

Domaća naloga 2

Neza Krzan, Tom Rupnik

Kazalo

1	Opis podatkov	3
2	Korelacijska matrika	3
3	Metoda glavnih komponent	5
4	Faktorska analiza	8
4.1	a)	8
4.2	b)	10
4.3	c)	10

Slike

Tabele

1 Opis podatkov

Podatki so bili zbrani med letoma 2016-2018 preko interaktivnega testa osebnosti na spletu.

Test osebnosti je bil zgrajen na podlagi “Big-Five Factor Markers” pridobljivih iz IPIP (International Personality Item Pool).

Lestvica uporabljena za odgovore je bila sledeča:

- 1 = se ne strinjam
- 3 = nevtrarno
- 5 = se strinjam

Za domočo nalogo sva si izbrala spremenljivke, ki so bile kodirane z “EXT1-10” in “EST1-10”. Te oznake predstavljajo trditev oz. osebnostno lastnost, katero so ankentiranci ovrednotili po zgornji lestvici. Pomen (s + oz. - je označeno ali je trditev bila označena kot pozitivna (velja za ta sklop/lastnost), ali negativno (ne velja za ta sklop/lastnost, lestvica je obrnjena)):

Energija (ektravertnost, surgentnost)

- EXT1...Sem zabavna oseba, ki je rada v središču dogajanja (+)
- EXT2...Ne govorim veliko (-)
- EXT3...Dobro se počutim v krogu ljudi (+)
- EXT4...Raje sem v ozadju (stran od pozornosti/neopazen) (-)
- EXT5...Jaz začenjam pogovore (+)
- EXT6...Nimam veliko za povedati/Sem redkobeseden (-)
- EXT7...Na zabavah se pogovarjam z veliko različnimi osebami (+)
- EXT8...Ne maram pritegniti pozornosti ostalih (-)
- EXT9...Ne moti me če sem v središču dogajanja (+)
- EXT10...V okolici neznancev ne govorim (-)

Čustvena stabilnost (nevroticizem)

- EST1...Hitro postanem živčen (-)
- EST2...Večino časa sem sproščen (+)
- EST3...Skrbi me za veliko stvari (-)
- EST4...Redko sem žalosten, melanholičen, emotično pretresen (+)
- EST5...Stvari me hitro zmotijo (-)
- EST6...Hitro se razjezim (-)
- EST7...Dosti spreminjam razpoloženje (-)
- EST8...Imam pogosta nihanja razpoloženja (-)
- EST9...Sem hitro vzkipljiv (-)
- EST10...Pogosto se počutim žalostno, melanholično, emotično pretreseno (-)

Vrednosti trditev, ki so označene z (-) bomo množili z -1 in s tem obrnili lestvico.

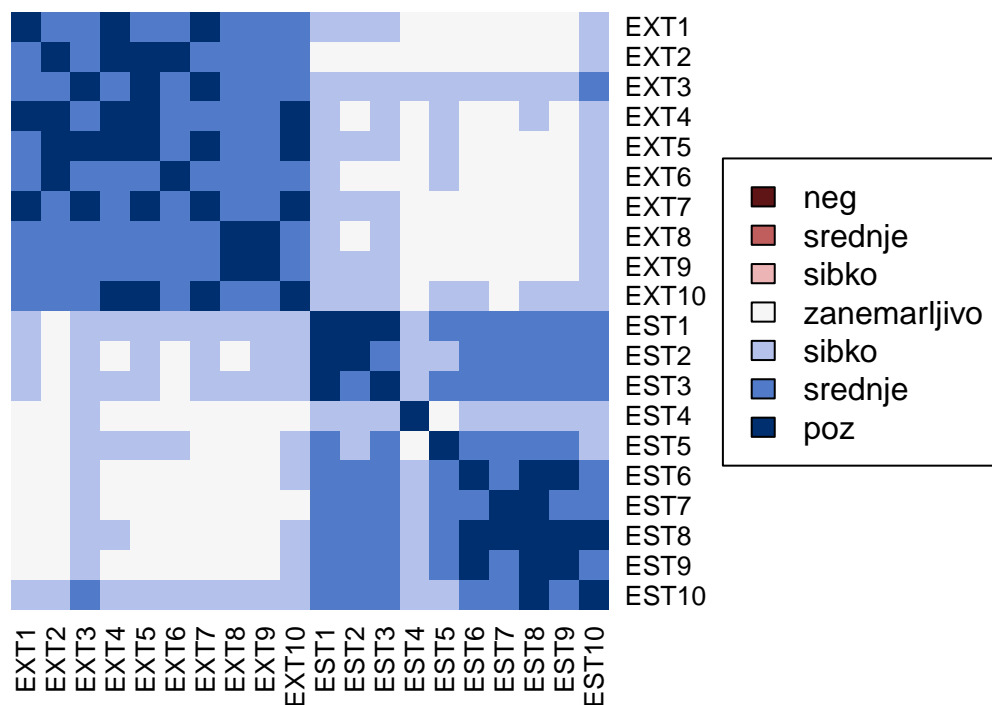
Iz podatkovnega okvirja sva odstranila vrstice, ki so vsebovale kakšno *NA* vrednost. Takih vrstic je bilo 280.

