ispitanika je probalo alkohol, a njih 64.23% ga piju samo na društvenim događanjima. Drogu je konzumiralo 23.32% ispitanika, tj. 163 osobe. Vidimo da je sličan postotak osoba koje su probale drogu kad uspoređujemo srednjoškolce (10.2%) i studente (13.%). Od njih 163, 117 ih je samo probalo, a 4 tvrdi da drogu konzumiraju svakodnevno. Iz Slike 2.8 frekvencijske tablica za varijablu Skupine.droga, zaključujemo da najviše njih ulazi pod kategoriju halucinogenih droga. U toj tablici, podatci koje nedostaju su osobe koje nisu nikad probale drogu jer nije bilo obavezno odgovoriti na to pitanje u anketi. Sa Slike 2.2 i 2.3 također vidimo raspodjelu srednjoškolaca i studenata po pitanjima jesu li ikad probali alkohol i drogu.

Za diskretne varijable koje ćemo koristiti u faktorskoj analizi podataka dana je Tablica 2.1 u kojoj imamo podatke o minimalnoj vrijednosti, medijanu, maksimalno vrijednosti, očekivanju i standardnoj devijaciji.

Tablica 2.1: Deskriptivna statistika (N, Minimum, Medijan, Maksimum, Aritmetička sredina, Standardna devijacija) za diskretne varijable koje ulaze u faktorsku analizu

Varijabla	N	Min	Medijan	Max	Arit sred	Std Dev
Dob	699	15.00	20.00	35.00	20.64	3.11
<b>MUZIKA</b>						
Volite li slušati glazbu	699	1.00	7.00	7.00	6.50	0.96
Pop	699	1.00	5.00	7.00	4.97	1.67
Rock	699	1.00	5.00	7.00	4.38	1.98
Metal	699	1.00	1.00	7.00	2.02	1.62
Hip hop, Rap	699	1.00	4.00	7.00	3.87	1.91
Jazz	699	1.00	3.00	7.00	2.92	1.73
Narodnjaci	699	1.00	3.00	7.00	3.53	2.20
Techno	699	1.00	3.00	7.00	3.48	1.98
Klasična glazba	699	1.00	3.00	7.00	3.16	1.80
Trash	699	1.00	4.00	7.00	4.01	2.13
House	699	1.00	3.00	7.00	3.07	1.91
<b>FILMOVI</b>						
Volite li gledati filmove	699	1.00	7.00	7.00	6.28	1.14
Horor	699	1.00	3.00	7.00	3.23	2.18
Triler	699	1.00	5.00	7.00	4.94	1.81

Komedija	699	1.00	6.00	7.00	5.71	1.41
Romantični	699	1.00	5.00	7.00	4.60	1.99
SCI-FI	699	1.00	5.00	7.00	4.33	2.09
Dokumentarni	699	1.00	5.00	7.00	4.49	1.80
Akcijski	699	1.00	5.00	7.00	4.92	1.73
Animirani	699	1.00	5.00	7.00	4.76	1.86
Western	699	1.00	2.00	7.00	2.61	1.68
<b>FOBIJE</b>						
Let avionom	699	1.00	1.00	7.00	2.26	1.68
Visina	699	1.00	3.00	7.00	3.21	1.93
Vremenske nepogode	699	1.00	1.00	7.00	2.03	1.51
Mrak	699	1.00	2.00	7.00	2.44	1.73
Pauci	699	1.00	3.00	7.00	3.22	2.15
Dizalo	699	1.00	1.00	7.00	1.89	1.57
Zmije	699	1.00	4.00	7.00	3.91	2.18
Mali prostori	699	1.00	1.00	7.00	2.42	1.86
Psi	699	1.00	1.00	7.00	1.65	1.22
Javni govor	699	1.00	3.00	7.00	3.67	2.00
Glodavci	699	1.00	2.00	7.00	3.02	2.04
Zubar	699	1.00	1.00	7.00	2.32	1.86
Doktorske igle	699	1.00	2.00	7.00	2.44	1.89
Krv	699	1.00	1.00	7.00	1.86	1.55
Bacili	699	1.00	1.00	7.00	2.20	1.64
<b>HOBLJI</b>						
Čitanje	699	1.00	4.00	7.00	3.99	1.83
Strani jezici	699	1.00	4.00	7.00	4.03	1.80
Likovna umjetnost	699	1.00	2.00	7.00	2.54	1.81
Gluma	699	1.00	1.00	7.00	2.01	1.64
Ples	699	1.00	2.00	7.00	2.78	2.02
Sviranje	699	1.00	1.00	7.00	2.32	2.00
Pjevanje	699	1.00	3.00	7.00	3.24	2.17
Koncerti	699	1.00	4.00	7.00	3.71	1.98
Kazalište	699	1.00	3.00	7.00	2.93	1.80
Video igrice	699	1.00	3.00	7.00	3.22	2.23
Automobili	699	1.00	2.00	7.00	2.76	2.08

Shopping	699	1.00	4.00	7.00	3.91	1.92
Profesionalno bavljenje sportom	699	1.00	1.00	7.00	2.18	2.03
Amatersko bavljenje sportom	699	1.00	4.00	7.00	4.10	2.26
<b>INTERESI</b>						
Povijest	699	1.00	4.00	7.00	3.65	2.06
Zemljopis	699	1.00	4.00	7.00	3.68	1.99
Psihologija	699	1.00	5.00	7.00	4.53	2.00
Politika	699	1.00	3.00	7.00	3.12	2.05
Matematika	699	1.00	4.00	7.00	4.14	2.28
Fizika	699	1.00	3.00	7.00	3.43	2.13
Informatika	699	1.00	4.00	7.00	3.89	2.16
Biologija	699	1.00	4.00	7.00	3.88	2.14
Kemija	699	1.00	3.00	7.00	3.15	2.10
Medicina	699	1.00	3.00	7.00	3.61	2.27
Pravo	699	1.00	2.00	7.00	2.50	1.91
Ekonomija	699	1.00	2.00	7.00	2.92	2.06
Elektrotehnika	699	1.00	2.00	7.00	2.53	1.89
Robotika	699	1.00	2.00	7.00	2.56	1.95
Religija	699	1.00	3.00	7.00	3.17	2.21

Tablica 2.1 nam govori da je srednja dob ispitanika 20 godina. Vidimo da većina ispitanika voli slušati glazbu i gledati filmove. Od glazbe su najpopularniji pop i rock, a najmanje ispitanika voli slušati metal. Među filmovima su najgledanije komedije, a najmanje gledani su westerni. Kod pitanja vezanih za fobije, ispitanici se najviše boje zmija, a najmanje leta avionom, vremenskih nepogoda, dizala, malih prostora, pasa, zubara, krvi i bacila. Hobiji kojima se ispitanici najviše bave su čitanje, odlasci na koncerte, shopping i amatersko bavljenje sportom, dok su gluma, sviranje i profesionalno bavljenje sportom manje popularni. Za psihologiju se interesira najveći broj ispitanika, potom za povijest, zemljopis, matematiku, informatiku i biologiju. Interesi s najmanjim brojem zainteresiranih ispitanika su pravo, ekonomija, elektrotehnika i robotika.

Tablica 2.2: Frekvencijska tablica za varijablu **Spol**

Spol	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
M	268	38.34	268	38.34
Ž	431	61.66	699	100.00

Tablica 2.3: Frekvencijska tablica za varijablu **Trenutno pohađam**

Trenutno pohađam	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Fakultet	367	52.50	367	52.50
Srednja škola	332	47.50	699	100.00

Tablica 2.4: Frekvencijska tablica za varijablu **Pušenje**

Pušenje	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Da	164	23.46	164	23.46
Ne	535	76.54	699	100.00

Tablica 2.5: Frekvencijska tablica za varijablu **Koliko dnevno**

Koliko dnevno	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
1-5	70	10.01	70	10.01
11-20	33	4.72	103	14.74
6-10	52	7.44	155	22.17
Ne pušim cigarete	535	76.54	690	98.71
Više od kutije dnevno	9	1.29	699	100.00

Tablica 2.6: Frekvencijska tablica za varijablu **Alkohol**

Alkohol	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Da	553	79.11	553	79.11
Ne	146	20.89	699	100.00

Tablica 2.7: Frekvencijska tablica za varijablu **Koliko često**

Koliko često	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Na društvenim događanjima	449	64.23	449	64.23
Nikad	135	19.31	584	83.55
Samo vikendom	95	13.59	679	97.14
Skoro svakodnevno	20	2.86	699	100.00

Tablica 2.8: Frekvencijska tablica za varijablu **Droga**

Droga	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Da	163	23.32	163	23.32
Ne	536	76.68	699	100.00

Tablica 2.9: Frekvencijska tablica za varijablu **Koliko učestalo**

Koliko učestalo	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Nikad	536	76.68	536	76.68
Par puta tjedno	5	0.72	541	77.40
Probao/la	117	16.74	658	94.13
Skoro svakodnevno	4	0.57	662	94.71
U prosjeku jednom tjedno	37	5.29	699	100.00

Tablica 2.10: Frekvencijska tablica za varijablu **Skupine droge**

Skupine droge	Frekvencija	Postotak	Kumulativna frekvencija	Kumulativni postotak
Halucinogene droge (marihuana, hašiš, LSD...)	156	95.71	156	95.71
Opojne droge (morfij, kodein, heroin, medatod...)	1	0.61	157	96.32
Stimulativne droge (kokain i crack, MDMA-ecstasy...)	6	3.68	163	100.00
Nedostaje frekvencija: 536				

## 2.2 Faktorska analiza podataka

Faktorsku analizu provodimo na diskretnim varijablama koje su navedene u Tablici 2.1. U anketi postoji 66 diskretnih varijabli na kojima ćemo raditi faktorsku analizu.

