

Metoda glavnih komponent in faktorska analiza

Domača naloga 3 - Multivariatna analiza

Urh Peček in Alen Kahteran

20. 3. 2021

Kazalo

Cilji naloge	4
Metoda glavnih komponent	5
Primernost podatkov	5
Odločitev o številu komponent	6
Uteži za izbrani glavni komponenti	6
Delež pojasnjene variabilnosti	7
Biplot	7
Poimenovanje glavnih komponent	8
Faktorska analiza	9
Primernost podatkov	9
Odločitev o številu faktorjev	10
Dodatni pristopi h odločanju o številu faktorjev	10
Delež pojasnjene variabilnosti	11
Korelacije na podlagi modela	12
Faktorske uteži	13
Biplot	14
Poimenovanje faktorjev	15
Analiza korelacij	16
Povezanost faktorjev z ostalimi spremenljivkami	16
Povezanost faktorjev z nadzorom	16
Povezanost faktorjev z razmerjem	17
Povezanost faktorjev s prebivališčem	18
Enote v prostoru	19
Enote v prostoru glavnih komponent	20
Enote v prostoru faktorjev	21
Zaključek - Alen	21

Slike

Tabele

Cilji naloge

Naloga se deli na 2 dela, v prvem delu je cilj naloge pojasniti čim več variabilnosti spremenljivk, v drugem delu pa bomo z nekaj faktorji poskusili čim bolje pojasniti korelacije med spremenljivkami vrednot. Naloga tako temelji na metodi glavnih komponent in faktorski analizi. Analizirali bomo vsebinske Schwartzove spremenljivke, ki smo jih opisali v prvi nalogi (Pregled in priprava podatkov).

Metoda glavnih komponent

Cilj metode glavnih komponent je s prvimi nekaj nekoreliranimi glavnimi komponentami pojasniti kar največ variabilnosti originalnih spremenljivk. Glavne komponente so linearne kombinacije merjenih spremenljivk in so med seboj neodvisne. Urejene so od tiste, ki pojasni največ variabilnosti osnovnih podatkov pa do tiste, ki pojasne najmanj variabilnosti. Z analizo glavnih komponent zmanjšamo razsežnost podatkov tako, da izgubimo čim manj informacij.

Primernost podatkov

Preden se lotimo metode glavnih komponent in v nadaljevanju faktorske analize preglejmo ali so podatki sploh primerni. Zgornji metodi je smiselno delati v primeru, ko so spremenljivke (v našem primeru so to Schwartzove spremenljivke) med seboj korelirane.

Na spodnjem grafu, ki prikazuje Pearsonov koeficient korelacije med pari spremenljivk (pomnožen z 10) opazimo dva sklopa spremenljivk glede na vzorec korelacij. Prvi sklop bi lahko poimenovali odprtost (drugačnost, užitek, vznemerljivost, zabava, kreativnost, svoboda), drugi sklop pa konservativnost (varnost, ponižnost, obramba, tradicija, ubogljivost, sprejemljivost). Korelacije znotraj obeh sklopov so pozitivne in srednje oziroma šibke, edino povezanost med vznemerljivostjo in drugačnostjo je močna. Spremenljivke med sklopoma so v večini zanemarljivo povezane, edino zabava, je nekoliko bolj povezana s konservativnimi spremenljivkami.

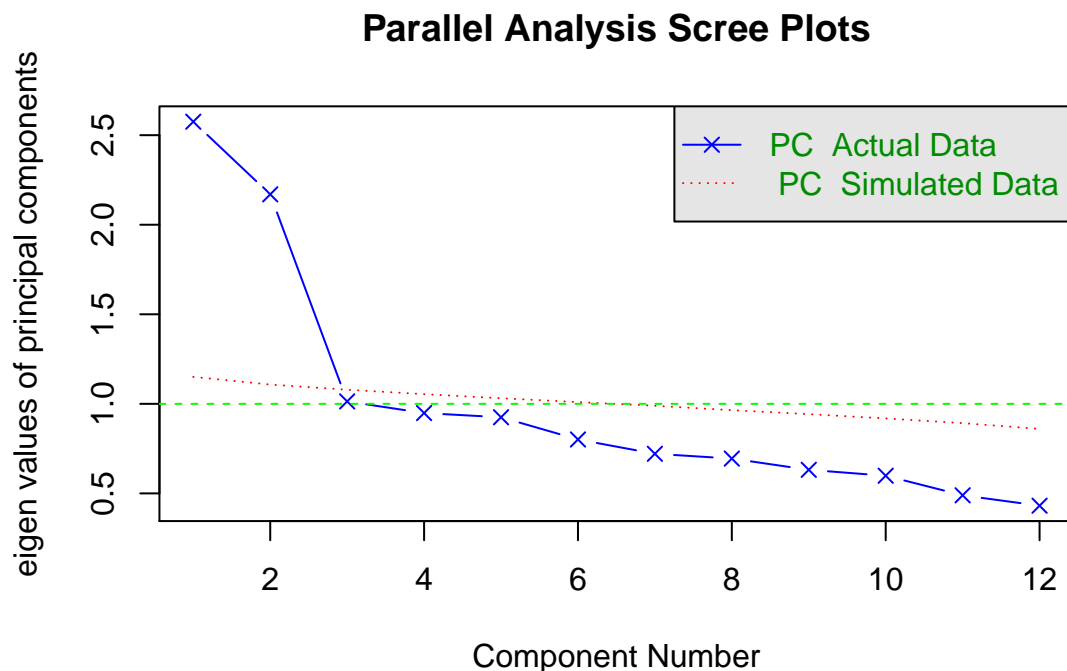
	drugačnost	užitek	vznemerljivost	zabava	kreativnost	svoboda	varnost	ponižnost	obramba	tradicija	ubogljivost	sprejemljivost
drugačnost	10	4	5	3	3	2	-1	0	1	0	0	0
užitek	4	10	4	5	2	2	1	0	2	2	0	1
vznemerljivost	5	4	10	3	3	2	-1	-1	1	-1	0	-1
zabava	3	5	3	10	1	2	1	0	1	1	0	0
kreativnost	3	2	3	1	10	3	0	-1	1	0	0	0
svoboda	2	2	2	2	3	10	0	0	0	0	-1	0
varnost	-1	1	-1	1	0	0	10	3	3	2	2	3
ponižnost	0	0	-1	0	-1	0	3	10	2	1	2	3
obramba	1	2	1	1	1	0	3	2	10	2	3	2
tradicija	0	2	-1	1	0	0	2	1	2	10	2	2
ubogljivost	0	0	0	0	0	-1	2	2	3	2	10	3
sprejemljivost	0	1	-1	0	0	0	3	3	2	2	3	10

