FAKTORSKA ANALIZA

Vježba 12

Faktorska analiza

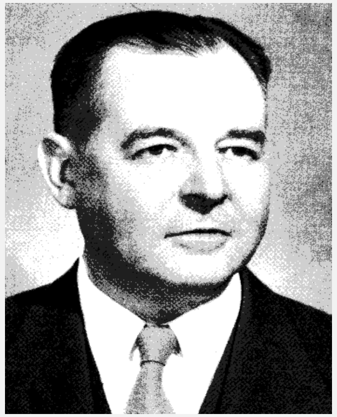
Faktorska analiza je zajedničko ime za više metoda kojima je cilj kondenzacija većeg broja **manifestnih varijabli**, među kojima postoji povezanost (korelacija), na manji broj **latentnih dimenzija (faktora)** koje su izvor te povezanosti.

Manifestne varijable su varijable, odnosno obilježja koja se mogu izravno mjeriti postojećim mjernim instrumentima (npr. skok udalj s mjesta, bacanje medicinke iz ležanja na leđima, bacanje kugle).

Latentne dimenzije su varijable, odnosno obilježja koja nisu izravno mjerljiva postojećim mjernim instrumentima, već se izračunavaju kao linearna kombinacija manifestnih varijabli (npr. eksplozivna snaga).

Komponentni model faktorske analize

Komponentni model ili **metoda glavnih komponenata** je metoda ekstrakcije, odnosno izračunavanja latentnih dimenzija koju je 1933. godine predložio američki ekonomist i statističar **Harold Hotelling.**



Harold Hotelling
(1895. – 1973.)

Metodom glavnih komponenata se iz skupa od *m* manifestnih varijabli na temelju nereducirane korelacijske matrice izračuna *m* latentnih dimenzija koje su međusobno linearno nezavisne, a nazivaju se **glavne komponente**.

Glavne komponente su linearne kombinacije manifestnih varijabli izračunate na način da prva glavna komponenta objašnjava maksimalan moguć dio ukupne varijance manifestnih varijabli te da druga, kao i svaka sljedeća glavna komponenta, objašnjava najveći dio preostale varijance manifestnih varijabli, odnosno najveći dio varijance manifestnih varijabli koji nije objašnjen prethodnim glavnim komponentama.

Komponentni model faktorske analize

Glavne komponente izračunavaju se na sljedeći način:

Neka su u matrici B podaci skupa $E = \{e_i ; i = 1,\ldots,n\}$ entiteta koji su opisani skupom $V = \{v_j ; j = 1,\ldots,m\}$ manifestnih varijabli. Operacijom

$$\boldsymbol{Z} = \boldsymbol{B_c}\, \boldsymbol{V}^{-1}$$

gdje je

- $\boldsymbol{B_c}$ - matrica centriranih podataka dobivenih operacijom $\boldsymbol{B_c} = \boldsymbol{B}^{-1}\boldsymbol{m}^T$
- $\boldsymbol{m}$ - vektor aritmetičkih sredina varijabli matrice $\boldsymbol{B}$
- $\boldsymbol{V}^{-1}$ - dijagonalna matrica standardnih devijacija varijabli iz matrice $\boldsymbol{B}$

izračuna se matrica standardiziranih podataka ($\boldsymbol{Z}$).

Komponentni model faktorske analize

Operacijom

$$\boldsymbol{R} = \boldsymbol{Z}^{T}\boldsymbol{Z}n^{-1} = \begin{vmatrix} 1 & r_{12} & . & . & r_{1m} \\ r_{21} & 1 & . & . & r_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{m1} & r_{m2} & . & . & 1 \end{vmatrix}$$

izračuna se matrica korelacija manifestnih varijabli (***R***) u čijoj su glavnoj dijagonali jedinice, odnosno varijance standardiziranih varijabli. Zbroj varijanci svih standardiziranih varijabli, odnosno ukupna varijanca skupa standardiziranih varijabli jednaka je *m*, odnosno broju manifestnih varijabli.