Prvi korak u faktorskoj analizi je ekstrakcija faktora. To je moguće na dva načina: metodom glavnih komponenti i metodom maksimalne vjerodostojnosti. Ako na početku ne definiramo koju metodu želimo koristiti, SAS kao zadanu metodu ima metodu glavnih komponenti te će se ekstrakcija izvršiti pomoću nje.

Kaiserova mjera adekvatnosti je sažetak [3], za svaku varijablu posebno i sve varijable skupa, koliko su manje parcijalne korelacije od izvornih korelacija. Na podacima koji su dobiveni iz anketa Kaiserova cjelokupna mjera adekvatnosti iznosi  $MSA=0.76514$ , što je približno 0.8 pa možemo zaključiti da su podatci dovoljno dobri za faktorsku analizu. U Tablica 2.11 vidimo da sve varijable imaju dobru mjeru adekvatnosti (MSA), svi su preko 0.6. U istoj toj tablici nalaze se višestruke kvadratne korelacije (SMC) koje su prikazane kao prethodne procjene komunaliteta. Dijagonala originalne matrice opaženih korelacija zamjenjuje se sa SMC vrijednostima, a s obzirom da su one sve manje od 1 nastaje reducirana korelacijska matrica.

Tablica 2.11: Mjera adekvatnosti varijabli i višestruke kvadratne korelacije

Varijable	Mjera adekvatnosti	Višestruke kvadratne korelacije
Dob	0.62	0.43
Volite li slušati glazbu	0.72	0.20
Pop	0.71	0.23
Rock	0.77	0.54
Metal	0.71	0.39
Hip hop, Rap	0.77	0.29
Jazz	0.81	0.33
Narodnjaci	0.82	0.36
Techno	0.69	0.37
Klasična glazba	0.83	0.37
Trash	0.66	0.21
House	0.70	0.30
Volite li gledati filmove	0.74	0.24
Horor	0.68	0.21
Triler	0.65	0.26
Komedija	0.69	0.16
Romantični	0.79	0.33

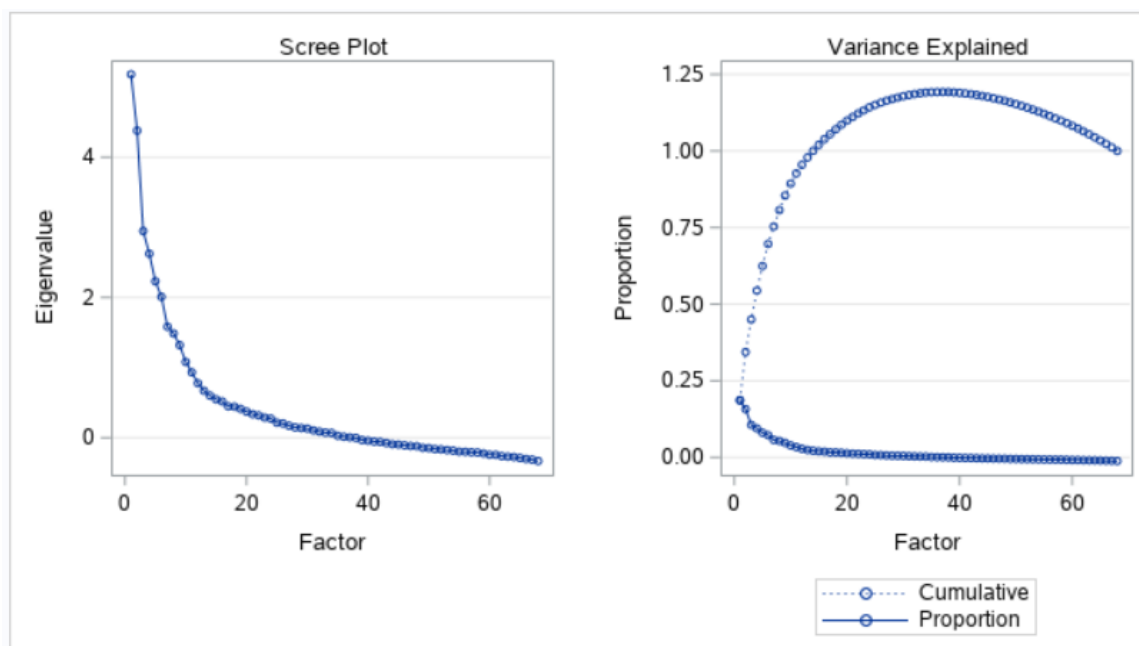
SCI-FI	0.80	0.27
Dokumentarni	0.84	0.30
Akcijski	0.77	0.28
Animirani	0.73	0.16
Western	0.82	0.30
Let avionom	0.69	0.34
Visina	0.73	0.29
Vremenske nepogode	0.86	0.30
Mrak	0.81	0.31
Pauci	0.79	0.22
Dizalo	0.75	0.34
Zmije	0.83	0.33
Mali prostori	0.81	0.29
Psi	0.70	0.14
Javni govor	0.80	0.24
Glodavci	0.86	0.40
Zubar	0.74	0.16
Doktorske igle	0.73	0.30
Krv	0.76	0.29
Bacili	0.83	0.19
Čitanje	0.80	0.34
Strani jezici	0.74	0.27
Likovna umjetnost	0.76	0.30
Gluma	0.77	0.38
Ples	0.78	0.38
Sviranje	0.77	0.41
Pjevanje	0.75	0.41
Koncerti	0.68	0.35
Kazalište	0.74	0.43
Video igrice	0.79	0.38
Automobili	0.77	0.29
Shopping	0.80	0.33
Profesionalno bavljenje sportom	0.75	0.17
Amatersko bavljenje sportom	0.75	0.13
Povijest	0.72	0.49

Zemljopis	0.74	0.36
Psihologija	0.80	0.30
Politika	0.73	0.45
Matematika	0.72	0.51
Fizika	0.83	0.64
Informatika	0.84	0.50
Biologija	0.77	0.64
Kemija	0.78	0.60
Medicina	0.78	0.57
Pravo	0.78	0.35
Ekonomija	0.72	0.31
Elektrotehnika	0.80	0.63
Robotika	0.794	0.61
Religija	0.75	0.19

Tablica 2.12: Svojevne vrijednosti faktora

Faktor	Svojevna vrijednost	Razlika	Proporcija	Kumulativno
1	5.18	0.80	0.19	0.19
2	4.38	1.43	0.16	0.34
3	2.95	0.32	0.11	0.45
4	2.62	0.39	0.09	0.54
5	2.23	0.22	0.08	0.62
6	2.01	0.43	0.07	0.70
7	1.58	0.10	0.06	0.75
8	1.49	0.17	0.05	0.81
9	1.32	0.24	0.05	0.85
10	1.08	0.15	0.04	0.89
11	0.93	0.16	0.03	0.93
...				