* all values in cells were multiplied by 10

Z Bartlettovim testom lahko preverimo ničelno domnevo, da obravnavane spremenljivke v populaciji niso povezane, kar bi pomenilo, da je korelacijska matrika spremenljivk identiteta. Na podlagi vrednosti p ($p < 0.0001$) pri 66 stopinjah prostosti pri stopnji značilnosti 0.05 zavrnilo ničelno domnevo in sprejememo sklep, da so Schwartzove spremenljivke v populaciji korelirane. To pomeni, da so metode, ki jih bomo delali v nadaljevanju smiselne, tako pri glavnih komponentah kot pri številu faktorjev.

Odločitev o številu komponent

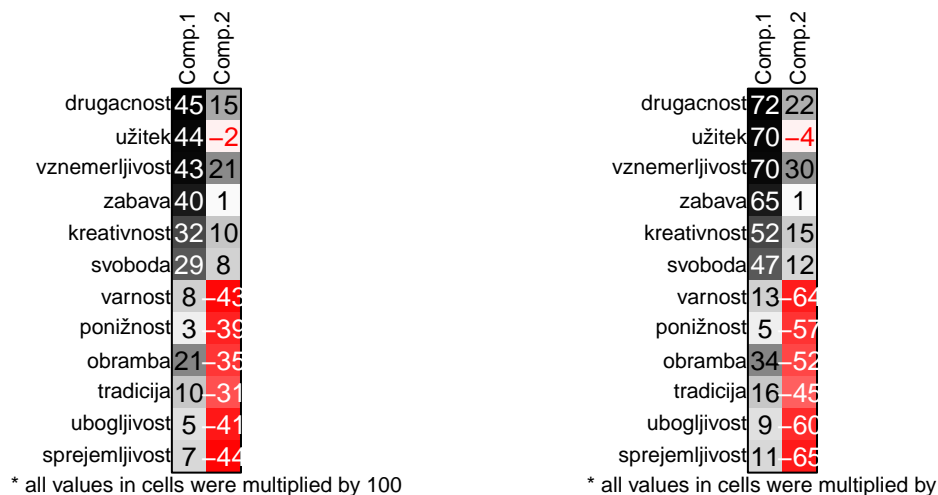
Glede števila glavnih komponent se lahko odločimo na podlagi scree diagrama, ki grafično predstavi lastne vrednosti. Izbrano število glavnih komponent je tam, kjer se graf lomi oziroma ima koleno, pri čemer obdržimo komponente nad kolenom. Število komponent je lahko tudi takšno, da so lastne vrednosti komponent večje od povprečne variabilnosti standardiziranih spremenljivk, ki je enaka 1. Lahko se odločimo tudi tako, da primerjamo lastne vrednosti komponent s tistimi na podlagi podatkov generiranih iz nekoreliranih spremenljivk.



Na podlagi kolena scree diagrama se odločimo za dve glavni komponenti, na podlagi lastnih vrednosti, ki morajo biti večje od 1 se odločimo za tri lastne vrednosti (ali 2, ni popolnoma razvidno) in na podlagi primerjav lastnih vrednosti se odločimo za 2 glavni komponenti. Ob upoštevanju vseh treh kriterijev je končna odločitev, da izberemo 2 glavni komponenti. Velja, da so vrednosti med glavnimi komponentami neodvisne. Opomnimo lahko še, da kot kriterij, ki sicer ni glavni, lahko vzamemo delež pojasnjene variabilnosti z glavnimi komponentami, na primer vsaj 50%, za kar bi pri nas morali vzeti 4 glavne komponente.

Uteži za izbrani glavni komponenti

Ponavadi uteži dobljenih glavnih komponent reskaliramo. Nereskalirane uteži predstavljajo regresijske koeficiente. Reskalirana utež i -te glavne komponente pri j -ti spremenljivki pa je korelacija med i -to komponento in j -to spremenljivko. S pomočjo teh korelacij lahko interpretiramo glavne komponente.