Komponentni model faktorske analize

Operacijom spektralne dekompozicije matrice korelacija

$$\boldsymbol{R} = \boldsymbol{X} \quad \lambda \quad \boldsymbol{X^T}$$

$$\begin{vmatrix} 1 & r_{12} & . & . & r_{1m} \\ r_{21} & 1 & . & . & r_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{m1} & r_{m2} & . & . & 1 \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix} \cdot \begin{vmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & . & & \\ & & & . & \\ & & & & \lambda_m \end{vmatrix} \cdot \begin{vmatrix} x_{11} & x_{21} & . & . & x_{m1} \\ x_{12} & x_{22} & . & . & x_{m2} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{1m} & x_{2m} & . & . & x_{mm} \end{vmatrix}$$

izračunaju se **matrica svojstvenih vektora** (X) i **matrica svojstvenih vrijednosti** (λ).

Faktorska analiza

Komponentni model faktorske analize

Matrica svojstvenih vektora (**X**) je kvadratna matrica reda $m \times m$ za koju vrijedi da je $\mathbf{X}^T\mathbf{X} = \mathbf{X}\mathbf{X}^T = \mathbf{I}$ tj. da su vektori matrice linearno nezavisni, a u čijim su stupcima ponderi za izračunavanje glavnih komponenata iz standardiziranih varijabli.

Matrica svojstvenih vrijednosti (**λ**) je dijagonalna matrica reda m×m za koju vrijedi da je $\lambda_1 + \lambda_2 + \ldots + \lambda_m = m$, tj. da je zbroj svojstvenih vrijednosti jednak ukupnoj varijanci standardiziranih varijabli, te da je $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, tj. da se svojstvene vrijednosti nižu od najveće prema najmanjoj.

Komponentni model faktorske analize

Operacijom

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

$$\begin{vmatrix} k_{11} & k_{12} & . & . & k_{1m} \\ k_{21} & k_{22} & . & . & k_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ k_{n1} & k_{n2} & . & . & k_{nm} \end{vmatrix} = \begin{vmatrix} z_{11} & z_{12} & . & . & z_{1m} \\ z_{21} & z_{22} & . & . & z_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ z_{n1} & z_{n2} & . & . & z_{nm} \end{vmatrix} \cdot \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix}$$

izračuna se **matrica glavnih komponenata** (**K**).

Faktorska analiza

Komponentni model faktorske analize

Matrica glavnih komponenata (**K**) je matrica reda n×m koju čine rezultati n ispitanika u m linearno nezavisnih (ortogonalnih) glavnih komponenata. Varijanca prve glavne komponente jednaka je λ_1, varijanca druge glavne komponente λ_2, itd.

Pošto vrijedi da je $\lambda_1 + \lambda_2 + \ldots + \lambda_m = m$ te da je $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, može se zaključiti da je ukupna varijanca m standardiziranih varijabli raspodijeljena tako da prva glavna komponenta objašnjava najveći dio ukupne varijance, druga glavna komponenta najveći dio preostale varijance, itd.

