# Faktorska analiza

**Magdić, Lucija**

**SVEUČILIŠTE U ZAGREBU**

**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**

**MATEMATIČKI ODSJEK**

Lucija Magdić

# FAKTORSKA ANALIZA

Diplomski rad

Voditelj rada:
doc. dr. sc. Snježana
Lubura Strunjak

Zagreb, veljača 2024.

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. _____

*Zahvaljujem svojoj mentorici doc. dr. sc. Snježani Luburi Strunjak na pruženoj prilici, stručnom vodstvu, prenesenom znanju te potpori i strpljenju prilikom izrade ovog diplomskog rada.*

*Veliko hvala mojoj obitelji, posebno mojim roditeljima, na bezuvjetnoj ljubavi, podršci i razumijevanju tijekom studiranja. Zahvaljujući Vama sam tu gdje jesam. Hvala mom zaručniku na potpori.*

*Hvala mojim dragim kolegama i kolegicama koji su me učili ustrajnosti i hrabrosti kroz sve trenutke studija. Hvala Vam što ste mi uljepšali i olakšali studiranje, bez Vas bi sve ovo bilo iznimno teško. Neizmjerno sam sretna što sam stekla prijatelje za cijeli život.*

*Najveće hvala dragom Bogu koji me uvijek pratio i davao mi snage za dalje. AMDG*

# Sadržaj

# Uvod

Faktorska analiza je statistička disciplina koja otkriva i uspostavlja korelaciju među opaženim slučajnim varijablama. Njena glavna svrha je opisati, ako je to moguće, vezu između promatranih varijabli i potencijalno manjeg broja latentnih veličina koje zovemo *faktorima*. Motivirana je činjenicom da se promatrane varijable mogu grupirati s obzirom na njihovu korelaciju. Odnosno, sve varijable unutar jedne grupe su jako korelirane ali njihova korelacija s varijablama iz drugih grupa je relativno malena. Na takav način svaka grupa varijabli predstavlja formaciju koju zovemo *faktor* i upravo on je odgovoran za promatrane korelacije. Cilj faktorske analize je smanjiti broj parametara tako da varijablu odaziva možemo opisati jednako dobro kao i s većim brojem parametra ili čak bolje jer ćemo ukloniti podudaranja. Tako se, umjesto nad velikim brojem koreliranih izvornih varijabli, analiza provodi nad manjim brojem nekoreliranih faktora, od kojih je svaki faktor linearna kombinacija nekoliko promatranih varijabli. Primjena faktorske analize je u psihologiji, sociologiji, marketingu i strojnom učenju. Glavna motivacija korištenja faktorske analize je smislenija interpretacija podataka. Naime, mnogi faktori koji se koriste u analizi se ne mogu promatrati i mjeriti pošto su oni nešto što je zamišljeno u ljudskom umu i nema direktne veličine kojom bismo ih izmjerili. Svaki puta kada si nešto želimo pojasniti, želju za točnošću i preciznošću ćemo ovdje zamijeniti s potrebom za jednostavnošću i shvaćanjem pojmova koji su u pozadini problema. Postoje dvije vrste faktorske analize: eksplorativna faktorska analiza i konfirmatorna faktorska analiza. Dok eksplorativnom faktorskom analizom želimo steći općenito bolje shvaćanje o samim faktorima u konfirmatornoj želimo potvrditi postojeće hipoteze, ideje, mjerenja i istraživanja.

U ovom radu bavit ćemo se teorijskom podlogom i koracima faktorske analize, a na samome kraju primijeniti iste korake na primjerima. Rad započinjemo samom definicijom modela, promatramo faktore i matricu težina koji tvore model. U drugom poglavlju zanimat će nas na koje načine možemo odrediti parametre, odnosno težine, u samom modelu. Postoje razne metode pri određivanju a mi ćemo se detaljnije pozabaviti metodom glavnih komponenti, metodom glavnih faktora, iterativnom metodom glavnih faktora i metodom maksimalne vjerodostojnosti.
Nakon što smo u prethodnoj cjelini odredili težine, u trećoj cjelini promatrat ćemo na koje sve načine možemo odabrati broj faktora potrebnih za model. Zatim u četvrtoj cjelini ćemo

promatrati mogućnost rotacija faktorskog modela. Rotacijama modela uvidjet ćemo da će dobiveni model odnosno faktori biti očitiji time što ćemo rotacijama osi prostora pomicati tako da su dobivene varijable što bliže samim osima. Petu cjelinu posvetit ćemo izučavanju faktorskih score-ova i analizom valjanosti samog modela koji smo dobili. U šestoj cjelini diskutirat ćemo kako provjeriti valjanost dobivenog modela faktorske analize.
Naš posljednji zadatak bit će sve dobiveno i naučeno u prethodnim cjelinama primijeniti na primjeru u sedmoj cjelini. Sve navedene korake primijenit ćemo na konkretnim primjerima i donijeti zaključke na temelju rezultata.

# Poglavlje 1

# Model faktorske analize

## 1.1 Definicija modela faktorske analize

Neka je $X = (X_1, X_2..., X_p)^T$ vektor $p$ opaženih slučajnih varijabli koji ima očekivanje $\mathbb{E}[X] = \mu$ i kovarijacijsku matricu $\Sigma$. Model faktorske analize tvrdi da je $X$ linearno zavisan o nekoliko neopaženih slučajnih varijabli $F_1, F_2, ..., F_m$ koje zovemo *zajednički faktori* i $p$ dodatnih izvora varijacije $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ koje zovemo *specifični faktori* ili *slučajne greške*. Za opservacije $X_1, X_2, ..., X_p$ model faktorske analize možemo zapisati u obliku

$$\begin{aligned} X_1 - \mu_1 &= a_{11}F_1 + a_{12}F_2 + ... + a_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= a_{21}F_1 + a_{22}F_2 + ... + a_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= a_{p1}F_1 + a_{p2}F_2 + ... + a_{pm}F_m + \varepsilon_p \end{aligned} \tag{1.1}$$

U idealnom slučaju, veličina $m$ bi trebala biti puno manja od veličine $p$, u suprotnom nismo postigli da se varijable mogu zapisati kao funkcija što manje faktora. Na $F_i$ u gornjem modelu možemo gledati kao na slučajne varijable koje uzrokuju $X_i$, $i \in 1, ..., p$. Koeficijente $a_{ij}$ zovemo *težine* te nam one ukazuju koliko $j$-ti faktor $F_j$ utječe na $i$-tu varijablu $X_i$ te nam služe za interpretaciju faktora $F_j$. Kada bismo tako željeli opisati ili interpretirati faktor $F_2$ onda bismo proučavali njegove koeficijente (težine) $a_{12}, a_{22}, ..., a_{p2}$. Veće vrijednosti težina $a_{i2}$ povezuju npr. faktor $F_2$ sa odgovarajućim $X_i$-om te od istih tih težina onda možemo protumačiti značenje faktora $F_2$. Nakon što procijenimo koeficijente (težine) očekujemo da će doći do razdvajanja varijabli u grupe koje odgovaraju faktorima. Slučajna varijabla $\varepsilon_i$ sadrži grešku mjerenja, individualan učinak svake varijable $X_i$ na grešku te grešku uzrokovanja.

Kada bismo htjeli direktno koristiti model (1.1) naišli bismo na problem pretjeranog broja nepromatranih varijabli. No, ako dodamo dodatne uvjete na slučajne vektore zajedničkih faktora $F = (F_1, F_2, ..., F_m)$ i slučajnih grešaka $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_p)$, model (1.1) će sugerirati određenu kovarijacijsku vezu.
Osnovne pretpostavke faktorskog modela su da za $i = 1, 2, ..., p,\ j = 1, 2, ..., m$ vrijedi

1. $E[F_j] = 0,$
2. $Var(F_j) = 1,$
3. $Cov(F_J, F_k) = 0 \quad \text{za} \quad j \neq k,$
4. $E[\varepsilon_i] = 0,$
5. $Cov(\varepsilon_i, \varepsilon_k) = 0 \quad \text{za} \quad i \neq k,$
6. $Cov(\varepsilon_i, F_j) = 0.$

Dodatno, moramo dopustiti da svaki $\varepsilon_i$ ima drugačiju varijancu pošto oni predstavljaju rezidualni dio od $X_i$ koji nije zajednički s ostalim varijablama pa je $Var(\varepsilon_i) = \psi_i$. Veličinu $\psi_i$ nazivamo *specifična varijanca*.
Gore navedene pretpostavke su prirodna posljedica bazičnog modela (1.1) i ciljeva faktorske analize. Pošto vrijedi $E[X_i - \mu_i] = 0$ trebamo $E[F_j] = 0$ za sve $j = 1, 2, ..., m$. Pretpostavka $Cov(F_j, F_k) = 0$ proizlazi iz cilja prikazivanja varijabli $X_i$ kao funkcije što je manje faktora moguće, odnosno razdvajanja varijabli u odvojene "grupe" faktora.
Uvjeti $Var(F_j) = 0$, $Var(\varepsilon_i) = \psi_i$, $Cov(F_j, F_k) = 0$ i $Cov(\varepsilon_i, F_j) = 0$ vode k jednostavnom izrazu za varijancu varijable $X_i$:

$$Var(X_i) = a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2 + \psi_i \tag{1.2}$$

koji ima bitnu ulogu u izgradnji našeg modela. Dodatno, pretpostavka $Cov(\varepsilon_i, \varepsilon_k) = 0$ implicira da faktori sadrže svu korelaciju između $X$-eva, odnosno, sve što $X$ imaju zajedničko. Od tuda dolazi i naglasak u faktorskoj analizi na modeliranju kovarijanci ili korelacija između $X$-eva.
Model (1.1) možemo zapisati i u matričnom obliku kao

$$X - \mu = AF + \varepsilon, \tag{1.3}$$

gdje su $X = (X_1, X_2, ..., X_P)'$, $\mu = (\mu_1, \mu_2, ..., \mu_p)'$, $F = (F_1, F_2, ..., F_m)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_p)'$ i

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix}. \tag{1.4}$$

Navest ćemo sada primjer modela (1.1) i (1.3) za parametre $p = 5, m = 2$. Model faktorske analize u tom slučaju možemo zapisati kao:

$$\begin{aligned} X_1 - \mu_1 &= a_{11}F_1 + a_{12}F_2 + \varepsilon_1 \\ X_2 - \mu_2 &= a_{21}F_1 + a_{22}F_2 + \varepsilon_2 \\ X_3 - \mu_3 &= a_{31}F_1 + a_{32}F_2 + \varepsilon_3 \\ X_4 - \mu_4 &= a_{41}F_1 + a_{42}F_2 + \varepsilon_4 \\ X_5 - \mu_5 &= a_{51}F_1 + a_{52}F_2 + \varepsilon_5 \end{aligned} \tag{1.5}$$

odnosno u matričnom obliku (1.3) postaje

$$\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ X_3 - \mu_3 \\ X_4 - \mu_4 \\ X_5 - \mu_5 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}, \tag{1.6}$$

odnosno $X - \mu = AF + \varepsilon$.
Osnovne pretpostavke koje smo naveli između (1.1) i (1.3) mogu pomoću matrica i vektora biti zapisane na sljedeći način:
$E[F_j] = 0,\ j = 1, 2, ..., m$ postaje

$$E[F] = 0, \tag{1.7}$$

$Var(F_j) = 1, j = 1, 2, ..., m$ i $Cov(F_j, F_k) = 0,\ j \neq k$ postaje

$$Cov(F) = I, \tag{1.8}$$

$E[\varepsilon_i], i = 1, 2, ..., p$, postaje

$$E[\varepsilon] = 0, \tag{1.9}$$

a iz $Var(\varepsilon_i) = \psi_i$ i $Cov(\varepsilon_i, \varepsilon_k) = 0,\ i \neq k$ dobivamo

$$Cov(\varepsilon) = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}, \tag{1.10}$$

te naposlijetku iz $Cov(\varepsilon_i, F_j) = 0$ za sve $i, j$ imamo

$$Cov(F, \varepsilon) = 0. \tag{1.11}$$

## 1.2 Kovarijacijska struktura faktorskog modela

Model faktorske analize sugerira kovarijacijsku strukturu vektora X. Naime, iz modela faktorske analize (1.1) i

$$\begin{aligned}(X-\mu)(X-\mu)^T &= (AF+\varepsilon)(AF+\varepsilon)^T \\ &= (AF+\varepsilon)((AF)^T+\varepsilon^T) \\ &= AF(AF)^T+\varepsilon(AF)^T+AF\varepsilon^T+\varepsilon\varepsilon^T\end{aligned} \tag{1.12}$$

dobivamo kovarijacijsku matricu $\Sigma$ vektora X

$$\begin{aligned}\Sigma = Cov(X) &= E[(X-\mu)(X-\mu)^T] \\ &= AE[FF^T]A^T+E[\varepsilon F^T]A^T+AE[\varepsilon^T F]+E[\varepsilon\varepsilon^T] \\ &= AA^T+\Psi.\end{aligned} \tag{1.13}$$

Ukoliko $A$ ima mali broj stupaca, recimo dva ili tri stupca, onda jednakost $\Sigma = AA^T + \Psi$ predstavlja pojednostavljenu strukturu za $\Sigma$ u kojoj su kovarijance modelirane pomoću $a_{ij}$-ova budući da je $\Psi$ dijagonalna matrica.
Dijagonalni elementi od $\Sigma$ mogu lagano biti modelirani prilagođavajući dijagonalne elemente od $\Psi$, dok uz pomoć $AA^T$ dolazimo do nedijagonalnih elemenata. U rijetkim slučajevima se populacijska kovarijacijska matrica može prikazati u obliku $\Sigma = AA^T + \Psi$ gdje je $\Psi$ dijagonalna matirica, a $A$ matrica dimenzija $p \times m$, uz relativno mali m. U praksi rijetko imamo uzoračke kovarijacijske matrice koje zadovolje taj idealni model, no tu pretpostavku ne izostavljamo jer je struktura $\Sigma = AA^T + \Psi$ esencijalna za procjenu $A$.
Kovarijacijsku matricu od $X$

$$\Sigma = AA^T + \Psi \tag{1.14}$$

možemo još zapisati kao

$$\begin{aligned}Var(X_i) &= a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2 + \psi_i = \sum_{j=1}^{m} a_{ij}^2 + \psi_i, \\ Cov(X_i, X_k) &= a_{i1}a_{k1} + a_{i2}a_{k2} + ... + a_{im}a_{km},\end{aligned} \tag{1.15}$$

te iz $Cov(X, F) = A$ imamo

$$Cov(X_i, F_j) = a_{ij}. \tag{1.16}$$

Sumu kvadrata $X_i$-evih kvadrata težina $h_i^2 = a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2$ nazivamo *komunalitet*. Komunalitet je dio varijance varijable koji je dobiven kao rezultat djelovanja $m$ zajedničkih

faktora, odnosno zajednička varijanca s ostalim varijablama dobivena kao rezultat djelovanja zajedničkih faktora. Dio varijance $\psi_i$ je rezultat djelovanja specifičnih faktora koji su karakteristični isključivo za tu varijablu $X_i$. Ako označimo $Var(X_i) = \sigma_i$ dobivamo zapis:

$$\underbrace{\sigma_i}_{\textit{varijanca}} = \underbrace{a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2}_{\textit{komunalitet}} + \underbrace{\psi_i}_{\text{specifična varijanca}}, \tag{1.17}$$

odnosno

$$\sigma_i = h_i^2 + \psi_i. \tag{1.18}$$

Naglasak faktorske analize je jednostavnije objašnjenje kovarijance u vektoru X sa što manje paramatera. Naime, faktorski model pretpostavlja da se $\frac{p(p+1)}{2}$ varijanci i kovarijanci u vektoru X može reproducirati na temelju $p \cdot m$ parametara težine $a_{i,j}$ i $p$ specifičnih varijanci. Ukoliko uzmemo $m = p$ onda naš model (1.14) se svodi na $\Sigma = AA^T$ pošto dijagonalna matrica $\Psi$ postaje nulmatrica. No, upravo kada je $m$ malen u odnosu na $p$ je slučaj od najvećeg interesa u faktorskoj analizi pošto se kovarijanca u $X$ može objasniti s manje parametara nego $\frac{p(p+1)}{2}$ parametara u $\Sigma$.

## 1.3 Nejedinstvenost težina faktora

U slučaju kada je $m > 1$ uvijek dolazi do nasljeđivanja nejedinstvenosti faktorskog modela. Kako bismo to promotrili, uzmimo $T$ $m \times m$ ortogonalnu matricu, za koju vrijedi $TT^T = T^TT = I$. Ponovno promotrimo raspis modela (1.1)

$$X - \mu = AF + \varepsilon = ATT^TF + \varepsilon = A^*F^* + \varepsilon,$$

gdje je

$$A^* = AT \quad \text{i} \quad F^* = T^TF.$$

Provjerom mi ponovno dobivamo da vrijede uvjeti kao i za $F$:

$$\begin{aligned} &(1) \quad E[F^*] = T^TE[F] = 0; \\ &(2) \quad Cov(F^*) = T^TCov(F)T = TT^T = I. \end{aligned}$$

Samim time je nemoguće na temelju opservacija $X$ napraviti razliku između težina iz $A$ i težina iz $A^*$. Faktori $F$ i $F^* = T^TF$ imaju iste statističke karakteristike. Iako se težine iz $A$ i $A^*$ razlikuju, obje generiraju istu kovarijacijsku matricu $\Sigma$:

$$\begin{aligned} \Sigma &= AA^T + \psi \\ &= ATT^TA^T + \psi \\ &= (A^*)(A^*)^T + \psi. \end{aligned}$$

Komunaliteti u dijagonalnim elementima od $AA^T = (A^*)(A^*)^T$ ne ovise o odabiru ortogonalne matrice $T$. Ova pojava zapravo odgovara samom pojmu rotacije faktora, naime množenje ortogonalnom matricom korespondira rotaciji koordinatnog sustava za $X$.
Samim time dobivamo da iako $A^* = AT$ i $A$ daju različite vrijednosti težina, one će ponovno dati istu reprezentaciju. U četvrtom poglavlju diskutirati ćemo kako nam upravo ova mogućnost rotacija može pojednostavniti početni problem i pomoći nam da s većom sigurnošću donosimo zaključke.
Jedna od prednosti modela faktorske analize je da ukoliko model ne odgovara podatcima, to se jasno vidi u procjeni od $A$. U takvim situacijama, dva su moguća problema: nejasno je koliko faktora je potrebno i nejasno je koji su faktori.
Nakon što smo prodiskutirali osnovni model faktorske analize, u narednim poglavljima bavit ćemo se procjenama faktora. Uz sada diskutiranu mogućnost rotacije samih podataka, ukoliko nam neće biti na prvu jednostavno diferencirati faktore na temelju podataka moći ćemo pribjeći mogućnosti rotacije koordinatnog sustava. Rotacijom koordinatnog sustava podatci s kojima radimo omogućit će nam jasniju interpretaciju skupa podataka i određivanje samih faktora.

# Poglavlje 2

# Metode procjene faktora

Postoje mnoge metode procjene faktora modela faktorske analize kao što su analiza glavnih komponenata, analiza zajedničkih faktora, metoda maksimalne vjerodostojnosti, image ekstrakcija, alfa ekstrakcija i metoda neponderiranih i ponderiranih najmanjih kvadrata. U ovome radu pobliže ćemo opisati metodu glavnih komponenti, metodu glavnih faktora i metodu maksimalne vjerodostojnosti. Svim metodama procjene zajedničko je da računaju skup ortogonalnih komponenti, odnosno faktora.

## 2.1 Metoda glavnih komponenata

Prva metoda procjene težina koju promatramo naziva se *metoda glavnih komponenti*. Iz slučajnog uzorka $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ izračunamo uzoračku kovarijacijsku matricu $S$ i želimo pronaći procjenitelj $\hat{A}$ koji aproksimira temeljnu faktorsku jednadžbu $\Sigma = AA^T + \Psi$ uz zamjenu $\Sigma$ sa $S$:

$$S \cong \hat{A}\hat{A}^T + \hat{\Psi}. \tag{2.1}$$

Kako bismo faktorizirali matricu $S$ koristit ćemo spektralnu dekompoziciju

$$S = CDC^T, \tag{2.2}$$

gdje je $C$ ortogonalna matrica koja se sastoji od normaliziranih svojstvenih vektora (vektori kojima je norma jednaka 1) matrice $S$, a $D$ je dijagonalna matrica koja na dijagonali ima svojstvene vrijednosti $\theta_1, \theta_2, ..., \theta_p$ matrice $S$:

$$D = \begin{bmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_p \end{bmatrix}. \tag{2.3}$$

Kako bismo faktorizaciju od $CDC^T$ doveli do oblika $AA^T$, iskoristit ćemo da su svojstvene vrijednosti $\theta_i$ pozitivno semidefinitne matrice $S$ sve pozitivne ili nula, pa matricu $D$ možemo faktorizirati kao

$$D = D^{\frac{1}{2}} D^{\frac{1}{2}}, \tag{2.4}$$

gdje je

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\theta_1} & 0 & \dots & 0 \\ 0 & \sqrt{\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\theta_p} \end{bmatrix}. \tag{2.5}$$

Sa faktorizacijom (2.4) matrice $D$ dobivamo

$$\begin{aligned} S &= CDC^T = CD^{\frac{1}{2}} D^{\frac{1}{2}} C^T \\ &= (CD^{\frac{1}{2}})(CD^{\frac{1}{2}})^T. \end{aligned} \tag{2.6}$$

Pogledamo li sada (2.6) vidimo da je zadan u obliku $\hat{A}\hat{A}^T$ no nećemo definirati da je $\hat{A} = CD^{\frac{1}{2}}$ pošto je $CD^{\frac{1}{2}}$ dimenzije $p \times p$, a mi tražimo $\hat{A}$ dimenzije $p \times m$ gdje $m < p$. Zbog toga ćemo definirati matricu $D_1 = diag(\theta_1, \theta_2, ..., \theta_m)$ koja sadrži $m$ najvećih svojstvenih vrijednosti $\theta_1 > \theta_2 > ... > \theta_m$ a matrica $C_1 = (c_1, c_2, ..., c_m)$ pripadne svojstvene vektore. Na takav način mi $A$ procjenjujemo s prvih $m$ stupaca matrice $CD^{\frac{1}{2}}$,

$$\hat{A} = C_1 D_1^{\frac{1}{2}} = (\sqrt{\theta_1} c_1, \sqrt{\theta_2} c_2, ..., \sqrt{\theta_m} c_m), \tag{2.7}$$

gdje je $\hat{A}$ sada dimenzije $p \times m$, $C_1$ dimenzije $p \times m$ i $D^{\frac{1}{2}}$ dimenzije $m \times m$.
Strukturu matrice $\hat{A}$ iz (2.7) prikazat ćemo za $p = 5$ i $m = 2$:

$$\begin{aligned} \begin{bmatrix} \hat{a_{11}} & \hat{a_{12}} \\ \hat{a_{21}} & \hat{a_{22}} \\ \hat{a_{31}} & \hat{a_{32}} \\ \hat{a_{41}} & \hat{a_{42}} \\ \hat{a_{51}} & \hat{a_{52}} \end{bmatrix} &= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \end{bmatrix} \begin{bmatrix} \sqrt{\theta_1} & 0 \\ 0 & \sqrt{\theta_2} \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\theta_1} c_{11} & \sqrt{\theta_2} c_{12} \\ \sqrt{\theta_1} c_{21} & \sqrt{\theta_2} c_{22} \\ \sqrt{\theta_1} c_{31} & \sqrt{\theta_2} c_{32} \\ \sqrt{\theta_1} c_{41} & \sqrt{\theta_2} c_{42} \\ \sqrt{\theta_1} c_{51} & \sqrt{\theta_2} c_{52} \end{bmatrix}. \end{aligned} \tag{2.8}$$

U prethodnom matričnom zapisu možemo vidjeti otkud dolazi sam naziv metode. Stupci matrice $\hat{A}$ su proporcionalni svojstvenim vektorima matrice $S$ pa su i težine na $j$-tom faktoru proporcionalne koeficijentima uz $j$-tu glavnu komponentu. Faktori su stoga povezani

sa prvih m glavnih komponenti te je za očekivati da je interpretacija faktora ista kao interpretacija glavnih komponenti. No, nakon rotacije težina dobivamo najčešće drugačiju interpretaciju faktora.
Znamo da je $i$-ti dijagonalni element od $\hat{A}\hat{A}^T$ suma kvadrata $i$-tog retka matrice $\hat{A}$ odnosno $\hat{a}_i^T\hat{a}_i = \sum_{j=1}^m \hat{a}_{ij}^2$. Kako bismo dovršili aproksimaciju od $S$, definiramo

$$\hat{\psi}_i = s_{ii} - \sum_{j=1}^{m} \hat{a}_{ij}^2 \tag{2.9}$$

te imamo

$$S \cong \hat{A}\hat{A}^T + \hat{\Theta}, \tag{2.10}$$

gdje $\hat{\Psi} = diag(\hat{\psi}_1, \hat{\psi}_2, ..., \hat{\psi}_p)$. U (2.10) su varijance na dijagonali matrice $S$ modelirane egzaktno dok su nedijagonalne varijance aproksimirane. Ovo je jedan od izazova faktorske analize.
U ovoj metodi procjene, sume kvadrata redaka i stupaca od $\hat{A}$ su jednake komunalitetima i svojstvenim vrijednostima, respektivno. Naime, iz (2.9) $i$-ti komunalitet je procijenjen s

$$\hat{h}_i^2 = \sum_{j=1}^{m} \hat{a}_{ij}^2, \tag{2.11}$$

što je suma kvadrata $i$-tog retka matrice $\hat{A}$. Suma kvadrata $j$-tog stupca matrice $\hat{A}$ je $j$-ta svojstvena vrijednost matrice $S$:

$$\begin{aligned} \sum_{i=1}^{p} \hat{a}_{ij}^2 &= \sum_{i=1}^{p} (\sqrt{\theta_j} c_{ij})^2 \\ &= \theta_j \sum_{i=1}^{p} c_{ij}^2 \\ &= \theta_j, \end{aligned} \tag{2.12}$$

pošto normalizirani svojstveni vektori (stupci matrice $C$) imaju duljinu 1.
Iz (2.9) i (2.11) slijedi da je varijanca $i$-te varijable podijeljena na dio koji dolazi od faktora i dio koji dolazi od same te varijable:

$$\begin{aligned} s_{ii} &= \hat{h}_i^2 + \hat{\psi}_i \\ &= \hat{a}_{i1}^2 + \hat{a}_{i2}^2 + ... + \hat{a}_{im}^2 + \hat{\Psi}_i. \end{aligned} \tag{2.13}$$