2 Korelacijska matrika

Iz matrike korelacije vidimo, da se korelacija znotraj spremenljivk **EXT** giblje približno med 0.3-0.6.

Pri spremenljivkah **EST** je razpon vrednosti nekoliko večji in sicer 0.1-0.8. Pri tem z nizkimi vrednostmi izstopa predvsem spremenljivka **EST4**, ki predstavlja trditev “Redko sem žalosten, melanholičen, emotično pretresen”.

Opaziti je, da so korelacije znotraj skupin spremenljivk (EXT in EST) večje, zato lahko na spodnjem prikazu pričakujemo dva očitna sklopa.



Res vidimo, da sta tvorjeni dve skupini spremenljivk, in sicer **EXT1-10** in **EST1-10**.

Da se zares prepričamo, da spremenljivke med seboj niso neodvisne lahko izvedemo Bartlettov test, ki ima ničelno hipotezo “vse spremenljivke so neodvisne”. Ko izvedemo test dobimo p-vrednost enako 0, kar pomeni da ovržemo ničelno hipotezo. Torej spremenljivke med seboj niso neodvisne.

Preverimo ali so podatki primerni za faktorsko analizo. V ta namen izvedemo Kaiser-Meyer-Olkin test.

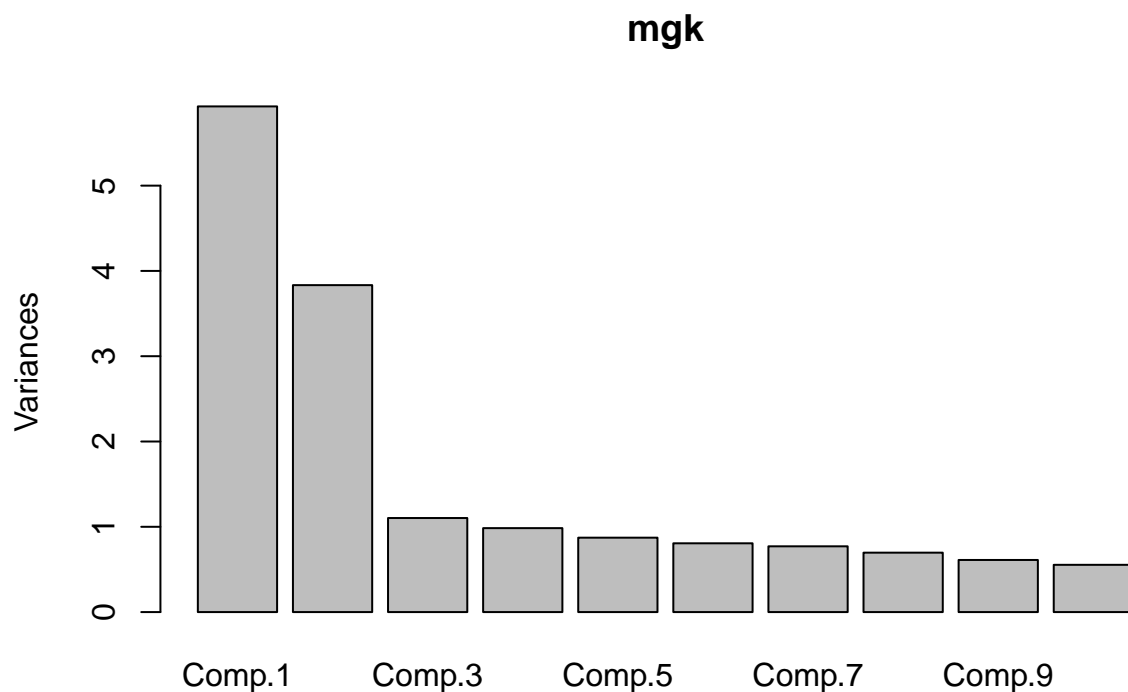
Test nam vrne naslednje vrednosti:

EXT1	EXT2	EXT3	EXT4	EXT5	EXT6	EXT7	EXT8	EXT9	EXT10
0.93	0.92	0.95	0.94	0.93	0.92	0.92	0.89	0.91	0.95
EST1	EST2	EST3	EST4	EST5	EST6	EST7	EST8	EST9	EST10
0.88	0.89	0.91	0.87	0.96	0.88	0.80	0.80	0.87	0.94

EXT1	EXT2	EXT3	EXT4	EXT5	EXT6	EXT7	EXT8	EXT9	EXT10
0.93	0.92	0.95	0.94	0.93	0.92	0.92	0.89	0.91	0.95
EST1	EST2	EST3	EST4	EST5	EST6	EST7	EST8	EST9	EST10
0.88	0.89	0.91	0.87	0.96	0.88	0.80	0.80	0.87	0.94

Vse vrednosti so ≥ 0.8 kar pomeni da so podatki zelo primerni za faktorsko analizo.

3 Metoda glavnih komponent



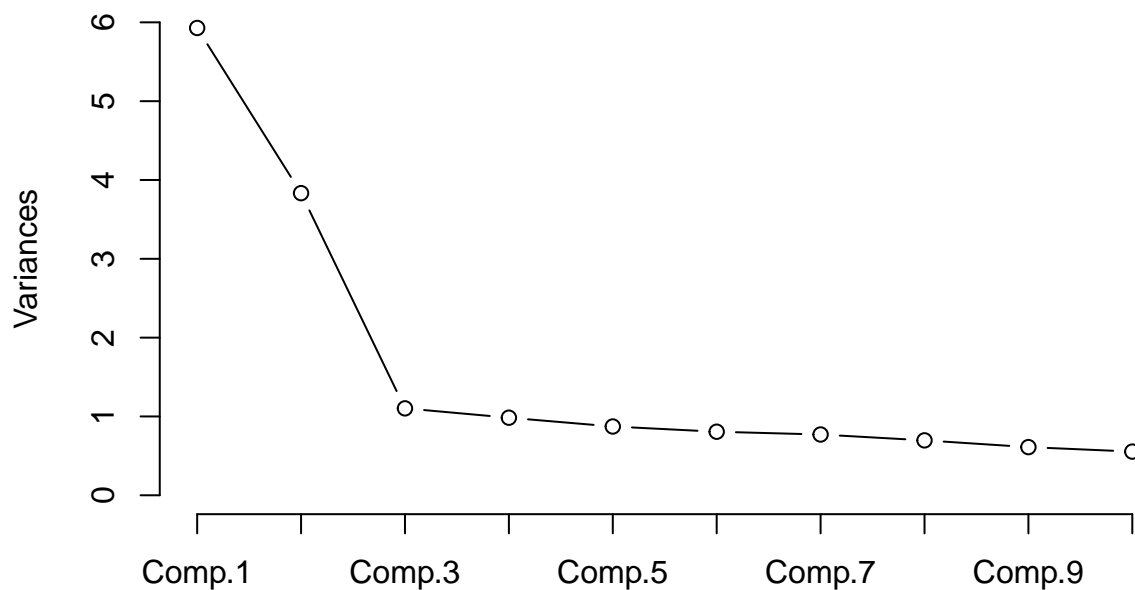
Prikazan je graf variance v odvisnosti od števila komponent (izvedeno na standardiziranih podatkih). Odločamo se med vrednostmi, ki so nad 1, zato bi se odločili za 2 komponenti. Tudi vrednost pri 3 komponentah je nekoliko nad 1 vendar jo vrednost le nekoliko preseže, zato se raje odločimo za 2 komponenti.

Podrobnejša analiza izbire števila glavnih komponent

st. komponent	lastna vrednost	varianca	standardni odklon
1	5.9288	5.9288	2.4349
2	3.8332	3.8332	1.9578
3	1.1027	1.1027	1.0501
4	0.9842	0.9842	0.9921
5	0.8724	0.8724	0.9340