Faktorska analiza

Komponentni model faktorske analize

Operacijom

$$\mathbf{H} = \mathbf{X} \quad \lambda^{1/2}$$

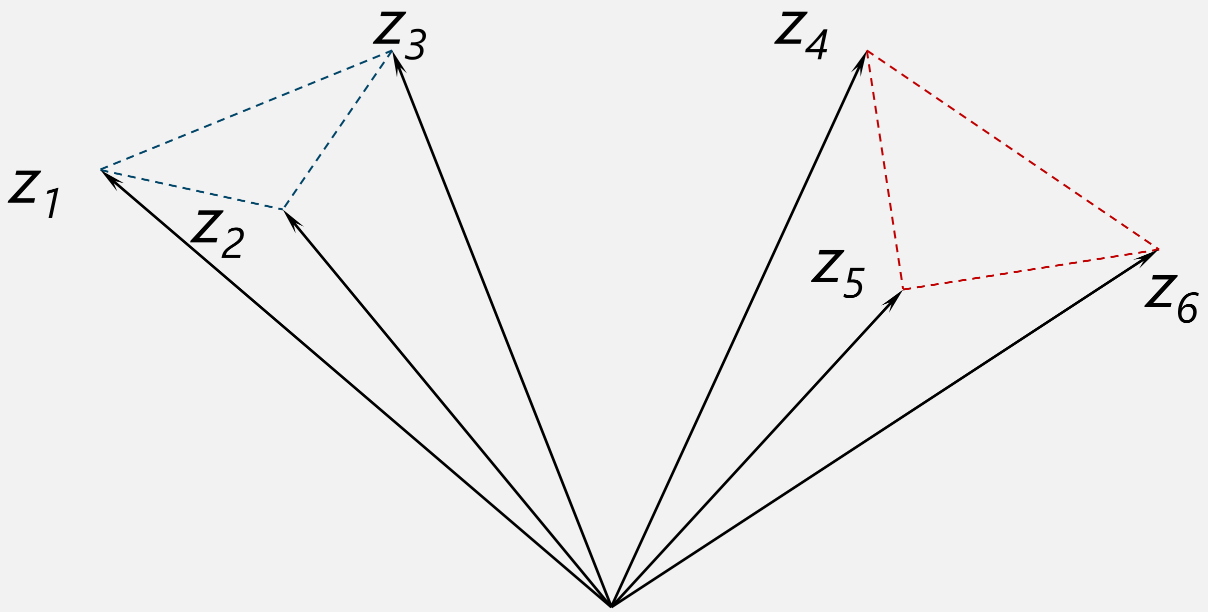
$$\begin{vmatrix} h_{11} & h_{12} & . & . & h_{1m} \\ h_{21} & h_{22} & . & . & h_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ h_{m1} & h_{m2} & . & . & h_{mm} \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & x_{2m} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{m1} & x_{m2} & . & . & x_{mm} \end{vmatrix} \cdot \begin{vmatrix} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & \\ & & . & & \\ & & & . & \\ & & & & \sqrt{\lambda_m} \end{vmatrix}$$

izračuna se **matrica glavnih osovina** (H). Matrica glavnih osovina je matrica korelacija manifestnih varijabli i glavnih komponenata. Elementi ove matrice ukazuju koliki je doprinos svake pojedine manifestne varijable pri formiranju svake pojedine glavne komponente.