Dakle, $j$-ti faktor doprinosi $\hat{h}_{ij}^2$ u izrazu od $s_{ii}$. Doprinos koji $j$-ti faktor ima u ukupnoj uzoračkoj varijanci $tr(S) = s_{11} + s_{22} + ... + s_{pp}$ tako iznosi

$$\text{Varijanca } j\text{-tog faktora} = \sum_{i=1}^{p} \hat{a}_{ij}^2 = \hat{a}_{1j}^2 + \hat{a}_{2j}^2 + ... + \hat{a}_{pj}^2, \tag{2.14}$$

što je suma kvadrata težina u $j$-tom stupcu matrice $\hat{A}$. Odnosno, po (2.12), to je upravo jednako $j$-toj svojstvenoj vrijednosti $\theta_j$. Dobivamo na kraju da vrijedi

$$\frac{\sum_{i=1}^{p} \hat{a}_{ij}^2}{tr(S)} = \frac{\theta_j}{tr(S)}. \tag{2.15}$$

Moguće je da su varijable s kojima radimo dane u različitim skalama (mjernim jedinicama) pa se može dogoditi da jedna od njih dominira nad rješenjem faktorske analize te prikrije pravu strukturu podataka. Na primjer, ako imamo varijablu koja mjeri visinu u centimetrima i varijablu koja mjeri težinu u kilogramima dobivamo da će varijabla visine imati puno veću varijancu a i veće težine na faktorima nego što će težine na faktorima imati varijabla težine. U tom slučaju kada varijable nisu proporcionalne možemo koristiti u zamjenu standardizirane varijable i raditi s matricom korelacija $R$. U tom slučaju se koriste svojstvene vrijednosti i svojstveni vektori matrice $R$ umjesto onih matrice $S$ u (2.7) kako bismo dobili procjene težina. U praksi, češće se koristi matrica $R$ nego $S$. Pošto je veći naglasak u faktorskoj analizi na reprodukciji kovarijanci i korelacija nego na varijancama, korištenje $R$ je prikladnije u faktorskoj analizi. Također, u primjeni $R$ daje bolje rezultate nego $S$.
Ukoliko se odlučimo za faktorizaciju matrice $R$, proporcija u (2.15) tada glasi

$$\frac{\sum_{i=1}^{p} \hat{a}_{ij}^2}{tr(R)} = \frac{\theta_j}{p}, \tag{2.16}$$

gdje je $p$ broj varijabli.
Kako bismo provjerili adekvatnost faktorskog modela, dovoljno je da usporedimo lijevu i desnu stranu od (2.10). Matrica grešaka $E$, dana s

$$E = S - (\hat{A}\hat{A}^T + \hat{\Psi}), \tag{2.17}$$

na dijagonali ima nule dok su nedijagonalni elementi različiti od nule.
Sljedeća nejednakost daje gornju ogradu na veličinu elemenata matrice $E$:

$$\sum_{ij} e_{ij}^2 \leq \theta_{m+1}^2 + \theta_{m+2}^2 + ... + \theta_p^2. \tag{2.18}$$

Odnosno, suma kvadrata elemenata matrice $E$ je najviše jednaka sumi kvadrata "zanemarenih" svojstvenih vrijednosti matrice S. Ukoliko su svojstvene vrijednosti malene, reziduali u matrici grešaka $S - (\hat{A}\hat{A}' + \hat{\Psi})$ su maleni i model je dobar.

## 2.2 Metoda glavnih faktora

Prilikom korištenja metode glavnih komponenti u svrhu procjene vrijednosti težina, zanemarili smo $\Psi$ i faktorizirali matrice $S$ ili $R$. *Metoda glavnih faktora* koristi početne procjene

$\hat{\Psi}$ i faktore $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ kako bi dobila

$$S - \hat{\Psi} \cong \hat{A}\hat{A}^T, \tag{2.19}$$

$$R - \hat{\Psi} \cong \hat{A}\hat{A}^T, \tag{2.20}$$

gdje je $\hat{A}$ matrica dimenzije $p \times m$ i računa se kao u (2.7) koristeći svojstvene vrijednosti i svojstvene vektore od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$.
U $S - \hat{\Psi}$, $i$-ti dijagonalni element je dan s $s_{ii} - \hat{\psi}_i$ što je upravo $i$-ti komunalitet, $\hat{h}_i^2 = 1 - \hat{\psi}_i$.
S ovakvim dijagonalnim vrijednostima, $S - \hat{\Psi}$ i $R - \hat{\Psi}$ imaju sljedeće forme

$$S - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & \hat{h}_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & \hat{h}_p^2 \end{bmatrix}, \tag{2.21}$$

$$R - \hat{\Psi} = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \dots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & \hat{h}_p^2 \end{bmatrix}. \tag{2.22}$$

Popularna početna procjena komunaliteta u $R - \hat{\Psi}$ je $\hat{h}_i^2 = R_i^2$, odnosno kvadrat višetruke korelacije izmedu $X_i$ i drugih $p - 1$ varijabli. To možemo još pisati kao

$$\hat{h}_i^2 = R_i^2 = 1 - \frac{1}{r^{ii}}, \tag{2.23}$$

gdje je $r^{ii}$ $i$-ti dijagonalni element od $R^{-1}$.
U slučaju $S - \hat{\Psi}$, početna procjena komunaliteta potpuno analogno kao u (2.23) je

$$\hat{h}_i^2 = s_{ii} - \frac{1}{s^{ii}}, \tag{2.24}$$

gdje je $s_{ii}$ $i$-ti dijagonalni element od $S$, a $s^{ii}$ $i$-ti dijagonalni element od $S^{-1}$. Može se pokazati da se izraz u (2.24) može ekvivalentno zapisati kao

$$\hat{h}_i^2 = s_{ii} - \frac{1}{s^{ii}} = s_{ii} R_i^2. \tag{2.25}$$

Kako bismo koristili (2.23) i (2.24), matrice $R$ ili $S$ ne smiju biti singularne jer kao takve nemaju inverz. Ukoliko je $R$ singularna, možemo koristiti apsolutnu vrijednost ili kvadrat najveće korelacije u $i$-tom retku matrice $R$ kao procjenu komunaliteta.

Nakon što smo odredili procjene komunaliteta, računamo svojstvene vrijednosti i svojstvene vektore od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ i koristimo (2.7) kako bismo dobili procjenu težina faktora, $\hat{A}$. Nakon toga, retci i stupci od $\hat{A}$ mogu se koristiti kako bismo dobili nove svojstvene vrijednosti (objašnjeni dio varijance) i komunalitete.
Suma kvadrata $j$-tog stupca matrice $\hat{A}$ je $j$-ta svojstvena vrijednost od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$, a suma kvadrata $i$-tog retka matrice $\hat{A}$ je komunalitet od $X_i$. Proporcija varijance koja je objašnjena $j$-tim faktorom je

$$\frac{\theta_j}{tr(S - \hat{\Psi})} = \frac{\theta_j}{\sum_{i=1}^{p} \theta_i}, \tag{2.26}$$

ili

$$\frac{\theta_j}{tr(R - \hat{\Psi})} = \frac{\theta_j}{\sum_{i=1}^{p} \theta_i}. \tag{2.27}$$

gdje je $\theta_j$ $j$-ta svojstvena vrijednost od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$. Matrice $S - \hat{\Psi}$ i $R - \hat{\Psi}$ nisu nužno pozitivno semidefinitne i često će imati neke malene negativne svojstvene vrijednosti. U tom slučaju, kumulativna proporcija varijance u (2.26) i (2.27) će preći vrijednost 1 i zatim pasti prema 1 kako se nadodaju negativne svojstvene vrijednosti.

## 2.3 Iterativna metoda glavnih faktora

Metoda glavnih faktora može se jednostavno iterirati kako bi se poboljšale procjene komunaliteta. Nakon što dobijemo $\hat{A}$ iz $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ u (2.19) i (2.20) koristeći početne procjene komunaliteta, možemo dobiti nove procjene komunaliteta pomoću težina u $\hat{A}$ koristeći (2.11),

$$\hat{h_i^2} = \sum_{j=1}^{m} \hat{a_{ij}^2}.$$

Vrijednosti od $\hat{h}_i^2$ se supstituiraju na dijagonalu od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$, iz kojih dobivamo novu vrijednost od $\hat{A}$ koristeći (2.7). Ovaj proces provodimo sve dok procijenjeni komunaliteti ne konvergiraju. Tada se svojstvene vrijednosti i svojstveni vektori finalnog oblika $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ koriste u (2.7) kako bismo dobili težine. Nakon što imamo konvergenciju, koristimo svojstvene vrijednosti i svojstvene vektore od $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ u (2.7) kako bi izračunali težine.
Metoda glavnih faktora i iterativna metoda glavnih faktora će najčešće davati rezultate koji su blizu rezultatima metode glavnih komponenti ukoliko je bilo koja od sljedećih situacija istinita:

1. Korelacije su poprilično velike, što rezultiram malom vrijednosti od $m$.

2. Broj varijabli $p$ je velik.

Jedna od loših strana iterativnog pristupa je ta što nas ponekad vodi k tome da procjene komunaliteta $\hat{h}_i^2$ prelaze vrijednost 1 (prilikom faktorizacije matrice $R$). Takav rezultat je još poznat pod nazivom *Heywood case* (Heywood 1931.). Ukoliko je $\hat{h}_i^2 > 1$, onda po (2.9) i (2.11) slijedi $\hat{\Psi}_i < 0$ što je očito netočno pošto se ne može dogoditi da je specifična varijanca negativna. U takvim situacijama, kada komunalitet prijeđe vrijednost 1, iterativni proces bi trebao stati te bi program trebao javiti da se ne može doći do rješenja. Neki statistički programski paketi imaju opciju nastavka s iteracijama tako što se komunaliteti postave na vrijednoost 1 u svim narednim iteracijama. Finalno rješenje za koje je $\hat{\Psi}_i = 0$ je upitno pošto implicira na egzaktnu ovisnost varijabli o faktorima što je moguć ali iznimno rijedak ishod.

## 2.4 Metoda maksimalne vjerodostojnosti

Ukoliko pretpostavimo da opservacije $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ čine slučajan uzorak iz $N_p(\mu, \Sigma)$, onda $A$ i $\Psi$ mogu biti procijenjeni koristeći metodu maksimalne vjerodostojnosti. Lako se pokaže da procjenitelji $\hat{A}$ i $\hat{\Psi}$ dobiveni ovom metodom zadovoljavaju

$$S\hat{\Psi}\hat{A} = \hat{A}(I + \hat{A}^T\hat{\Psi}^{-1}\hat{A}), \tag{2.28}$$

$$\hat{\Psi} = diag(S - \hat{A}\hat{A}^T), \tag{2.29}$$

$$\hat{A}\hat{\Psi}^{-1}\hat{A} \quad \text{dijagonalna.} \tag{2.30}$$

Navedene jednadžbe moraju biti riješene iterativno i u praksi procedura možda ne iskonvergira ili ishodi rezultatom poznatim kao *Heywood case*. Također, proporcije varijance objašnjene faktorima, kao što je u (2.15) i (2.16), neće nužno biti u padajućem poretku u ovoj metodi kao što je za faktore dobivene metodom glavnih komponenata ili metodom glavnih faktora. Za više detalja pogledati [2].

# Poglavlje 3

# Određivanje broja faktora

Cilj svih istraživanja koja koriste faktorsku analizu je smanjiti veliki broj varijabli na manji broj faktora. Uz odabir broja faktora veže se niz pitanja. Koliko se pouzdanih i interpretabilnih faktora nalazi u promatranom skupu podataka? Da li je dobiven broj faktora pouzdan i postoji li možda još pouzdanih faktora? Uključivanje većeg broja faktora u rješenje poboljšava sličnost između promatrane i reproducirane matrice korelacija, stoga je adekvatnost ekstrakcije vezana uz odabir broja faktora. S druge strane, što je veći broj faktora ekstrahiran to je rješenje slabije. Kako bi objasnili čitavu kovarijancu u skupu podataka trebali bi imati jednak broj faktora kao i broj promatranih varijabli. Dakle, jasno je da je potrebno učiniti kompromis, odnosno želimo zadržati dovoljan broj faktora za adekvatno odgovaranje modela podacima, ali ne i previše. Odabir broja faktora je obično delikatniji od odabira tehnike za ekstrakciju i rotaciju ili vrijednosti komunaliteta. Postoji mnogo kriterija po kojima možemo odabrati parametar $m$, broj faktora u faktorskoj analizi. Proučit ćemo četiri takva kriterija.

1. Odaberimo parametar $m$ jednak broju faktora koji je potreban da postignemo određeni postotak objašnjene varijance, recimo 80% ukupne varijance $tr(S)$ ili $tr(R)$.

2. Odabiremo parametar $m$ tako da je broj svojstvenih vrijednosti veći od prosjeka svojstvenih vrijednosti. Za $R$ prosjek iznosi 1, dok za $S$ iznosi $\frac{\sum_{j=1}^{p} \theta_j}{p}$.

3. Koristeći *scree test* koji je baziran na grafičkom prikazu svojstvenih vrijednosti od $S$ ili $R$. Ukoliko u jednom trenutku vidimo da na grafu dolazi do naglog pada nakon kojeg slijedi ravna linija malenog nagiba, odabiremo $m$ koji je jednak broju svojstvenih vrijednosti koje se nalaze prije ravne linije blagog nagiba.

4. Testiramo hipotezu da je $m$ odgovarajući faktor, odnosno testiramo nultu hipotezu $H_0 : \Sigma = AA^T + \Psi$, gdje je $A$ matrica dimenzija $p \times m$.

Metoda 1 se primjenjuje u metodi glavnih komponenti pa proporcija ukupne uzoračke varijance uzrokovane $j$-tim faktorom od $S$ iznosi $\frac{\sum_{i=1}^{p} \hat{a}_{ij}^2}{tr(S)}$. Odgovarajuća proporcija od $R$ iznosi $\frac{\sum_{i=1}^{p} \hat{a}_{ij}^2}{p}$. Slijedi da ukupan utjecaj $m$ faktora u $tr(S)$ ili $p$ tako iznosi $\sum_{i=1}^{p} \sum_{j=1}^{m} \hat{a}_{ij}^2$, što je zapravo suma kvadrata svih elemenata iz $\hat{A}$. Za metodu glavnih komponenti iz (2.11) i (2.12) vidimo da je ta suma također jednaka sumi prvih $m$ svojstvenih vrijednosti ili sumi svih $p$ komunaliteta:

$$\sum_{i=1}^{p} \sum_{j=1}^{m} \hat{a_{ij}^2} = \sum_{i=1}^{p} \hat{h_i^2} = \sum_{j=1}^{m} \theta_j. \tag{3.1}$$

Zato biramo $m$ dovoljno velik tako da suma komunaliteta ili suma svojstvenih vrijednosti čini relativno veliki dio od $tr(S)$ ili $p$.
Metoda 1 može se proširiti do metode glavnih faktora gdje se onda prethodne procjene komunaliteta koriste u formiranju $S - \hat{\Psi}$ ili $R - \hat{\Psi}$. No, $S - \hat{\Psi}$ ili $R - \hat{\Psi}$ će često imati neke negativne svojstvene vrijednosti. Kako je vrijednost od $m$ u rangu od 1 do $p$, kumulativne proporcije svojstvenih vrijednosti $\frac{\sum_{j=1}^{m} \theta_j}{\sum_{j=1}^{p} \theta_j}$ će prijeći vrijednost od 1, ali zbrajanjem negativnih svojstvenih vrijednosti će se reducirati na vrijednost 1. Zato će postotak od npr. 80% biti dostignut za manju vrijednost od $m$ nego što bi to bilo kod $R$ ili $S$ i bolja strategija bi mogla biti odabir $m$ koji je jednak vrijednosti za koju postotak prvi put prijeđe 100%.
U iterativnoj metodi $m$ se bira prije nego što se provede iteriranje i $\sum_i \hat{h}_i^2$ se dobije nakon iteracije kao $\sum_i \hat{h}_i^2 = tr(S - \hat{\Psi})$. Kako bismo izabrali $m$ prije provedbe iteracija možemo se poslužiti prethodno navedenim metodama ili svojstvenim vrijednostima od $S$ ili $R$, kao što to radimo u metodi glavnih komponenti.
Metoda 2 je popularan kriterij u mnogim statističkim programskim paketima. Iako je heuristički bazirana, ponekad zadovolji u praksi. Predložena je također varijacija metode 2 kada se koristi u slučaju $R - \hat{\Psi}$ u kojoj se $m$ bira tako da je on jednak broju pozitivnih svojstvenih vrijednosti. Nedostatak navedene varijacije je to što ta metoda nerijetko ishodi velikom broju faktora, budući da će suma pozitivih svojstvenih vrijednosti biti veća od sume komunaliteta.
Metoda 3, još zvana i *metoda lakta*, pokazala se vrlo dobrom u praksi.
U metodi 4 želimo testirati hipoteze:

$$\begin{aligned} H_0 &: \Sigma = AA^T + \Psi, \\ H_1 &: \Sigma \neq AA^T + \Psi. \end{aligned} \tag{3.2}$$

Ukoliko $\Sigma$ nema neki specijalan oblik, maksimum funkcije vjerodostojnosti je proporcionalan s

$$|S_n|^{\frac{-n}{2}} e^{\frac{-np}{2}}, \tag{3.3}$$

gdje je $S_n = \frac{n-1}{n}S$. No, ukoliko smo pod nultom hipotezom onda $\Sigma$ mora imati oblik kao u $H_0$ u (3.2). U tom slučaju, maksimum funkcije vjerodostojnosti je proporcionalan s

$$|\hat{\Sigma}|^{-\frac{n}{2}} exp\left(-\tfrac{1}{2}tr\left[\hat{\Sigma}^{-1}\left(\textstyle\sum_{j=1}^{n}(x_j-\overline{x})(x_j-\overline{x})^T\right)\right]\right) = |\hat{A}\hat{A}^T+\hat{\Psi}|^{-\frac{n}{2}} exp\left(-\tfrac{1}{2}n\cdot tr[(\hat{A}\hat{A}^T+\hat{\Psi})^{-1}S_n]\right). \quad (3.4)$$

Koristeći (3.3) i (3.4) dobivamo da je najmanje moguće odstupanje između opaženih i prediktivnh vrijednosti odnosno devijacija dana s

$$\begin{aligned} -2ln\Lambda &= -2ln\left[\frac{\text{maksimizirana vjerodostojnost pod } H_0}{\text{maksimizirana vjerodostojnost}}\right] \\ &= -2ln\left(\frac{|\hat{\Sigma}|}{|S_n|}\right)^{-\frac{n}{2}} + n[tr(\hat{\Sigma}^{-1}S_n) - p], \end{aligned} \quad (3.5)$$

gdje su stupnjevi slobode $\frac{1}{2}[(p-m)^2-p-m]$. Kako vrijedi $tr(\hat{\Sigma}^{-1}S_n)-p=0$, slijedi da je $\hat{\Sigma}=\hat{A}\hat{A}^T+\hat{\Psi}$ procjenitelj maksimalne vjerodostojnosti od $\Sigma = AA^T+\Psi$. Zbog toga dobivamo da (3.5) još možemo zapisati kao

$$-2ln\Lambda = n\cdot ln\left(\frac{|\hat{\Sigma}|}{|S_n|}\right). \quad (3.6)$$

Bartlett [3] je pokazao kako se hi-kvadrat aproksimacija distribucije od $-2ln\Lambda$ može poboljšati na način da $n$ zamjenimo multiplikativnim faktorom $(n-1-\frac{(2p+4m+5)}{6})$. Dakle, koristeći Bartlettov kriterij dobivamo da je testna statistika koju koristimo

$$\left(n-\frac{2p+4m+11}{6}\right)\ln\left(\frac{|\hat{A}\hat{A}^T+\hat{\Psi}|}{|S|}\right), \quad (3.7)$$

koja aproksimativno ima $\chi^2_\nu$ razdiobu kada je nulta hipoteza istinita, gdje je $\nu=\frac{1}{2}[(p-m)^2-p-m]$, a $\hat{A}$ i $\hat{\Psi}$ procjenitelji maksimalne vjerodostojnosti. Odbacivanjem $H_0$ impliciramo da je parametar $m$ premalen i da je potrebno više faktora.

U praksi, kada je $n$ velik, pokazalo se da metoda 4 nerijetko pokazuje da je više faktora signifikantno nego što to pokažu ostale tri diskutirane metode. Zbog toga vrijednost $m$ koju dobijemo pomoću metode 4 možemo promatrati kao gornju među za stvarni broj potrebnih faktora prilikom praktične primjene.

Nakon što smo skup podataka uspješno opisali modelom faktorske analize, prve tri navedene metode će gotovo uvijek dati istu vrijednost $m$ pa nekako nećemo niti proispitivati da li je ta vrijednost $m$ validna ili ne.

# Poglavlje 4

# Rotacije faktora

U faktorskoj analizi, inicijalni skup težina je samo jedan od beskonačno mnogo mogućih rješenja koja mogu jednako opisati skup podataka. Ponekad je inicijalni skup težina teško interpretirati pošto svaki faktor može sadržavati popriličan broj težina za mnoge varijable i to nam otežava preciznije određivanje faktora. Idealno bi bilo kada bismo mogli reći da su određene varijable jako korelirane s određenim faktorom dok ostale gotovo da nisu korelirane s istim faktorom. Kada bismo imali jak kontrast između velike i malene vrijednosti težine faktora, onda bi naše zaključivanje bilo olakšano. *Rotacije faktora* se upravo bave tim problemom. Sa željom da se skup težina faktora minimizira i maksimizira, rotacije za cilj imaju producirati ograničen broj velikih težina i velik broj malih vrijednosti težina za svaki faktor. Takva kombinacija vrijednosti težina za određen faktor nam pomažu da što jednostavnije uočimo varijable koje imaju jaču korelaciju s dotičnim faktorom odnosno velik broj varijabli koje uopće nisu korelirane s tim istim faktorom.
Faktorska analiza krene s računanjem težina za određene faktore gdje primjenom jedne od opisanih metoda u poglavlju 2, matrica $\hat{A}$ je matrica težina faktora. Iz prvog poglavlja znamo da su težine faktora iz populacijskog modela jedinstvene do na množenje ortogonalnom matricom $T$. Promatrali smo matricu težina $\hat{A}$ te rotiranu matricu težina $\hat{A}^* = \hat{A}T$. Statističke karakteristike prilikom rotacije su očuvane, tako nove težine iz $\hat{A}^*$ imaju ista svojstva kao i težine iz matrice $\hat{A}$. Kao što smo i vidjeli, rotirana matrica težina $\hat{A}^*$ i $\hat{A}$ dovode do iste procjene kovarijacijske matrice:

$$S \cong \hat{A}^*(\hat{A}^*)^T + \hat{\psi} = \hat{A}TT^T\hat{A}^T + \hat{\psi} = \hat{A}(\hat{A})^T + \hat{\psi}. \tag{4.1}$$

Geometrijski, težine u i-tom redu matrice $\hat{A}$ sudjeluju u kreiranju koordinata točke u prostoru težina koja odgovara $X_i$. Rotacijom $p$ točaka ostaju nam iste koordinate točaka samo u odnosu na nove koordinatne osi, a sve ostale geometrijske karakteristike istih točaka ostaju nepromijenjene.
Cilj je postaviti koordinatne osi što je bliže moguće što većem broju točaka. Ukoliko

vidimo da točke tvore očite nakupine u koordinatnom sustavu, željet ćemo pomicati koordinatne osi tako da osi prolaze kroz ili jako blizu tih nakupina. Na takav način, svaka ta nakupina varijabli pridružuje se nekom faktoru koji je u grafičkom slučaju reprezentiran koordinatnom osi. Kao što smo i komentirali na početku poglavlja, ukoliko možemo postići rotaciju u kojoj je svaka od točaka blizu jedne od koordinatne osi onda svaka od varijabli će imati veliku težinu upravo na faktoru koji odgovara koordinatnoj osi odnosno malene težine na ostalim faktorima.
Promatrat ćemo dva tipa rotacija: ortogonalne i kose rotacije. Gore spomenuta rotacija u kojoj se primjenjuje ortogonalna matrica je ortogonalna rotacija. Originalne okomite osi se rotiraju i ostaju okomite i nakon rotacije. U ortogonalnoj rotaciji kutevi i udaljenosti ostaju sačuvani te komunaliteti nepromijenjeni. S druge strane, u kosoj rotaciji nema zahtjeva da osi moraju nakon rotacije ostati okomite te su time slobodne prolaziti bliže nakupinama točaka.