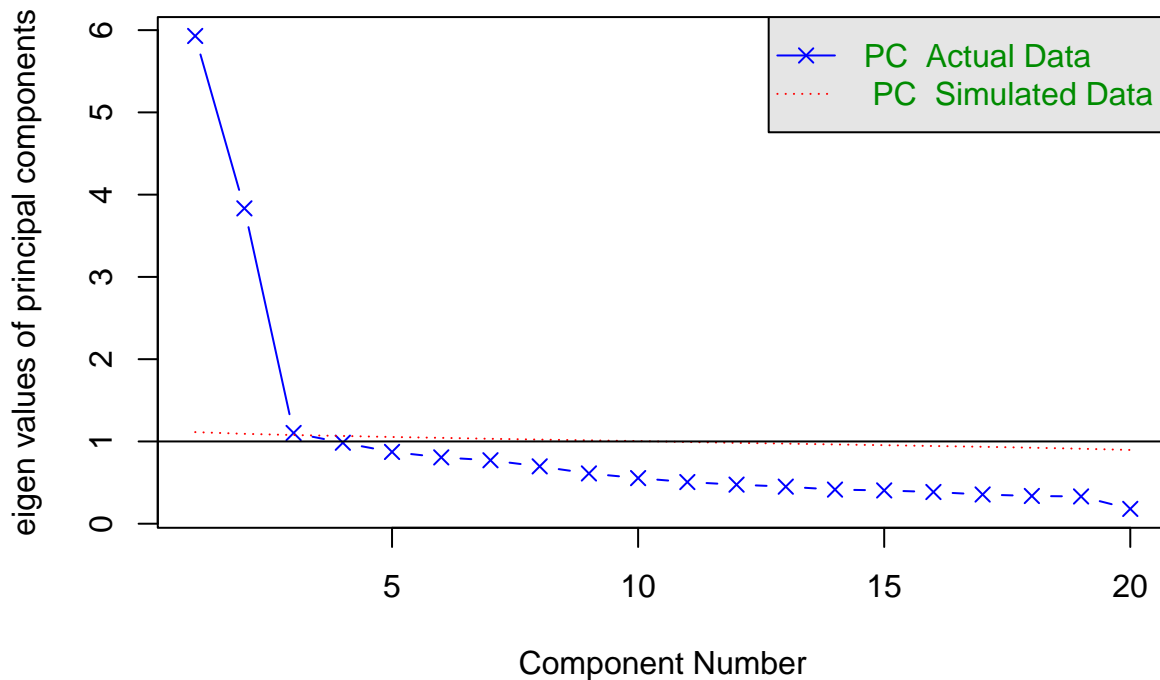
Podobno kot zgoraj s pomočjo grafa pridemo do ugotovitve, da je najbolj primerno število glavnih komponent 2.

mgk



Najbolj očitno koleno je pri 3 komponentah. Vendar pa za število glavnih komponent izberemo vrednost (st. komponent), ki je nad kolenom. Torej število glavnih komponent v našem primeru je 2.

Parallel Analysis Scree Plots



S paralelno analizo smo dobili nekoliko drugačen rezultat kot v prejšnjih primerih. Predlagano število glavnih komponent je 3 vendar je ta vrednost tik na meji (je tudi težko razvidno iz grafa), zato bomo obržali število glavnih komponent 2.

Celotna variabilnost

Koliko prispeva k celotni variabilnosti posamezna glavna komponenta:

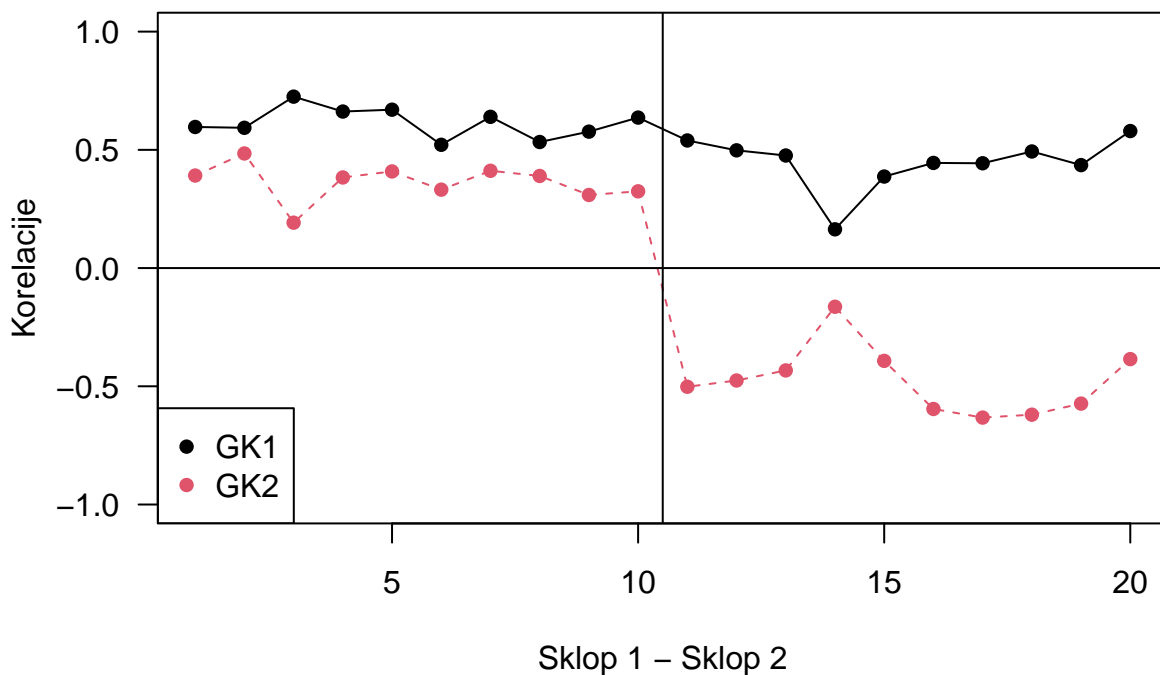
število g. komponent	1	2	3	4	5	...
proporcij varinace	0.2964	0.1917	0.0551	0.0492	0.0436	...