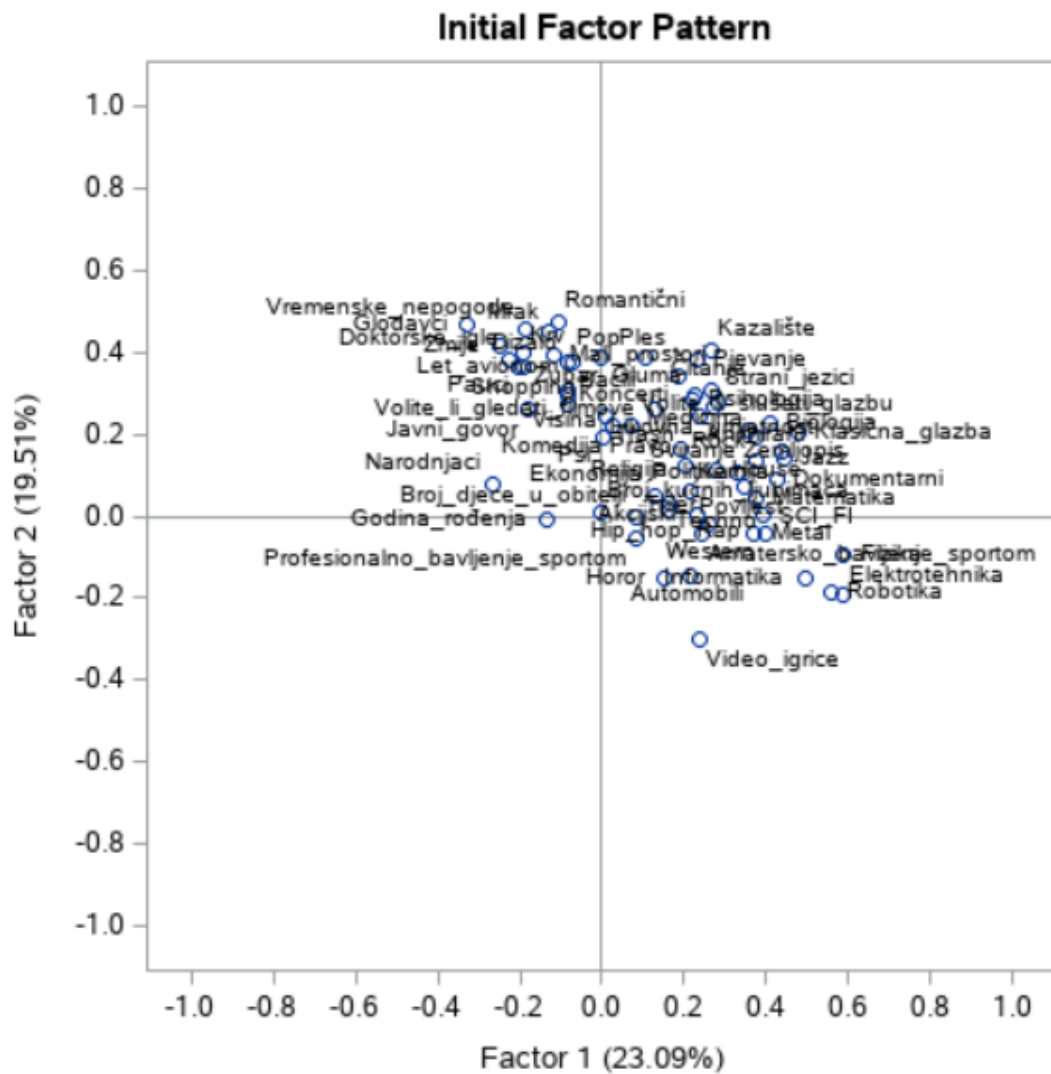


Slika 2.6: Grafički prikaz svojstvenih vrijednosti

Na Slici 2.6, lijevi graf, vidimo grafički prikaz svojstvenih vrijednosti reducirane matrice korelacija. Po pravilu lakta zaključujemo da se otprilike na desetom faktoru mijenja nagib. Na desnom grafu su prikazani kumulativni proporcionalni udjeli zajedničke varijance objašnjeni faktorima. Donja krivulja se počne zaglađivati oko desetog faktora.

Kad u kodu nemamo konkretno zadani broj faktora, program će sam odredi broj faktora za ekstrakciju pomoću neke od navedenih metoda u teoriji. Ako dodamo funkciju NFACTORS, onda sami možemo zadati koliki broj faktora želimo. U outputu koji smo dobili u SAS-u vidimo da nam je dano 10 faktora kao rješenje. Unatoč tome proučeni su rezultati kad je stavljeno da postoji 8 faktora. Prolazeći kroz varijable koje su dodijeljene pojedinom faktoru, možemo zaključiti da gledanje 8 faktora ima smisla. Zbog toga koristimo funkciju NFACTORS i kao zadani broj faktora stavljamo 8. Daljnja analiza će se provoditi na tih osam faktora.

S obzirom da 66 diskretnih varijabli ulazi u faktorsku analizu, sa grafičkog prikaza nerotiranih varijabli, Slika 2.7 i Slika 2.8, ne možemo donijeti zaključak oko broja faktora i grupiranja varijabli.



Slika 2.7: Grafički prikaz varijabli za prva dva faktora



Idući korak je napraviti rotaciju. Prva napravljena rotacija je varimax. Kako bi dobili rotirani uzorak faktora početna matrica faktora je pomnožena (eng. *postmultiplied*) s matricom ortogonalne transformacije. Tablica 2.13 nam pokazuje kako izgleda matrica ortogonalne rotacije.

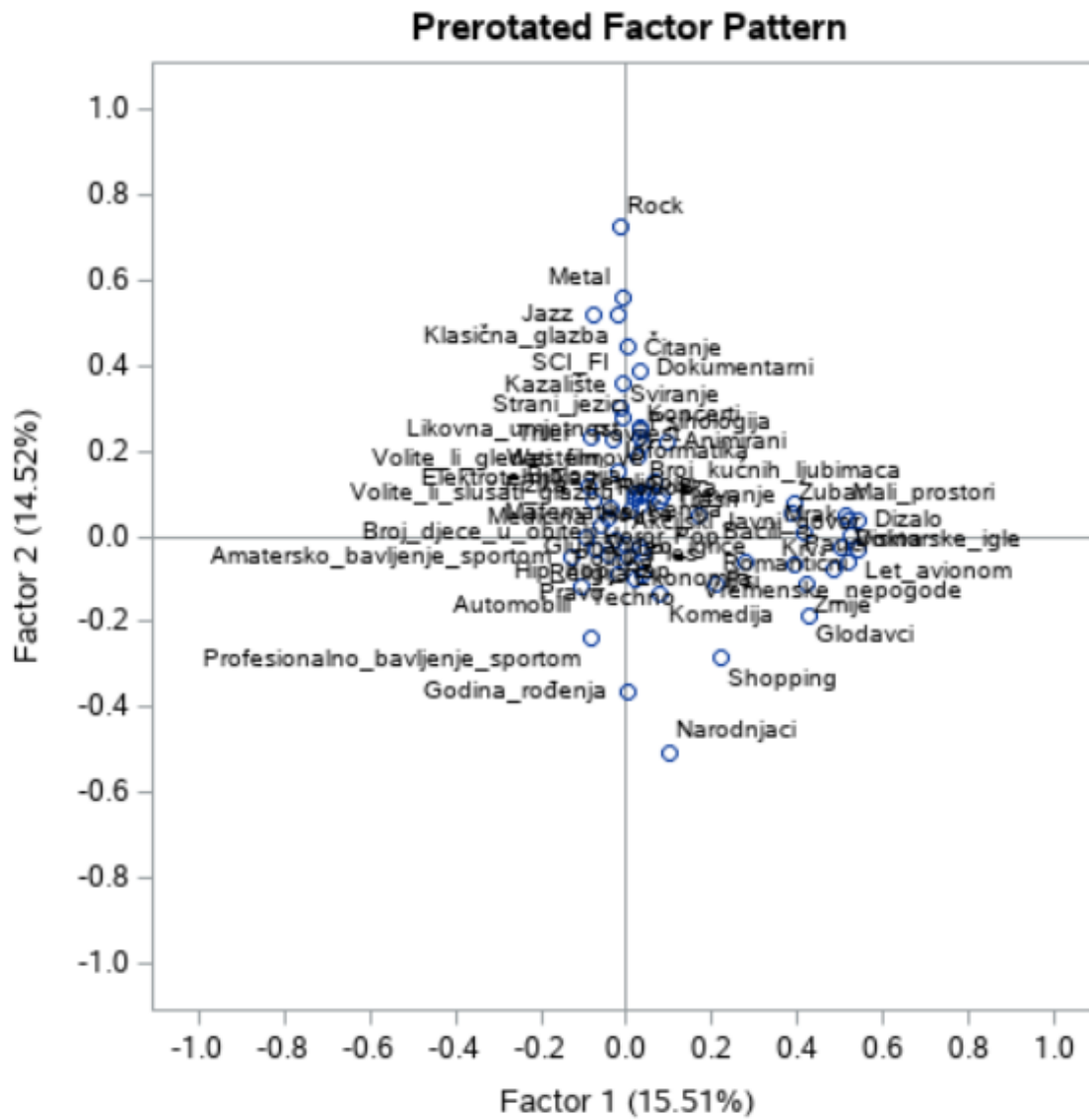
Tablica 2.13: Matrica ortogonalne transformacije