Faktorska analiza

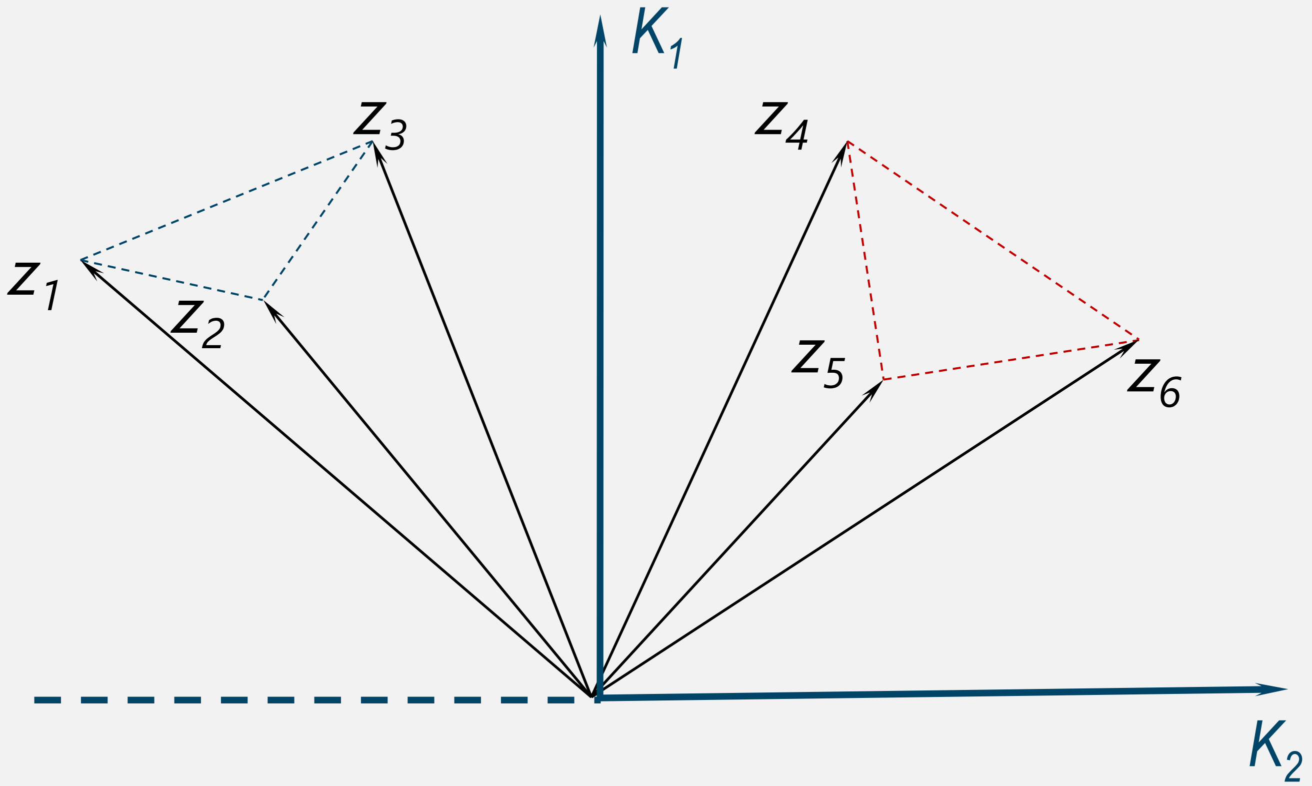
Komponentni model faktorske analize

Neka je skup od *n* entiteta opisan sa 6 manifestnih varijabli (v_1,...,v_6). Na temelju korelacijske matrice može se utvrditi odnos odgovarajućih standardiziranih vektora u *n*-dimenzionalnom prostoru.



Komponentni model faktorske analize

Na temelju matrice glavnih osovina može se utvrditi prostorni odnos standardiziranih manifestnih vektora i standardiziranih glavnih komponenata.



Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Pošto je osnovni cilj faktorske analize kondenzacija većeg broja manifestnih varijabli na manji broj latentnih dimenzija, potrebno je izvršiti **redukciju broja glavnih komponenata**.

Redukcija broja glavnih komponenata odnosno određivanje broja značajnih glavnih komponenata vrši se putem različitih kriterija kao što su **GK-kriterij**, **PB-kriterij** i **Scree-test**.

Prema **GK-kriteriju** glavna komponenta j je značajna ako je njena varijanca (λ_j) veća ili jednaka 1.

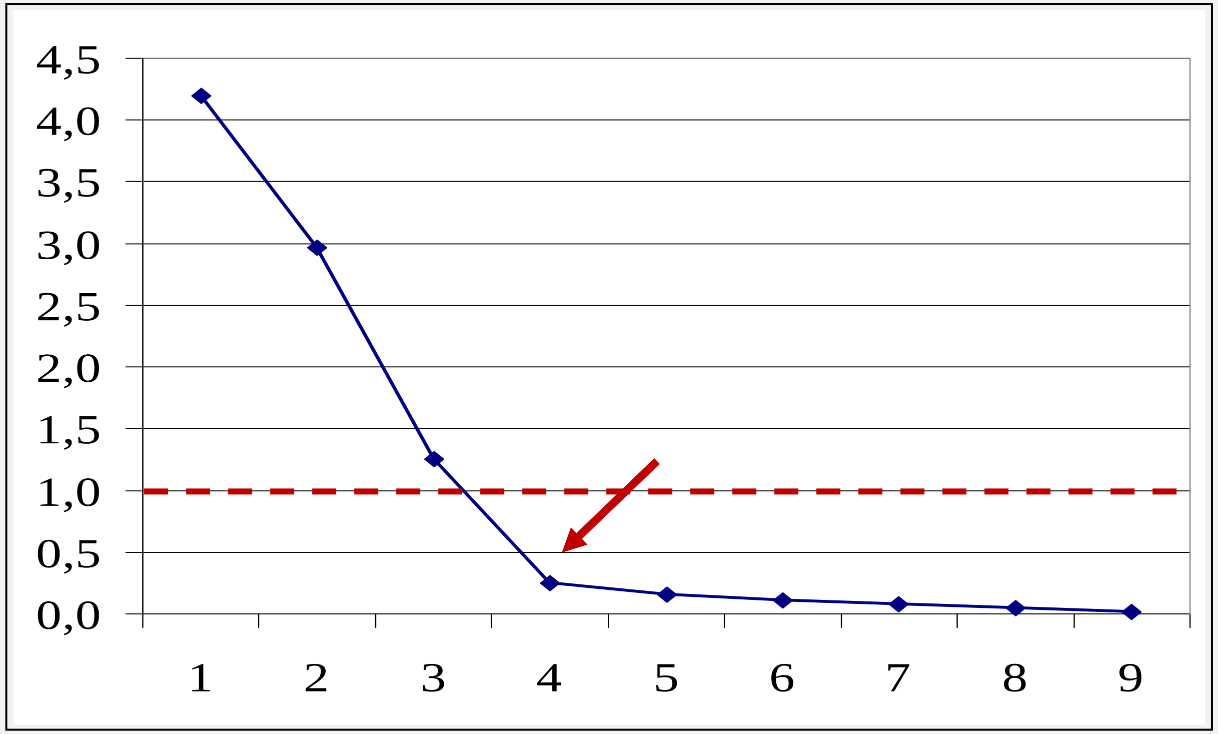
Prema **PB-kriteriju** broj značajnih glavnih komponenata jednak je broju svojstvenih vrijednosti poredanih po veličini čiji zbroj ne prelazi Σsmc (sumu kvadrata multiplih korelacija).