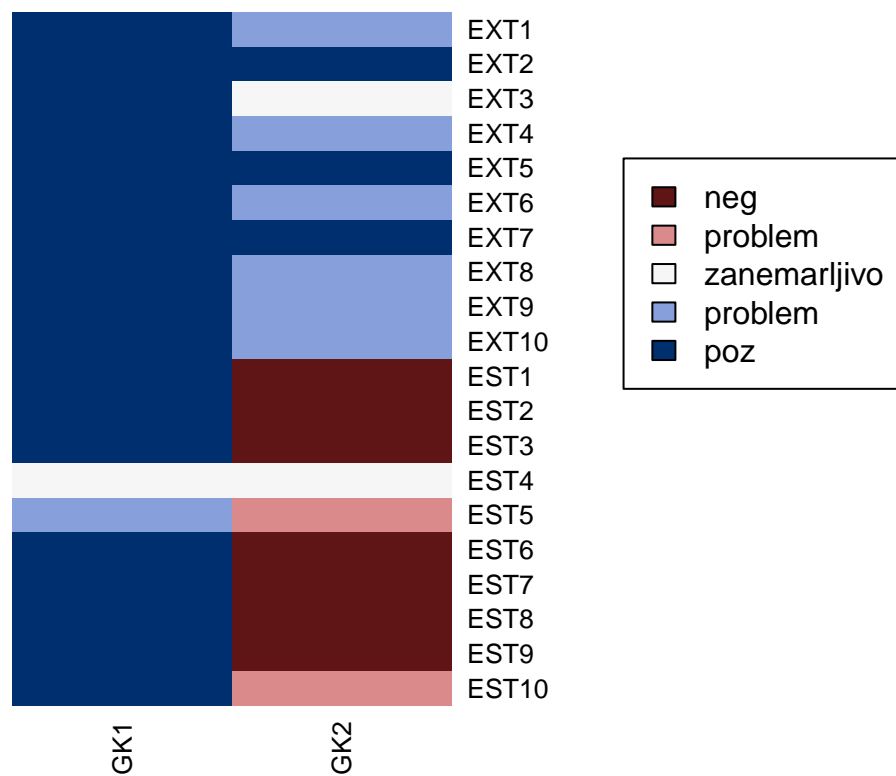
Z dvema glavnima komponentama nam je uspelo pojasniti 0.488% variabilnosti.

spremenljivka	EST4	EST5	EXT6	EST3	EXT9	EXT8	EST2	EST10	EXT1	EXT10
	0.0537	0.3039	0.3817	0.4140	0.4284	0.4359	0.4739	0.4840	0.5093	0.5104
	EST9	EST1	EST6	EXT3	EXT7	EXT4	EXT2	EST7	EXT5	EST8
	0.5182	0.5432	0.5529	0.5625	0.5784	0.5851	0.5868	0.5964	0.61570	.6276

Opazimo da je le **EST4** slabo pojasnjen, saj ima vrednost $0.054 < 0.2$ (mejo 0.2 vzamemo kot vrednost pri kateri lahko indikator povzroča probleme).