	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5	Faktor6	Faktor7	Faktor8
Faktor1	-0.25	0.53	0.55	0.33	0.29	0.26	0.30	-0.09
Faktor2	0.69	0.15	-0.16	0.11	0.050	0.44	0.20	0.48
Faktor3	0.22	-0.59	0.31	0.34	0.59	-0.14	-0.16	0.02
Faktor4	0.43	-0.02	0.60	-0.37	-0.29	-0.38	0.29	-0.06
Faktor5	0.34	0.38	-0.01	0.53	-0.26	-0.31	-0.50	-0.22
Faktor6	0.18	0.35	0.04	-0.58	0.50	0.10	-0.48	-0.15
Faktor7	0.09	0.24	-0.44	0.03	0.42	-0.57	0.48	-0.10
Faktor8	0.27	-0.17	-0.15	0.03	-0.04	0.38	0.22	-0.83

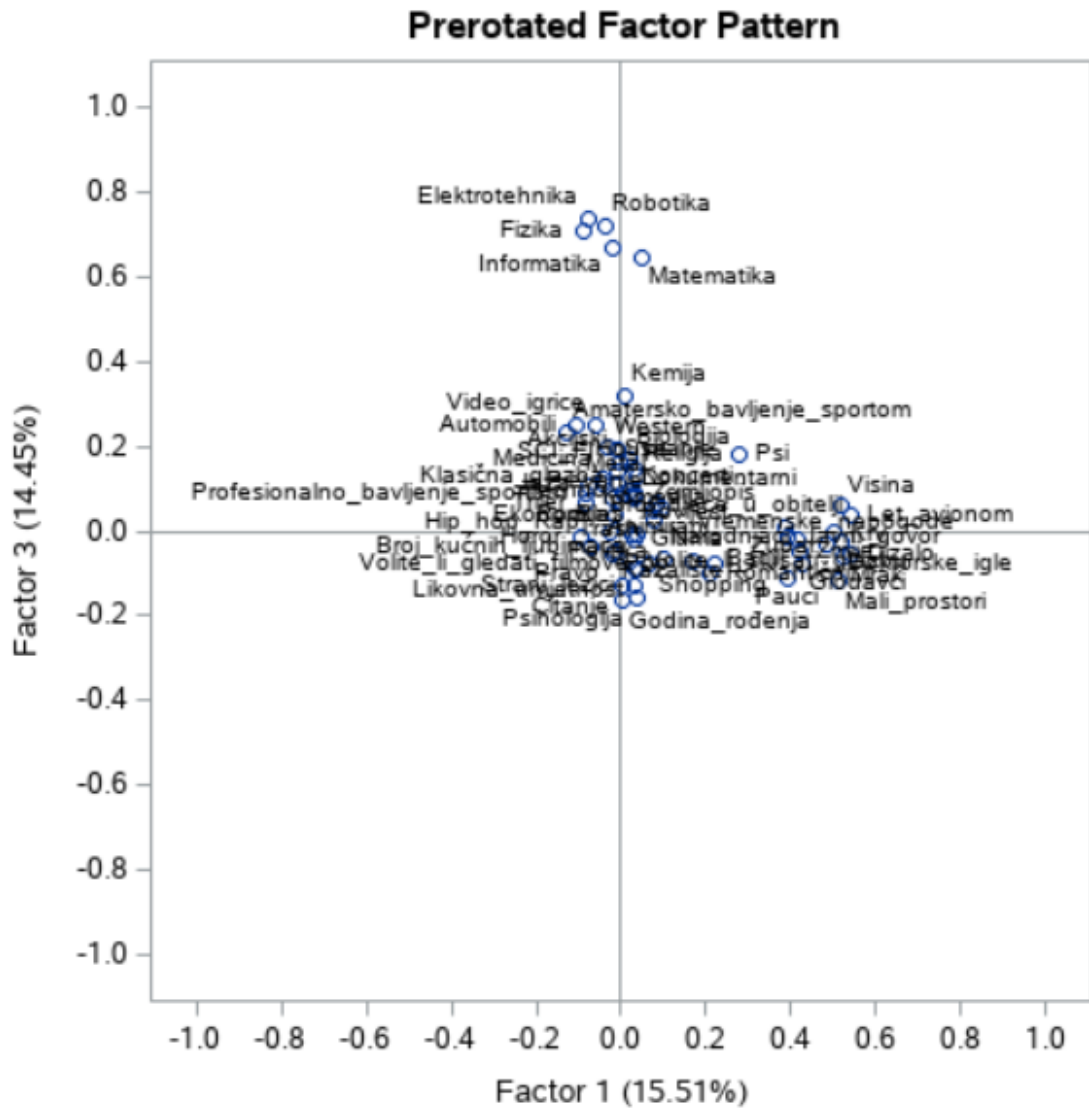
Varijanca koju objašnjavaju faktori ravnomjernije je raspoređena nakon ortogonalne rotacije, nego varijanca koja je bila objašnjena faktorima prije rotacije. To možemo vidjeti u Tablica 2.14. Suma varijance koja je objašnjena prije i poslije rotacije ostaje nepromijenjena.

Tablica 2.14: Varijanca objašnjena faktorima prije i poslije rotacije

Faktori	Varijanca prije rotacije	Varijanca poslije rotacije
Faktor1	5.18	3.48
Faktor2	4.38	3.26
Faktor3	2.95	3.24
Faktor4	2.62	2.65
Faktor5	2.23	2.59
Faktor6	2.01	2.59
Faktor7	1.65	2.26
Faktor8	1.49	2.22

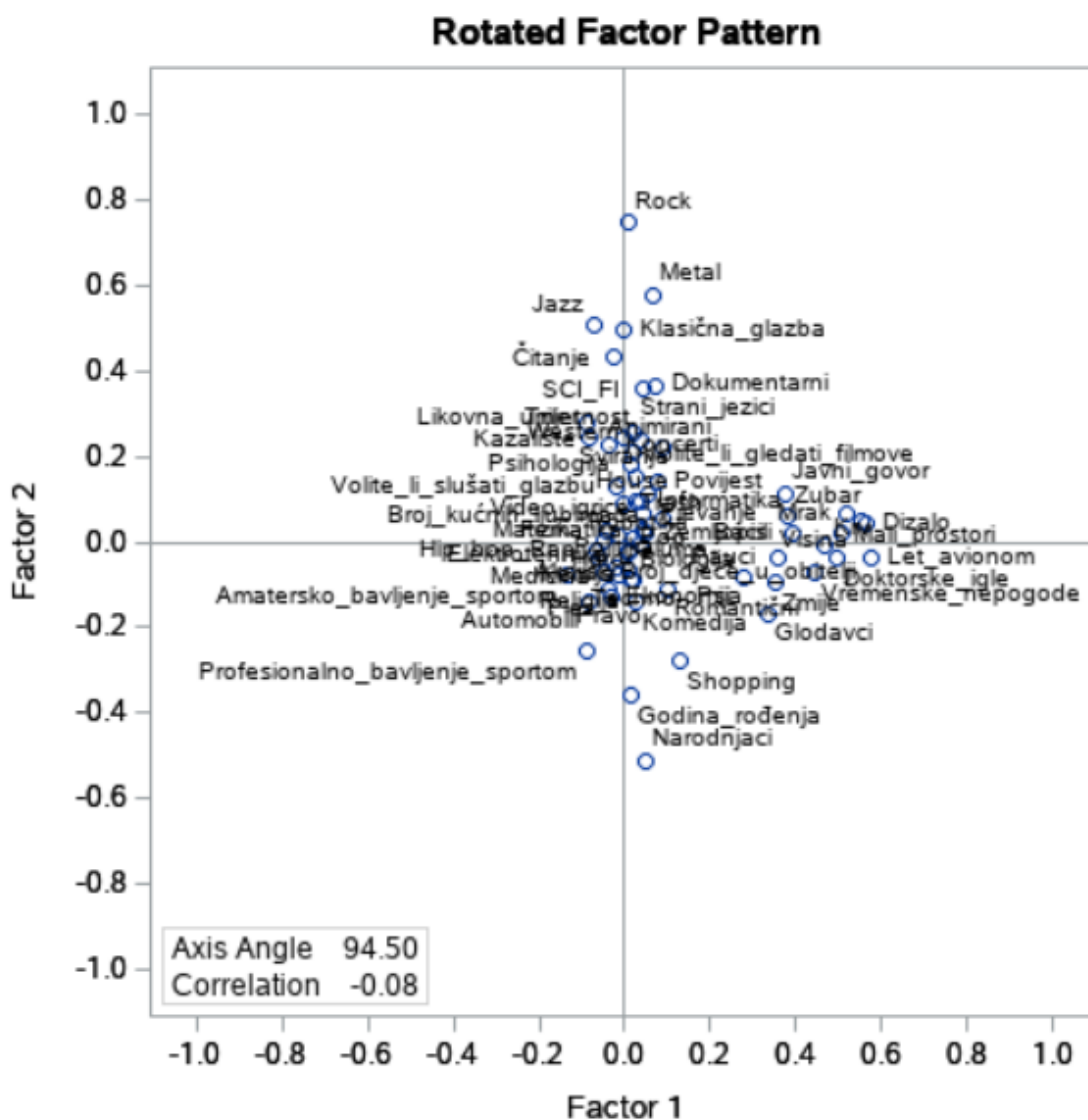


Slika 2.9: Grafički prikaz varijabli prva dva faktora nakon varimax rotacije



Slika 2.10: Grafički prikaz varijabli prvog i trećeg faktora nakon varimax rotacije

Nakon ortogonalne rotacije provest ćemo i kosu kako bi imali što jasniju sliku o faktorima. U tu svrhu koristimo promax rotaciju.