## 4.1 Ortogonalna rotacija

Ortogonalna rotacija je vrsta rotacije u kojoj okomite koordinatne osi i nakon rotacije ostaju okomite. Također, ortogonalne rotacije ne mijenjaju komunalitete. Naime, ako je $\hat{A}$ matrica težina faktora procijenjena jednom od metoda u drugom poglavlju, slijedi

$$(\hat{A})(\hat{A})^T + \hat{\psi} = (\hat{A})TT^T(\hat{A}) + \hat{\psi} = (\hat{A}^*)(\hat{A}^*)^T + \hat{\psi}.$$

Matrica reziduala ostaje nepromijenjena a s time i specifične varijance $\psi_i$ i komunalitet. U ortogonalnoj rotaciji faktori su nekorelirani, a rješenja dobivena tom rotacijom su jednostavna za interpretaciju ali ne odražavaju pravu stvarnost osim ako smo sigurni da su latentni procesi gotovo nezavisni. Za više detalja pogledati u [2]. U narednim poglavljima promatrat ćemo dva pristupa ortgonalnoj rotaciji: grafička metoda i varimax metoda.

### 4.1.1 Grafička metoda

Ukoliko radimo sa slučajem kada $m = 2$, odnosno razmatramo dva faktora, onda se često transformacija na jednostavnijoj strukturi provodi grafički. U ovom slučaju, retci matrice težina faktora $\hat{A}$ su parovi težina. Svaki od uređenih parova $(\hat{a}_{i,1}, \hat{a}_{i,2})$ za $i \in 1, ..., p$ određuje $p$ točaka u koordinatnom sustavu, a svaka točka korespondira varijabli $X_1, ..., X_p$. Koordinatne osi nakon toga mogu biti rotirane za kut $\Phi$ i nove rotirane težine $\hat{a}^*_{i,j}$ dobivamo iz

$$\hat{A}^* = \hat{A}T,$$

gdje vrijedi

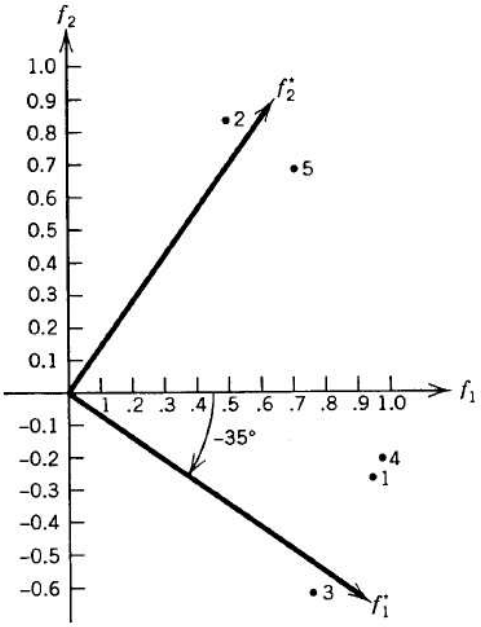
$$T = \begin{bmatrix} \cos\Phi & \sin\Phi \\ -\sin\Phi & \cos\Phi \end{bmatrix}, \tag{4.2}$$

za rotaciju u smjeru kazaljke na satu, odnosno

$$T = \begin{bmatrix} \cos\Phi & -\sin\Phi \\ \sin\Phi & \cos\Phi \end{bmatrix}, \tag{4.3}$$

za rotaciju suprotnu od smjera kazaljke na satu.

**Primjer 4.1.1.** Na sljedećoj slici prikazano je pet parova težina $(\hat{a}_{i1}, \hat{a}_{i2})$ koje odgovaraju nekih pet varijabli. Ortogonalna rotacija za kut $\Phi = -35°$ rotira osi, u našem slučaju faktore, bliže dvama nakupinama točaka (varijabli). Nakon rotacije obje nakupine varijabli odgovaraju puno više novim faktorima.




Slika 4.1: Ortogonalna rotacija

U slučaju $m > 2$ nije više lako vizualizirati orijentaciju pa se okrećemo analitičkoj metodi kako bi pronašli smislenu interpretaciju originalnih podataka.

### 4.1.2 Varimax metoda

Nedostatak grafičke metode je što je ograničen na analizu dva faktora odnosno slučajeve kada je $m = 2$. U slučaju kada je $m > 2$ privrženiji smo korištenju neke od analitičkih metoda. Neke od analitičkih metoda su *varimax*, *quartimax* i *equimax*. Mi ćemo promotriti

varimax kriterij.
Definiramo $\tilde{a}^*_{ij} = \tilde{a}^*_{ij}/\hat{h}_i$ rotirane koeficijente težine koji su skalirani drugim korijenom od komunaliteta. Tada varimax procedura odabire ortogonalnu transformaciju $T$ koja će maksimizirati

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{a}^{*\,4}_{i,j} - \frac{\left(\sum_{i=1}^{p}\tilde{a}^{*\,2}_{i,j}\right)^2}{p}\right]. \tag{4.4}$$

Skaliranje rotiranih koeficijenata $\tilde{a}^*_{ij}$ ima efekt davanja varijablama s malenim komunalitetom veću težinu u svrhu jednostavnije strukture. Nakon što odaberemo transformaciju $T$, težine $\tilde{a}^*_{i,j}$ su pomnožene s $\hat{h}_i$ kako bi se očuvali originalni komunaliteti. Efektivno, maksimiziranje $V$ korespondira raspršivanju kvadrata težina na svaki od faktora što je više moguće. Odnosno, ona traži rotirane težine koje maksimiziraju varijancu kvadrata težina u svakom stupcu matrice $\hat{A}^*$. Zato i očekujemo grupiranje koeficijenata u svakom stupcu matrice težina $\hat{A}^*$.
Varimax tehnika ne garantira da će sve varijable imati veliku težinu na samo jednom faktoru. Zapravo, niti jedna metoda ovo ne može postići za sve moguće podatkovne skupove. Konfiguracija točaka u prostoru težina ostaje fiksirana, sve što mi radimo je rotiranje osi (faktora) kako bi bili bliže točkama što je više moguće. U mnogim slučajevima točke nisu u lijepim nakupinama te je nemoguće zarotirati osi da budu blizu svim točkama. Ovaj je problem sakriven u odabiru $m$. Ako je $m$ promijenjen, koordinate se mijenjaju te se time mijenjaju i relativne pozicije točaka. Postoje računalni algoritmi koji maksimiziraju $V$ ali i popularni računalni programi koji provode faktorsku analizu (SAS, SPSS i MINITAB). Nakon primjene metode izlaz koji dobijemo u većini statističkih programa je rotirana matrica težina $\hat{A}^*$, udio objašnjene varijabilnosti (suma kvadrata pojedinih stupaca matrice $\hat{A}^*$), komunalitete (suma kvadrata pojedinih redaka matrice $\hat{A}^*$), te ortogonalna matrica $T$ koja služi kako bismo dobili $\hat{A}^*$, odsnosno $\hat{A}^* = \hat{A}T$.

## 4.2 Kosa rotacija

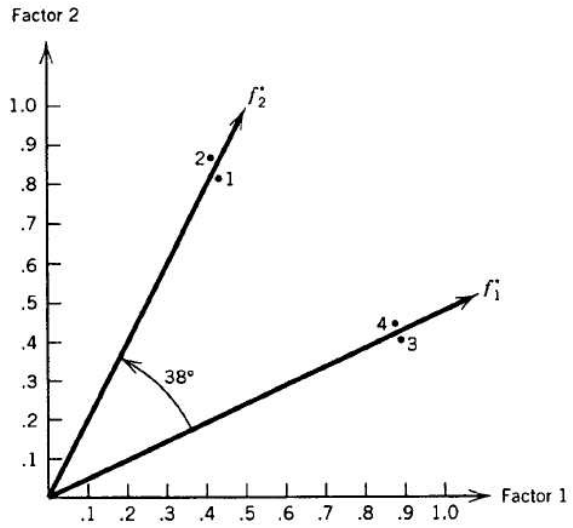
Kose rotacije u terminima faktorske analize su one rotacije u kojima osi ne ostaju okomite. S time slijedi da niti kutevi ni udaljenosti nisu sačuvane, a kao posljedica niti komunalitet. Naime, umjesto ortogonalne matrice koju smo koristili u slučaju ortogonalne rotacije, kosa rotacija koristi nesingularnu matricu $Q$. Prilikom izračuna $F^* = Q^T F$ dobivamo

$$Cov(F^*) = Q^T I Q = Q^T Q \neq I. \tag{4.5}$$

Dakle, novi faktori su korelirani. Zbog činjenice da udaljenosti i kutevi nisu očuvani vrijedit će da niti komunaliteti od $F^*$ nisu jednaki komunalitetima od $F$.
Ukoliko promatramo $m$ faktora kao koordinatne osi, točka $(\hat{a}_{i,1}, \hat{a}_{i,2}, ..., \hat{a}_{i,m})$ predstavlja poziciju $i$-te varijable u faktorskom prostoru. Uz pretpostavku da su varijable grupirane u

disjunktne nakupine, kosa rotacija korespondira proizvoljnim rotacijama koordinatnog sustava tako da koordinatne osi (koje više nisu okomite) prolaze kroz ili u blizini navedenih nakupina.

**Primjer 4.2.1.** Na sljedećoj slici prikazana su četiri para težina $(\hat{a}_{i,1}, \hat{a}_{i,2})$ koje odgovaraju nekim četirima varijablama. Kose koordinatne osi koje se nalaze pod kutem od $\Phi = 38°$ će proći puno bliže točkama nego što bi to originalne koordinatne osi. Nove kose koordinatne osi bi tako rezultirale da su navedene težine po apsolutnoj vrijednosti puno bliže vrijednostima 0 (u slučaju malene korelacije varijable s faktorom) ili 1(u slučaju jake korelacije varijable s faktorom).




Slika 4.2: Kosa rotacija

U praksi kose rotacije daju slične rezultate kao i ortogonalne ako su faktori nekorelirani. No, ako mi primijenimo ortogonalnu rotaciju na koreliranim faktorima to će vidno utjecati na naše rezultate. Postoje mnoge analitičke metode za postizanje kosih rotacija. Izlaz koji se dobije kao rezultat tih metoda su najčešće: *matrica uzoraka, matrica strukture* i *matrica korelacija medu kosim faktorima* (osima). Za interpretaciju najčešće se koristi matrica uzoraka. Težine u retcima matrice uzoraka su prirodne koordinate točaka (varijabli) na kosim osima i služe kao koeficijenti u modelu koji povezuje varijable s faktorima.
Jedna od koristi kose rotacije je da provjerava ortogonalnost faktora. Ortogonalnost u originalnim faktorima je nametnuta modelom i održavana ortogonalnim rotacijama. Ako kosa rotacija rezultira matricom korelacije koja je gotovo dijagonalna, možemo biti i više nego sigurni da su faktori ortogonalni. Za više detalja pogledati u [2].

## 4.3 Interpretacija faktora

U ovom dijelu ćemo ukratko objasniti kako interpretirati faktore na temelju matrice težina faktora.
Kao što smo i diskutirali na početku, cilj faktorske analize i same primjene rotacija je da postignemo što jednostavniju strukturu. Dakle, željeli bismo da svaka varijabla ima veliku težinu na samo jednom faktoru dok će na ostalima imati malene težine. To je često teško postići pa tu i priskačemo primjeni rotacija koje nam pomažu da dobijemo težine koje su bliže u vrijednostima težinama koje bismo željeli u jednostavnoj strukturi.
Nakon što smo primjenom svih prethodnih koraka dobili matricu težina faktora, na nama je da interpretiramo faktore. Za svaku od $p$ varijabli, praksa je da istaknemo težine koje su signifikantne na način da ih podcrtamo odnosno zaokružimo, a ako je moguće one koje nisu signifikantne odstranimo. Postoje dva načina na koje možemo mjeriti značajnost: praktična značajnost i statistička značajnost.
Praktična značajnost provjerava da li je težina faktora dovoljno velika u smislu da bi imala značajan utjecaj na varijable. Hair [5] predlaže sljedeće smjernice prilikom određivanja praktične značajnosti:

$\pm$ 0.3 minimalna značajnost;
$\pm$ 0.4 osrednja značajnost;
$\pm$ 0.5 praktična značajnost.

Statistička značajnost također teži da su težine faktora značajno različite od 0. Stevens [6] predlaže sljedeću skalu statističke značajnosti s obzirom na veličinu uzorka:

| n | težina |
|---|---|
| 50 | 0.722 |
| 100 | 0.512 |
| 200 | 0.384 |
| 300 | 0.298 |
| 600 | 0.210 |
| 1000 | 0.162 |

Tablica 4.1: Statistička značajnost težine u odnosu ne veličinu uzorka.

Ako radimo sa skupom podataka veličine 50, koristeći tablicu 4.1, uzimat ćemo da su značajne težine upravo one koje su po apsolutnoj vrijednosti veće od 0.72 za određen faktor. Potpuno analogno, ako radimo sa skupom podataka veličine 600 onda će značajna težina za određen faktor biti upravo ona koja ima apsolutnu vrijednostu veću od 0.21.

Razmotrit ćemo neke generalne smjernice za interpretaciju faktora ispitivajući matricu rotiranih težina. U svakom retku matrice težina, krenuvši slijeva te se kretajući nadesno detektiramo, po apsolutnoj vrijednosti, najveću vrijednost, odnosno težinu. Ukoliko je najveća težina značajne veličine zaokružimo ju. Ovo radimo za svaku od $p$ varijabli. Ukoliko u jednom retku ima više značajnih težina koje će biti uzete u obzir, interpretacija postaje manje jednostavna. S druge strane moguće je i da postoje varijable s malim komunalitetima da se ne pojavljuje niti jedna značajna težina na nekom faktoru. U ovom slučaju bismo povećali broj faktora i ponovno pokrenuli program kako bi se takve varijable pridružile nekom novom faktoru. Da bi procjenili značajnost težine faktora $\hat{a}_{i,j}$ najčešće se uzima vrijednost 0.3 kao donja granica. No, za uspješniju primjenu, kritična vrijednost od 0.3 je premalena i rezultirat će u varijablama čija će kompleksnost biti veća od 1. Zbog navedenog, ciljane vrijednosti od 0.5 i 0.6 su većinom korisnije. Nakon što smo identificirali potencijalno značajne težine faktora za svaku od varijabli, nastojimo otkriti značenje faktora te ih u idealnim slučajevima imenovati. U mnogim situacijama, gdje grupiranja nisu logična, analizu možemo provesti ponovno, mijenjajući $m$, prilagođavajući razinu značajnosti težina, koristeći neku drugu metodu za procjenu težina ili korištenjem druge vrste rotacije.

# Poglavlje 5

# Faktorski score-ovi

U faktorskoj analizi cilj istraživanja je najčešće provjera da li dobiveni model faktorske analize odgovara danim podatcima i identificiranje faktora. No, postoje primjene u kojima želimo odrediti procijenjene vrijednosti zajedničkih faktora, odnosno *faktorske score-ove* u oznaci $\hat{F}_i = (\hat{F}_{i1}, \hat{F}_{i2}, ..., \hat{F}_{im})$, $i = 1, 2, ..., n$. Faktorske score-ove definiramo kao procjene od pozadinskih faktorskih vrijednosti $F_i$, $i = 1, 2, ..., n$ za svaku opservaciju. Postoje dvije primjene za definirane faktorske score-ove:

1. Želimo analizirati ponašanje opservacija u terminima faktora.
2. Želimo iskoristiti faktorske score-ove za neku drugu analizu, kao na primjer multivarijatnu analizu varijance.

Kako $F$-ovi nisu izmjereni moramo ih procijeniti kao funkcije opaženih $X$-eva. Analizirat ćemo metodu procjene faktora na bazi regresije. Budući da je $E[F_i] = 0$, $i = 1, 2, ..., n$, povezat ćemo $F$-ove i $X$-ove pomoću centralnog regresijskog modela

$$\begin{aligned} F_1 &= \beta_{11}(X_1 - \overline{X}_1) + \beta_{12}(X_2 - \overline{X}_2) + ... + \beta_{1p}(X_p - \overline{X}_p) + \varepsilon_1 \\ F_2 &= \beta_{21}(X_1 - \overline{X}_1) + \beta_{22}(X_2 - \overline{X}_2) + ... + \beta_{2p}(X_p - \overline{X}_p) + \varepsilon_2 \\ &\vdots \\ F_m &= \beta_{m1}(X_1 - \overline{X}_1) + \beta_{m2}(X_2 - \overline{X}_2) + ... + \beta_{mp}(X_p - \overline{X}_p) + \varepsilon_m, \end{aligned} \tag{5.1}$$

koji može biti zapisan u matričnom obliku

$$F = B_1^T(X - \overline{X}) + \epsilon. \tag{5.2}$$

Važno je uvidjeti da greške $\epsilon$ u (5.2) nisu jednake greškama $\varepsilon$ u modelu (1.2).
Naš pristup će biti prvo procjeniti $B_1$ pa iskoristiti predviđenu vrijednost $\hat{F} = \hat{B_1^T}(X - \overline{X})$

kako bismo procijenili $F$.
Model (5.2) po komponentama ima oblik

$$F_i = B_1^T(X_i - \overline{X}_i) + \epsilon_i, \quad i = 1, 2, ..., n. \tag{5.3}$$

Transponiranjem komponente modela (5.2) postaju

$$F_i^T = (X_i - \overline{X}_i)^T B_1 + \epsilon_i^T, \quad i = 1, 2, ..., n \tag{5.4}$$

odnosno ovih $n$ jednadžbi možemo ukomponirati u jedan model

$$\begin{aligned} F = \begin{bmatrix} F_1^T \\ F_2^T \\ \vdots \\ F_n^T \end{bmatrix} &= \begin{bmatrix} (X_1 - \overline{X}_1)^T B_1 \\ (X_2 - \overline{X}_2)^T B_1 \\ \vdots \\ (X_n - \overline{X}_n)^T B_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1^T \\ \epsilon_2^T \\ \vdots \\ \epsilon_n^T \end{bmatrix} \\ &= \begin{bmatrix} (X_1 - \overline{X}_1)^T \\ (X_2 - \overline{X}_2)^T \\ \vdots \\ (X_n - \overline{X}_n)^T \end{bmatrix} B_1 + \Xi \\ &= X_c B_1 + \Xi \end{aligned} \tag{5.5}$$

Model (5.5) izgleda kao model centrirane multivarijatne višestruke regresije. Procjenitelj za $B_1$ bi bio

$$\hat{B}_1 = (X_C^T X_C)^{-1} X_C^T F. \tag{5.6}$$

No, $F$ nije opservirana. Kako bismo odredili $\hat{B}_1$ usprkos tome, prvo ćemo model (5.6) zapisati u terminima kovarijacijskih matrica

$$\hat{B}_1 = S_{XX}^{-1} S_{XF} \tag{5.7}$$

gdje je $S_{XX}$ reprezentirano sa $S$ a $S_{XF}$ s $\hat{A}$ pošto $\hat{A}$ procjenjuje $Cov(X, F) = A$. Na temelju pretpostavki za model faktorske analize koje smo komentirali u prvom poglavlju slijedi da možemo (5.7) zapisati kao

$$\hat{B}_1 = S^{-1}\hat{A}. \tag{5.8}$$

Primjenom (5.5) slijedi da su procijenjene vrijednosti od $F$

$$\hat{F} = \begin{bmatrix} \hat{F}_1^T \\ \hat{F}_2^T \\ \vdots \\ \hat{F}_n^T \end{bmatrix} = X_C \hat{B}_1 = X_C S^{-1} \hat{A}. \tag{5.9}$$

Na isti način, ukoliko umjesto $S$ faktoriziramo $R$ prethodne jednadžbe postaju

$$\hat{B}_1 = R^{-1}\hat{A},$$
$$\hat{F} = X_S R^{-1}\hat{A},$$

gdje je $X_S$ matrica standardnih varijabli $\frac{X_{ij}-\overline{X_j}}{s_j}$. Uobičajeno je da se faktorski score-ovi računaju za rotirane a ne originalno dobivene faktore pa zato sve $\hat{A}$ u gornjim izračunima i jednadžbama možemo zamijeniti s $\hat{A}^*$.
Kako bi izračunali faktorske score-ove zahtjev je da matrice $S$ ili $R$, ovisno koju koristimo, nisu singularne. Ukoliko je $S$ ili $R$ singularna možemo izračunati faktorske score-ove primijenjujući jednostavnu metodu direktno na rotirane težine. Grupiramo varijable u grupe (faktore) sukladno težinama i pronađemo score za svaki faktor uprosječavanjem varijabli koje su dodijeljene tom faktoru. Ukoliko varijable nisu razmjerne trebale bi biti standardizirane prije uprosječavanja. Za više detalja pogledati u [2].

# Poglavlje 6

# Valjanost modela faktorske analize

Za mnoge statističare faktorska analiza nije legitimna multivarijatna metoda. Razlozi ove odbojnosti su mnogi: težina odabira broja faktora $m$, mnoge metode za ekstrahiranje faktora, previše metoda rotacije ali i prevelika subjektivnost prilikom interpretacija podataka i rezultata. Zapravo je mogućnost rotacije ta koja daje faktorskoj analizi nekakvu korist.
Osnovno je pitanje da li faktori uopće postoje. Model za kovarijacijsku matricu glasi $\Sigma = AA^T + \Psi$ pri čemu je $AA^T$ ranga $m$. Problem je što mnoge populacije nemaju ovakav uzorak u pogledu kovarijacijske matrice osim ako $m$ nije dovoljno velik. U takvim populacijama model onda neće odgovarati podatcima kada pokušamo nametnuti malu vrijednost za $m$. No, ako i imamo populacije u kojima je $\Sigma$ dovoljno blizu $AA^T + \Psi$ i za malu vrijednost od $m$, procedura uzrokovanja koja nas dovodi do matrice $S$ može narušiti taj uzorak. U mnogim slučajevima temeljni je problem to što $S$ ili $R$ sadrže model i grešku, a koraci faktorske analize nisu u mogućnosti odvojiti navedeno.
Prilikom provedbe faktorske analize moramo biti strpljivi u samoj interpretaciji, ukoliko se pronađeni faktori podudaraju s podatcima na nama je da provodimo analizu sve dok zaista ne potvrdimo postojanje faktora. Ukoliko pri ponovnom uzrokovanju iz iste populacija nailazimo na iste faktore, možemo biti sigurni da smo primjenom modela zaista otkrili neke od pravih faktora. Dakle, poželjno je da u praksi ponovimo svaki eksperiment kako bismo dodatno provjerili i na takav način bili sigurni u stabilnost faktora.
Jedna od mogućnosti je i da skup podataka, ukoliko je velik, prepolovimo i primijenimo faktorsku analizu na svaku od navedenih polovica. Dobivene rezultate, u ovom slučaju dva rezultata, zatim možemo međusobno usporediti ali i s rezultatom koji dobijemo nakon primjene faktorske analize na čitavom podatkovnom skupu.
Postoje razne preporuke čijom primjenom bi se olakšala faktorska analiza i osigurao model koji više odgovara podatcima. Jedan od prijedloga je da matrica $R^{-1}$ mora biti približno dijagonalna kako bismo dobili model faktorske analize koji uspješno odgovara podatcima. Kako bismo odredili koliko je $R^{-1}$ blizu dijagonalne matrice, Kaiser [4] predlaže

korištenje *mjere uzoračke adekvatnosti* (MSA) koja se još naziva Kaiser-Meyer-Olkinova mjera (KMO):

$$MSA = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} q_{ij}^2}, \tag{6.1}$$

gdje su $r_{ij}^2$ kvadrirani elementi matrice $R$, a $q_{ij}^2$ kvadrirani elementi matrice $Q$ gdje

$$Q = DR^{-1}D \quad \text{i} \quad D = [(diagR^{-1})^{\frac{1}{2}}]^{-1}. \tag{6.2}$$

Kako se matrica $R^{-1}$ približava dijagonalnoj matrici to se veličina MSA približava vrijednosti 1. Kaiser i Rice (1974) predlažu da bi vrijednost MSA trebala iznositi barem 0.8 kako bismo mogli očekivai zadovoljavajuće rezultate.
Na kraju, postoje i mnoge podatkovne skupine na kojima se faktorska analiza ne bi trebala primjenjivat. Jedan pokazatelj da matrica $R$ nije prikladna za faktorizaciju je neuspjeh metoda u Poglavlju 3 prilikom jasnog i objektivnog odabira vrijednosti $m$. Ukoliko scree graf nema jasno izražen pregib ili svojstvene vrijednosti nemaju veliku udaljenost od 1, onda je vrlo vjerojatno da $R$ nije prikladan za faktorizaciju. Dodatno, procjene komunaliteta bi nakon faktorizacije trebale biti dovoljno velike.

# Poglavlje 7

# Primjena faktorske analize

U ovom poglavlju provest ćemo faktorsku analizu nad podatcima. Prilikom analize određivat ćemo korelacijske matrice i određivati broj potrebnih faktora koristeći različite metode te iste brojeve onda uspoređivati. Nakon što odredimo s kojim brojem faktora radimo, analizirat ćemo originalni model težina s rotiranim modelom koji ćemo dobiti primjenom varimax i promax metoda. Na temelju dobivenih modela analiziramo grupiranje varijabli oko faktora te interpretiramo značenje tih istih faktora.