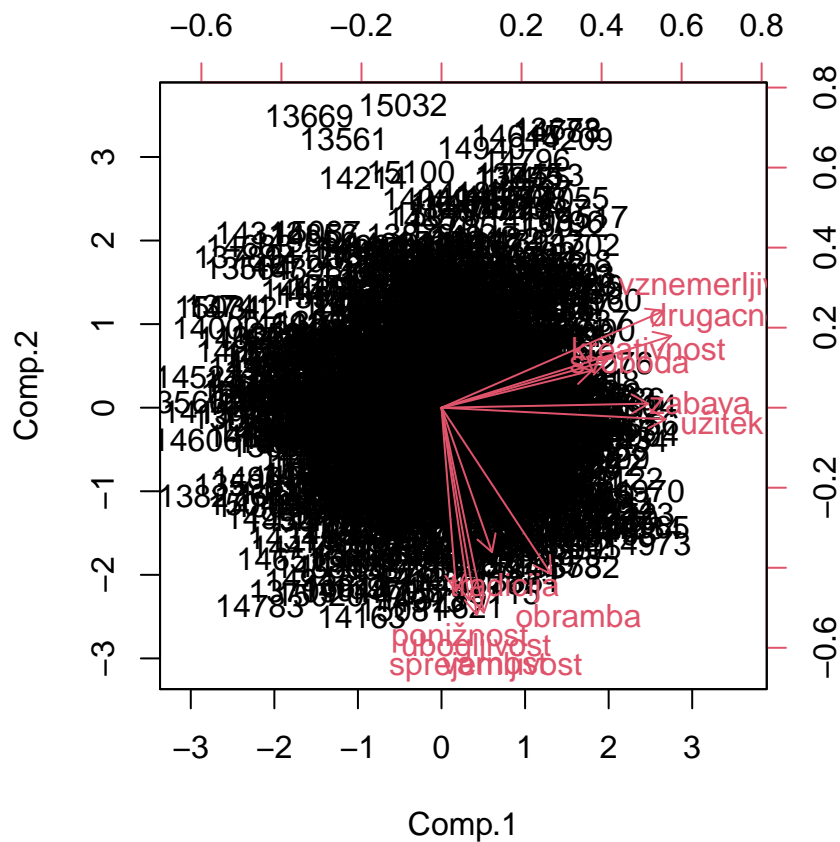
Na levem grafu vidimo nereskalirane in na desnem reskalirane uteži. Interpretiramo reskalirane uteži. Velja, da so spremenljivke odprtosti močno pozitivno korelirane s prvo glavno komponento, medtem ko so spremenljivke konservativnosti močno negativno korelirane z drugo glavno komponento. Pri odprtosti nekoliko odstopa vznemerljivost in morda tudi drugačnost, ki sta šibko pozitivno korelirani z drugo glavno komponento, pri konservativnosti pa je obramba srednje pozitivno korelirana s prvo glavno komponento. Če strnemo, prva glavna komponenta povzema predvsem pomembnost odprtosti in druga glavna komponenta povzema predvsem pomembnost nekonvativnosti (negativne korelacije).

Delež pojasnjene variabilnosti

Velja, da je vsota kvadratov reskaliranih uteži po spremenljivkah varianca spremenljivke. V našem primeru pri izbranih dveh glavnih komponentah s prvo komponento pojasnimo 21% celotne variabilnosti in z drugo komponento 18%, skupaj nekaj manj kot 40% celotne variabilnosti. Z izbranimi prvima glavnima komponentama pojasnimo največji delež vznemerljivosti (57%), drugačnosti (56%) in užitka (49%). Najmanjši delež variabilnosti pojasnimo pri spremenljivkah svobode in tradiciji (23%).

Biplot

Biplot nam v dvodimenzionalnem prostoru prikaže položaj posameznih enot glede na vrednosti vseh obravnavanih spremenljivk (Schwartzove spremenljivke), hkrati pa so narisane tudi vrednosti reskaliranih uteži po komponentah.



Iz grafa vidimo, da imajo spremenljivke odprtosti podobne, visoke uteži pri prvi glavni komponenti, medtem ko imajo relativno nizke uteži ostale spremenljivke. Na drugi glavni komponenti pa imajo nizke in negativne uteži spremenljivke konservativnosti, spremenljivke odprtosti pa imajo v večini pozitivne in nizke vrednosti.

Poimenovanje glavnih komponent

Za konec metode glavnih komponent pa še poimenujmo izbrani komponenti. Za prvo komponento smo ugotovili, da povzema predvsem pomembnost odprtosti, za drugo pa nekonsevativnosti. V skladu z zgornjim prvo komponento poimenujemo odprtost in drugo nekonsevativnost.

Faktorska analiza

Faktorska analiza poizkuša poenostaviti kompleksnost povezav med opazovanimi spremenljivkami. Gre za študij povezav med spremenljivkami, pri čemer skušamo najti novo množico spremenljivk, ki predstavlja to kar je skupnega opazovanim spremenljivkam. Načeloma jo uporabimo v primeru, ko nekaterih pojmov ne moremo meriti neposredno. V tem primeru izberemo nekaj direktno merjenih spremenljivk, ki so indikatorji teoretičnega pojma, ki ga želimo meriti. Tako prek nekaj faktorjem poizkušamo čim bolj pojasniti korelacije oziroma kovariance med merjenimi spremenljivkami tj. skupne variabilnosti spremenljivk. Merjene spremenljivke so tako lineane kombinacije skupnih in specifičnih faktorjev. Merljive spremenljivke imenujemo manifestne, nemerljive pa latentne spremenljivke.

Primernost podatkov

Že pri preverjanju primernosti podatkov za metodo glavnih komponent smo ugotovili, da so spremenljivke korelirane, lahko smo zavrnili ničelno domnevo, da je populacijska korelacijska matrika identiteta z manj kot 1 odstotna stopnjo tveganja. Na vzorčnih korelacijah smo videli, da se oblikujeta 2 sklopa spremenljivk, sklop odprtosti in sklop konservativnosti kar nakazuje na dvofaktorsko strukturo.

Da pogledamo ali so podatki primerni za analizo oziroma kako dobre rezultate lahko pričakujemo ugotovimo tako, da pogledamo parcialne korelacije med spremenljivkami (anti image korelacije), ki so prikazane na spodnji anti image korelacijski matriki.

	drugacnost	užitek	vznemerljivost	zabava	kreativnost	svoboda	varnost	ponižnost	obramba	tradicija	ubogljivost	sprejemljivost
drugacnost	74	-12	41	-8	22	-3	3	1	-4	-1	1	1
užitek	-12	75	-17	37	-3	-2	-1	0	-8	-14	1	-3
vznemerljivost	41	-17	73	-11	-7	-10	14	0	-5	7	0	4
zabava	-8	37	-11	74	2	-12	12	-1	2	0	3	
kreativnost	22	-3	-7	2	77	-19	-1	7	-5	1	0	-4
svoboda	-3	-2	-10	12	-19	79	-5	0	1	1	7	-3
varnost	3	-1	14	-12	-1	-5	71	-16	21	-10	-3	16
ponižnost	1	0	0	2	7	0	-16	73	-9	6	-14	18
obramba	-4	-8	-5	-1	-5	1	-21	-9	77	-10	-17	-4
tradicija	-1	-14	7	2	1	1	-10	6	-10	72	-9	-13
ubogljivost	1	1	0	0	0	7	-3	-14	17	-9	74	-20
sprejemljivost	1	-3	4	3	-4	-3	-16	18	-4	-13	20	74