Scree-test je grafički kriterij. Na **scree plotu** se subjektivnom procjenom odredi točka nakon koje se svojstvene vrijednosti smanjuju u skladu s blagim linearnim trendom. Značajnima se smatraju sve prethodne glavne komponente.

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Scree plot



GK - kriterij

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Komunaliteti i unikviteti

$$\mathbf{H} = \left|\begin{array}{cccc|cccc} \multicolumn{4}{c|}{\overbrace{\hspace{6em}}^{\text{značajne glavne komponente}}} & \multicolumn{4}{c}{\overbrace{\hspace{6em}}^{\text{glavne komponente koje nisu značajne}}} \\ h_{11} & . & . & h_{1k} & . & . & . & h_{1m} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ h_{m1} & . & . & h_{mk} & . & . & . & h_{mm} \end{array}\right|$$

(Matrica **H** prije i nakon redukcije broja glavnih komponenata)

Faktorska analiza

Kriteriji za odabir značajnih glavnih komponenata

Komunaliteti i unikviteti

Varijancu svake manifestne varijable moguće je dekomponirati na **komunalitet** (h^2) i **unikvitet** (u^2) pri čemu je

$$s_j^2 = 1 = h_j^2 + u_j^2$$

Komunalitet je dio varijance manifestne varijable j koji je moguće objasniti s k značajnih glavnih komponenata.