Grafični prikaz glavnih komponent

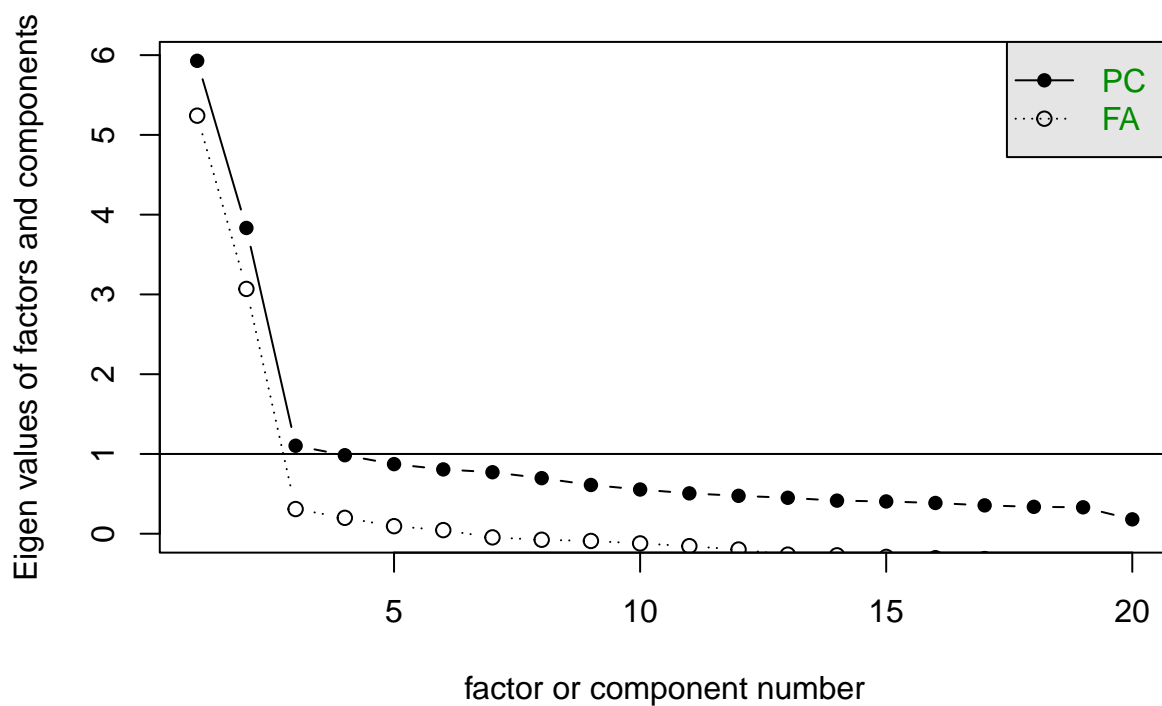




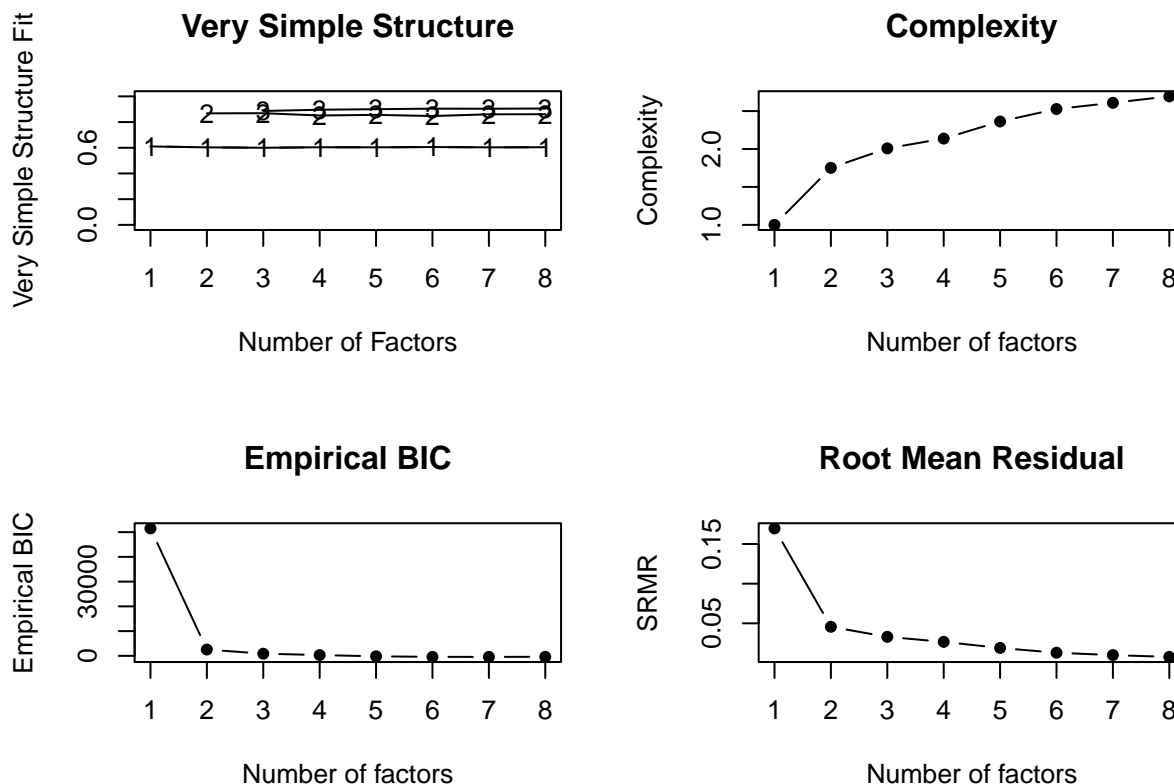
4 Faktorska analiza

4.1 a)