## 7.1 Primjer 1-zadovoljstvo aviokompanijom

Podatci nad kojima ćemo primijeniti faktorsku analizu odnose se na istraživanja zadovoljstva putnika aviokompanije koji su preuzeti s [8].
Raspolažemo s 25976 podataka, odnosno, 25976 osoba je ocijenilo odredene usluge koje nudi aviokompanija brojevima od 0 do 5. Navedene usluge su varijable koje će sudjelovati u faktorskoj analizi:

| Varijabla | Opis varijable |
|---|---|
| $X1$ | Udobnost sjedala |
| $X2$ | Točnost vremena polaska/odlaska |
| $X3$ | Hrana i piće |
| $X4$ | Lokacija gate-a |
| $X5$ | WIFI usluga tokom leta |
| $X6$ | Animacije tokom leta |
| $X7$ | Online podrška |
| $X8$ | Lakoća online rezerviranja |
| $X9$ | Usluga tokom leta |
| $X10$ | Usluga prostora za noge |
| $X11$ | Rukovanje prtljagom |
| $X12$ | Čistoća |
| $X13$ | Online ukrcavanja |

Tablica 7.1: Varijable koje sudjeluju u faktorskoj analizi

Cilj provedbe faktorske analize je vidjeti možemo li 13 ulaznih varijabli sažeti u manji broj latentnih varijabli, odnosno faktora. Faktorsku analiza ćemo provesti koristeći RStudio. Za početak, u sljedećoj tablici prikazana je deskriptivna statistika varijabli X1-X13 s kojima ćemo raditi.

| varijabla | mean | sd | median | trimmed | min | max | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|
| X1 | 2.84 | 1.39 | 3 | 2.84 | 0 | 5 | -0.09 | -0.94 |
| X2 | 2.99 | 1.53 | 3 | 3.05 | 0 | 5 | -0.25 | -1.09 |
| X3 | 2.85 | 1.44 | 3 | 2.87 | 0 | 5 | -0.12 | -0.99 |
| X4 | 2.99 | 1.31 | 3 | 2.99 | 0 | 5 | -0.05 | -1.09 |
| X5 | 3.25 | 1.32 | 3 | 3.31 | 0 | 5 | -0.19 | -1.12 |
| X6 | 3.38 | 1.35 | 4 | 3.51 | 0 | 5 | -0.6 | -0.53 |
| X7 | 3.52 | 1.31 | 4 | 3.65 | 0 | 5 | -0.58 | -0.81 |
| X8 | 3.47 | 1.31 | 4 | 3.59 | 0 | 5 | -0.49 | -0.91 |
| X9 | 3.47 | 1.27 | 4 | 3.58 | 0 | 5 | -0.51 | -0.79 |
| X10 | 3.49 | 1.29 | 4 | 3.6 | 0 | 5 | -0.5 | -0.84 |
| X11 | 3.7 | 1.16 | 4 | 3.82 | 1 | 5 | -0.74 | -0.24 |
| X12 | 3.71 | 1.15 | 4 | 3.83 | 0 | 5 | -0.76 | -0.21 |
| X13 | 3.35 | 1.3 | 4 | 3.44 | 0 | 5 | -0.37 | -0.94 |

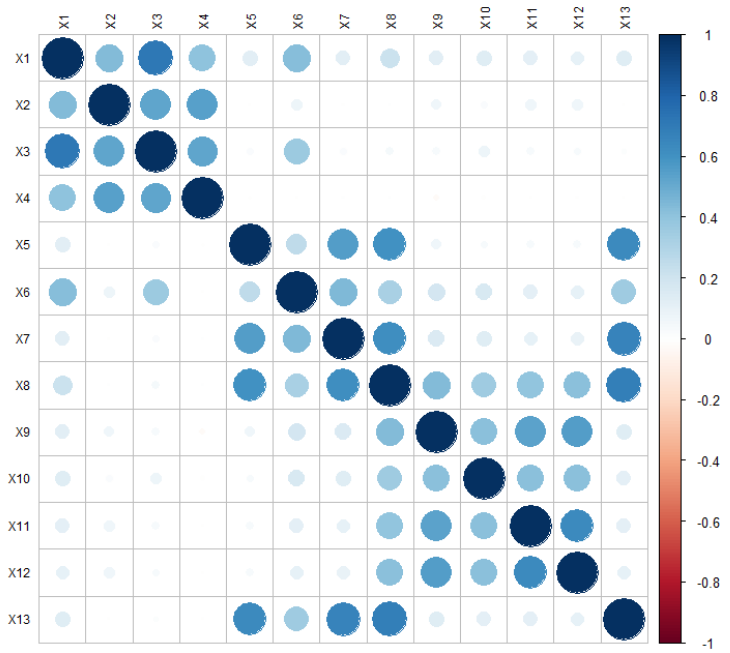
Tablica 7.2: Deskriptivna statistika varijabli

Matrica korelacija varijabli je:

| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1 | 0.43 | 0.72 | 0.41 | 0.13 | 0.43 | 0.12 | 0.21 | 0.12 | 0.14 | 0.12 | 0.11 | 0.13 |
| X2 | 0.43 | 1 | 0.53 | 0.54 | 0 | 0.08 | 0 | 0 | 0.06 | 0.03 | 0.07 | 0.07 | 0 |
| X3 | 0.72 | 0.53 | 1 | 0.52 | 0.03 | 0.37 | 0.03 | 0.04 | 0.04 | 0.07 | 0.04 | 0.03 | 0.01 |
| X4 | 0.41 | 0.54 | 0.52 | 1 | 0 | 0 | 0 | 0 | -0.03 | -0.01 | 0 | 0 | 0 |
| X5 | 0.13 | 0 | 0.03 | 0 | 1 | 0.25 | 0.56 | 0.6 | 0.06 | 0.03 | 0.04 | 0.04 | 0.63 |
| X6 | 0.43 | 0.08 | 0.37 | 0 | 0.25 | 1 | 0.44 | 0.32 | 0.18 | 0.16 | 0.12 | 0.11 | 0.36 |
| X7 | 0.12 | 0 | 0.03 | 0 | 0.56 | 0.44 | 1 | 0.62 | 0.16 | 0.14 | 0.1 | 0.1 | 0.67 |
| X8 | 0.21 | 0 | 0.04 | 0 | 0.6 | 0.32 | 0.62 | 1 | 0.44 | 0.36 | 0.4 | 0.42 | 0.68 |
| X9 | 0.12 | 0.06 | 0.04 | -0.03 | 0.06 | 0.18 | 0.16 | 0.44 | 1 | 0.41 | 0.53 | 0.55 | 0.14 |
| X10 | 0.14 | 0.03 | 0.07 | -0.01 | 0.03 | 0.16 | 0.14 | 0.36 | 0.41 | 1 | 0.41 | 0.41 | 0.11 |
| X11 | 0.12 | 0.07 | 0.04 | 0 | 0.04 | 0.12 | 0.1 | 0.4 | 0.53 | 0.41 | 1 | 0.63 | 0.11 |
| X12 | 0.11 | 0.07 | 0.03 | 0 | 0.04 | 0.11 | 0.1 | 0.42 | 0.55 | 0.41 | 0.63 | 1 | 0.11 |
| X13 | 0.13 | 0 | 0.01 | 0 | 0.63 | 0.36 | 0.67 | 0.68 | 0.14 | 0.11 | 0.11 | 0.11 | 1 |

Tablica 7.3: Matrica korelacija varijabli X1-X13

U korelacijskoj matrici vidimo da izmedu nekih varijabli imamo jake korelacije, ali i jako male korelacije. Vidimo da su slijedeći parovi varijabli jako korelirani: X1 i X3, X1 i X2, X1 i X6, X4 i X1, X4 i X2, X4 i X3, X5 i X7, X5 i X13, X6 i X7, X7 i X8, X7 i X13, X8 i X12, X8 i X13, X9 i X11, X9 i X12, X10 i X11, X10 i X12, X11 i X12. Budući da je dovoljan broj jako koreliranih varijabli zaključujemo da su podatci pogodni za provođenje faktorske analize. Promotrimo dodatno vizualnu reprezentaciju matrice korelacija varijabli na kojoj možemo uočiti grupacije varijabli s obzirom na jačinu koreliranosti.




Slika 7.1: Vizualna reprezentacija matrice korelacija

Kao što smo prethodno uočili jake korelacije u tablici korelacija, tako iste možemo sada uočiti na vizualnoj reprezentaciji matrice korelacija. Tamnija boja naglašava da je korelacija među navedenim varijablama jača. Vidimo tako da se stvaraju 3 grupice unutar kojih su varijable koje su pozitivno korelirane, to nas navodi na mišljenje kako će zajednički faktori zaista i postojati.
Slijedeći korak je analiza vrijednosti MSA, analiza svojstvenih vrijednosti matrice korelacija te odabir broja faktora pomoću istih.

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.76 | 0.79 | 0.71 | 0.73 | 0.84 | 0.74 | 0.83 | 0.79 | 0.85 | 0.89 | 0.82 | 0.8 | 0.83 |

Tablica 7.4: Kaiser-Meyer-Olkinova mjera ili Mjera uzoračke adekvatnosti (MSA)

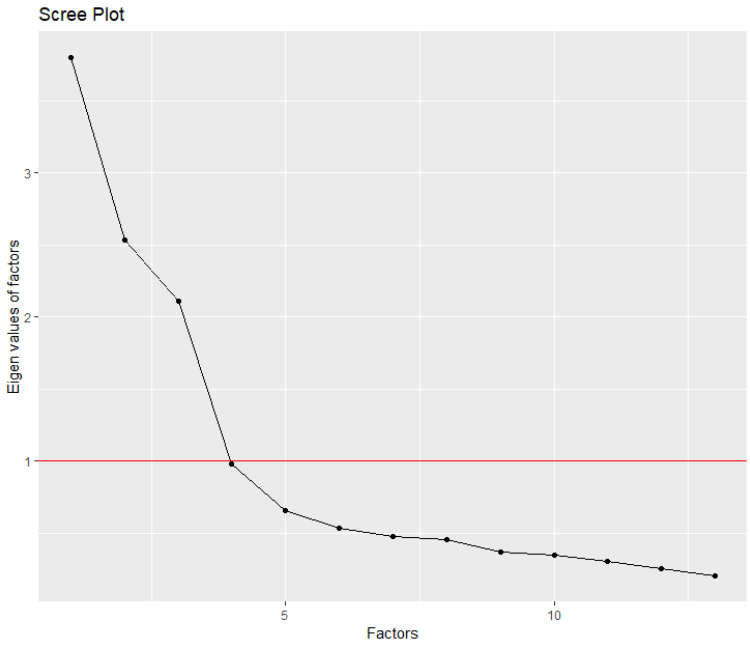
U poglavlju 6 smo razmatrali valjanost modela te pretpostavke koje bismo trebali provjeriti prije provedbe faktorske analize. Ukoliko je vrijednost od MSA manja od 0.5 to se smatra neprihvatljivim te je potrebno više koreliranih varijabli za analizu. Kaiser i Rice [4] predlažu da bi vrijednost MSA trebala iznositi barem 0.8 kako bismo mogli očekivati zadovoljavajuće rezultate, a vidimo da smo u našem slučaju jako blizu te vrijednosti pa možemo očekivati uspješnu sprovedbu analize. Sljedeći korak je odrediti broj faktora. Odabir broja faktora moguće je vršiti na više načina. U poglavlju 3 diskutirali smo razne načine odabira

broja faktora, mi ćemo proučiti dva načina. Prvi način je da odaberemo onoliko faktora koliko je svojstvenih vrijednosti većih od jedan (Kaiserov kriterij). Drugi način je promatrajući scree plot i koristeći metodu lakta.

| Svojstvene vrijednosti | |
|---|---|
| 3.8014341 | 0.4513068 |
| 2.5314125 | 0.3670326 |
| 2.1120241 | 0.344626 |
| 0.9762237 | 0.3023937 |
| 0.6532794 | 0.2523368 |
| 0.5329805 | 0.199599 |
| 0.4753509 | |

Tablica 7.5: Svojstvene vrijednosti

Iz tablice vidimo da imamo 3 svojstvene vrijednosti koje su veće od 1 pa pomoću Kaiserovog kriterija možemo zaključiti da su nam potrebna 3 faktora. Navodimo grafički prikaz svojstvenih vrijednosti.



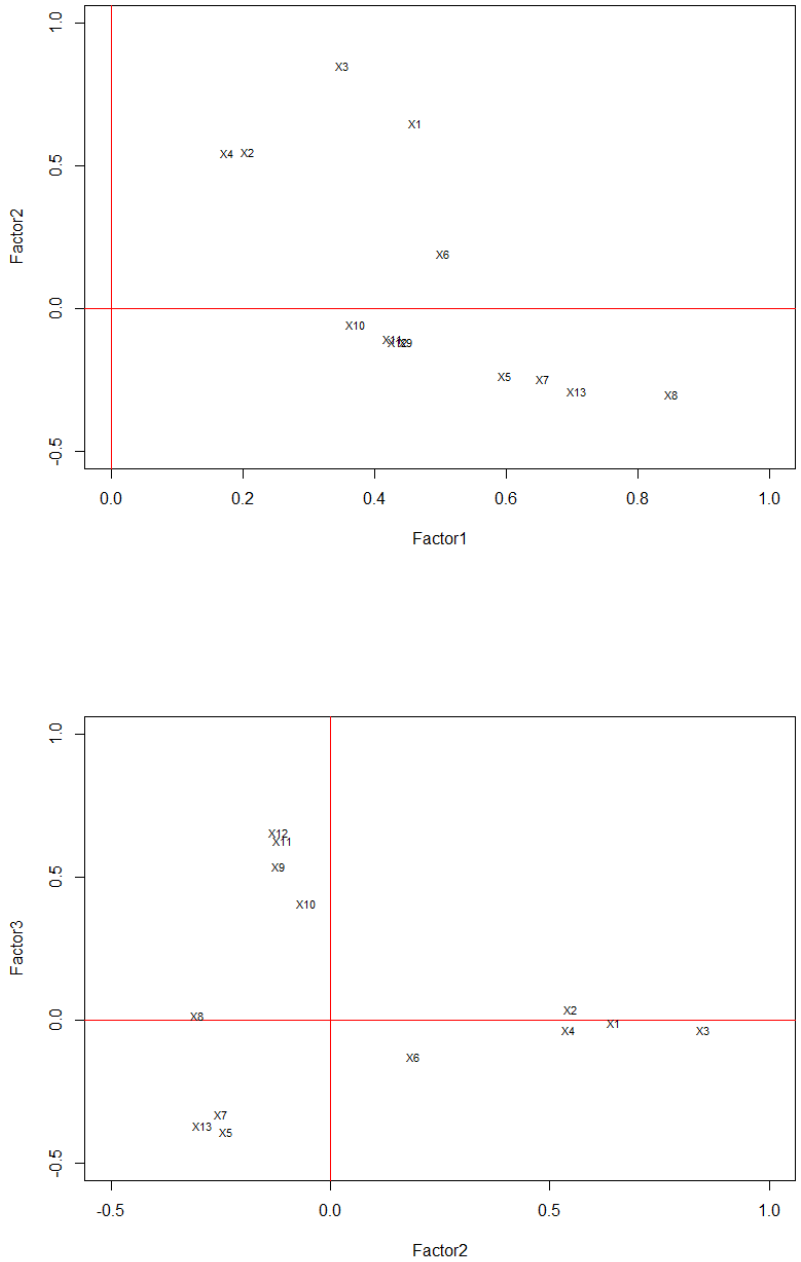

Slika 7.2: Grafički prikaz svojstvenih vrijednosti

Iz 7.2 vidimo kako bi metodom lakta očitali broj faktora. Budući da je oštri pad nakon 3. faktora, s grafa očitavamo da su potrebna 3 faktora.

| | Faktor1 | Faktor2 | Faktor3 | Komunaliteti |
|---|---|---|---|---|
| X5 | 0.6 | -0.24 | -0.39 | 0.5678416 |
| X6 | 0.51 | 0.19 | -0.13 | 0.3083785 |
| X7 | 0.66 | -0.25 | -0.33 | 0.6014316 |
| X8 | 0.85 | -0.3 | 0.02 | 0.8173054 |
| X13 | 0.71 | -0.29 | -0.37 | 0.7229755 |
| X1 | 0.46 | 0.65 | -0.01 | 0.6346274 |
| X2 | 0.21 | 0.55 | 0.04 | 0.3464377 |
| X3 | 0.35 | 0.85 | -0.03 | 0.849647 |
| X4 | 0.18 | 0.54 | -0.04 | 0.3287948 |
| X9 | 0.45 | -0.12 | 0.54 | 0.5045962 |
| X11 | 0.43 | -0.11 | 0.63 | 0.5896825 |
| X12 | 0.44 | -0.12 | 0.65 | 0.6329775 |
| X10 | 0.37 | -0.06 | 0.41 | 0.3079918 |
| Svojstvene vrijednosti | 3.400836 | 2.113375 | 1.698476 | 7.212688 |

Tablica 7.6: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti

Faktor 1 smo nazvali Udobnost, faktor 2 Pogodnost, a faktor 3 Usluge. Iz tablice 7.6 očitavamo da prvom faktoru Udobnosti pripadaju varijable X5, X6, X7, X8 i X13. Drugom faktoru Pogodnosti pripadaju varijable X1, X2, X3 i X4. Trećem faktoru Usluge pripadaju varijable X9, X10, X11 i X12. Ovo dodjeljivanje varijabli faktorima rađeno je na temelju najveće vrijednosti (gledajući apsolutnu vrijednost) u svakom retku matrice faktora. Kao što smo prije objasnili kako bi procijenili značajnost težine na nekom faktoru, vrijednosti od 0.5 su se pokazale dobrima u praksi. Također iz 7.6 vidimo procjene komunaliteta za svaku od varijabli. Komunalitet varijable nam govori koliko je varijance te varijable objašnjeno zajedničkim faktorima. Na temelju dobivenih vrijednosti možemo reći da su komunaliteti prihvatljivi. Može se vidjeti i da su svojstvene vrijednosti, odnosno varijance objašnjene pojedinim faktorom, jednake zbroju kvadrata pripadnih faktorskih težina. Vidimo i da varijabla X10 - Usluga prostora za noge te varijabla X12 - Usluga check-ina imaju vrlo bliske težine na faktorima Udobnost i Usluge. Na slijedećim grafičkim prikazima vidjet ćemo početnih 13 varijabli prikazane u odnosu na faktore.




Slika 7.3: Grafički prikaz faktora

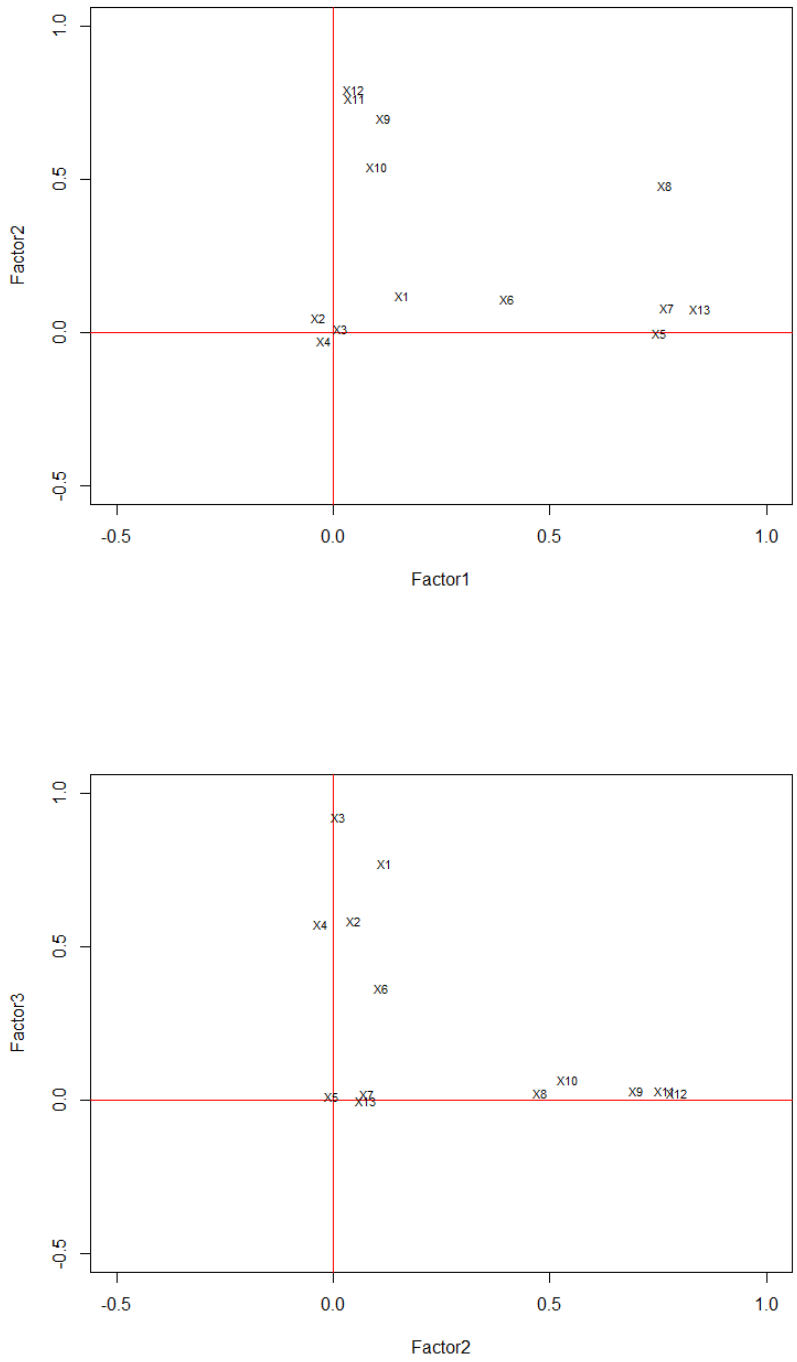
Pogledajmo sada što dobivamo rotacijama. Prvo ćemo analizirati ortogonalnu rotaciju faktora. Koristit ćemo varimax metodu, vrstu ortogonalne rotacije. U sljedećoj tablici

navedene su nove težine, komunaliteti i svojstvene vrijednosti nakon ortogonalne rotacije.

| | Faktor1 | Faktor2 | Faktor3 | Komunalitet |
|---|---|---|---|---|
| X5 | 0.75 | 0.00 | 0.01 | 0.5678416 |
| X7 | 0.77 | 0.08 | 0.02 | 0.6014316 |
| X8 | 0.77 | 0.48 | 0.02 | 0.8173054 |
| X13 | 0.85 | 0.08 | 0.00 | 0.7229755 |
| X9 | 0.12 | 0.70 | 0.03 | 0.5045962 |
| X10 | 0.10 | 0.54 | 0.07 | 0.3079918 |
| X11 | 0.05 | 0.77 | 0.03 | 0.5896825 |
| X12 | 0.05 | 0.79 | 0.02 | 0.6329775 |
| X1 | 0.16 | 0.12 | 0.77 | 0.6346274 |
| X2 | -0.03 | 0.05 | 0.59 | 0.3464377 |
| X3 | 0.02 | 0.01 | 0.92 | 0.8496470 |
| X4 | -0.02 | -0.03 | 0.57 | 0.3287948 |
| X6 | 0.40 | 0.11 | 0.36 | 0.3083785 |
| Svojstvene vrijednosti | 2.685730 | 2.271532 | 2.255425 | 7.2126875 |

Tablica 7.7: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon ortogonalne rotacije

Prilikom rotacije došlo je do promjena u faktorima. Promjene su u varijablama X10 i X12 koje su prije pripadale faktoru Udobnost, a sada su pridružene faktoru Usluga. Faktor Pogodnosti nema promjena. Vidimo da se komunaliteti nisu promijenili nakon ortogonalne rotacije što smo i očekivali. Naime, prilikom ortogonalne rotacije dolazi do rotacije redaka matrice težina po faktorima. Varijance su se nakon rotacije po faktorima međusobno približno izjednačile u odnosu na varijance koje smo imali na nerotiranim faktorima. Također, vidimo da se nakon ortogonalne rotacije ukupna varijabilnost nije promijenila. Na slijedećim grafičkim prikazima vidjet ćemo 13 varijabli prikazanih u odnosu na faktore nakon primjene ortogonalne rotacije.