Slika 2.11: Grafički prikaz varijabli za prva dva faktora nakon promax rotacije





Nakon promax rotacije, faktori postaju korelirani.

Tablica 2.15: Korelacije između faktora nakon promax rotacije

	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5	Faktor6	Faktor7	Faktor8
Faktor1	0,10	0,08	0,11	0,03	0,02	0,08	0,01	0,21
Faktor2	0,08	0,10	0,17	0,14	0,03	0,11	0,17	0,01
Faktor3	0,11	0,17	0,10	0,13	0,15	0,05	0,08	0,18
Faktor4	0,03	0,14	0,13	0,10	0,16	0,11	0,01	0,02
Faktor5	0,02	0,03	0,15	0,16	0,10	0,14	0,01	0,14
Faktor6	0,08	0,11	0,05	0,11	0,14	0,10	0,17	0,13
Faktor7	0,01	0,17	0,08	0,01	0,01	0,17	0,10	0,04
Faktor8	0,21	0,01	0,18	0,02	0,14	0,13	0,04	0,10

Komunaliteti su isti set vrijednosti na početku, nakon varimax rotacije i nakon promax rotacije. Osnovno svojstvo faktorske rotacije je da ona samo redistribuira varijancu koja je objašnjena faktorima. Ukupna varijanca objašnjena faktorima za svaku varijablu (tj. komunalitet varijable) ostaje nepromijenjen. Tablica 2.16 nam prikazuje krajnje procjene komunaliteta prije rotacije, komunalitete nakon varimax rotacije i komunalitete nakon promax rotacije.

Tablica 2.16: Procjene komunaliteta prije rotacije, nakon varimax rotacije te nakon promax rotacije

<b>Varijable</b>	Procjena komunaliteta prije rotacije	Varimax procjene komunaliteta	Promax procjene komunaliteta
Dob	0.43	0.43	0.43
Volite li slušati glazbu	0.20	0.20	0.20
Pop	0.23	0.23	0.23
Rock	0.54	0.54	0.54
Metal	0.39	0.39	0.39
Hip hop Rap	0.29	0.29	0.29
Jazz	0.33	0.33	0.33
Narodnjaci	0.36	0.36	0.36
Techno	0.37	0.37	0.37
Klasična glazba	0.37	0.37	0.37

Trash	0.21	0.21	0.21
House	0.30	0.30	0.30
Volite li gledati filmove	0.24	0.24	0.24
Horor	0.21	0.21	0.21
Triler	0.26	0.26	0.26
Komedija	0.16	0.16	0.16
Romantični	0.33	0.33	0.33
SCI FI	0.27	0.27	0.27
Dokumentarni	0.30	0.30	0.30
Akcijski	0.28	0.28	0.28
Animirani	0.16	0.16	0.16
Western	0.30	0.30	0.30
Let avionom	0.34	0.34	0.34
Visina	0.29	0.29	0.29
Vremenske nepogode	0.29	0.29	0.29
Mrak	0.31	0.31	0.31
Pauci	0.22	0.22	0.22
Dizalo	0.34	0.34	0.34
Zmije	0.33	0.33	0.33
Mali prostori	0.29	0.29	0.29
Psi	0.14	0.14	0.14
Javni govor	0.24	0.24	0.24
Glodavci	0.40	0.40	0.40
Zubar	0.16	0.16	0.16
Doktorske igle	0.30	0.30	0.30
Krv	0.29	0.29	0.29
Bacili	0.19	0.19	0.19
Čitanje	0.34	0.34	0.34
Strani jezici	0.27	0.27	0.27
Likovna umjetnost	0.30	0.30	0.30
Gluma	0.38	0.38	0.38
Ples	0.38	0.38	0.38
Sviranje	0.41	0.41	0.41
Pjevanje	0.41	0.41	0.41

Koncerti	0.35	0.35	0.35
Kazalište	0.43	0.43	0.43
Video igrice	0.38	0.38	0.38
Automobili	0.29	0.29	0.29
Shopping	0.33	0.33	0.33
Profesionalno bavljenje sportom	0.17	0.17	0.17
Amatersko bavljenje sportom	0.13	0.13	0.13
Povijest	0.49	0.49	0.49
Zemljopis	0.36	0.36	0.36
Psihologija	0.30	0.30	0.30
Politika	0.45	0.45	0.45
Matematika	0.51	0.51	0.51
Fizika	0.64	0.64	0.64
Informatika	0.50	0.50	0.50
Biologija	0.64	0.64	0.64
Kemija	0.60	0.60	0.60
Medicina	0.57	0.57	0.57
Pravo	0.35	0.35	0.35
Ekonomija	0.32	0.32	0.32
Elektrotehnika	0.63	0.63	0.63
Robotika	0.61	0.61	0.61
Religija	0.19	0.19	0.19

Za razliku od ortogonalne rotacije, varijanca koju objašnjavaju faktori nakon kose rotacije ne daje u zbroju cijelu procjenu komunaliteta. Izračunava se varijanca koja je objašnjena zajedničkim faktorom, ali zanemarujemo doprinos ostalih faktora. To možemo vidjeti u Tablici 2.17.

Tablica 2.17: Varijanca objašnjena faktorima nakon promax rotacije

Varijanca objašnjena svakim od faktora bez doprinosa ostalih faktora							
Faktor1	Faktor2	Faktor3	Faktor4	Faktor5	Faktor6	Faktor7	Faktor8
3.56	3.60	3.68	2.92	2.81	2.94	2.74	2.87

Nadalje tražimo i uzorak faktorskih opterećenja, što se ujedno naziva i jednostavna struktura (eng. *simple structures*). Što se opterećenja više razlikuju to je jednostavnije objasniti faktore.

Kako bi ispitali korelaciju između varijabli i faktora promatramo matricu faktorske strukture. Samo pomoću ove matrice možemo interpretirati korelaciju.

Tablica 2.18: Faktorska struktura (faktorska opterećenja); C-komunaliteti za svaku od varijabli

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Faktor 8	C
<b>FOBIJE</b>									
Let avionom	<b>0.52</b>	0.02	0.00*	0.00*	0.10	-0.09	0.02	0.04	0.29
Dizalo	<b>0.53</b>	-0.07	0.03	0.03	-0.02	-0.01	0.17	0.01	0.32
Visina	<b>0.49</b>	0.04	-0.05	0.03	0.07	-0.02	-0.05	-0.03	0.26
Mali prostori	<b>0.51</b>	-0.14	0.07	-0.01	0.01	0.04	0.09	-0.05	0.30
Mrak	<b>0.53</b>	-0.11	0.17	-0.07	-0.11	0.04	0.13	-0.06	0.36
Doktorske igle	<b>0.53</b>	-0.09	0.09	-0.16	-0.03	0.00*	-0.10	-0.14	0.35
Krv	<b>0.51</b>	-0.08	0.14	-0.15	-0.04	-0.01	-0.10	-0.17	0.35
Vremenske nepogode	<b>0.51</b>	-0.11	0.18	-0.19	-0.07	-0.04	0.09	-0.19	0.39
Bacili	<b>0.42</b>	-0.06	0.16	-0.07	0.04	0.08	0.00*	-0.12	0.23
Javni govor	<b>0.40</b>	-0.08	-0.03	-0.03	-0.20	-0.02	0.02	-0.21	0.25
Zubar	<b>0.38</b>	-0.03	0.06	-0.01	0.01	-0.05	0.01	-0.15	0.18
Pauci	<b>0.42</b>	-0.17	0.14	-0.17	-0.14	0.09	0.04	-0.10	0.29
Zmije	<b>0.46</b>	-0.15	0.16	-0.29	0.00*	0.03	-0.08	-0.32	0.45
Glodavci	<b>0.48</b>	-0.21	0.26	-0.39	-0.11	-0.04	-0.06	-0.31	0.60
Psi	<b>0.27</b>	0.14	0.08	-0.06	0.01	-0.12	0.07	-0.02	0.12
<b>STEM</b>									
Elektrotehnika	-0.13	<b>0.78</b>	-0.07	0.21	0.22	0.22	0.06	0.12	0.79
Robotika	-0.10	<b>0.77</b>	-0.07	0.22	0.16	0.19	0.09	0.20	0.76
Fizika	-0.13	<b>0.72</b>	0.01	0.22	0.07	0.04	0.38	0.04	0.73
Informatika	-0.07	<b>0.69</b>	-0.10	0.24	0.15	0.13	0.03	-0.01	0.59
Matematika	0.03	<b>0.59</b>	0.07	0.09	-0.03	-0.09	0.24	-0.19	0.46
Automobili	-0.13	<b>0.29</b>	-0.02	-0.04	0.31	0.27	-0.25	0.10	0.34
Amatersko bavljenje sportom	-0.14	<b>0.24</b>	0.09	-0.03	0.14	0.21	0.02	0.01	0.15