Scree plot

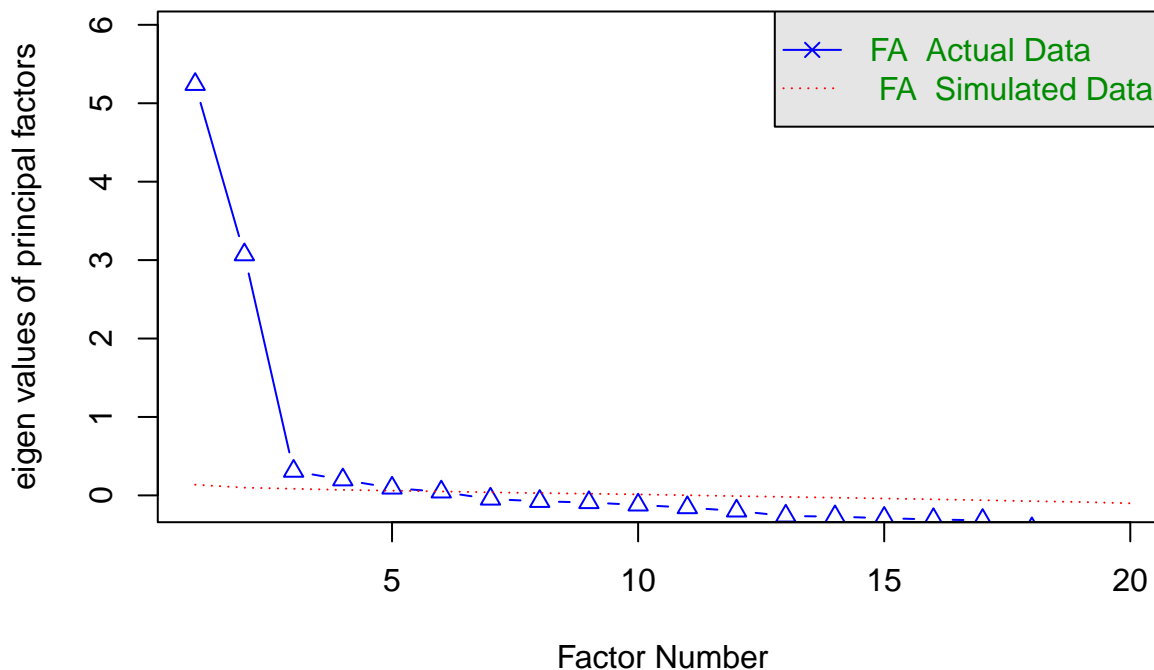


Pri scree grafu gledamo koliko komponent je ≥ 1 , torej je v našem primeru število faktorjev 2. Pri 3 je tudi prisotno največje koleno zato izberemo vrednost 2.



VSS: če upoštevamo le uteži 1 in 2, predlaga 3 faktorje (utež 2 pri vrednosti 3 doseže najvišjo vrednost) eBIC: doseže minimum pri 7 faktorjih

Parallel Analysis Scree Plots



Pri paralelni gledamo koliko vrednosti je večjih od simuliranih podatkov. Predlagano število faktorjev je 5.

4.2 b

Obdržimo 2 faktorja

Loading required namespace: GPArotation

Ker je korelacija manjša od 0.2 naredimo še pravokotno rotacijo.