Slika 7.4: Grafički prikaz faktora

Iz 7.4 vidljiv je raspored varijabli po faktorima. Vidljivo je da varijable X14, X7, X5, X8, X6 i X13 imaju najveće težine na faktoru 1, varijable X11, X9, X10 i X12 na faktoru 2, dok preostale varijable X1, X2, X3 i X4 na faktoru 3.
Pogledajmo sada što dobivamo kosom rotacijom. Matrica međufaktorske korelacije je:

| | Faktor1 | Faktor2 | Faktor3 |
|---|---|---|---|
| Faktor1 | 1.00 | -0.17 | 0.28 |
| Faktor2 | -0.17 | 1.00 | -0.14 |
| Faktor3 | 0.28 | -0.14 | 1.00 |

Tablica 7.8: Međufaktorska korelacija

Vidimo da su nakon kose rotacije faktori korelirani budući da oni nisu više ortogonalni.

| | Faktor1 | Faktor2 | Faktor3 | Komunaliteti |
|---|---|---|---|---|
| X5 | 0.78 | -0.02 | -0.14 | 0.6336980 |
| X7 | 0.79 | -0.02 | -0.05 | 0.6306715 |
| X8 | 0.74 | -0.04 | 0.36 | 0.6786723 |
| X13 | 0.87 | -0.04 | -0.07 | 0.7696721 |
| X1 | 0.07 | 0.77 | 0.05 | 0.6021887 |
| X2 | -0.10 | 0.59 | 0.02 | 0.3643263 |
| X3 | -0.08 | 0.94 | -0.04 | 0.8836387 |
| X4 | -0.08 | 0.58 | -0.06 | 0.3510827 |
| X9 | 0.04 | -0.01 | 0.70 | 0.4912260 |
| X10 | 0.04 | 0.03 | 0.54 | 0.2908456 |
| X11 | -0.04 | -0.01 | 0.78 | 0.6090871 |
| X12 | -0.04 | -0.02 | 0.81 | 0.6573824 |
| X6 | 0.37 | 0.35 | 0.02 | 0.2578183 |
| Svojstvene vrijednosti | 2.723404 | 2.291747 | 2.205159 | 7.2203097 |

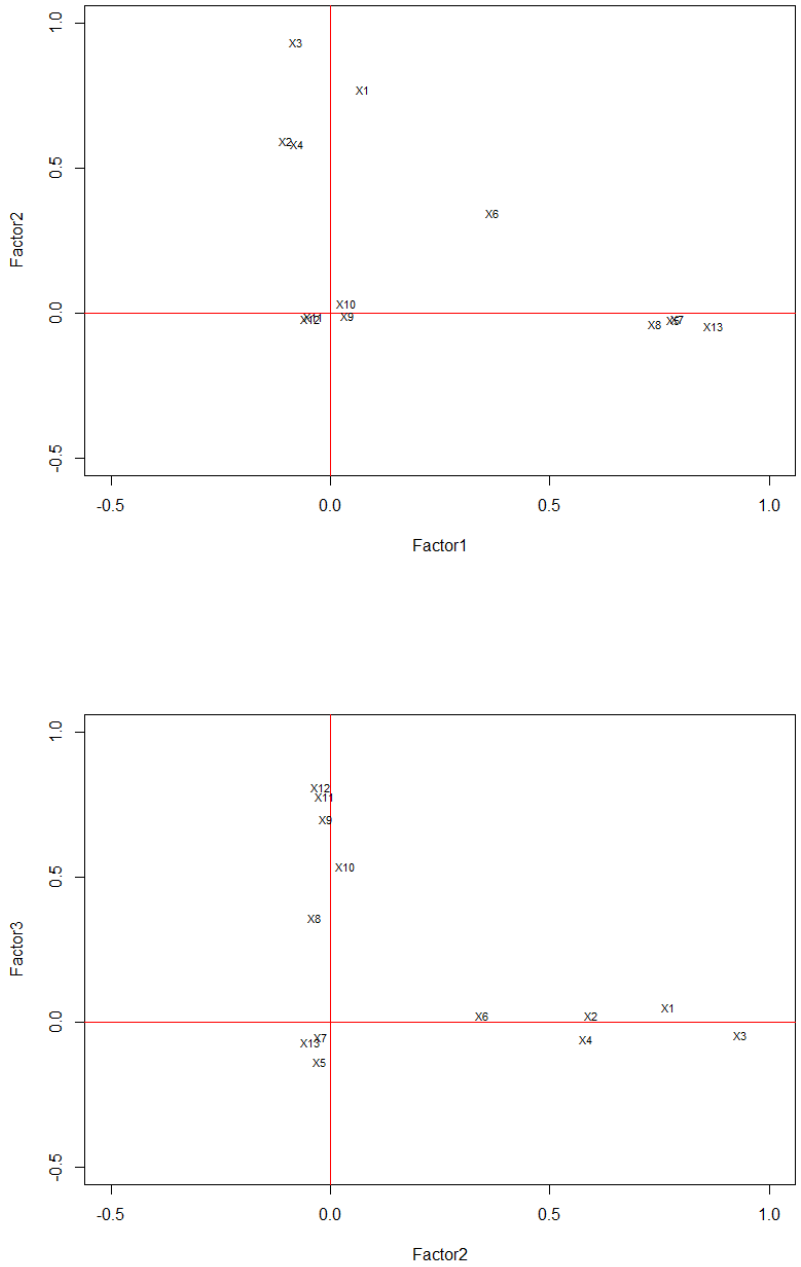
Tablica 7.9: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon kose rotacije

Za razliku od ortogonalnih faktorskih rješenja gdje se faktorske težine interpretiraju kao korelacije izmedu varijabli i faktora, kod kosokutih faktorskih rješenja kao što je ovdje promatrano promax rješenje, potrebno je gledati matricu faktorske strukture kako bi se ispitale korelacije izmedu varijabli i faktora.

| | Faktor1 | Faktor2 | Faktor3 |
|---|---|---|---|
| X1 | -0.04 | 0.75 | -0.04 |
| X2 | -0.20 | 0.61 | -0.09 |
| X3 | -0.25 | 0.96 | -0.20 |
| X4 | -0.19 | 0.61 | -0.16 |
| X5 | 0.75 | -0.14 | 0.09 |
| X6 | 0.32 | 0.28 | 0.08 |
| X7 | 0.78 | -0.15 | 0.17 |
| X8 | 0.85 | -0.21 | 0.57 |
| X9 | 0.24 | -0.11 | 0.71 |
| X10 | 0.18 | -0.05 | 0.54 |
| X11 | 0.18 | -0.12 | 0.77 |
| X12 | 0.19 | -0.13 | 0.80 |
| X13 | 0.86 | -0.18 | 0.18 |
| Svojstvene vrijednosti | 3.028814 | 2.456996 | 2.513333 |

Tablica 7.10: Matrica faktorske strukture

Glavna razlika izmedu tablica 7.9 i 7.10 je što je korelacijsku interpretaciju moguće dobiti koristeći se matricom faktorske strukture. Matricu strukture dobivamo iz matrice standardiziranih regresijskih koeficijentata. Ukoliko se matrica standardiziranih regresijskih koeficijentata pomnoži sa međufaktorskom korelacijskom matricom 7.8 dobiva se matrica faktorske strukture 7.10. Vidimo da je raspored varijabli po faktorima nakon kose rotacije ostao isti kao i raspored varijabli po faktorima nakon ortogonalne rotacije, odnosno faktoru Udobnost pripadaju varijable X5, X6, X7, X8 i X13, faktoru Usluge pripadaju X9, X1, X11, X12, a trećem faktoru Pogodnost pripadaju X1, X2, X3 i X4. Uočavamo da se suma svojstvenih vrijednosti, odnosno ukupna varijabilnost nakon kose rotacije promijenila te je ona sada veća i iznosi 7.22031. Vidimo i da sada za razliku od originalnog i ortogonalnog rješenja suma svojstvenih vrijednosti nije jednaka sumi komunaliteta. Komunaliteti su ostali isti kao i kod faktora bez rotacije i kod faktora nakon ortogonalne rotacije. Ovo je temeljna činjenica o faktorskim rotacijama; rotacije samo redistribuiraju varijancu objašnjenu faktorima dok varijanca objašnjena faktorima za pojedinu varijablu (komunalitet) ostaje nepromijenjen.
Na sljedećim grafičkim prikazima vidjet ćemo 13 varijabli prikazanih u odnosu na faktore nakon primjene kose rotacije.




Slika 7.5: Grafički prikaz faktora

Vidimo da je na ovom primjeru uspješno provedena faktorska analiza budući da su se varijable lijepo rasporedile po faktorima i komunalitet je visok. Uspješno smo od 13

varijabli dobili manji broj, odnosno 3 latentna faktora. Prvi faktor (Udobnost) čine varijable WIFI usluga tokom leta, Animacija tokom leta, Online podrška, Lakoća online rezervacije i Online ukrcavanje. Drugi faktor (Usluge) čine varijable Usluga tokom leta, Usluga prostora za noge, Rukovanje prtljagom i Čistoća. Treći faktor (Pogodnost) čine varijable udobnost sjedala, Točnost vremena polaska/dolaska, Hrana i piće i Lokacija gate-a .

## 7.2 Primjer 2-kvaliteta graška

Podatci nad kojima ćemo primijeniti faktorsku analizu odnose se na istraživanje promjene okusa graška koji su preuzeti s [7]. Različite vrste graška su nakon blanširanja brzo smrznute te zapakirane stavljene u hladnjak na tri mjeseca.
Raspolažemo s 60 podataka, odnosno, 60 osoba je ocijenilo karakteristike graška kao što su okus i boja brojevima od 1 do 9 te je dodatno provedena i kemijska analiza uzorka.
Navedene karakteristike su varijable koje će sudjelovati u faktorskoj analizi:

| Varijabla | Opis varijable |
|---|---|
| $X1$ | Tenderometrija |
| $X2$ | Postotak suhe tvari |
| $X3$ | Postotak suhe tvari nakon zamrzavanja |
| $X4$ | Saharoza |
| $X5$ | Glukoza 1 |
| $X6$ | Glukoza 2 |
| $X7$ | Okus |
| $X8$ | Slatkoća |
| $X9$ | Voćni okus |
| $X10$ | Bezukusnost |
| $X11$ | Brašnjavost |
| $X12$ | Tvrdoća |
| $X13$ | Bjelina |
| $X14$ | Boja 1 |
| $X15$ | Boja 2 |
| $X16$ | Boja 3 |

Tablica 7.11: Varijable koje sudjeluju u faktorskoj analizi

Cilj je vidjeti možemo li 16 ulaznih varijabli u 7.11 sažeti u manji broj latentnih varijabli, odnosno faktora.
U sljedećoj tablici prikazana je deskriptivna statistika varijabli X1-X16.

| varijabla | mean | sd | median | trimmed | min | max | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|
| X1 | 135.18 | 31.60 | 129.00 | 132.62 | 88.00 | 200.00 | 0.60 | -0.69 |
| X2 | 19.51 | 3.14 | 19.35 | 19.37 | 13.80 | 28.10 | 0.38 | -0.51 |
| X3 | 22.31 | 2.58 | 22.27 | 22.14 | 17.91 | 28.75 | 0.47 | -0.47 |
| X4 | 4.27 | 1.13 | 4.45 | 4.35 | 1.00 | 6.00 | -0.62 | -0.20 |
| X5 | 4.23 | 1.24 | 4.00 | 4.16 | 2.10 | 6.70 | 0.40 | -0.83 |
| X6 | 4.16 | 1.16 | 4.00 | 4.08 | 2.20 | 7.00 | 0.50 | -0.56 |
| X7 | 5.33 | 1.18 | 5.74 | 5.44 | 2.28 | 7.09 | -0.82 | -0.52 |
| X8 | 5.42 | 1.19 | 5.70 | 5.54 | 2.23 | 7.03 | -0.75 | -0.33 |
| X9 | 3.55 | 1.04 | 3.77 | 3.60 | 1.29 | 5.18 | -0.39 | -0.97 |
| X10 | 2.89 | 1.04 | 2.47 | 2.73 | 1.74 | 6.45 | 1.45 | 1.74 |
| X11 | 4.39 | 1.36 | 4.26 | 4.39 | 2.09 | 6.70 | 0.10 | -1.23 |
| X12 | 4.76 | 1.45 | 4.46 | 4.67 | 2.60 | 7.83 | 0.42 | -0.96 |
| X13 | 4.54 | 0.57 | 4.46 | 4.53 | 3.50 | 5.68 | 0.23 | -1.02 |
| X14 | 5.37 | 0.73 | 5.25 | 5.32 | 4.00 | 7.30 | 0.51 | -0.24 |
| X15 | 5.79 | 0.53 | 5.74 | 5.80 | 4.36 | 6.80 | -0.11 | -0.51 |
| X16 | 5.34 | 0.98 | 5.35 | 5.34 | 3.25 | 7.18 | -0.05 | -0.95 |

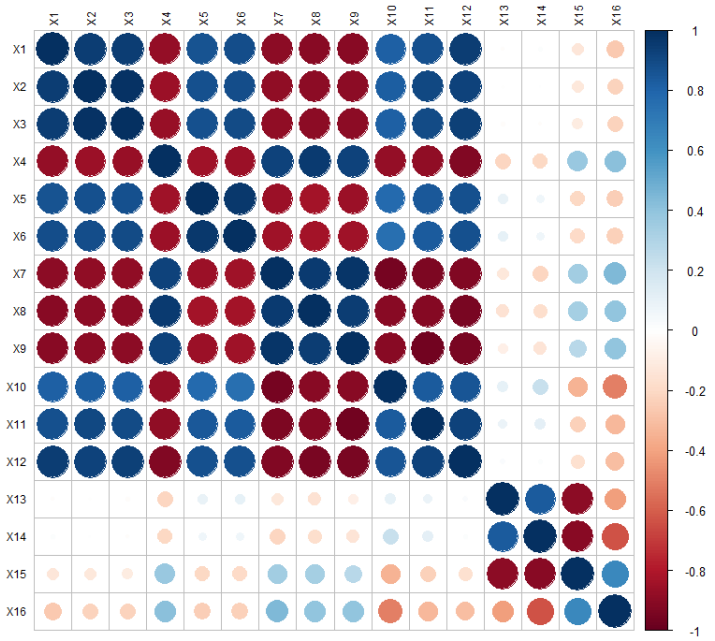
Tablica 7.12: Deskriptivna statistika varijabli

Pošto radimo s velikim brojem varijabli, matrica korelacija imate će dimenzije $16 \times 16$. Mi ćemo u svrhu shvaćanja oblika matrice prikazati tablično samo prvih 10 redaka i 10 stupaca matrice korelacije.

| | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X1 | 1.00 | 0.94 | 0.95 | -0.87 | 0.86 | 0.88 | -0.90 | -0.91 | -0.90 | 0.82 |
| X2 | 0.94 | 1.00 | 0.98 | -0.85 | 0.87 | 0.88 | -0.89 | -0.89 | -0.90 | 0.82 |
| X3 | 0.95 | 0.98 | 1.00 | -0.86 | 0.88 | 0.90 | -0.88 | -0.89 | -0.90 | 0.82 |
| X4 | -0.87 | -0.85 | -0.86 | 1.00 | -0.85 | -0.86 | 0.92 | 0.96 | 0.93 | -0.88 |
| X5 | 0.86 | 0.87 | 0.88 | -0.85 | 1.00 | 0.96 | -0.85 | -0.84 | -0.86 | 0.77 |
| X6 | 0.88 | 0.88 | 0.90 | -0.86 | 0.96 | 1.00 | -0.84 | -0.84 | -0.85 | 0.75 |
| X7 | -0.90 | -0.89 | -0.88 | 0.92 | -0.85 | -0.84 | 1.00 | 0.95 | 0.98 | -0.95 |
| X8 | -0.91 | -0.89 | -0.89 | 0.96 | -0.84 | -0.84 | 0.95 | 1.00 | 0.95 | -0.9 |
| X9 | -0.90 | -0.90 | -0.90 | 0.93 | -0.86 | -0.85 | 0.98 | 0.95 | 1.00 | -0.9 |
| X10 | 0.82 | 0.82 | 0.82 | -0.88 | 0.77 | 0.75 | -0.95 | -0.90 | -0.90 | 1 |

Tablica 7.13: Matric korelacije varijabli X1-X10

Kako bismo si predstavili korelacije među varijablama, vizualno ćemo reprezentirati matricu korelacija na sljedećoj slici.




Slika 7.6: Vizualna reprezentacija matrice korelacija

Tamnija boja naglašava da je korelacija među navedenim varijablama jača. Vidimo jasno da se stvaraju 2 grupice unutar kojih su varijable koje su korelirane, to nas navodi na mišljenje kako će zajednički faktori zaista i postojati.
Slijedeći korak je analiza vrijednosti MSA, analiza svojstvenih vrijednosti matrice korelacija te odabir broja faktora pomoću istih.

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | X16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.94 | 0.92 | 0.93 | 0.94 | 0.92 | 0.9 | 0.88 | 0.95 | 0.93 | 0.88 | 0.92 | 0.95 | 0.68 | 0.81 | 0.76 | 0.81 |

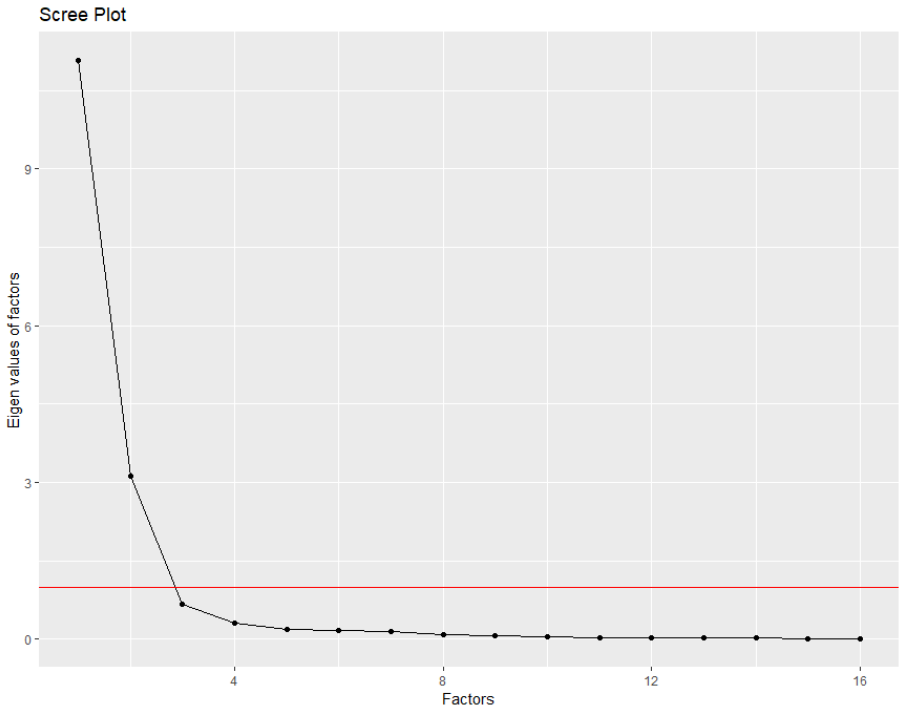
Tablica 7.14: Kaiser-Meyer-Olkinova mjera ili Mjera uzoračke adekvatnosti (MSA)

Sukladno diskusiji u Poglavlju 6 te primjenom u Primjeru 1, željeli bismo da vrijednost KMO iznosi barem 0.8 kako bismo mogli očekivati zadovoljavajuće rezultate. Vidimo da smo u našem slučaju jako blizu te vrijednosti pa možemo očekivati uspješnu provedbu analize.
Sljedeći korak je odrediti broj faktora. Promatramo prvo svojstvene vrijednosti.

| Svojstvene vrijednosti | |
|---|---|
| 11.06729004 | 0.060114 |
| 3.11576 | 0.049258 |
| 0.67882 | 0.033228 |
| 0.30139 | 0.027823 |
| 0.19347 | 0.026531 |
| 0.16740 | 0.017904 |
| 0.14743 | 0.01394 |
| 0.09119 | 0.00845 |

Tablica 7.15: Svojstvene vrijednosti

Iz 7.15 vidimo da imamo 2 svojstvene vrijednosti koje su veće od 1 pa pomoću Kaiserovog kriterija možemo zaključiti da su nam potrebna 2 faktora. Navodimo grafički prikaz svojstvenih vrijednosti.



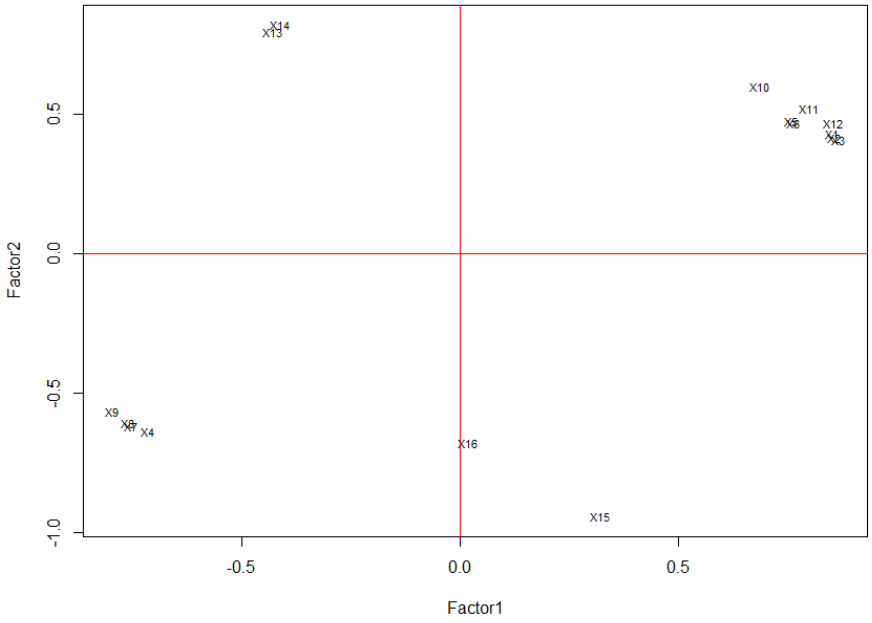

Slika 7.7: Grafički prikaz svojstvenih vrijednosti

Iz 7.7 vidimo kako bi metodom lakta očitali broj faktora. Budući da je oštri pad nakon 2. faktora, s grafa očitavamo da su potrebna 2 faktora.

| | Faktor1 | Faktor2 | Komunaliteti |
|---|---|---|---|
| X1 | 0.85 | 0.43 | 0.91275 |
| X2 | 0.86 | 0.42 | 0.90906 |
| X3 | 0.87 | 0.41 | 0.91661 |
| X4 | -0.71 | -0.64 | 0.91418 |
| X5 | 0.76 | 0.48 | 0.80247 |
| X6 | 0.76 | 0.47 | 0.80413 |
| X7 | -0.75 | -0.62 | 0.94967 |
| X8 | -0.76 | -0.61 | 0.94624 |
| X9 | -0.80 | -0.57 | 0.95329 |
| X10 | 0.69 | 0.60 | 0.83578 |
| X11 | 0.80 | 0.52 | 0.90950 |
| X12 | 0.86 | 0.47 | 0.95464 |
| X13 | -0.43 | 0.80 | 0.81965 |
| X14 | -0.41 | 0.82 | 0.84427 |
| X15 | 0.32 | -0.94 | 0.99502 |
| X16 | 0.02 | -0.68 | 0.46159 |
| Svojstvene vrijednosti | 11.06729004 | 3.11575902 | 13.92884 |

Tablica 7.16: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti

Faktor 1 smo nazvali Sastav, a faktor 2 Boja. Iz tablice 7.16 očitavamo da prvom faktoru Sastava pripadaju varijable X1, X2, X3, X4, X5, X6, X10, X11 i X12. Drugom faktoru Pogodnosti pripadaju varijable X13, X14, X15 i X16. Dodjeljivanje varijabli faktorima rađeno je na temelju najveće vrijednosti (gledajući apsolutnu vrijednost) u svakom retku matrice faktora. Kao što smo i prije komentirali, kako bi procijenili značajnost težine na nekom faktoru, vrijednosti od 0.5 su se pokazale dobrima u praksi.
Takoder iz 7.16 vidimo procjene komunaliteta za svaku od varijabli. Na temelju dobivenih vrijednosti možemo reći da su komunaliteti prihvatljivi. Vidimo da su svojstvene vrijednosti (varijance objašnjene pojedinim faktorom) jednake zbroju kvadrata pripadnih faktorskih težina. Na slijedećem grafičkom prikazu vidjet ćemo početnih 16 varijabli prikazane u odnosu na faktore.