$$h_j^2 = \sum_{p=1}^{k} h_{jp}^2 = h_{j1}^2 + h_{j2}^2 + \ldots + h_{jk}^2$$

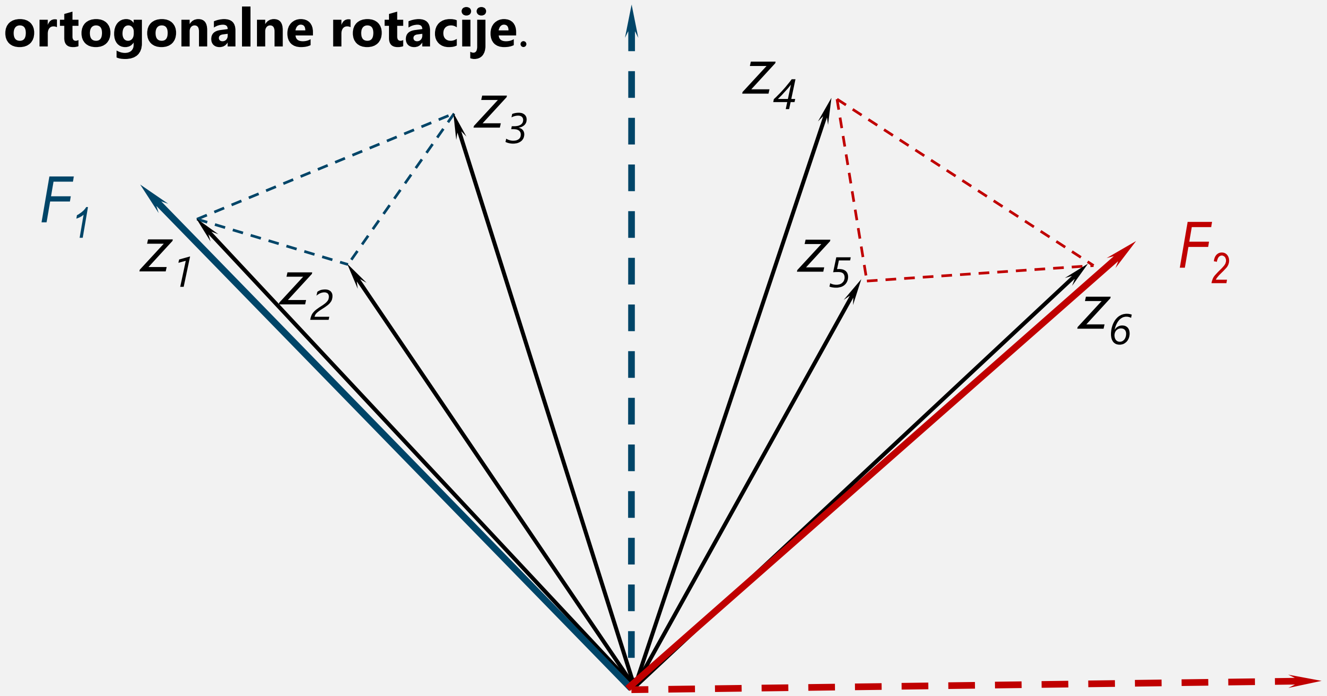
Unikvitet je dio varijance manifestne varijable j koji nije moguće objasniti s k značajnih glavnih komponenata

$$u_j^2 = 1 - h_j^2$$

Rotacije

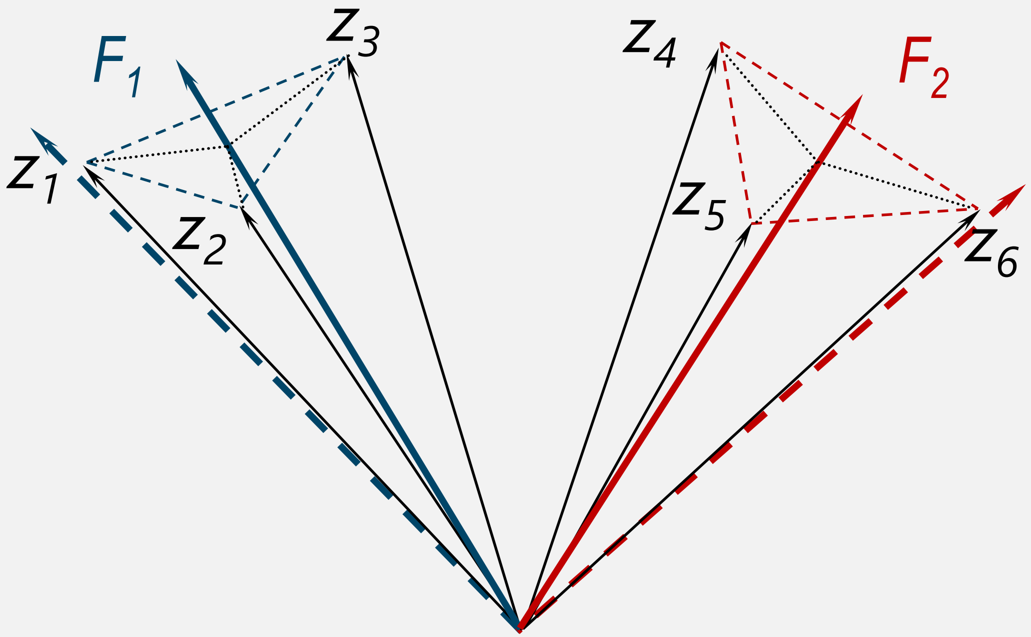
Pravi uvid u strukturu međusobnih odnosa manifestnih varijabli često nije moguće steći putem glavnih komponenata. U takvim se slučajevima koriste transformacije glavnih komponenata čija je svrha postizanje jednostavne faktorske strukture, a koje se nazivaju **rotacije**.