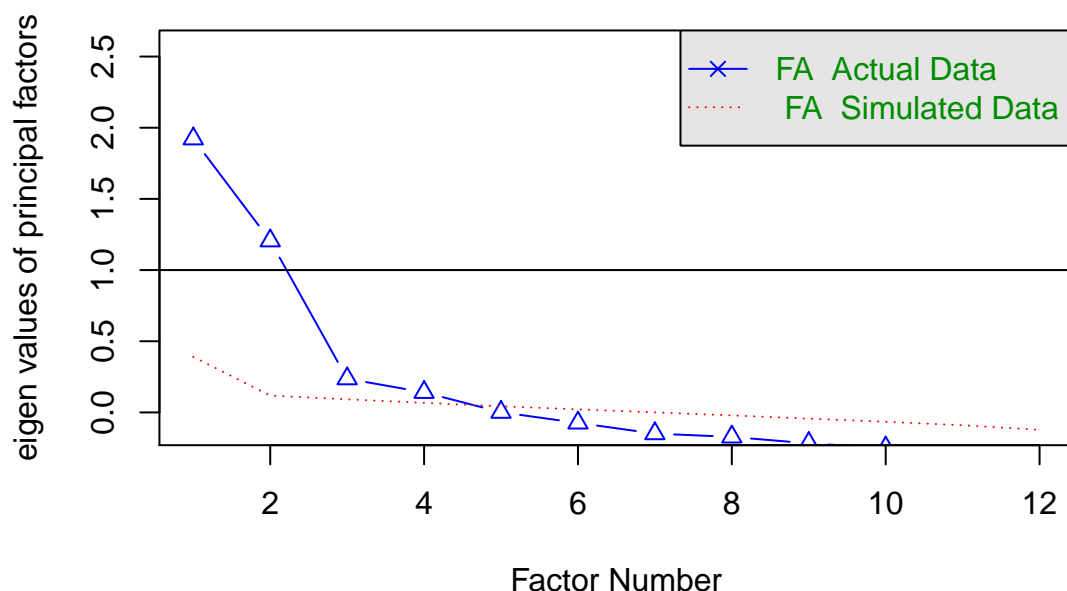
* all values in cells were multiplied by 100

Izven diagonale anti image korelacijske matrike so parcialni korelacijski koeficienti pomnoženi z -1. Če si spremenljivke delijo skupne faktorje, so njihovi parcialni korelacijski koeficienti po absolutni vrednosti majhni. Pri nas v večini primerov so, odstopata zgolj parcialni korelaciji med vznemerljivostjo in drugačnostjo ter zabavo in užitkom. Na diagonalni matrike pa so vrednosti Kaiser-Meyer-Olkinove mere, ki predstavlja delež spremenljivke, ki bi lahko bil pojasnjen s skupnimi faktorji. Vrednosti KMO so pri vseh spremenljivkah med 0.7 in 0.8, kar pomeni, da so spremenljivke dovolj korelirane in tako primerne za faktorsko analizo. Opazimo tudi, da so korelacije znotraj sklopov nekoliko večje kot med sklopi. Povprečna oziroma skupna Keiser-Meyer-Olkinova mera je enaka 0.734 kar dodatno nakazuje na primernost podatkov.

Odločitev o številu faktorjev

Za določitev števila faktorjev uporabljamo podobne kriterije kot pri metodi glavnih komponent. Odločimo se lahko na podlago kolena scree diagrama na podlagi lastnih vrednosti faktorjev ali pa s primerjavo lastnih vrednosti faktorjev s tistimi na podlagi podatkov generiranih iz nekoreliranih spremenljivk. Predpostavimo, da se spremenljivke porazdeljujejo normalno in tako lahko uporabimo faktorsko metodo največjega verjetja.