<b>UMJETNICI</b>									
Pjevanje	0.11	0.04	<b>0.58</b>	0.07	0.05	0.01	0.15	0.00*	0.38
Sviranje	-0.01	0.15	<b>0.47</b>	0.30	0.07	-0.04	0.09	0.12	0.36
Ples	0.09	-0.06	<b>0.56</b>	-0.16	-0.05	0.14	0.20	-0.05	0.42
Kazalište	0.03	-0.09	<b>0.57</b>	0.16	0.19	-0.02	0.16	-0.30	0.51
Koncerti	0.07	0.02	<b>0.50</b>	0.09	0.08	0.00*	-0.03	-0.30	0.36
Gluma	0.05	0.00*	<b>0.49</b>	0.00*	0.19	0.14	0.13	0.17	0.35
Likovna umjetnost	0.04	-0.06	<b>0.40</b>	0.26	0.10	0.08	0.19	0.12	0.30
Volite li slušati glazbu	0.06	-0.08	<b>0.32</b>	0.00*	-0.06	0.29	-0.03	-0.04	0.20
Romantični	0.28	-0.19	<b>0.41</b>	-0.30	-0.06	0.13	0.16	-0.28	0.50
Animirani	0.11	0.06	<b>0.28</b>	0.18	0.00*	0.20	0.15	-0.01	0.19
<b>ALTERNATIVCI</b>									
Rock	-0.02	0.06	0.19	<b>0.63</b>	0.03	0.01	0.04	-0.24	0.49
Metal	-0.05	0.18	0.01	<b>0.59</b>	0.03	0.08	-0.01	0.06	0.40
Klasična glazba	-0.02	0.11	0.27	<b>0.49</b>	0.23	0.03	0.26	-0.14	0.47
Jazz	-0.08	0.14	0.26	<b>0.46</b>	0.13	0.10	0.15	-0.15	0.38
SCI-FI	-0.03	0.25	0.06	<b>0.39</b>	0.02	0.28	0.09	0.12	0.32
Čitanje	0.03	-0.13	0.33	<b>0.33</b>	0.15	-0.04	0.27	-0.26	0.40
Dokumentarni	0.00*	0.15	0.03	<b>0.41</b>	0.37	0.16	0.18	-0.05	0.38
Komedija	0.12	0.00*	0.22	<b>-0.21</b>	0.09	0.25	0.04	-0.07	0.18
Shopping	0.28	-0.16	0.25	<b>-0.43</b>	0.00*	0.23	0.01	-0.23	0.45
Narodnjaci	0.13	-0.10	-0.02	<b>-0.54</b>	0.09	0.20	-0.11	0.00*	0.37
<b>DRUŠTVENJACI</b>									
Politika	-0.04	0.04	0.02	0.01	<b>0.65</b>	0.04	0.02	-0.05	0.43
Povijest	0.00*	0.08	0.06	0.26	<b>0.67</b>	0.04	0.02	0.11	0.55
Zemljopis	0.01	0.14	0.08	0.11	<b>0.59</b>	0.11	0.09	0.02	0.41
Pravo	-0.01	-0.01	0.11	-0.08	<b>0.53</b>	0.22	0.10	0.03	0.36
Ekonomija	0.01	0.09	0.00*	-0.09	<b>0.48</b>	0.11	-0.11	-0.17	0.30
Religija	0.04	0.12	0.21	-0.04	<b>0.34</b>	-0.04	0.10	-0.04	0.19

Western	-0.07	0.27	0.02	0.28	<b>0.41</b>	0.27	-0.09	0.10	0.41
<b>MAIN- STREAM</b>									
Techno	0.02	0.14	-0.01	-0.09	0.03	<b>0.58</b>	0.06	0.05	0.38
House	0.03	0.18	0.06	0.05	0.07	<b>0.51</b>	0.12	-0.10	0.33
Hip hop, Rap	-0.07	0.02	0.09	-0.02	0.04	<b>0.51</b>	0.03	0.17	0.30
Akcijski	-0.01	0.21	0.05	0.00*	0.20	<b>0.47</b>	-0.10	0.13	0.33
Volite li gledati filmove	0.11	-0.07	0.28	0.02	0.04	<b>0.40</b>	0.03	-0.09	0.26
Triler	-0.08	0.11	-0.08	0.18	0.18	<b>0.34</b>	0.04	-0.20	0.25
Horor	-0.12	0.07	-0.15	0.08	0.13	<b>0.37</b>	-0.01	0.21	0.24
Strani jezici	0.05	-0.09	0.33	0.20	0.23	<b>0.31</b>	0.15	-0.02	0.33
Profesionalno bavljenje sportom	-0.09	0.11	0.04	-0.18	0.08	<b>0.28</b>	0.00*	0.18	0.17
<b>PRIRODO- SLOVCI</b>									
Biologija	0.03	0.19	0.16	0.13	0.07	0.04	<b>0.79</b>	-0.03	0.70
Medicina	-0.03	0.14	0.14	0.06	0.09	0.10	<b>0.74</b>	-0.07	0.61
Kemija	0.00*	0.33	0.14	0.11	-0.02	0.08	<b>0.71</b>	0.01	0.66
Psihologija	0.05	-0.14	0.23	0.16	0.24	0.00*	<b>0.41</b>	-0.12	0.35
<b>OSTALO</b>									
Dob	0.00*	-0.13	0.14	-0.26	-0.14	0.22	0.02	<b>0.60</b>	0.54
Video igrice	-0.12	0.35	-0.22	0.20	0.16	0.27	-0.26	<b>0.37</b>	0.53
Pop	0.22	-0.13	0.29	-0.12	-0.05	0.21	0.16	<b>-0.30</b>	0.33
Trash	0.11	-0.03	0.15	-0.09	-0.06	0.06	0.12	<b>-0.46</b>	0.28
<b>SVOJSTVE- NE VRI- JEDNOSTI</b>	3.77	3.76	3.48	3.54	2.98	2.75	2.85	2.10	
* vrijednosti manje od 0.005									

Ako je korelacija u Tablici 2.18 pozitivna, onda pokazuje u kojoj mjeri određena varijabla doprinosi tom faktoru, a ako je negativna, pokazuje koliko varijable ne učestvuju u danom faktoru.

Nakon napravljene faktorske analize imamo 8 faktora koje smo nazvali prema varijabla koje su se učitale na njih. To su: Fobije, STEM, Umjetnici, Alternativci, Društvenjaci, Mainstream, Prirodoslovci te Ostalo.