Transformacije glavnih komponenata koje se provode uz uvjet zadržavanja njihove linearne nezavisnosti nazivaju se **ortogonalne rotacije**.



Rotacije

Transformacije glavnih komponenata koje se provode bez uvjeta zadržavanja njihove linearne nezavisnosti nazivaju se **neortogonalne** ili **kosokutne rotacije**.

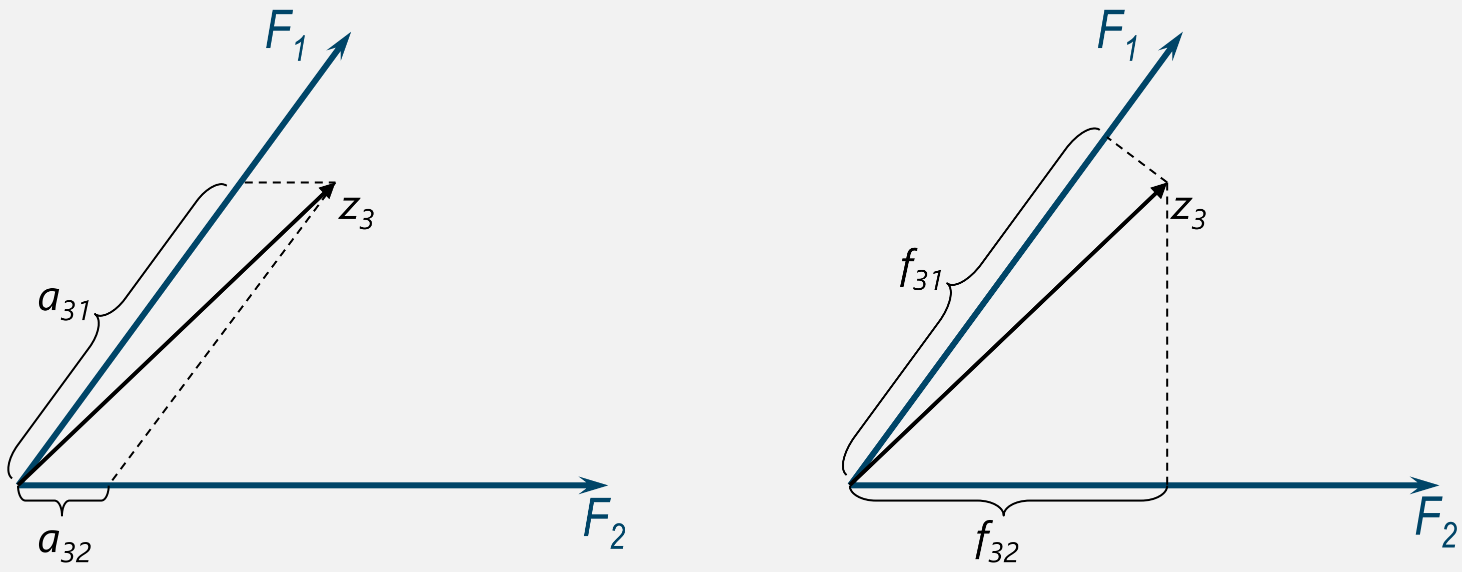


Rotacije

Interpretacija faktora nakon rotacije vrši se putem **matrice strukture** (**F**), **matrice sklopa** (**A**), i **matrice korelacija među faktorima** (**M**).

Matricu strukture čine korelacije manifestnih varijabli s faktorima, a **matricu sklopa** čine paralelne projekcije manifestnih varijabli na faktore.