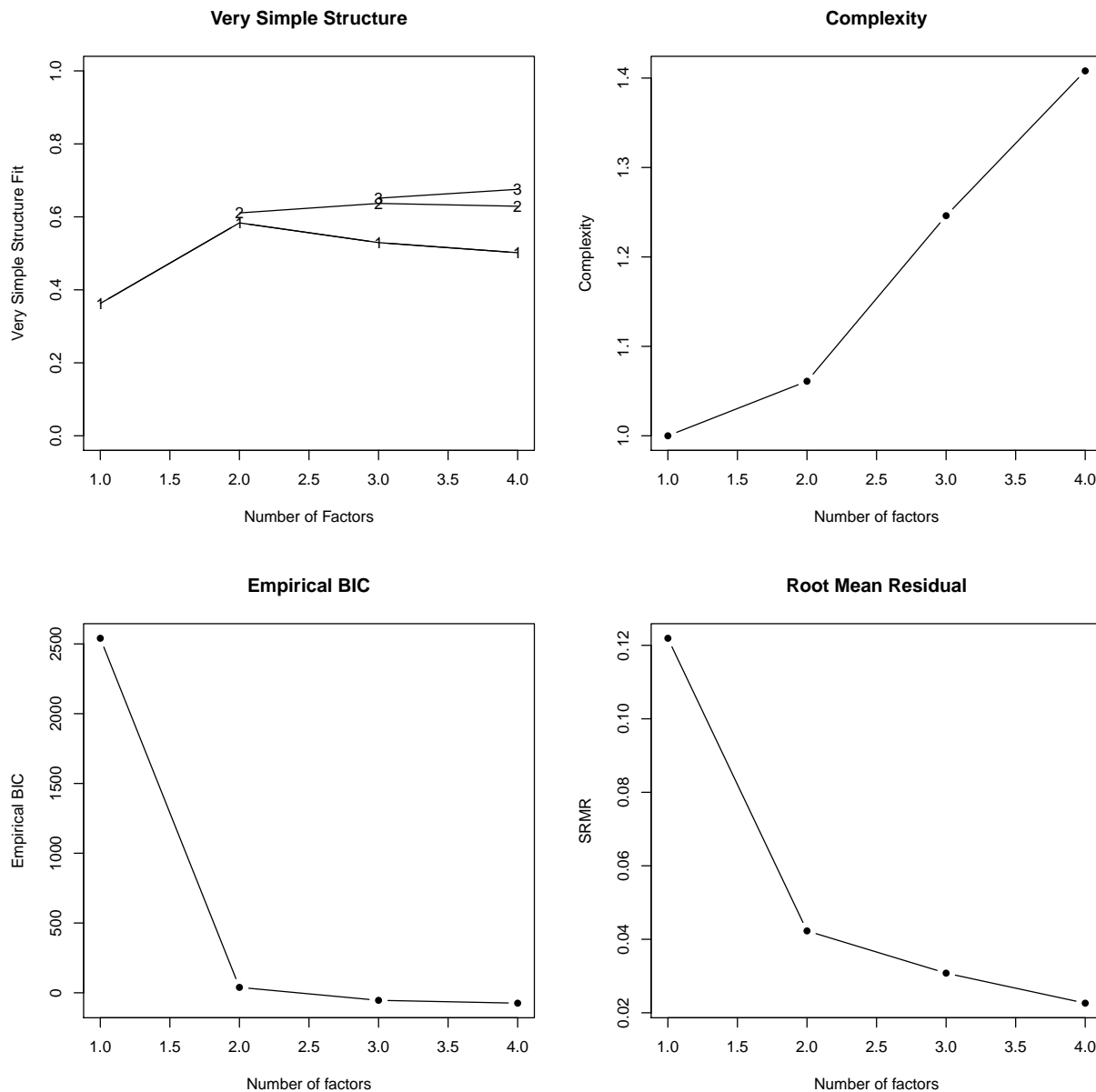
Parallel Analysis Scree Plots



Na podlagi kolena scree diagrama se odločimo za dva faktorja (tiste pri koleno izpustimo), na podlagi lastnih vrednosti, ki morajo biti večje od 1 se prav tako odločimo za dva faktorja in na podlagi primerjav lastnih vrednosti se odločimo za štiri faktorje. Ob upoštevanju vseh treh kriterijev prevladata koleno in vrednost lastnih vrednosti in končna odločitev je, da izberemo dva faktorja.

Dodatni pristopi h odločanju o številu faktorjev

Za določanje števila faktorjev lahko uporabimo tudi nekatere dodatne pristope. Pri pristopu VSS (very simple structure) izberemo število, kjer je vrh grafa, torej tam kjer so uteži pri spremenljivkah največje in s tem največ pojasnijo (upoštevane so samo najvišje uteži pri spremenljivkah). Pristop EBIC (empirical bayesian information criterion) nam števila faktorjev določi tam, kjer je koleno, pri čemer koleno tudi upoštevamo. EBIC nam pove kako dobro je prileganje modela pri določenem številu faktorjev, najboljša je najnižja vrednost. Temelji na odklonih korelacijske matrike, popravljenih za število enot in število faktorjev. S pristopom RMSE (root mean residual), kjer upoštevamo koren povprečnega odklona modelskih korelacij od pravih korelacij pa število faktorjev določimo kot najmanjše kjer je RMSE manjše od 0.08. Za izračun uporabimo metodo rotacije varimax (pravokotna rotacija), kjer predpostavimo, da med faktorji ni korelacij.



Na podlagi VSS se za najboljše število faktorjev odločimo za 2 faktorja pri čemer upoštevamo obe faktorski uteži, saj po dveh faktorjih vrednost VSS zelo počasi narašča in ob upoštevanju preprostosti ne predstavlja boljše rešitve. Majhna razlika med vrednostjo pri 1 ali 2 upoštevanih utežev nam pove tudi to, da ni spremenljivk, ki bi imele visoko utež pri obeh faktorjih. Complexity nam pove kako čisto faktorsko strukturo dobimo, torej na koliko faktorjih ima spremenljivka visoko utež, se pa na podlagi teh vrednosti ponavadi ne odločamo o številu faktorjev. EBIC nam pove, da je najboljše število faktorjev dva faktorja, saj je pri večjem številu faktorjev prileganje zanemarljivo boljše. Odločimo se za 2 faktorja. Na podlagi RMSE se prav tako odločimo za dva faktorja. Končno, tudi na podlagi nadaljnjih pristopov ocen najboljšega števila faktorjev se odločimo za dva faktorja, kar je tako naša končna odločitev o številu faktorjev.

Delež pojasnjene variabilnosti

Na podlagi izbranega števila faktorjev tj. dveh faktorjev bomo izvedli faktorsko analizo. Po izvedeni poševni rotaciji (oblimin) vidimo, da je korelacija med faktorjema zanemarljiva ($r = 0.01$) in posledično se odločimo

za pravokotno rotacijo (varimax), kjer so korelacije med faktorji enake 0.

Hoffmanov indeks kompleksnosti nam pove, kako dobro spremenljivka ločuje med faktorjema in je v večini primerov enak 1.0 oziroma 1.1, višji je le pri užitku (1.2) in pri obrambi (1.3) kar pomeni, da je utež nekoliko višja pri obeh faktorjih, vendar zanemarljivo. Torej ima večina spremenljivk visoko utež le na enem faktorju. S prvim faktorjem pojasnimo 16% z drugim pa 12% variabilnosti med spremenljivkami, skupno torej pojasnimo 28% variabilnosti med spremenljivkami. Izmed vse pojasnene variabilnosti z izbranim števiom faktorjev s prvim faktorjem pojasnimo 57% variabilnosti, z drugim pa 43%.