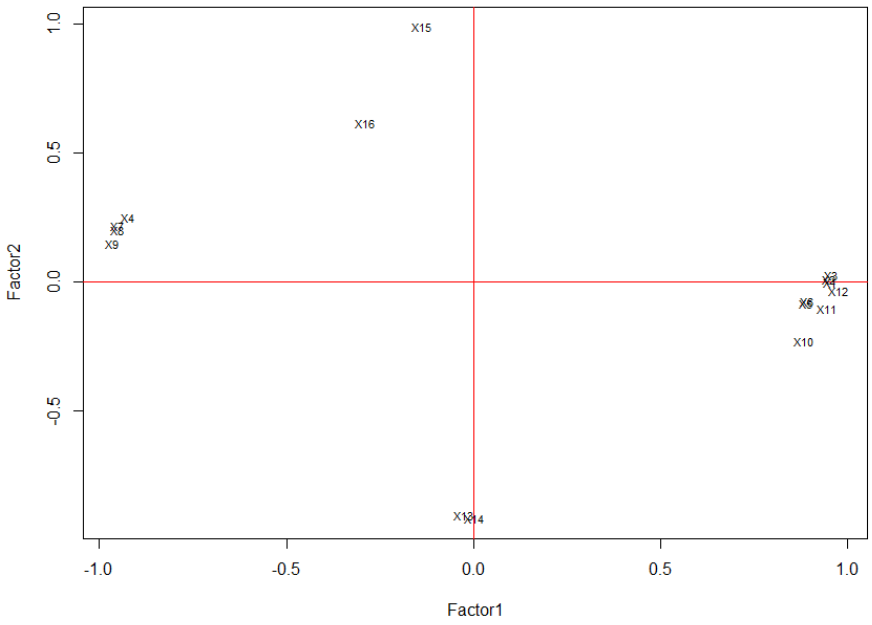

Slika 7.8: Grafički prikaz faktora

Kao što smo i uvidjeli u prethodnoj tablici, varijable se otprilike grupiraju oko faktora uz velike vrijednosti težina. Kako bismo sa sigurnoću mogli donijeti zaključke, provest ćemo rotacije koje će nam pomoći pri jednostavnijem shvaćanju grupiranja varijabli oko faktora. Pogledajmo sada što dobivamo rotacijama. Prvo ćemo analizirati ortogonalnu rotaciju faktora. Koristit ćemo varimax metodu, vrstu ortogonalne rotacije. U sljedećoj tablici navedene su nove težine, komunaliteti i svojstvene vrijednosti nakon primjene ortogonalne rotacije.

| | Faktor1 | Faktor2 | Komunaliteti |
|---|---|---|---|
| X1 | 0.96 | 0 | 0.9127462 |
| X2 | 0.95 | 0.01 | 0.909063 |
| X3 | 0.96 | 0.02 | 0.9166071 |
| X4 | -0.92 | 0.25 | 0.9141766 |
| X5 | 0.89 | -0.08 | 0.80 |
| X6 | 0.89 | -0.07 | 0.80 |
| X7 | -0.95 | 0.22 | 0.9496673 |
| X8 | -0.95 | 0.20 | 0.9462371 |
| X9 | -0.97 | 0.15 | 0.9532888 |
| X10 | 0.88 | -0.23 | 0.8357802 |
| X11 | 0.95 | -0.11 | 0.90950 |
| X12 | 0.98 | -0.04 | 0.95464 |
| X13 | -0.02 | -0.91 | 0.81965 |
| X14 | 0.00 | -0.92 | 0.84427 |
| X15 | -0.14 | 0.99 | 0.99502 |
| X16 | -0.29 | 0.61 | 0.46159 |
| Svojstvene vrijednosti | 10.662069 | 3.266772 | 13.92884 |

Tablica 7.17: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon ortogonalne rotacije

Prilikom ortogonalne rotacije nije došlo do promjena u faktorima. Vidimo da se komunaliteti nisu promijenili nakon ortogonalne rotacije što smo i očekivali. Varijance su se nakon rotacije po faktorima medusobno približno izjednačile u odnosu na varijance koje smo imali na nerotiranim faktorima. Takoder, vidimo da se nakon ortogonalne rotacije ukupna varijabilnost nije promijenila. Na slijedećem grafičkom prikazu vidjet ćemo 16 varijabli prikazanih u odnosu na faktore nakon primjene ortogonalne rotacije.




Slika 7.9: Grafički prikaz faktora nakon ortogonalne rotacije

Pogledajmo sada što dobivamo kosom rotacijom. Matrica međufaktorske korelacije je:

| | Faktor1 | Faktor2 |
|---|---|---|
| Faktor1 | 1.00 | -0.22 |
| Faktor2 | -0.22 | 1.00 |

Tablica 7.18: Matrica međufaktorske korelacije

Pogledajmo matricu faktorske strukture kako bi se ispitala korelacija između vrijabli i faktora.

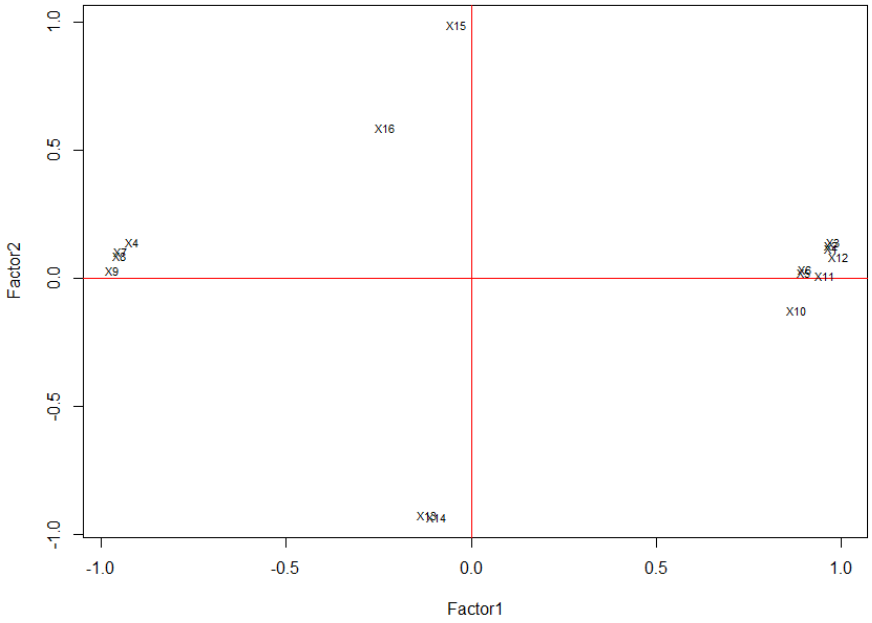
| | Faktor1 | Faktor2 | Komunaliteti |
|---|---|---|---|
| X1 | 0.97 | 0.12 | 0.9127462 |
| X2 | 0.97 | 0.13 | 0.909063 |
| X3 | 0.98 | 0.14 | 0.9166071 |
| X4 | -0.92 | 0.14 | 0.9141766 |
| X5 | 0.90 | 0.02 | 0.8024746 |
| X6 | 0.90 | 0.03 | 0.8041278 |
| X7 | -0.95 | 0.10 | 0.9496673 |
| X8 | -0.95 | 0.09 | 0.9462371 |
| X9 | -0.97 | 0.03 | 0.9532888 |
| X10 | 0.88 | -0.13 | 0.8357802 |
| X11 | 0.96 | 0.01 | 0.9094977 |
| X12 | 0.99 | 0.08 | 0.9546401 |
| X13 | -0.12 | -0.92 | 0.8196533 |
| X14 | -0.09 | -0.93 | 0.8442746 |
| X15 | -0.04 | 0.99 | 0.9950178 |
| X16 | -0.23 | 0.59 | 0.4615884 |
| Svojstvene vrijednosti | 10.802122 | 3.165976 | 13.9681 |

Tablica 7.19: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon kose rotacije

| | Faktor1 | Faktor2 |
|---|---|---|
| X1 | 0.95 | -0.09 |
| X2 | 0.95 | -0.09 |
| X3 | 0.95 | -0.07 |
| X4 | -0.95 | 0.34 |
| X5 | 0.90 | -0.17 |
| X6 | 0.90 | -0.16 |
| X7 | -0.97 | 0.31 |
| X8 | -0.97 | 0.30 |
| X9 | -0.98 | 0.25 |
| X10 | 0.91 | -0.32 |
| X11 | 0.95 | -0.20 |
| X12 | 0.97 | -0.13 |
| X13 | 0.08 | -0.90 |
| X14 | 0.11 | -0.91 |
| X15 | -0.26 | 1.00 |
| X16 | -0.36 | 0.64 |
| Svojstvene vrijednosti | 10.91595 | 3.64939 |

Tablica 7.20: Matrica faktorske strukture

Vidimo da je raspored varijabli po faktorima nakon kose rotacije ostao isti kao i raspored varijabli po faktorima nakon ortogonalne rotacije, odnosno faktoru Sastav pripadaju varijable X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11 i X12, a faktoru Boje pripadaju X13, X14, X15, X12. Uočavamo da se suma svojstvenih vrijednosti, odnosno ukupna varijabilnost, nakon kose rotacije promijenila te je ona sada veća i iznosi 14.56534. Vidimo i da sada, za razliku od originalnog i ortogonalnog rješenja, suma svojstvenih vrijednosti nije jednaka sumi komunaliteta. Komunaliteti su naime ostali isti kao kod faktora bez rotacije i kod faktora nakon ortogonalne rotacije.




Slika 7.10: Grafički prikaz faktora nakon kose roatcije

Vidimo da je na ovom primjeru uspješno provedena faktorska analiza budući da su se varijable lijepo rasporedile po faktorima i komunalitet je visok. Uspješno smo od 16 varijabli dobili manji broj, odnosno 2 latentna faktora. Prvi faktor (Struktura) čine varijable Tenderometrija, Postotak suhe tvari, Postotak suhe tvari nakon zamrzavanja, Saharoza, Glukoza 1, Glukoza 2, Okus, Slatkoća, Voćni okus, Bezukusnost, Brašnjavost i Tvrdoća. Drugi faktor (Boja) čine varijable Bjelina, Boja 1, Boja 2 i Boja 3.

## 7.3 Primjer 3-Humor Styles

Podatci nad kojima ćemo primijeniti faktorsku analizu odnose se na istraživanje postojanja različitih vrsta humora koji su preuzeti s [9]. Skup podataka s kojim radimo sadrži 1072 podataka, odnosno 1072 osoba je ispunilo test odgovaranjem na 32 pitanja koja se odnose na generalni odnos prema humoru i drugim ljudima te pitanja vezana uz spol, godine i slično.

Prilikom analize odlučili smo koristiti isključivo 32 varijable koje su ispitane u ovom istraživanju koje se odnose na ispitivanje postojanja raznih vrsta humora. Test se sastoji od 32 pitanja, odnosno ta pitanja će predstavljati varijable na kojima ćemo primijeniti faktorsku analizu. Detaljnije objašnjenje varijabli na kojima primjenjujemo faktorsku analizu nalazi se u 8.1 u Dodatak A.
Cilj provedbe faktorske analize je vidjeti možemo li 32 ulazne varijable sažeti u manji broj latentnih varijabli, odnosno faktora. Za početak, pošto radimo s velikim brojem varijabli, u sljedećoj tablici prikazat ćemo samo isječak tablice vrijednosti deskriptivne statistike varijabli.

| | mean | sd | median | trimmed | min | max | skew | kurtosis |
|---|---|---|---|---|---|---|---|---|
| Q1 | 2.03 | 1.08 | 2.00 | 1.88 | -1 | 5 | 0.86 | 0.26 |
| Q2 | 3.34 | 1.11 | 3.00 | 3.39 | -1 | 5 | -0.43 | -0.13 |
| Q3 | 3.08 | 1.17 | 3.00 | 3.10 | -1 | 5 | -0.17 | -0.62 |
| Q4 | 2.83 | 1.16 | 3.00 | 2.81 | -1 | 5 | 0.11 | -0.74 |
| Q5 | 3.60 | 1.06 | 4.00 | 3.68 | -1 | 5 | -0.63 | 0.23 |
| Q6 | 4.15 | 0.98 | 4.00 | 4.3 | -1 | 5 | -1.26 | 1.75 |
| Q7 | 3.28 | 1.1 | 3.00 | 3.31 | -1 | 5 | -0.4 | -0.23 |
| Q8 | 2.54 | 1.23 | 2.00 | 2.48 | -1 | 5 | 0.22 | -0.62 |
| Q9 | 2.58 | 1.22 | 2.00 | 2.5 | -1 | 5 | 0.4 | -0.56 |
| Q10 | 2.87 | 1.21 | 3.00 | 2.86 | -1 | 5 | -0.04 | -0.72 |

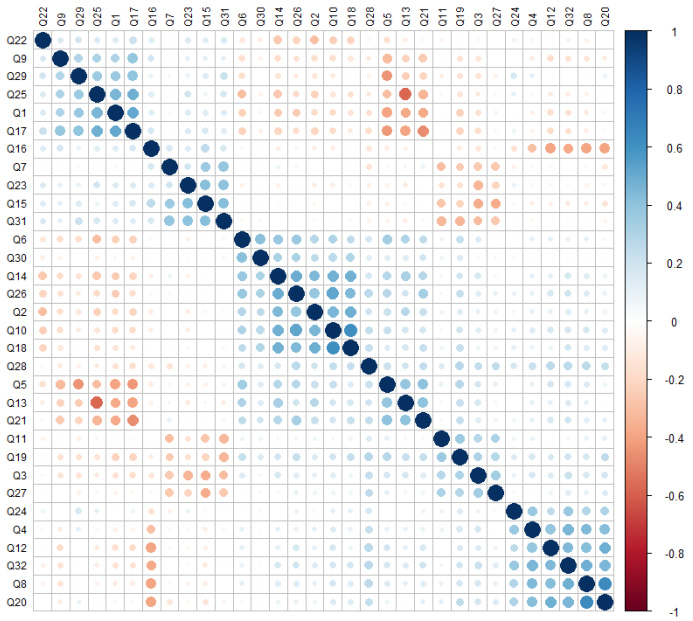
Tablica 7.21: Svojstvene vrijednosti

Također, kako radimo s 32 varijable, matrica korelacije je dimenzije $32 \times 32$. Iz tog razloga navodimo samo isječak matrice korelacija.

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Q1 | 1 | -0.18 | -0.15 | -0.08 | -0.39 | -0.24 | -0.04 | -0.11 | 0.31 | -0.19 |
| Q2 | -0.18 | 1.00 | 0.12 | 0.10 | 0.22 | 0.29 | 0.01 | 0.12 | -0.16 | 0.45 |
| Q3 | -0.15 | 0.12 | 1.00 | 0.18 | 0.18 | 0.15 | -0.26 | 0.17 | -0.13 | 0.12 |
| Q4 | -0.08 | 0.10 | 0.18 | 1.00 | 0.09 | 0.11 | -0.05 | 0.44 | -0.11 | 0.15 |
| Q5 | -0.39 | 0.22 | 0.18 | 0.09 | 1 | 0.34 | 0.04 | 0.15 | -0.32 | 0.23 |
| Q6 | -0.24 | 0.29 | 0.15 | 0.11 | 0.34 | 1 | -0.03 | 0.11 | -0.18 | 0.31 |
| Q7 | -0.04 | 0.01 | -0.26 | -0.05 | 0.04 | -0.03 | 1 | -0.04 | 0.02 | 0 |
| Q8 | -0.11 | 0.12 | 0.17 | 0.44 | 0.15 | 0.11 | -0.04 | 1 | -0.13 | 0.19 |
| Q9 | 0.31 | -0.16 | -0.13 | -0.11 | -0.32 | -0.18 | 0.02 | -0.13 | 1 | -0.17 |
| Q10 | -0.19 | 0.45 | 0.12 | 0.15 | 0.23 | 0.31 | 0 | 0.19 | -0.17 | 1 |

Tablica 7.22: Korelacijska matrica

Kako smo u prethodnoj tablici prikazali isključivo jedan malen dio originalne matrice korelacija, iskoristit ćemo vizualnu reprezentaciju matrice korelacija kako bismo stekli dojam o korelacijama među svim varijablama.




Slika 7.11: Vizualna reprezentacija matrice korelacija

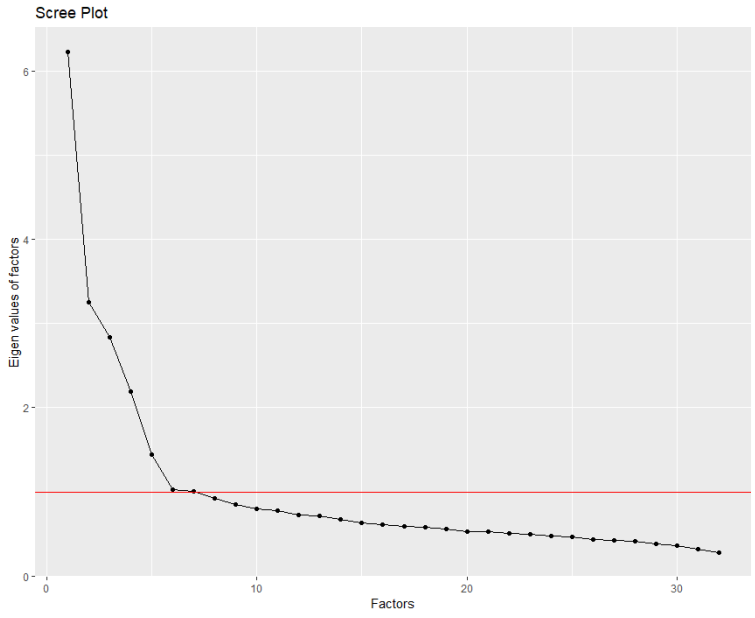
Tamnija boja naglašava da je korelacija među navedenim varijablama jača. Vidimo tako da se stvara 4 do 5 grupica unutar kojih su varijable koje su korelirane, to nas navodi na mišljenje kako će zajednički faktori zaista i postojati.

Slijedeći korak je analiza vrijednosti MSA, analiza svojstvenih vrijednosti matrice korelacija te odabir broja faktora pomoću istih.

| Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 |
|---|---|---|---|---|---|---|---|
| 0.92 | 0.92 | 0.9 | 0.89 | 0.9 | 0.88 | 0.82 | 0.86 |
| Q9 | Q10 | Q11 | Q12 | Q13 | Q14 | Q15 | Q16 |
| 0.95 | 0.86 | 0.8 | 0.9 | 0.86 | 0.92 | 0.83 | 0.86 |
| Q17 | Q18 | Q19 | Q20 | Q21 | Q22 | Q23 | Q24 |
| 0.88 | 0.83 | 0.89 | 0.84 | 0.87 | 0.84 | 0.81 | 0.85 |
| Q25 | Q26 | Q27 | Q28 | Q29 | Q30 | Q31 | Q32 |
| 0.83 | 0.89 | 0.83 | 0.93 | 0.85 | 0.82 | 0.81 | 0.9 |

Tablica 7.23: Kaiser-Meyer-Olkinova mjera ili Mjera uzoračke adekvatnosti (MSA)

Vidimo da su u našem slučaju sve vrijednosti veće od 0.8 pa možemo očekivati uspješnu sprovedbu analize. Sljedeći korak je odrediti broj faktora u faktorskoj analizi.




Slika 7.12: Grafički prikaz svojstvenih vrijednosti

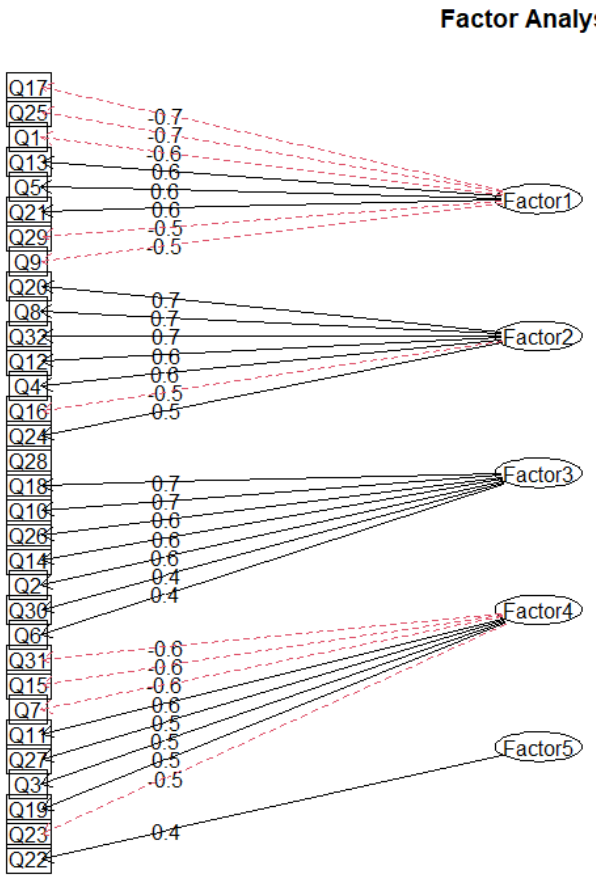
Iz 7.12 vidimo kako bi metodom lakta očitali broj faktora. Budući da je oštri pad nakon 5. faktora, s grafa očitavamo da je potrebno 5 faktora. Analizirat ćemo prvo model s 5 faktora da vidimo da li je ovaj broj faktora prevelik. Pogledajmo prvo koje vrijednosti

dobijemo ukoliko primijenjujemo faktorsku analizu s 5 faktora.
Slijedi isječak tablice 8.2 u Dodatku A koja sadrži težine po faktorima i procjene komunaliteta.

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Faktor5 | Komunaliteti |
|---|---|---|---|---|---|---|
| Q9 | -0.41 | 0.11 | 0.12 | 0.21 | 0.11 | 0.2543933 |
| Q22 | -0.3 | 0.15 | 0.03 | -0.12 | 0.35 | 0.2502252 |
| Q24 | 0.19 | 0.4 | 0.24 | 0 | 0.16 | 0.2800136 |
| Q28 | 0.44 | 0.14 | 0.04 | 0.07 | 0.22 | 0.2684765 |
| Q30 | 0.29 | -0.2 | 0.19 | 0.22 | 0.13 | 0.2222783 |

Tablica 7.24: Težine po faktorima i procjene komunaliteta

Željeli bismo da svaka varijabla s kojom radimo ima komunalitet veći od 0.3, no u našem slučaju imamo 5 varijabli s komunalitetom manjim od 0.3. Nastavit ćemo dalje s analizom kako bi uvidjeli koje zaista od ovih varijabli nisu povezane niti s jednim faktorom pa time nisu ni relevantne u našoj analizi. Kako bismo došli do rješenja, moramo modificirati naše rješenje tako što ćemo prvo rotirati varijable. Na takav način ćemo povećati težine svake varijable na jednom faktoru a smanjiti težine na ostalim faktorima. Primjenom ortogonalne rotacije dobivamo sljedeće grupacije varijabli po faktorima.




Slika 7.13: Grafički prikaz svojstvenih vrijednosti

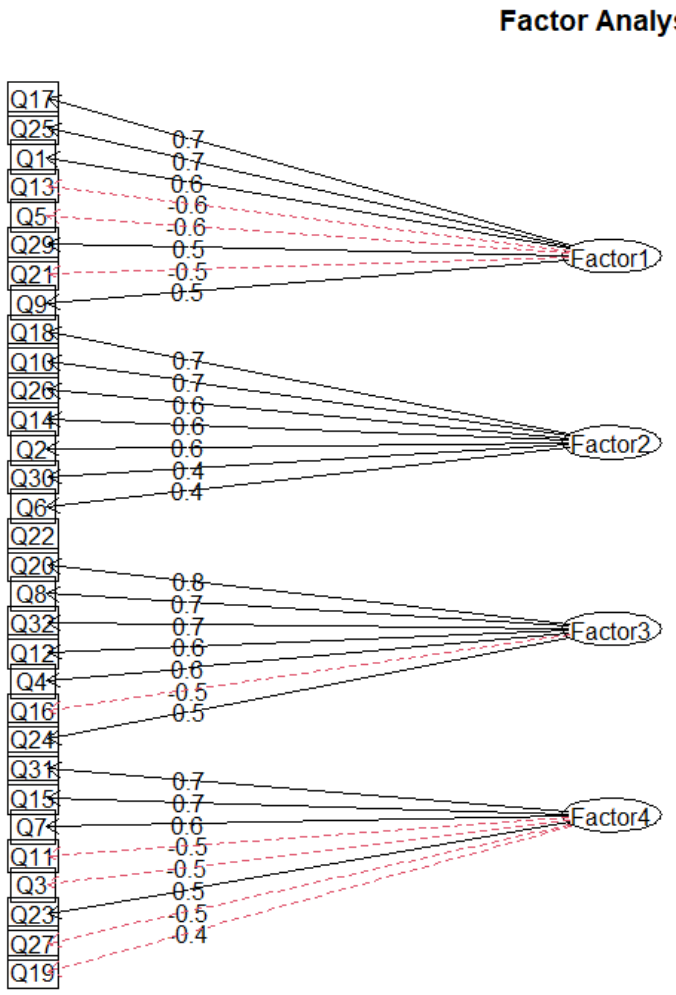
Na 7.13 možemo uočiti par problema. Vidimo da varijabla Q28 nije povezana niti s jednim faktorom, a Faktor 5 je povezan isključivo s jednom varijablom što ne želimo (željeli bismo da je po svakom faktoru podjednako varijabli). Također, na ovakav način ispada da nam uopće nije potreban model s 5 faktora (pošto radimo s velikim brojem varijabli a samo jedna od njih je povezana s faktorom 5). Iz navedenih razloga varijablu Q28 odstranjujemo iz podataka te ponavljamo analizu na modelu s 4 faktora. Sljedeće je isječak tablice 8.3 iz Dodatak A.

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Komunaliteti |
|---|---|---|---|---|---|
| Q22 | -0.3 | 0.14 | 0.02 | -0.12 | 0.1278526 |

Tablica 7.25: Težine po faktorima i procjene komunaliteta

Vidimo da je komunalitet varijable Q22 malen pa ćemo ponovno grafički provjeriti

je li nam relevantna u analizi. Primijenit ćemo ortogonalnu rotaciju, dobivene rezultate promatramo na sljedećem grafičkom prikazu faktorske analize s 4 faktora.