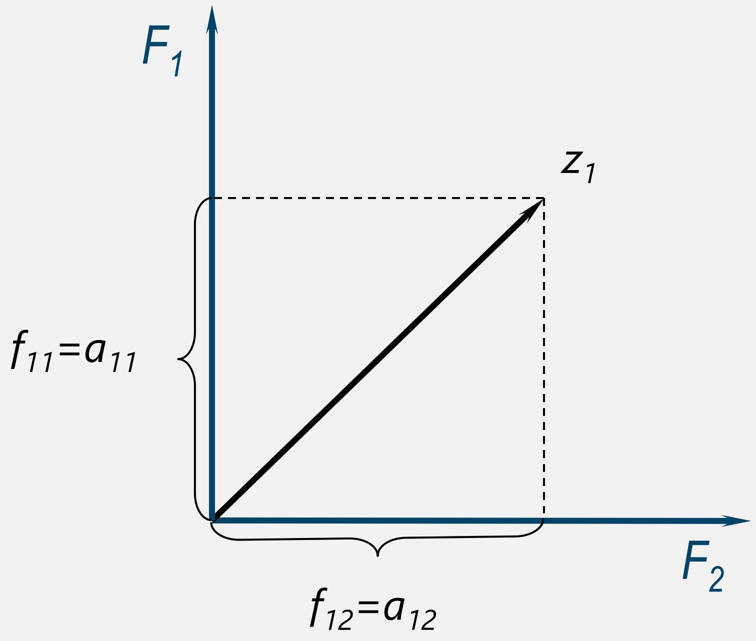
Matrica sklopa često jasnije pokazuje koje varijable određuju pojedine faktore nego matrica strukture.



Faktorska analiza

Rotacije

Paralelne projekcije i ortogonalne projekcije manifestnih varijabli na faktore bit će sličnije što su korelacije među faktorima manje. Matrica strukture bit će jednaka matrici sklopa ako su faktori potpuno nezavisni.



RStudio – Factor Analysis

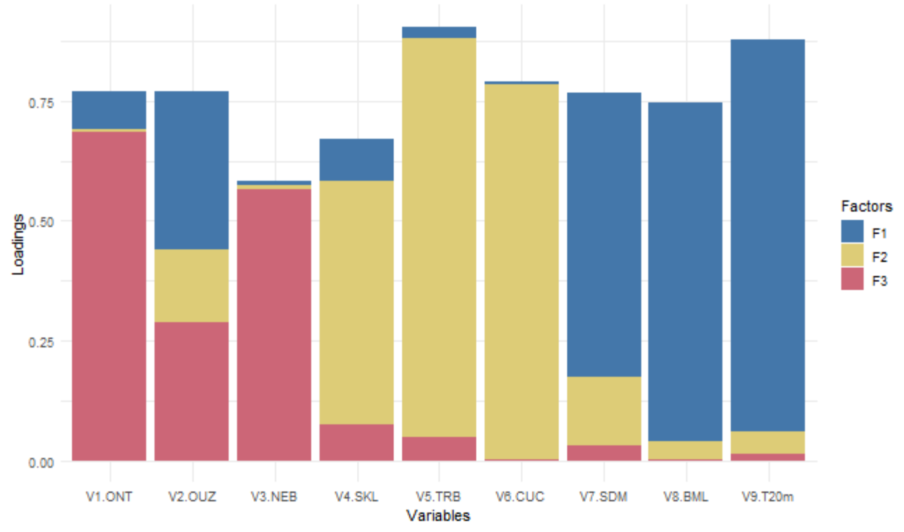
Zadatak: Na varijablama matrice *JUDO3F.csv* provedite komponentni model faktorske analize uz *GK-kriterij* redukcije i neortogonalnu *oblimin* rotaciju! Izračunajte matricu faktorske strukture i matricu rezultata entiteta na faktorima!

Eigenvalue

Copy

	Eigenvalue	Cum.Eign	Percentage	Cum.Per
1	**3.991**	**3.991**	44.345	44.345
2	**1.732**	5.723	19.245	63.590
3	**1.151**	6.874	12.784	76.374
4	0.807	7.680	8.963	85.337
5	0.481	8.161	5.339	90.677
6	0.342	8.503	3.797	94.473
7	0.250	8.753	2.779	97.252
8	0.132	8.884	1.464	98.716
9	0.116	9.000	1.284	100.000

Sum SMC = 5.602



Graph Factor Loadings (Factors)