U faktor Fobije učitale su se sve varijable, tj. pitanja koja su bila povezana sa strahovima. Sve korelacije varijabli s faktorom su pozitivne. Na faktor STEM s visokom korelacijom su se učitale varijable elektrotehnika, robotika, fizika, informatika i matematika, a s nižom korelacijom varijable automobili i amatersko bavljenje sportom. Sve varijable su pozitivno korelirane s faktorom. Ujedno vidimo se da te iste varijable negativno učitavaju na faktor Fobije što bi značilo da uglavnom osobe koje imaju interes za STEM nemaju strahove. Za faktor Umjetnici vezale su se varijable pjevanje, sviranje, ples, kazalište, koncerti, gluma, likovna umjetnost, volite li slušati glazbu, romantični i animirani, svi s pozitivnim korelacijama. Faktor Alternativci se sastoji od varijabli rock, metal, klasična glazba, jazz, sci-fi, čitanje, dokumentarni, komedija, shopping i narodnjaci. Jedino se varijable komedija, shopping i narodnjaci učitavaju s negativnim predznakom što nam govori da su te varijable suprotne ostalima koje se nalaze u faktoru. Iz tablice se isto tako vidi da varijable koje pozitivno koreliraju s faktorom Alternativci, negativno koreliraju s faktorom Fobije. Na idući faktor, Društvenjaci, učitale su se varijable politika, povijest, zemljopis, pravo, ekonomija, religija i western, svi s pozitivnom korelacijom. U faktor Mainstream ušle su varijable techno, house, hip hop, rap, akcijski, volite li gledati filmove, triler, horor, strani jezici i profesionalno bavljenje sportom, također svi s pozitivnim korelacijama. Faktor Prirodoslovci sadrži varijable biologija, medicina, kemija, psihologija. Sve varijable imaju pozitivnu, visoku korelaciju i faktor ih dobro opisuje. Pod faktor Ostalo su se učitale varijable godina rođenja, video igrice, pop i trash. Pop i trash imaju negativnu korelaciju s faktorom. Te četiri varijable nemaju neku logičku povezanost jedna s drugom, ali ih je faktorska analiza spojila skupa zato što se ne učitavaju s visokom korelacijom ni na koji drugi faktor.

Kroz analizu svih 8 faktora možemo zaključiti da ovakvo grupiranje manifestnih varijabli ima smisla. Idući korak bi bila daljnja analiza podataka kroz dobivene faktore ili konfirmatorna faktorska analiza.

# Poglavlje 3

## Prilog

### 3.1 Anketa

Pitanja iz online ankete:

Osobni podatci

1. Spol: M Ž
2. Dob (brojkama):
3. Regija Hrvatske u kojoj ste odrasli
  - Centralna Hrvatska
  - Istočna Hrvatska
  - Planinska Hrvatska
  - Sjeverna Hrvatska obala
  - Južna Hrvatska Južna Hrvatska obala
4. Mjesto u kojem ste odrasli:
5. Mjesto trenutnog školovanja (npr. Zagreb):
6. Trenutno pohađam:
  - Srednja škola-gimnazija
  - Srednja škola-strukovna



- Veleučilište
- Sveučilište

7. Pohađate li privatnu školufakultet? DA NE

8. Ako ste student, naznačite kojim se područjem znanosti bavite:

- Prirodne Znanosti
- Tehničke znanosti i tehnologija
- Medicinske znanosti
- Biotehnološke znanosti
- Društvene znanosti
- Umjetničke znanosti
- Humanističke znanosti

9. Status veze roditelja:

- Vjenčani
- Razvedeni
- Vanbračna zajednica
- Samohrani roditelj
- Udovaica
- Bez roditelja

10. Broj djece u obitelji (odgovor napišite brojem):

11. Broj kućnih ljubimaca (odgovor napišite brojem):

12. Jeste li zaposleni dok se školujete? DA NE

Sljedeći niz pitanja ispituje Vaše interese. Molimo za svako pitanje označite odgovarajućim brojem koliko uživate u određenoj vrsti glazbe i filmova, pri čemu je: 1 - ne slušam/gledam uopće, 7 - jako volim slušati/gledati.

**Glazba**

13. Volite li slušati glazbu? 1 2 3 4 5 6 7

14. Pop: 1 2 3 4 5 6 7

15. Rock: 1 2 3 4 5 6 7

16. Metal: 1 2 3 4 5 6 7

17. Hip hop, Rap: 1 2 3 4 5 6 7

18. Jazz: 1 2 3 4 5 6 7

19. Narodnjaci: 1 2 3 4 5 6 7

20. Techno: 1 2 3 4 5 6 7

21. Klasična glazba: 1 2 3 4 5 6 7

22. Trash: 1 2 3 4 5 6 7

23. House: 1 2 3 4 5 6 7

**Filmovi**

24. Volite li gledati filmove? 1 2 3 4 5 6 7

25. Horor: 1 2 3 4 5 6 7

26. Triler: 1 2 3 4 5 6 7

27. Komedija: 1 2 3 4 5 6 7

28. Romantični: 1 2 3 4 5 6 7

29. SCIFI: 1 2 3 4 5 6 7

30. Dokumentarni: 1 2 3 4 5 6 7

31. Akcijski: 1 2 3 4 5 6 7

32. Animirani: 1 2 3 4 5 6 7

33. Western: 1 2 3 4 5 6 7

### **Strahovi**

Sljedeći niz pitanja ispituje Vaše fobije. Molimo, za svako pitanje označite Vaš odgovor brojem, pri čemu je:

1 - nemam strah od navedenog, 7 - imam veliki strah od navedenog.

34. Let avionom: 1 2 3 4 5 6 7

35. Visina: 1 2 3 4 5 6 7

36. Vremenske nepogode (grmljavina, oluja,...): 1 2 3 4 5 6 7

37. Mrak: 1 2 3 4 5 6 7

38. Pauci: 1 2 3 4 5 6 7

39. Dizalo: 1 2 3 4 5 6 7

40. Zmije: 1 2 3 4 5 6 7

41. Mali prostori: 1 2 3 4 5 6 7

42. Psi: 1 2 3 4 5 6 7

43. Javni govor: 1 2 3 4 5 6 7

44. Glodavci (štakori, miševi,...): 1 2 3 4 5 6 7

45. Zubar: 1 2 3 4 5 6 7

46. Doktorske igle: 1 2 3 4 5 6 7

47. Krv: 1 2 3 4 5 6 7

48. Bacili: 1 2 3 4 5 6 7

### **Hobiji i interesi**

Sljedeći niz pitanja ispituje Vaše hobije. Molimo, za svako pitanje označite odgovor brojem 1-7, ovisno o Vašoj uključenosti u sljedeće aktivnosti, pri čemu je:

1- nikad ovo ne radim, 7- redovito se ovim bavim.

49. Čitanje: 1 2 3 4 5 6 7

50. Strani jezici: 1 2 3 4 5 6 7

51. Umjetnost: 1 2 3 4 5 6 7
52. Likovna umjetnost: 1 2 3 4 5 6 7
53. Gluma: 1 2 3 4 5 6 7
54. Ples: 1 2 3 4 5 6 7
55. Sviranje: 1 2 3 4 5 6 7
56. Pjevanje: 1 2 3 4 5 6 7
57. Koncerti: 1 2 3 4 5 6 7
58. Kazalište: 1 2 3 4 5 6 7
59. Video igrice: 1 2 3 4 5 6 7
60. Automobili: 1 2 3 4 5 6 7
61. Shopping: 1 2 3 4 5 6 7
62. Profesionalno bavljenje sportom: 1 2 3 4 5 6 7
63. Amatersko bavljenje sportom: 1 2 3 4 5 6 7
64. Povijest: 1 2 3 4 5 6 7
65. Zemljopis: 1 2 3 4 5 6 7
66. Psihologija: 1 2 3 4 5 6 7
67. Politika: 1 2 3 4 5 6 7
68. Matematika: 1 2 3 4 5 6 7
69. Fizika: 1 2 3 4 5 6 7
70. Informatika: 1 2 3 4 5 6 7
71. Biologija: 1 2 3 4 5 6 7
72. Kemija: 1 2 3 4 5 6 7
73. Medicina: 1 2 3 4 5 6 7
74. Pravo: 1 2 3 4 5 6 7

75. Ekonomija: 1 2 3 4 5 6 7

76. Elektrotehnika: 1 2 3 4 5 6 7

77. Robotika: 1 2 3 4 5 6 7

78. Religija: 1 2 3 4 5 6 7

#### **Zdravstvene navike**

Sljedeći (ujedno i posljednji) niz pitanja ispituje Vašu svijest o zdravlju. Molimo odgovorite iskreno..

79. Pušenje: DA NE

80. U prosjeku, koliko dnevno cigareta popušite?

- Ne pušim cigarete
- 15
- 610
- 1120
- Više od kutije dnevno

81. Pijete li alkohol?: DA NE

82. Koliko često pijete alkohol?

- Nikad
- Samo vikendom
- Na društvenim događanjima
- Skoro svakodnevno

83. Jeste li ikad konzumirali drogu?: DA NE

84. Koliko učestalo konzumirate drogu?

- Nikad
- Probao/la
- U prosjeku jednom tjedno
- Par puta tjedno

85. Ako konzumirate, u koju skupinu najviše ulazite:

- Opojne droge(morfij, kodein, heroin, metadon...)
- Stimulativne droge (kokain i crack, MDMA-ecstasy...)
- Halucinogene droge (marihuana, hašiš, LSD...)