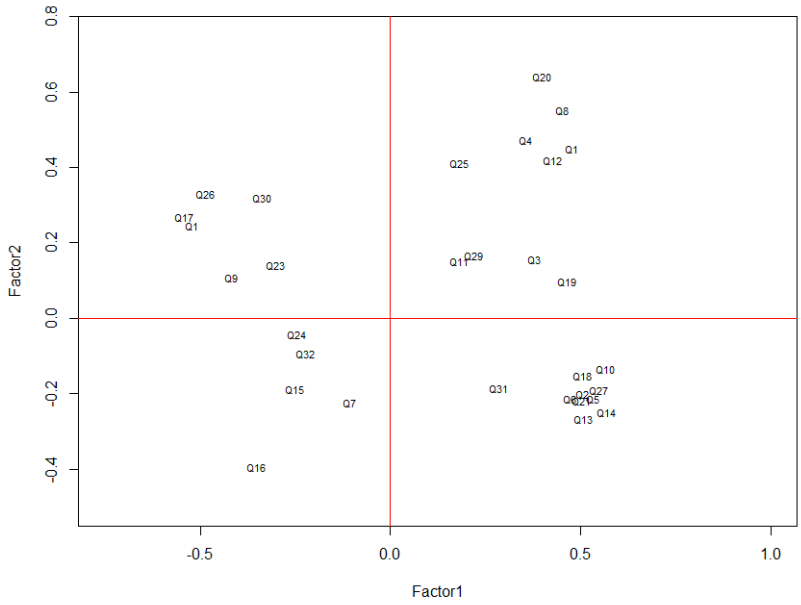

Slika 7.14: Grafički prikaz faktorske analize

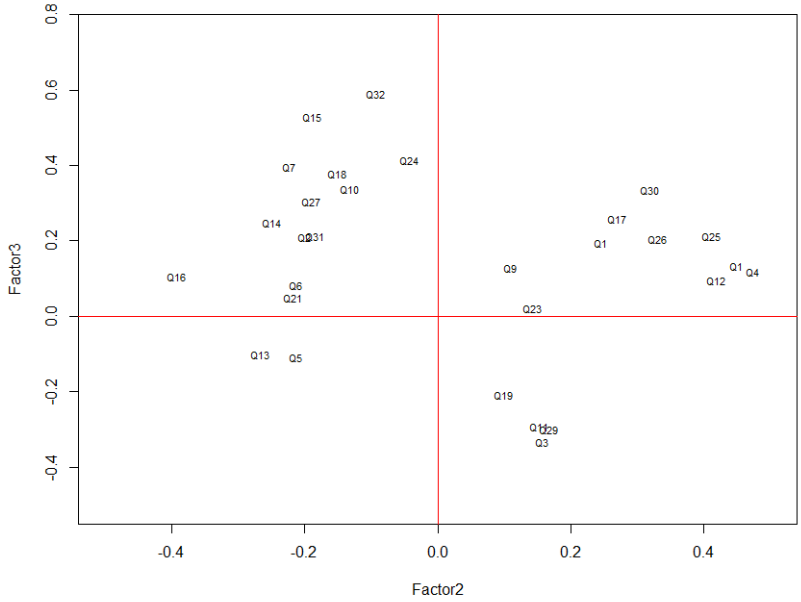
Vidimo iz 7.14 kako varijabla Q22 nije povezana niti s jednim od faktora pa ju možemo izbaciti pošto nije relevantna u analizi. Dalje nastavljamo faktorsku analizu na podatcima u kojima je izbačena i varijabla Q22, odnosno nastavljamo faktorsku analizu s 30 varijabli i s 4 faktora.

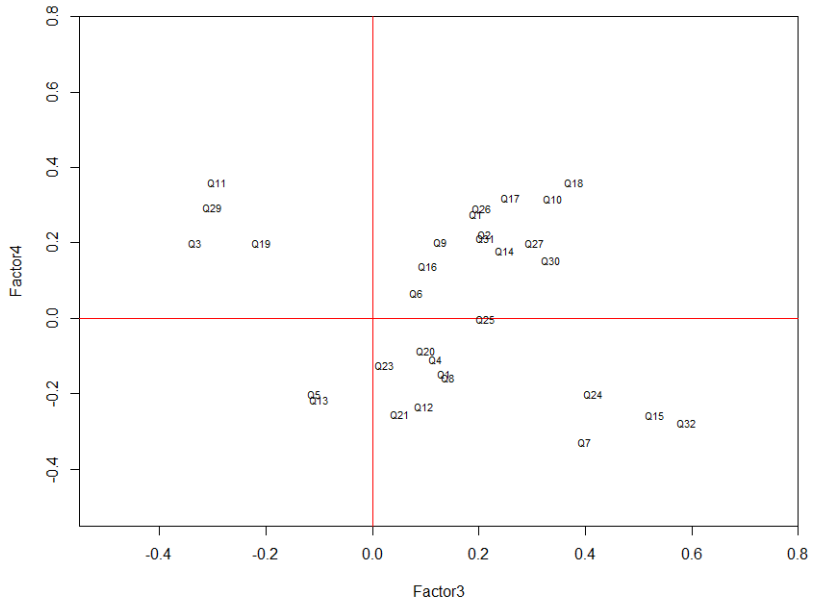
| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Komunaliteti |
|---|---|---|---|---|---|
| Q1 | -0.52 | 0.24 | 0.19 | 0.28 | 0.447031 |
| Q2 | 0.51 | -0.2 | 0.21 | 0.22 | 0.3898279 |
| Q3 | 0.38 | 0.16 | -0.33 | 0.20 | 0.3184788 |
| Q4 | 0.36 | 0.47 | 0.12 | -0.11 | 0.3763512 |
| Q5 | 0.53 | -0.21 | -0.11 | -0.20 | 0.3834022 |
| Q6 | 0.47 | -0.21 | 0.08 | 0.07 | 0.2805482 |
| Q7 | -0.11 | -0.22 | 0.4 | -0.33 | 0.3263348 |
| Q8 | 0.45 | 0.55 | 0.14 | -0.16 | 0.5539539 |
| Q9 | -0.42 | 0.11 | 0.13 | 0.2 | 0.2439201 |
| Q10 | 0.57 | -0.13 | 0.34 | 0.32 | 0.5532062 |
| Q11 | 0.18 | 0.15 | -0.29 | 0.36 | 0.2716209 |
| Q12 | 0.43 | 0.42 | 0.1 | -0.23 | 0.4208709 |
| Q13 | 0.51 | -0.27 | -0.1 | -0.22 | 0.3862897 |
| Q14 | 0.57 | -0.25 | 0.25 | 0.18 | 0.4792541 |
| Q15 | -0.25 | -0.19 | 0.53 | -0.26 | 0.4452244 |
| Q16 | -0.35 | -0.39 | 0.1 | 0.14 | 0.3091304 |
| Q17 | -0.54 | 0.27 | 0.26 | 0.32 | 0.5362626 |
| Q18 | 0.51 | -0.15 | 0.38 | 0.36 | 0.5509928 |
| Q19 | 0.46 | 0.1 | -0.21 | 0.2 | 0.3083244 |
| Q20 | 0.4 | 0.64 | 0.1 | -0.09 | 0.5900571 |
| Q21 | 0.5 | -0.22 | 0.05 | -0.25 | 0.3681084 |
| Q23 | -0.25 | -0.04 | 0.41 | -0.2 | 0.2739756 |
| Q24 | 0.18 | 0.41 | 0.21 | 0.00 | 0.2473512 |
| Q25 | -0.49 | 0.33 | 0.2 | 0.29 | 0.4715178 |
| Q26 | 0.55 | -0.19 | 0.31 | 0.20 | 0.4703384 |
| Q27 | 0.22 | 0.17 | -0.3 | 0.29 | 0.2531853 |
| Q29 | -0.34 | 0.32 | 0.34 | 0.15 | 0.3509442 |
| Q30 | 0.29 | -0.18 | 0.21 | 0.21 | 0.2059363 |
| Q31 | -0.22 | -0.09 | 0.59 | -0.28 | 0.4833411 |
| Q32 | 0.48 | 0.45 | 0.13 | -0.15 | 0.4689524 |
| Svojstvene vrijednosti | 5.422431 | 2.620935 | 2.226886 | 1.622332 | 11.77112 |

Tablica 7.26: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti

Na slijedećem grafičkom prikazu vidjet ćemo početnih 30 varijabli prikazanih u odnosu na sva 4 faktora.


Factor2
Factor1
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.5
0.0
0.5
1.0
Q20
Q8
Q4
Q1
Q12
Q25
Q26
Q30
Q17
Q1
Q9
Q23
Q11
Q29
Q3
Q19
Q24
Q32
Q15
Q7
Q16
Q18
Q10
Q31
Q27
Q14
Q13


Factor3
Factor2
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2
0.4
Q32
Q15
Q7
Q24
Q18
Q10
Q27
Q14
Q31
Q16
Q6
Q21
Q13
Q5
Q30
Q17
Q1
Q26
Q25
Q9
Q1
Q4
Q12
Q23
Q19
Q29
Q3



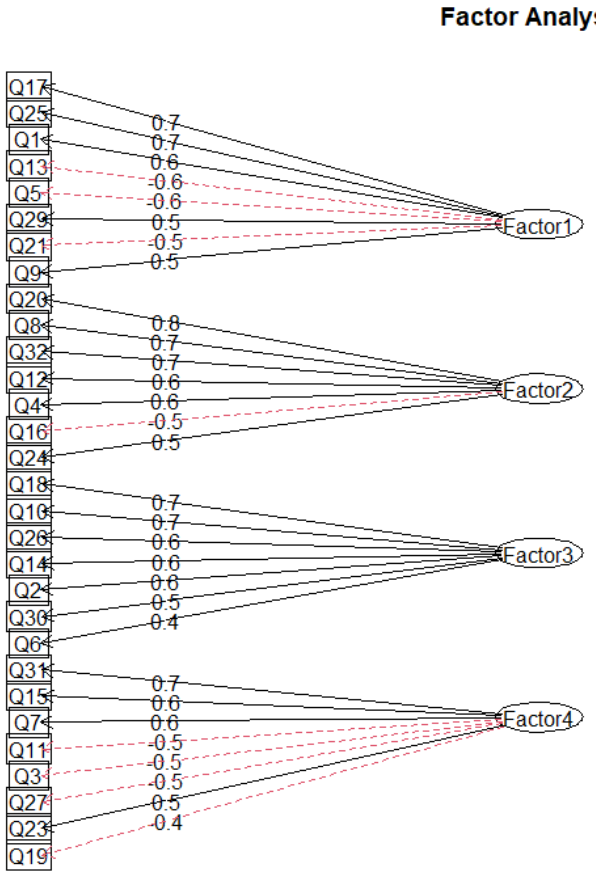

Slika 7.15: Grafički prikaz faktora

Iz 7.15 možemo vidjeti na koji način su se varijable smjestile u koordinatni sustav s obzirom na sva 4 faktora.
Pogledajmo sada što dobivamo rotacijama kako bismo lakše interpretirali povezanost varijabli s faktorima. Prvo ćemo analizirati ortogonalnu rotaciju faktora. U sljedećoj tablici navedene su nove težine, komunaliteti i svojstvene vrijednosti nakon ortogonalne rotacije.

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Komunaliteti |
|---|---|---|---|---|---|
| Q1 | 0.65 | -0.07 | -0.13 | 0.08 | 0.447815 |
| Q2 | -0.20 | 0.06 | 0.58 | -0.06 | 0.379544 |
| Q3 | -0.16 | 0.17 | 0.1 | -0.51 | 0.323465 |
| Q4 | -0.03 | 0.61 | 0.07 | -0.06 | 0.376884 |
| Q5 | -0.57 | 0.1 | 0.21 | -0.07 | 0.383295 |
| Q6 | -0.31 | 0.04 | 0.43 | -0.07 | 0.285366 |
| Q7 | -0.08 | -0.05 | 0.05 | 0.56 | 0.327671 |
| Q8 | -0.07 | 0.73 | 0.09 | -0.06 | 0.555355 |
| Q9 | 0.46 | -0.13 | -0.1 | 0.07 | 0.243162 |
| Q10 | -0.11 | 0.15 | 0.71 | -0.06 | 0.547135 |
| Q11 | 0.06 | 0.04 | 0.1 | -0.51 | 0.276304 |
| Q12 | -0.18 | 0.62 | 0.05 | -0.02 | 0.420533 |
| Q13 | -0.58 | 0.05 | 0.21 | -0.04 | 0.389553 |
| Q14 | -0.27 | 0.07 | 0.63 | -0.02 | 0.475062 |
| Q15 | 0.11 | -0.08 | 0.06 | 0.65 | 0.441650 |
| Q16 | 0.16 | -0.50 | 0.05 | 0.18 | 0.312054 |
| Q17 | 0.72 | -0.06 | -0.08 | 0.1 | 0.534225 |
| Q18 | -0.04 | 0.1 | 0.73 | -0.03 | 0.548514 |
| Q19 | -0.2 | 0.19 | 0.24 | -0.43 | 0.318838 |
| Q20 | 0.02 | 0.75 | 0.05 | -0.14 | 0.589424 |
| Q21 | -0.53 | 0.13 | 0.26 | 0.08 | 0.370666 |
| Q23 | 0.16 | 0 | -0.01 | 0.49 | 0.269487 |
| Q24 | 0.14 | 0.46 | 0.1 | 0 | 0.245559 |
| Q25 | 0.68 | 0.01 | -0.12 | 0.04 | 0.472905 |
| Q26 | -0.20 | 0.11 | 0.65 | 0.00 | 0.471859 |
| Q27 | 0 | 0.08 | 0.08 | -0.5 | 0.260163 |
| Q29 | 0.55 | 0.14 | -0.03 | 0.18 | 0.348079 |
| Q30 | -0.06 | -0.03 | 0.45 | -0.01 | 0.210278 |
| Q31 | 0.14 | 0.03 | 0.07 | 0.67 | 0.477266 |
| Q32 | -0.13 | 0.66 | 0.13 | -0.06 | 0.469009 |
| Svojstvene vrijednosti | 3.336688 | 2.968562 | 2.935068 | 2.530801 | 11.771119 |

Tablica 7.27: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon ortogonalne rotacije

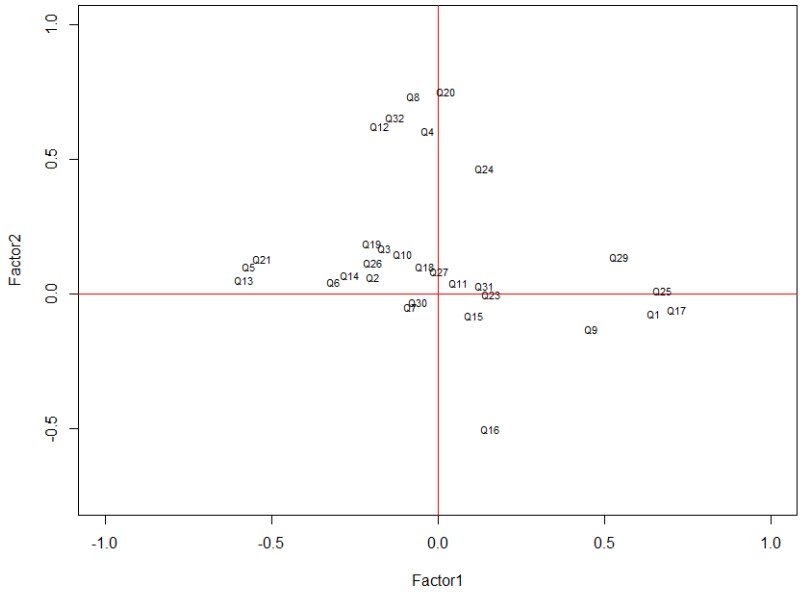
Sljedeći prikaz trebao bi olakšati interpretaciju grupiranja varijabli po faktorima koja su dana u prethodnoj tablici. Značajne težine koje su po apsolutnoj vrijednosti veće od 0.4 su naznačene za svaki od faktora i pripadnih mu varijabli.

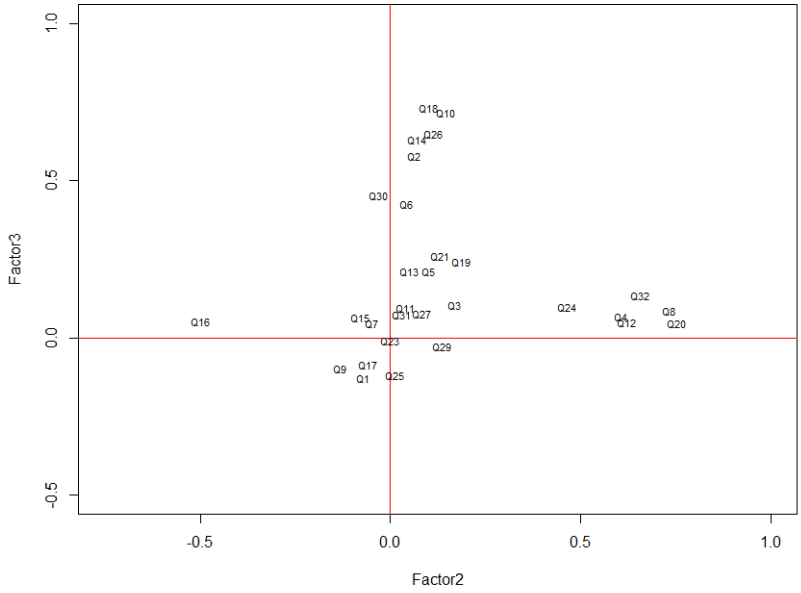


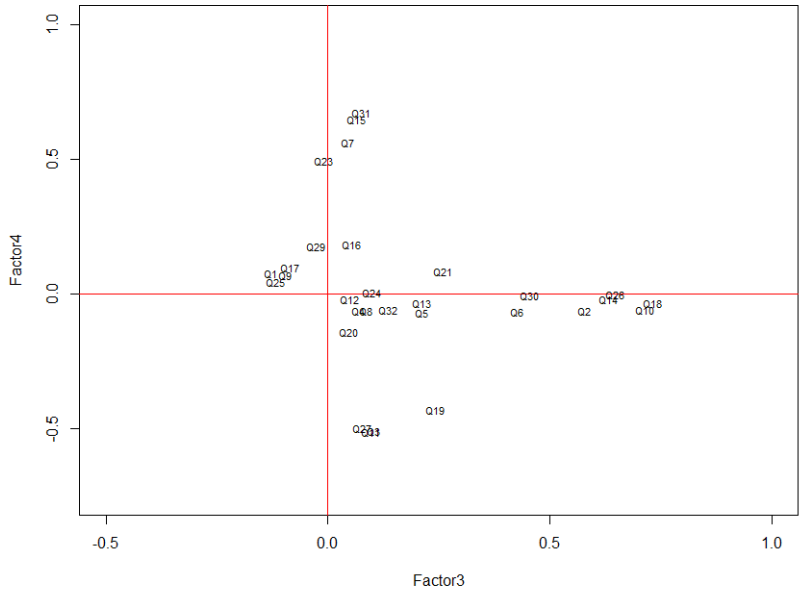

Slika 7.16: Grafički prikaz faktorske analize

Faktor 1 nazvat ćemo Afilijativni humor, faktor 2 Samoporažavajući humor, faktor 3 Agresivni humor te faktor 4 Samopoboljšavajući humor. Dodjeljivanje varijabli faktorima rađeno je na temelju najveće vrijednosti (gledajući apsolutnu vrijednost) u svakom retku matrice faktora. Kao što smo i prije komentirali, kako bi procijenili značajnost težine na nekom faktoru, vrijednosti od 0.4 su se pokazale dobrima u praksi. Također iz 7.27 vidimo procjene komunaliteta za svaku od varijabli. Na temelju dobivenih vrijednosti možemo reći da su komunaliteti prihvatljivi. Vidimo da su svojstvene vrijednosti (varijance objašnjene pojedinim faktorom) jednake zbroju kvadrata pripadnih faktorskih težina.
Također, komunaliteti se nisu promijenili nakon ortogonalne rotacije što smo i očekivali. Naime, prilikom ortogonalne rotacije dolazi do rotacije redaka matrice težina po faktorima. Varijance su se nakon rotacije po faktorima međusobno prbližno izjednačile u odnosu na varijance koje smo imali na nerotiranim faktorima. Također, vidimo da se nakon ortogonalne rotacije ukupna varijabilnost nije promijenila. Na slijedećim grafičkim prikazima

vidjet ćemo 30 varijabli prikazanih u odnosu na faktore nakon primjene ortogonalne rotacije.






Slika 7.17: Grafički prikaz faktora 3 i faktora 4

Iz 7.17 vidljiv je raspored varijabli po faktorima. Vidljivo je da varijable Q17, Q25, Q1, Q13, Q5, Q29, Q21 i Q9 imaju najveće težine na faktoru 1, a varijable Q20, Q8, Q32, Q12, Q4, Q16 i Q24 na faktoru 2. Varijable Q18, Q10, Q26, Q14, Q2, Q30 i Q6 imaju najveću težinu na faktoru 3 te varijable Q31, Q15, Q7, Q11, Q3, Q27, Q23 i Q29 na faktoru 4. Pogledajmo sada što dobivamo kosom rotacijom. Matrica međufaktorske korelacije je:

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 |
|---|---|---|---|---|
| Faktor1 | 1 | 0.2 | 0.22 | 0.47 |
| Faktor2 | 0.20 | 1.00 | 0.23 | 0.28 |
| Faktor3 | 0.22 | 0.23 | 1 | 0.15 |
| Faktor4 | 0.47 | 0.28 | 0.15 | 1 |

Tablica 7.28: Matrica međufaktorske korelacije

Vidimo da su nakon kose rotacije faktori korelirani budući da oni nisu više ortogonalni. Primijeniti ćemo kosu rotaciju primjenom metode promax na varijable.

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 |
|---|---|---|---|---|
| Q1 | 0.68 | 0.04 | -0.02 | 0.02 |
| Q2 | -0.06 | 0.59 | -0.03 | -0.01 |
| Q3 | -0.08 | 0.07 | 0.11 | -0.49 |
| Q4 | 0.03 | 0.01 | 0.62 | 0.01 |
| Q5 | -0.56 | 0.08 | 0.05 | -0.01 |
| Q6 | -0.23 | 0.39 | -0.03 | -0.01 |
| Q7 | -0.15 | 0.03 | -0.01 | 0.58 |
| Q8 | 0 | 0.01 | 0.75 | 0.03 |
| Q9 | 0.47 | 0.03 | -0.1 | 0.02 |
| Q10 | 0.08 | 0.76 | 0.04 | 0.01 |
| Q11 | 0.16 | 0.13 | -0.02 | -0.52 |
| Q12 | -0.14 | -0.05 | 0.64 | 0.07 |
| Q13 | -0.59 | 0.08 | 0 | 0.02 |
| Q14 | -0.13 | 0.63 | -0.03 | 0.04 |
| Q15 | 0.05 | 0.10 | -0.03 | 0.66 |
| Q16 | 0.13 | 0.15 | -0.51 | 0.12 |
| Q17 | 0.76 | 0.1 | -0.01 | 0.04 |
| Q18 | 0.15 | 0.8 | 0 | 0.02 |
| Q19 | -0.09 | 0.21 | 0.11 | -0.4 |
| Q20 | 0.11 | -0.01 | 0.77 | -0.06 |
| Q21 | -0.52 | 0.14 | 0.09 | 0.16 |
| Q23 | 0.12 | 0.03 | 0.05 | 0.5 |
| Q24 | 0.21 | 0.1 | 0.48 | 0.05 |
| Q25 | 0.72 | 0.05 | 0.06 | -0.01 |
| Q26 | -0.05 | 0.67 | 0.02 | 0.06 |
| Q27 | 0.09 | 0.09 | 0.03 | -0.5 |
| Q29 | 0.59 | 0.1 | 0.19 | 0.16 |
| Q30 | 0.05 | 0.5 | -0.1 | 0.02 |
| Q31 | 0.09 | 0.11 | 0.09 | 0.69 |
| Q32 | -0.06 | 0.06 | 0.66 | 0.03 |

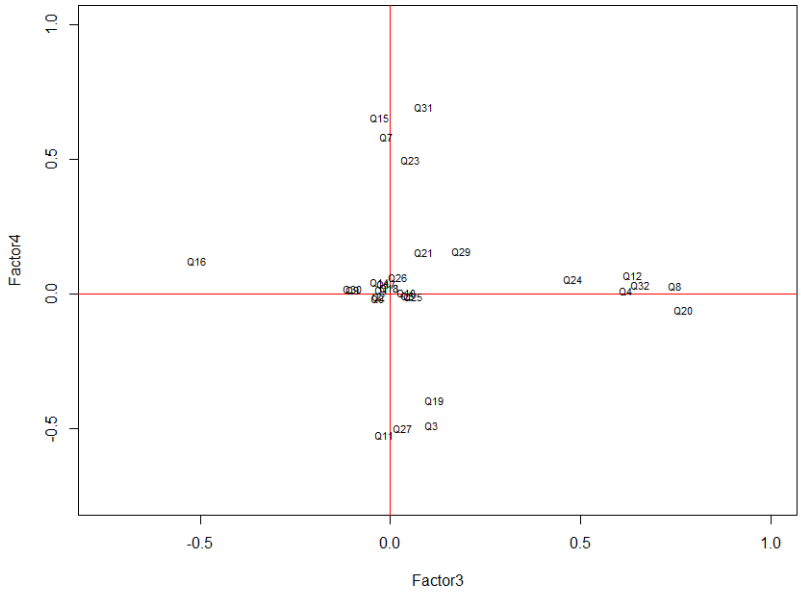
Tablica 7.29: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti nakon kose rotacije

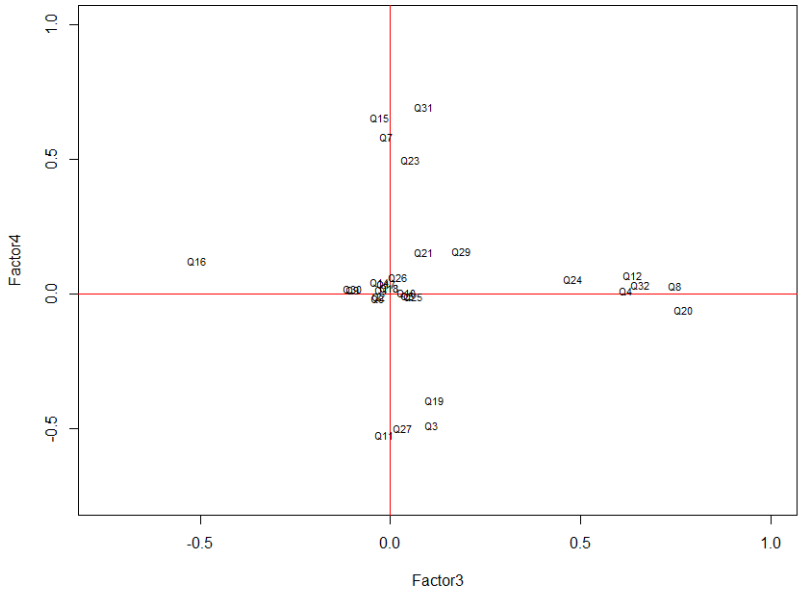
Pogledajmo matricu faktorske strukture kako bi se ispitala korelacija između varijabli i faktora.