Korelacije na podlagi modela

Na spodnji sliki vidimo korelacije ocenjene na podlagi modela in so precej podobne empiričnim. Na diagonali so komunalitete, ki odražajo del variabilnosti spremenljivke, ki je pojasnjen s skupnimi faktorji. V našem primeru je večina komunalitet nad 0.2, kar pomeni, da je spremenljivka dobro pojasnjena s skupnim faktorskim modelom. Odstopajo zgolj svoboda, kreativnost in tradicija pri katerih je vrednost komunalitet nižja od 0.2 in je njihova variabilnost slabše pojasnjena z modelom.

	drugacnost	užitek	vznemerljivost	zabava	kreativnost	svoboda	varnost	ponižnost	obramba	tradicija	ubogljivost	sprejemljivost
drugacnost	50	39	51	35	29	25	-5	-6	11	1	-4	-5
užitek	39	36	38	32	24	20	8	5	19	9	7	8
vznemerljivost	51	38	53	35	30	25	-9	-9	7	-2	-8	-9
zabava	35	32	35	28	21	18	6	3	16	7	5	6
kreativnost	29	24	30	21	17	14	-2	-2	7	1	-1	-2
svoboda	25	20	25	18	14	12	0	-1	7	2	0	0
varnost	-5	8	-9	6	-2	0	31	25	26	20	26	30
ponižnost	-6	5	-9	3	-2	-1	25	20	20	16	21	24
obramba	11	19	7	16	7	7	26	20	26	18	22	25
tradicija	1	9	-2	7	1	2	20	16	18	13	17	19
ubogljivost	-4	7	-8	5	-1	0	26	21	22	17	22	25
sprejemljivost	-5	8	-9	6	-2	0	30	24	25	19	25	29

* all values in cells were multiplied by 100

Na spodnjem grafu pa vidimo matriko odklonov korelacij na podlagi modela od empiričnih korelacij. Izven diagonale so razlike ocenjenih in empiričnih korelacij, na diagonali pa so unikvitete, ki povedo kolikšen delež variabilnosti spremenljivke ni pojasnen s skupnimi faktorji. Vidimo, da so unikvitete relativno visoke, njihove vrednosti pa želimo da so manjše od 0.8. Kot smo videli že pri komunalitetah na zgornjem grafu, spremenljivke kreativnost, svoboda in tradicija niso dobro pojasnene s skupnim faktorskim modelom. Pri ostalih spremenljivkah so vrednosti unikvitet pod 0.8 in njihove variabilnosti so tako dobro pojasnjene. Izven diagonal so vrednosti po absolutni vrednosti v večini blizu 0, kar je dobro, odstopajo le trije pari spremenljivk, ki pa so omenjeni že zgoraj.

	drugacnost	užitek	vznemerljivost	zabava	kreativnost	svoboda	varnost	ponižnost	obramba	tradicija	ubogljivost	sprejemljivost
drugacnost	50	-4	4	-5	5	-3	0	2	1	0	2	2
užitek	-4	64	-2	16	-5	-3	-1	-3	-1	7	-3	-2
vznemerljivost	4	-2	47	-3	-3	-1	-4	4	1	-3	3	1
zabava	-5	16	-3	72	-7	4	6	-3	-3	-1	-4	-5
kreativnost	5	-5	-3	-7	83	12	1	-4	2	-1	0	4
svoboda	-3	-3	-1	4	12	88	4	0	-2	-2	-6	2
varnost	0	-1	-4	6	1	4	69	1	3	0	-7	-1
ponižnost	2	-3	4	-3	-4	0	1	80	0	-10	4	4
obramba	1	-1	1	-3	2	-2	3	0	74	1	4	-5
tradicija	0	7	-3	-1	-1	-2	0	-10	1	87	1	2
ubogljivost	2	-3	3	-4	0	-6	-7	4	4	1	78	5
sprejemljivost	2	-2	1	-5	4	2	-1	4	-5	2	5	71

* all values in cells were multiplied by 100

Faktorske uteži

Pokažimo pattern uteži oziroma regresijske koeficiente ter strukturne uteži oziroma korelacijske koeficiente, ki so za pravokotno (varimax) rotacijo, ki je v našem primeru upoštevana, enake (v pedagoške namene pokažemo obe).

pattern uteži

	ML1	ML2
drugacnost	71	-6
užitek	57	17
vznemerljivost	71	-14
zabava	51	13
kreativnost	41	-1
svoboda	35	1
varnost	-2	56
ponižnost	-4	44
obramba	19	47
tradicija	4	36
ubogljivost	-2	47
sprejemljivost	-2	54

* all values in cells were multiplied by 100

strukturne uteži

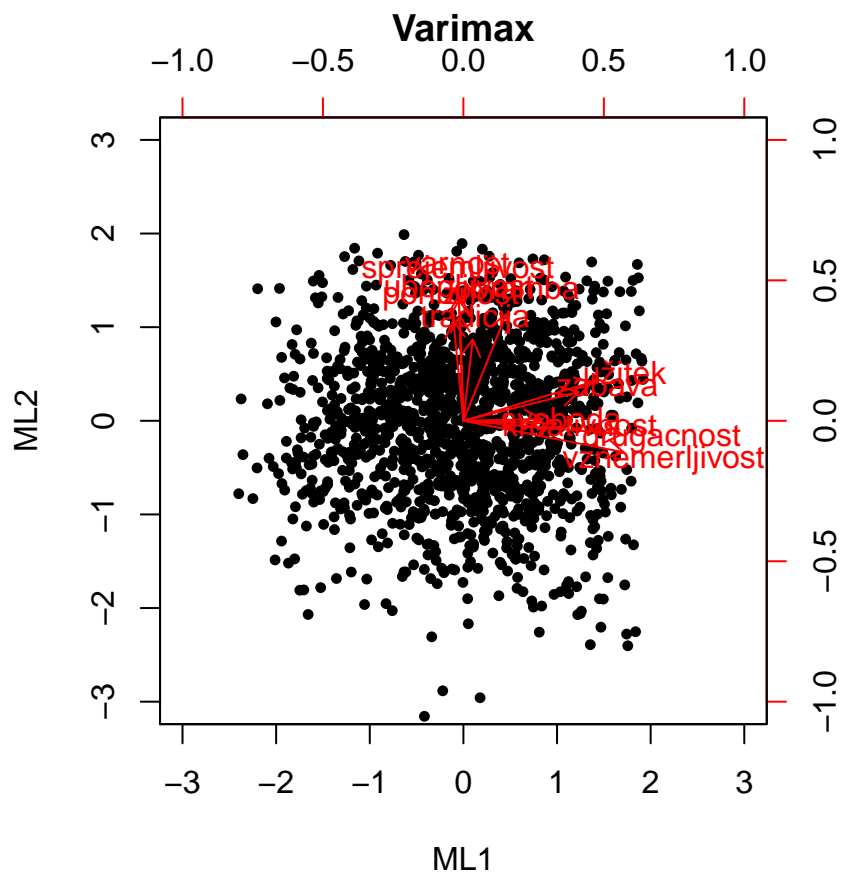
	ML1	ML2
drugacnost	71	-6
užitek	57	17
vznemerljivost	71	-14
zabava	51	13
kreativnost	41	-1
svoboda	35	1
varnost	-2	56
ponižnost	-4	44
obramba	19	47
tradicija	4	36
ubogljivost	-2	47
sprejemljivost	-2	54