4.3 c

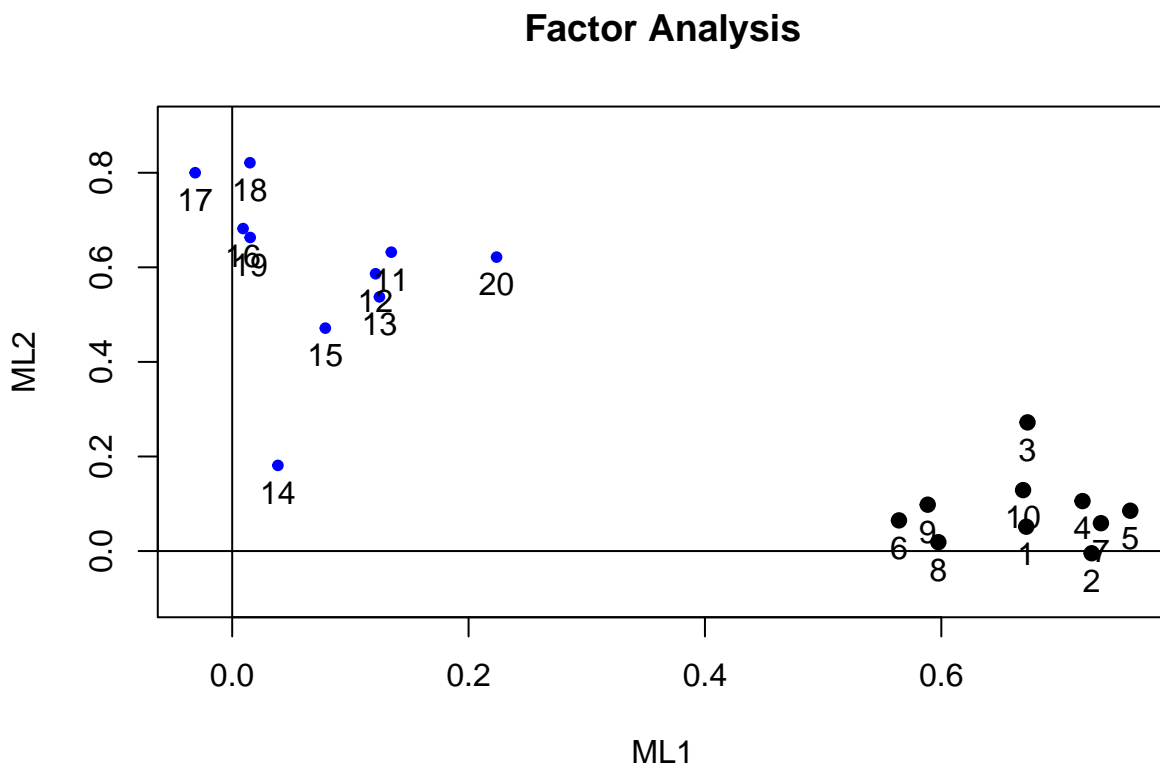
Komunalitete

Z izbranimi skupnimi faktorji smo pojasnili naslednje deleže posameznega indikatorja:

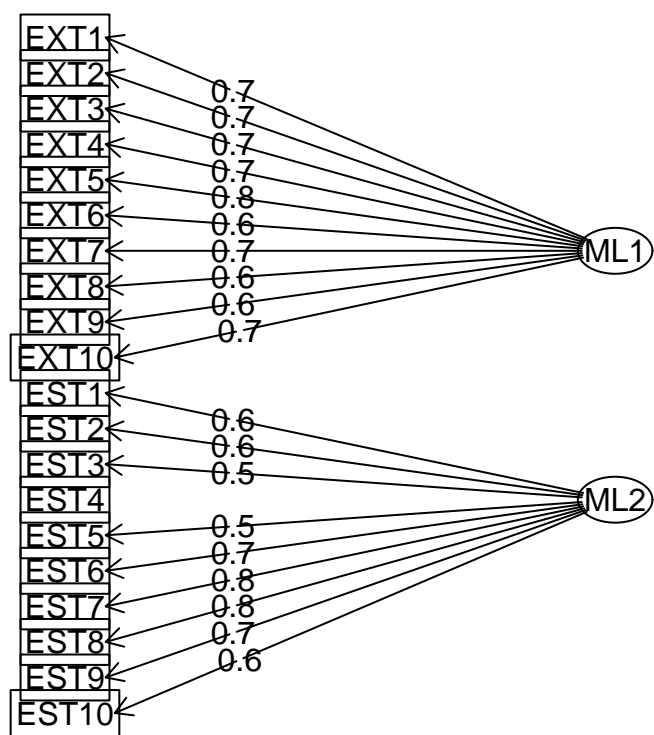
spremenljivka	EST4	EST5	EST3	EXT6	EXT9	EXT8	EST2	EST1	EST10	EST9
	0.0343	0.2285	0.3044	0.3225	0.3560	0.3574	0.3586	0.4176	0.4365	0.4400
	EXT1	EXT10	EST6	EXT3	EXT4	EXT2	EXT7	EXT5	EST7	EST8
	0.4541	0.4644	0.4651	0.5269	0.5288	0.5289	0.5438	0.5847	0.6411	0.6745

Opazimo da je le **EST4** slabo pojasnjen, saj ima vrednost $0.034 < 0.2$ (mejo 0.2 vzamemo kot vrednost pri kateri lahko indikator povzroča probleme).

Grafični prikaz



Factor Analysis



Pattern