## 3.2 Kod

```

/** Import an XLSX file. */
PROC IMPORT DATAFILE="/home/mateamijatovic200/sasuser.v94/
        Diplomski-anketa/Anketa.xlsx"
        OUT=WORK.anketa
        DBMS=XLSX
        REPLACE;

RUN;

/*SPOL*/
title "Spol";
proc freq data=anketa;
        table spol;
run;

goptions reset=all border;
title "Spol";
proc gchart data=anketa;
        pie spol / midpoints="M" "Ž"
                value=none
                percent=arrow
                slice=arrow
                noheading;

run;
quit;

/*SREDNJOŠKOLCI/STUDENTI*/
title "Srednjoškolci/studenti";
proc freq data=anketa;
        table Trenutno_pohadaj;
run;

goptions reset=all border;

```

```
title "Srednjoškolci/studenti";
proc gchart data=anketa;
  pie Trenutno_pohadam /
    value=none
    percent=arrow
    slice=arrow
    noheading;

run;
quit;

/*PUŠENJE*/
title "Pušenje";
proc freq data=anketa;
  table Pušenje Koliko_dnevno;
run;

goptions reset=all border;
title "Pušenje";
proc gchart data=anketa;
  pie Pušenje /
    value=none
    percent=arrow
    slice=arrow
    noheading;

run;
quit;

goptions reset=all border;
title "Pušenje";
proc gchart data=anketa;
  pie Trenutno_pohadam / detail=pušenje
  detail_percent=best
  detail_value=none
  detail_slice=best
  detail_threshold=2
  legend
;
run;
quit;
```

```
/*ALKOHOL*/  
title "Alkohol";  
proc freq data=anketa;  
    table Alkohol Koliko_cesto;  
run;  
  
goptions reset=all border;  
title "Alkohol";  
proc gchart data=anketa;  
    pie Alkohol /  
        value=none  
        percent=arrow  
        slice=arrow  
        noheading;  
run;  
quit;  
proc gchart data=anketa;  
  
goptions reset=all border;  
title "Alkohol";  
pie Trenutno_pohadajam / detail=Alkohol  
detail_percent=best  
detail_value=none  
detail_slice=best  
detail_threshold=2  
legend  
;  
run;  
quit;  
  
/*DROGA*/  
title "Droga";  
proc freq data=anketa;  
    table Jeste_li_ikad_konzumirali_drogu  
        Koliko_cestalo Skupine_droge;  
run;  
  
goptions reset=all border;
```



```

title "Droga";
proc gchart data=anketa;
  pie Jeste_li_ikad_konzumirali_drogu /
      value=none
      percent=arrow
      slice=arrow
      noheading;

run;
quit;

goptions reset=all border;
title "Droga";
proc gchart data=anketa;
pie Trenutno_pohadajam / detail=Jeste_li_ikad_konzumirali_drogu
detail_percent=best
detail_value=none
detail_slice=best
detail_threshold=2
legend
;
run;
quit;

/*Diskretne varijable*/
proc means data=anketa n min median max mean std;
  var  Godina_rodenja Volite_li_slusati_glazbu Pop Rock
      Metal Hip_hop_Rap Jazz Narodnjaci
      Techno Klasicna_glazba Trash House
      Volite_li_gledati_filmove Horor Triler Komediya
      Romantichni SCI_FI Dokumentarni Akcijski Animirani
      Western Let_avionom Visina Vremenske_nepogode Mrak
      Pauci Dizalo Zmije Mali_prostori Psi Javni_govor
      Glodavci Zubar Doktorske_igle Krv Bacili Citanje
      Strani_jezici Likovna_umjetnost Gluma Ples Sviranje
      Pjevanje Koncerti Kazaliste Video_igrice Automobili
      Shopping Profesionalno_bavljenje_sportom
      Amatersko_bavljenje_sportom Povijest Zemljopis
      Psihologija Politika Matematika Fizika
      Informatika Biologija Kemija Medicina Pravo

```

```

                                Ekonomija Elektrotehnika Robotika Religija;
run;
*options validvarname=v7;
/** Print the results. **/

*PROC PRINT DATA=WORK.anketa;
*RUN;

/*Faktorska analiza*/
proc factor data=Anketa
    priors= smc msa residuals score
    rotate= promax reorder
                                NFACTORS=8
    outstat=fact_all
    plots=(scree initloadings preloadings loadings) ;
run;

/*Proc Score*/
proc score data=anketa score=fact_all out=FScore;
    var  Godina_rodjenja Volite_li_slusati_glazbu Pop Rock
        Metal Hip_hop_Rap Jazz Narodnjaci
        Techno Klasična_glazba Trash House
        Volite_li_gledati_filmove Horror Triler Komediya
        Romantični SCI_FI Dokumentarni Akcijski Animirani
        Western Let_avionom Visina Vremenske_nepogode Mrak
        Pauci Dizalo Zmije Mali_prostori Psi Javni_govor
        Glodavci Zubar Doktorske_igle Krv Bacili Čitanje
        Strani_jezici Likovna_umjetnost Gluma Ples Sviranje
        Pjevanje Koncerti Kazalište Video_igrice Automobili
        Shopping Profesionalno_bavljenje_sportom
        Amatersko_bavljenje_sportom Povijest Zemljopis
        Psihologija Politika Matematika Fizika
        Informatika Biologija Kemija Medicina Pravo
        Ekonomija Elektrotehnika Robotika Religija;
run;
proc print data=FScore;
    title2 'Data Set from PROC SCORE';
run;

```



# Bibliografija

- [1] *Exploratory factor analysis*, 2004, [https://en.wikipedia.org/wiki/Exploratory\\_factor\\_analysis](https://en.wikipedia.org/wiki/Exploratory_factor_analysis).
- [2] *Factor analysis*, 2004, [https://en.wikipedia.org/wiki/Factor\\_analysis](https://en.wikipedia.org/wiki/Factor_analysis).
- [3] *SAS/STAT ®9.22 User's Guide, Example 33.2 Principal Factor Analysis*, 2010, [https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug\\_factor\\_sect029.htm](https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_factor_sect029.htm).
- [4] *Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy*, 2016, <https://www.statisticshowto.datasciencecentral.com/kaiser-meyer-olkin/>.
- [5] Anamarija Jazbec, *Faktorska analiza*, bilješke s predavanja, 2017.
- [6] Miroslav Sabo, *Young People Survey*, 2016, <https://www.kaggle.com/miroslavsabo/young-people-survey>.
- [7] Fidell L.S. Tabachnick B.G., *Using Multivariate Statistics*, Allyn & Bacon, 2001.



# Sažetak

U ovom radu opisane su faktorska analiza i eksplorativna faktorska analiza te njihova primjena na bazu podataka o srednjoškolicima i studentima. Cilj rada je smanjivanje dimenzije mjerelog prostora, tj. grupirati varijable u faktore. Prvi korak bio je napraviti analizu podataka koji su bili dobiveni putem internet ankete. Napravljena je deskriptivna statistika pitanja iz ankete, strukturni krugovi za kategorijske varijable te tablica za diskretne varijable u kojoj su podatci o minimumu, medijanu, maksimumu, aritmetičkoj sredini i standardnoj devijaciji. Eksplorativna faktorska analiza provedena je na 66 diskretnih varijabli i ekstrakcijom je dobiveno 8 faktora. Potom su napravljene ortogonalne i kose rotacije. Pomoću toga dobivena je faktorska struktura. Svaki od 8 faktora dobio je ime s obzirom koje su se varijable učitale u njega.



# Summary

This paper describes factor analysis and exploratory factor analysis and their application to database of high school and college student. The purpose of the paper was to reduce the dimension of the measured variables, to group them. The first step was to analyze the data that was obtained through an online survey. Descriptive statistics were created from the survey questions, pie charts for categorical variables and table for discrete variables with information about minimum, media, maximum, mean and standard deviation. Exploratory factor analysis was performed on 66 discrete variables and 8 factors were obtained by extraction. Then orthogonal and oblique rotations were implemented. That gave us factor structure. Every factor was given name considering the variables that load on to it.





# Životopis

Rođena sam dana 31. prosinca 1993. godine u Zadru. Završila sam osnovnu školu Šimuna Kožičića Benje, te gimnaziju Franje Petrića, matematički smjer. Godine 2012. upisujem preddiplomski sveučilišni studij Matematika na Prirodoslovno-matematičkom fakultetu u Zagrebu kojeg završavam 2017. godine. Te iste godine upisujem diplomski studij Matematičke statistike na istom fakultetu.