* all values in cells were multiplied by

Če se osredotočimo na prvi faktor, vidimo, da ima visoke pozitivne, srednje do močne, korelacije s sklopom spremenljivk, ki merijo pomembnost odprtosti in nizke korelacije s spremenljivkami, ki merijo pomembnost konservativnosti. Morda je za izpostaviti korelacijo prvega faktorja z obrambo, katera je nekoliko višja kot pri ostalih spremenljivkah odprtosti in sicer šibka. Drugi faktor ima visoke pozitivne srednje do močne korelacije s spremenljivkami, ki merijo pomembnost konservativnosti ter nekoliko višjo, šibko, korelacijo tudi s spremenljivkami odprtost in zabava (pozitivna korelacija) ter vznemerljivostjo (negativna korelacija).

Biplot

Tudi biplot pri faktorski analizi nam v dvodimenzionalnem prostoru prikaže položaj posameznih enot glede na vrednosti spremenljivk. Hkrati pa so narisane tudi vrednosti faktorskih uteži.



Iz grafa vidimo, da imajo spremenljivke odprtosti podobne, visoke uteži pri prvem faktorju, medtem ko imajo relativno nizke uteži ostale spremenljivke. Pri drugem faktorju pa imajo visoke in pozitivne uteži spremenljivke konservativnosti, spremenljivke odprtosti pa imajo nizke vrednosti.

Poimenovanje faktorjev

Na podlagi vsega do sedaj povedanega, prvi faktor lahko poimenujemo odprtost, meri odprtost človeka in drugega konservativnost, saj meri konservativnost človeka. Uporabili smo regresijsko metodo.

Analiza korelacij

S pomočjo spodnih korelacij si lahko pomagamo tudi pri preverjanju pravilnega poimenovanja Likartovih lestvic, faktorjev in glavnih komponent. Vidimo, da je Likartova lestvica konservativnosti močno negativno povezana z glavno komponento ne konservativnosti, obenem je šibko pozitivno povezana z glavno komponento odprtosti in močno pozitivno povezana s faktorjem konservativnosti ter zanemarljivo s faktorjem odprtosti. Podobno velja za ostale korelacije. Iz prve naloge vemo tudi, da je korelacija med Likartovimi spremenljivkami šibka (in negativna). Korelacija med obema glavnima komponentama je 0, saj sta pravokotni in podobna velja za faktorja. Vse skupaj nakazuje na pravilno poimenovanje Likartovih lestvic, faktorjev in glavnih komponent.

	L1_konservativnost	L2_odprtost	PC2_nekonservativnost	PC1_odprtost	F2_konservativnost	F1_odprtost
L1_konservativnost	10	0	-10	2	10	0
L2_odprtost	0	10	2	10	0	10
PC2_nekonservativnost	-10	2	10		-10	2
PC1_odprtost	2	10		10	2	10
F2_konservativnost	10	0	-10	2	10	
F1_odprtost	0	10	2	10		10

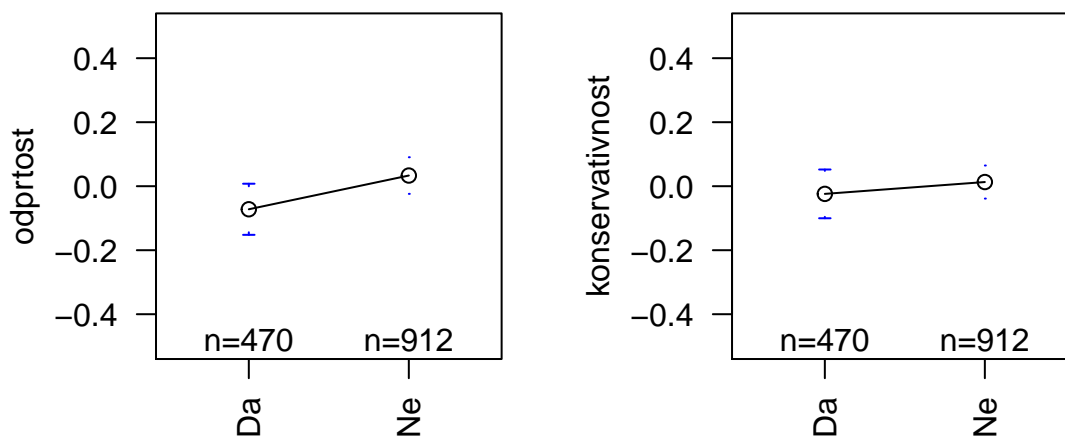
* all values in cells were multiplied by 10

Povezanost faktorjev z ostalimi spremenljivkami

Poglejmo si kako je s povezanostjo faktorjev z izbranimi spremenljivkami nadzora, prebivališča in razmerja, kjer ne upoštevamo manjkajočih vrednosti.