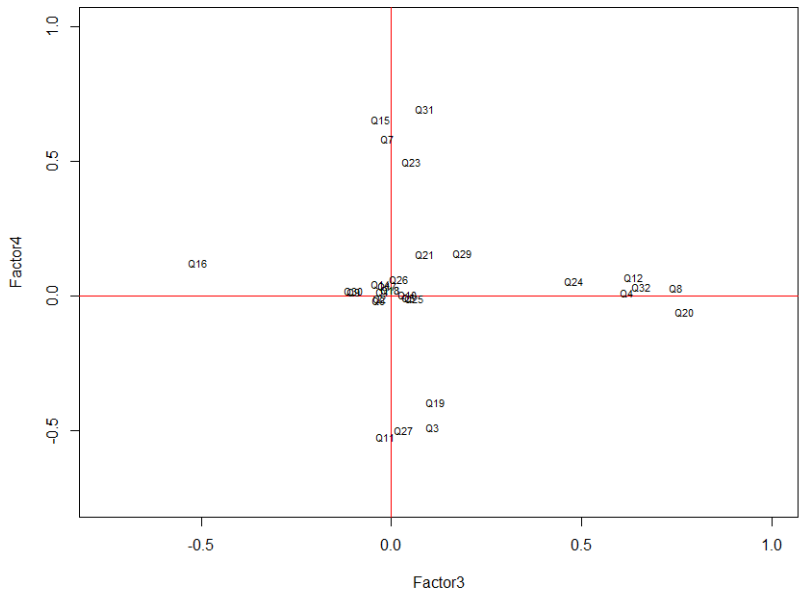
| | Faktor1 | Faktor2 | Faktor3 | Faktor4 |
|---|---|---|---|---|
| Q1 | 0.6720 | 0.1699 | 0.1360 | 0.3413 |
| Q2 | -0.1895 | 0.5703 | 0.0906 | 0.1229 |
| Q3 | -0.2969 | -0.0618 | 0.0334 | -0.4900 |
| Q4 | 0.1719 | 0.1659 | 0.6310 | 0.1244 |
| Q5 | -0.5735 | -0.0262 | -0.0613 | -0.2439 |
| Q6 | -0.3239 | 0.3328 | 0.0025 | -0.0191 |
| Q7 | 0.1134 | 0.1561 | 0.0490 | 0.5177 |
| Q8 | 0.1748 | 0.1897 | 0.7549 | 0.1448 |
| Q9 | 0.4519 | 0.1019 | 0.0125 | 0.2310 |
| Q10 | -0.0604 | 0.7898 | 0.2361 | 0.2633 |
| Q11 | -0.1141 | 0.0106 | -0.0295 | -0.4138 |
| Q12 | 0.0433 | 0.0898 | 0.6049 | 0.0869 |
| Q13 | -0.5948 | -0.0341 | -0.1124 | -0.2323 |
| Q14 | -0.2425 | 0.6118 | 0.0941 | 0.1551 |
| Q15 | 0.3316 | 0.2857 | 0.1026 | 0.7028 |
| Q16 | 0.0485 | 0.0890 | -0.4281 | 0.1496 |
| Q17 | 0.7567 | 0.2609 | 0.1848 | 0.4220 |
| Q18 | 0.0020 | 0.8396 | 0.2185 | 0.3187 |
| Q19 | -0.2941 | 0.1064 | 0.0832 | -0.3625 |
| Q20 | 0.2516 | 0.1731 | 0.7834 | 0.1058 |
| Q21 | -0.4603 | 0.0956 | 0.0261 | -0.0399 |
| Q23 | 0.3552 | 0.2006 | 0.1573 | 0.5672 |
| Q24 | 0.3231 | 0.2730 | 0.5589 | 0.2560 |
| Q25 | 0.7191 | 0.2014 | 0.2279 | 0.3515 |
| Q26 | -0.1465 | 0.6780 | 0.1714 | 0.2305 |
| Q27 | -0.1531 | -0.0266 | -0.0030 | -0.4273 |
| Q29 | 0.6836 | 0.3048 | 0.3626 | 0.4910 |
| Q30 | -0.0645 | 0.4892 | 0.0288 | 0.1670 |
| Q31 | 0.4159 | 0.3409 | 0.2375 | 0.7795 |
| Q32 | 0.0895 | 0.2051 | 0.6629 | 0.1198 |
| Svojstvene vrijednosti | 4.209346 | 3.594242 | 3.389846 | 3.693054 |

Tablica 7.30: Matrica faktorske strukture

Na slijedećim grafičkim prikazima vidjet ćemo 30 varijabli prikazanih u odnosu na faktore nakon primjene kose rotacije.

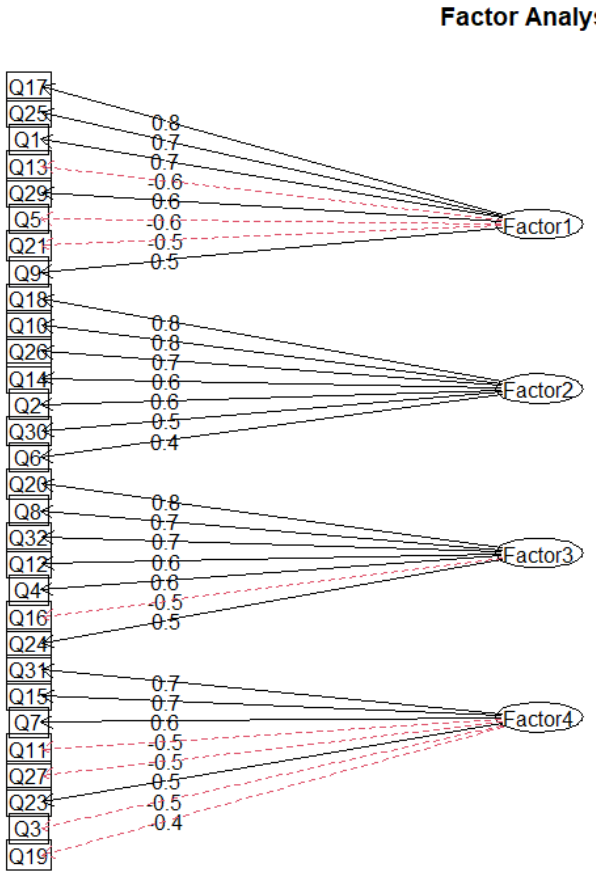

Factor4
Factor3
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
Q31
Q15
Q7
Q23
Q16
Q21
Q29
Q24
Q12
Q32
Q8
Q4
Q20
Q19
Q11
Q27
Q3


Factor4
Factor3
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
Q31
Q15
Q7
Q23
Q16
Q21
Q29
Q24
Q12
Q32
Q8
Q4
Q20
Q19
Q11
Q27
Q3




Slika 7.18: Grafički prikaz faktora

Radi lakše interpretacije podataka u 7.18 prikazat ćemo grafički odnose faktora i varijabli.




Vidimo da je na ovom primjeru uspješno provedena faktorska analiza budući da su se varijable lijepo rasporedile po faktorima i komunalitet je visok. Uspješno smo od 30 varijabli dobili manji broj, odnosno 4 latentna faktora. Prvi faktor (Afilijativni humor) čine varijable Q17, Q25, Q1, Q13, Q5, Q29, Q21 i Q9. Drugi faktor (Samoporažavajući humor) čine varijable Q20, Q8, Q32, Q12, Q4, Q16 i Q24. Treći faktor (Agresivni humor) čine varijable Q18, Q10, Q26, Q14, Q2, Q30 i Q6. Četvrti faktor (Samopoboljšavajući humor) čine varijable Q31, Q15, Q7, Q11, Q3, Q27, Q23 i Q29.

# Poglavlje 8

# Dodatak A

U ovom poglavlju nalaze se tablice podataka koje se odnose na Primjer 3.

| | |
|---|---|
| Q1 | I usually don't laugh or joke around much with other people. |
| Q2 | If I am feeling depressed, I can usually cheer myself up with humor. |
| Q3 | If someone makes a mistake, I will often tease them about it. |
| Q4 | I let people laugh at me or make fun at my expense more than I should. |
| Q5 | I don't have to work very hard at making other people laugh—I seem to be a naturally humorous person. |
| Q6 | Even when I'm by myself, I'm often amused by the absurdities of life. |
| Q7 | People are never offended or hurt by my sense of humor. |
| Q8 | I will often get carried away in putting myself down if it makes my family or friends laugh. |
| Q9 | I rarely make other people laugh by telling funny stories about myself. |
| Q10 | If I am feeling upset or unhappy I usually try to think of something funny about the situation to make myself feel better. |
| Q11 | When telling jokes or saying funny things, I am usually not very concerned about how other people are taking it. |
| Q12 | I often try to make people like or accept me more by saying something funny about my own weaknesses, blunders, or faults. |
| Q13 | I laugh and joke a lot with my closest friends. |
| Q14 | My humorous outlook on life keeps me from getting overly upset or depressed about things. |
| Q15 | I do not like it when people use humor as a way of criticizing or putting someone down. |
| Q16 | I don't often say funny things to put myself down. |
| Q17 | I usually don't like to tell jokes or amuse people. |
| Q18 | If I'm by myself and I'm feeling unhappy, I make an effort to think of something funny to cheer myself up. |
| Q19 | Sometimes I think of something that is so funny that I can't stop myself from saying it, even if it is not appropriate for the situation. |
| Q20 | I often go overboard in putting myself down when I am making jokes or trying to be funny. |
| Q21 | I enjoy making people laugh. |
| Q22 | If I am feeling sad or upset, I usually lose my sense of humor. |
| Q23 | I never participate in laughing at others even if all my friends are doing it. |
| Q24 | When I am with friends or family, I often seem to be the one that other people make fun of or joke about. |
| Q25 | I don't often joke around with my friends. |
| Q26 | It is my experience that thinking about some amusing aspect of a situation is often a very effective way of coping with problems. |
| Q27 | If I don't like someone, I often use humor or teasing to put them down. |
| Q28 | If I am having problems or feeling unhappy, I often cover it up by joking around, so that even my closest friends don't know how I really feel. |
| Q29 | I usually can't think of witty things to say when I'm with other people. |
| Q30 | I don't need to be with other people to feel amused – I can usually find things to laugh about even when I'm by myself. |
| Q31 | Even if something is really funny to me, I will not laugh or joke about it if someone will be offended. |
| Q32 | Letting others laugh at me is my way of keeping my friends and family in good spirits. |

Tablica 8.1: Objašnjenje varijabli

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Faktor5 | Komunaliteti |
|---|---|---|---|---|---|---|
| Q1 | -0.51 | 0.24 | 0.20 | 0.28 | 0.02 | 0.442827 |
| Q2 | 0.5 | -0.22 | 0.19 | 0.23 | -0.16 | 0.4157239 |
| Q3 | 0.39 | 0.18 | -0.33 | 0.18 | 0.12 | 0.3369458 |
| Q4 | 0.36 | 0.46 | 0.15 | -0.12 | 0 | 0.3763158 |
| Q5 | 0.53 | -0.22 | -0.11 | -0.2 | 0.17 | 0.4156529 |
| Q6 | 0.47 | -0.22 | 0.07 | 0.07 | 0.1 | 0.2899431 |
| Q7 | -0.12 | -0.26 | 0.39 | -0.3 | -0.02 | 0.3247016 |
| Q8 | 0.45 | 0.53 | 0.18 | -0.17 | -0.05 | 0.5510948 |
| Q9 | -0.41 | 0.11 | 0.12 | 0.21 | 0.11 | 0.2543933 |
| Q10 | 0.57 | -0.15 | 0.33 | 0.33 | -0.14 | 0.579338 |
| Q11 | 0.19 | 0.18 | -0.3 | 0.36 | 0.22 | 0.3391576 |
| Q12 | 0.43 | 0.4 | 0.13 | -0.24 | -0.04 | 0.4271884 |
| Q13 | 0.51 | -0.27 | -0.11 | -0.22 | 0.10 | 0.3987612 |
| Q14 | 0.56 | -0.27 | 0.23 | 0.18 | -0.07 | 0.4763377 |
| Q15 | -0.26 | -0.22 | 0.53 | -0.22 | 0.14 | 0.4629079 |
| Q16 | -0.35 | -0.4 | 0.07 | 0.17 | 0.32 | 0.4238561 |
| Q17 | -0.54 | 0.27 | 0.26 | 0.33 | 0.02 | 0.5361946 |
| Q18 | 0.51 | -0.17 | 0.36 | 0.37 | -0.07 | 0.5579981 |
| Q19 | 0.48 | 0.12 | -0.22 | 0.2 | 0.31 | 0.4277914 |
| Q20 | 0.40 | 0.62 | 0.15 | -0.10 | 0 | 0.5807991 |
| Q21 | 0.51 | -0.24 | 0.05 | -0.26 | 0.28 | 0.4655295 |
| Q22 | -0.3 | 0.15 | 0.03 | -0.12 | 0.35 | 0.2502252 |
| Q23 | -0.25 | -0.07 | 0.42 | -0.18 | 0.21 | 0.3179891 |
| Q24 | 0.19 | 0.4 | 0.24 | 0 | 0.16 | 0.2800136 |
| Q25 | -0.48 | 0.33 | 0.21 | 0.3 | 0.11 | 0.4843034 |
| Q26 | 0.55 | -0.21 | 0.29 | 0.21 | 0.05 | 0.4753302 |
| Q27 | 0.24 | 0.2 | -0.3 | 0.29 | 0.23 | 0.3208406 |
| Q28 | 0.44 | 0.14 | 0.04 | 0.07 | 0.22 | 0.2684765 |
| Q29 | -0.33 | 0.3 | 0.35 | 0.16 | 0.05 | 0.3508221 |
| Q30 | 0.29 | -0.2 | 0.19 | 0.22 | 0.13 | 0.2222783 |
| Q31 | -0.23 | -0.13 | 0.59 | -0.25 | 0.11 | 0.4944017 |
| Q32 | 0.48 | 0.43 | 0.17 | -0.16 | 0 | 0.4696181 |

Tablica 8.2: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti

| | Faktor1 | Faktor2 | Faktor3 | Faktor4 | Komunaliteti |
|---|---|---|---|---|---|
| Q1 | -0.52 | 0.24 | 0.19 | 0.28 | 0.447031 |
| Q2 | 0.51 | -0.2 | 0.21 | 0.22 | 0.3898279 |
| Q3 | 0.38 | 0.16 | -0.33 | 0.20 | 0.3184788 |
| Q4 | 0.36 | 0.47 | 0.12 | -0.11 | 0.3763512 |
| Q5 | 0.53 | -0.21 | -0.11 | -0.20 | 0.3834022 |
| Q6 | 0.47 | -0.21 | 0.08 | 0.07 | 0.2805482 |
| Q7 | -0.11 | -0.22 | 0.4 | -0.33 | 0.3263348 |
| Q8 | 0.45 | 0.55 | 0.14 | -0.16 | 0.5539539 |
| Q9 | -0.42 | 0.11 | 0.13 | 0.2 | 0.2439201 |
| Q10 | 0.57 | -0.13 | 0.34 | 0.32 | 0.5532062 |
| Q11 | 0.18 | 0.15 | -0.29 | 0.36 | 0.2716209 |
| Q12 | 0.43 | 0.42 | 0.1 | -0.23 | 0.4208709 |
| Q13 | 0.51 | -0.27 | -0.1 | -0.22 | 0.3862897 |
| Q14 | 0.57 | -0.25 | 0.25 | 0.18 | 0.4792541 |
| Q15 | -0.25 | -0.19 | 0.53 | -0.26 | 0.4452244 |
| Q16 | -0.35 | -0.39 | 0.1 | 0.14 | 0.3091304 |
| Q17 | -0.54 | 0.27 | 0.26 | 0.32 | 0.5362626 |
| Q18 | 0.51 | -0.15 | 0.38 | 0.36 | 0.5509928 |
| Q19 | 0.46 | 0.1 | -0.21 | 0.2 | 0.3083244 |
| Q20 | 0.4 | 0.64 | 0.1 | -0.09 | 0.5900571 |
| Q21 | 0.5 | -0.22 | 0.05 | -0.25 | 0.3681084 |
| Q22 | -0.3 | 0.14 | 0.02 | -0.12 | 0.1278526 |
| Q23 | -0.25 | -0.04 | 0.41 | -0.2 | 0.2739756 |
| Q24 | 0.18 | 0.41 | 0.21 | 0.00 | 0.2473512 |
| Q25 | -0.49 | 0.33 | 0.2 | 0.29 | 0.4715178 |
| Q26 | 0.55 | -0.19 | 0.31 | 0.20 | 0.4703384 |
| Q27 | 0.22 | 0.17 | -0.3 | 0.29 | 0.2531853 |
| Q29 | -0.34 | 0.32 | 0.34 | 0.15 | 0.3509442 |
| Q30 | 0.29 | -0.18 | 0.21 | 0.21 | 0.2059363 |
| Q31 | -0.22 | -0.09 | 0.59 | -0.28 | 0.4833411 |
| Q32 | 0.48 | 0.45 | 0.13 | -0.15 | 0.4689524 |

Tablica 8.3: Težine po faktorima, procjene komunaliteta i svojstvene vrijednosti

# Poglavlje 9

# Dodatak B

Kod u R-u za Primjer 1.

```
#deskriptivna statistika podataka
describe(data3)

#graficki prikaz matrice korelacija
corrplot(cor(data3), order = "original", tl.col='black', tl.cex=.75)
corrplot(cor(data3, use="complete.obs"), order = "hclust", tl.col='black
    ', tl.cex=.75)

cor_subtests3 <- cor(data3)
round(cor_subtests3, 2)

#KMO mjera
KMO(data3)

ev3 <- eigen(cor(data3));
ev3$values

#scree plot
qplot(c(1:16),ev3$values)+
  geom_line() +
  xlab("Factors") +
  ylab("Eigen values of factors") +
  ggtitle("Scree Plot")+
  geom_hline(yintercept =1,col="red")

#faktorska analiza bez primjene rotacije
Nfacs<-2;
fit1 <-factanal(data3,Nfacs, cor=TRUE,rotation="none")
print(fit1, digits=2, cutoff=0, sort=TRUE)

apply(fit1$loadings^2, 1, sum)
```

```
sum(apply(fit1$loadings^2, 1, sum))

#svojstvene vrijednosti
svr1<-c(sum(as.vector(fit1$loadings[,1])^2),sum(as.vector(fit1$loadings
    [,2])^2));
svr1

load1 <- fit1$loadings[,1:2]
plot(load1,type="n")
abline(h=0,col="red")
abline(v=0,col="red")
text(load1,labels=names(data3),cex=.7)

#faktorska analiza uz primjenu varimax metode
fit2 <- factanal(data3, Nfacs, rotation="varimax")
print(fit2, digits=2, cutoff=0, sort=TRUE)

apply(fit2$loadings^2, 1, sum)
sum(apply(fit2$loadings^2, 1, sum))

svr2<-c(sum(as.vector(fit2$loadings[,1])^2),sum(as.vector(fit2$loadings
    [,2])^2));
svr2

load2 <- fit2$loadings[,1:2]
plot(load2,type="n")
abline(h=0,col="red")
abline(v=0,col="red")
text(load2,labels=names(data3),cex=.7)

#faktorska analiza uz primjenu promax metode
fit3 <- factanal(data3, Nfacs, rotation="promax")
print(fit3, digits=2, cutoff=0, sort=TRUE)

apply(fit2$loadings^2, 1, sum)
sum(apply(fit3$loadings^2, 1, sum))
svr3<-c(sum(as.vector(fit3$loadings[,1])^2),sum(as.vector(fit3$loadings
    [,2])^2));
svr3

#racunanje matrice faktorske strukture
M11<-as.matrix(fit3$loadings[,1:2]);
M11

M21<- matrix(0,2,2);
M21[1,]<-c(1.0,-0.22);
M21[2,]<-c(-0.22,1.0);
```

```
M21

MF1<-M11%*%M21;
MF1

sv11<-sum(as.vector(MF1[,1])^2);
sv11

sv21<-sum(as.vector(MF1[,2])^2);
sv21

load3 <- fit3$loadings[,1:2]
plot(load3,type="n")
abline(h=0,col="red")
abline(v=0,col="red")
text(load3,labels=names(data3),cex=.7)

loadp <- MF1;
plot(loadp,type="n")
abline(h=0,col="red")
abline(v=0,col="red")
text(loadp,labels=names(data3),cex=.7)

#graficki prikaz raspodjele tezina izmedu varijabli i faktora
fa.diagram(load3)
```

# Bibliografija


[1] Richard Arnold Johnson, Dean W. Wichern, *Applied multivariate statistical analysis, 6th Edition*, Pearson Prentice Hall, 2007.

[2] Alvin C. Rencher, *Methods of multivariate analysis*, J. Wiley & Sons, 2002.

[3] Anderberg,M.R., *Cluster Analysis for Applications*, Academic Press, I973.

[4] Kaiser, H. F., Rice, J. *Educational and Psychological Measurement, Little Jiffy, Mark Iv.,* , 1974.

[5] Hair, J., Anderson, R., Tatham, R. and Black, W. *Multivariate Data Analysis, 5th Edition*, Prentice Hall, 1988.

[6] James P. Stevens, *Applied Multivariate Statistics for the Social Sciences, 5th Edition*, Routledge, 2009.

[7] https://openmv.net/info/peas, (pristupljeno: siječanj 2024.)

[8] https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction, (pristupljeno: siječanj 2024.)

[9] https://www.kaggle.com/datasets/lucasgreenwell/humor-styles-questionnaire-responses, (pristupljeno: siječanj 2024.)

# Sažetak

U ovom diplomskom radu proučavali smo teorijsku pozadinu faktorske analize te njenu primjenu na konkretnom skupu podataka. Faktorska analiza je statistička metoda koja kao cilj ima pronalaženje manjeg broja neopaženih varijabli (faktora) koje objašnjavaju što veći dio varijabilnosti početnog skupa podataka. Početne varijable koje su korelirane tada možemo prikazati kao linearne kombinacije nekoreliranih faktora, iz čijih težina možemo uvidjeti koliki je utjecaj određenog faktora na neku od varijabli.
Na samome početku smo definirali faktorski model te objasnili od kojih se sve komponenti sastoji. Zatim smo proučavali razne metode procjene faktora koje su sve imale isti cilj, a to je ekstrakcija maksimalne varijance iz skupa podataka sa svakim faktorom. Kod odabira broja faktora u modelu uvidjeli smo da postoje razni kriteriji pomoću kojih možemo odrediti broj potrebnih faktora, koji su na kraju davali identične rezultate. Nakon što smo došli do rješenje nerijetko je teško interpretirati dobivene rezultate pa smo se u takvim slučajevima oslanjali na rotacije koje su nam olakšale donošenje zaključaka na način da su poboljšavale samu interpretaciju našeg rješenja. Na samome kraju, dali smo tri primjera u kojima smo sproveli faktorsku analizu.

# Summary

In this master's thesis, we examined the theoretical background of factor analysis and its application to a specific dataset. Factor analysis is a statistical method aimed at finding a smaller number of unobserved variables (factors) that explain a larger portion of the variability in the original dataset. The initial correlated variables can then be represented as linear combinations of uncorrelated factors, from which we can determine the influence of a particular factor on each variable based on their weights.
At the outset, we defined the factor model and explained its components. We then explored various factor estimation methods, all with the same goal of extracting maximum variance from the dataset with each factor. When selecting the number of factors in the model, we found various criteria to determine the necessary number of factors, which ultimately yielded identical results. After obtaining the solution, interpreting the results can often be challenging, so in such cases, we relied on rotations to facilitate decision-making by improving the interpretation of our solution. Finally, we provided three examples where we conducted factor analysis.

# Životopis

Rođena sam 22. lipnja 1996. godine u Zagrebu. Osnovnu školu Dobriše Cesarića u Zagrebu krenula sam pohađati 2003. godine, a svoje srednjoškolsko obrazovanje započela sam 2011. godine u XV.gimnaziji u Zagrebu. Nakon završene srednje škole, 2015. godine upisujem preddiplomski sveučilišni studij Matematika na Prirodoslovno-matematičkom fakultetu u Zagrebu na kojem 2021. godine stječem titulu univ. bacc. math. Iste sam godine upisala diplomski studij Matematičke statistike, također na Prirodoslovno-matematičkom fakultetu u Zagrebu.