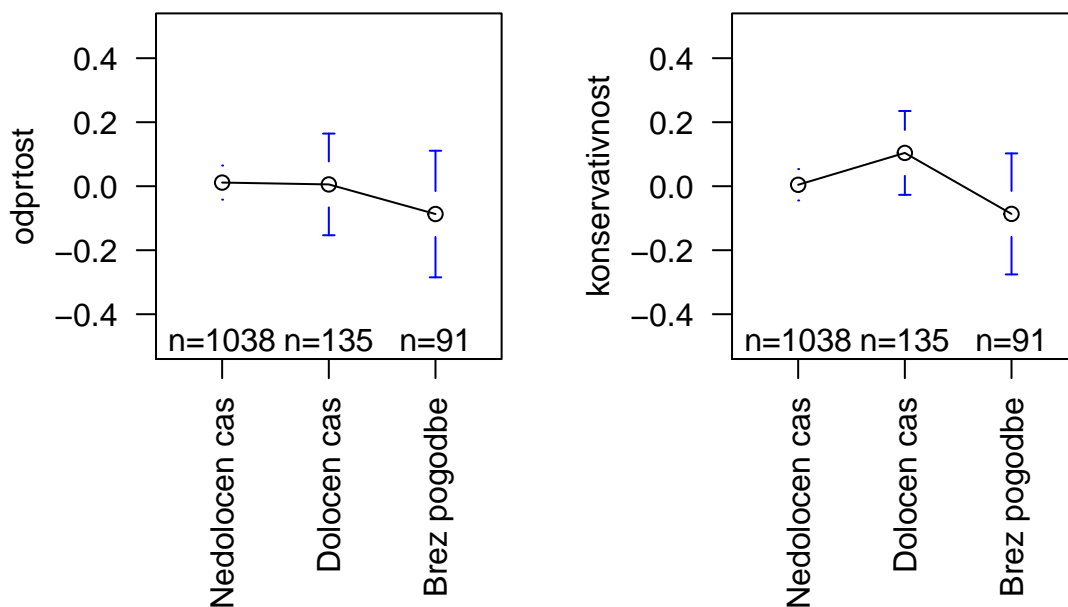
Povezanost faktorjev z nadzorom

Na spodnji sliki sta prikazani povprečji faktorjev odprtosti in konservativnosti glede spremenljivke Nadzor. Vidimo, da so zaposleni, ki niso odgovorni za nadzor nad ostalimi v povprečju nekoliko bolj odprti kot odgovorni za nadzor. Glede konservativnosti sta povprečji skupin spremenljivke Nadzor precej podobni, za odtenek višje vrednosti pa v povprečju ponovno dosegajo zaposleni, ki ne nadzorujejo sodelavcev. Na grafih so prikazane standarizirane vrednosti spremenljivke. S testom t na podlagi vzorca preverimo kako je s povprečjem skupin spremenljivke Nadzor glede faktorjev odprtosti in konservativnosti v populaciji. Na podlagi vrednosti p ($p = 0.035$) lahko s 95% zaupanjem zavrnemo ničelno domnevo, da sta skupini spremenljivke Nadzor enaki glede faktorja odprtosti. 95% interval zaupanja je enak $[0.01, 0.20]$ v korist ne nadzornikov. Na drugi strani pa pri stopnji značilnosti 0.05 ne moremo zavrniti ničelne domneve, da sta skupini spremenljivke Nadzor enaki glede konservativnosti in sprejmemo sklep o njuni enakosti.



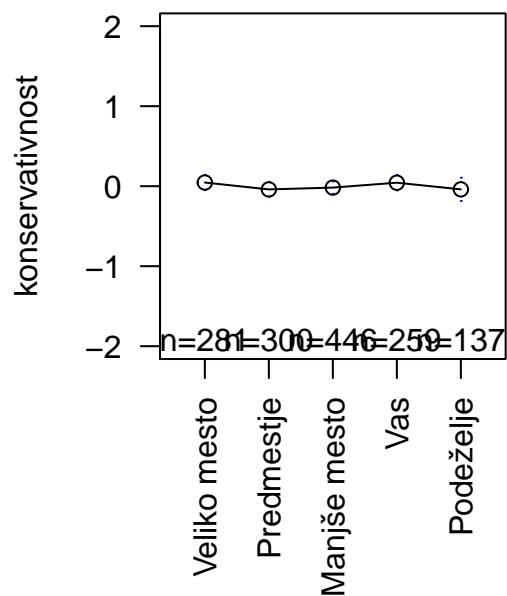
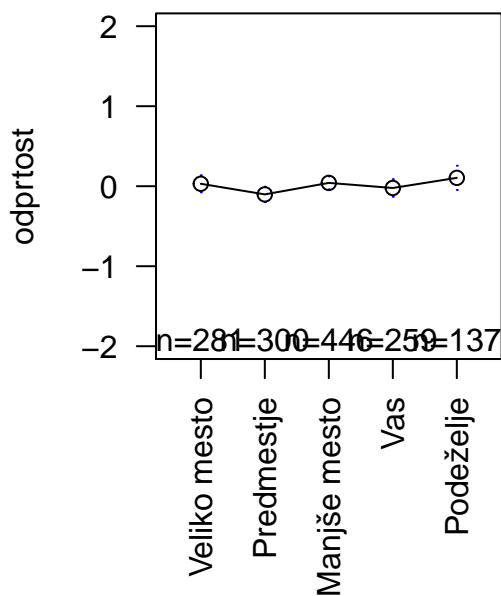
Povezanost faktorjev z razmerjem

Na grafu povezanosti faktorjev s spremenljivko Razmerje vidimo, da zaposleni brez pogodbe dosegajo najnižje standirizirane vrednosti tako pri odprtosti kot pri konservativnosti. Zaposleni s pogodbo imajo v povprečju enako vrednost faktorja odprtosti kot tisti s pogodbo za določen čas, pri faktorju konservativnostu pa v povprečju dosegajo nižje vrednosti. Z Levenovim testom preverimo enakost povprečij po skupinah v populaciji in na podlagi vrednosti p ($p = 0.21$ in $p = 0.31$) pri stopnji značilnosti 0.05 ne moremo zavrniti domneve o enakosti variabilnosti med skupinami pri nobenem izmed faktorjev. Z enosmerno anovo s predpostavko enakih varianc po skupinah, preverimo še domnevo o enakosti povprečij med skupinami v populaciji znotraj posameznega faktorja. Pri nobenem izmed faktorjev pri 95% zaupanju ne moremo zavrniti ničelne domneve in sprejmemo sklep, da so povprečja med skupinami spremenljivke Razmerje v populaciji enaka znotraj obeh faktorjev.



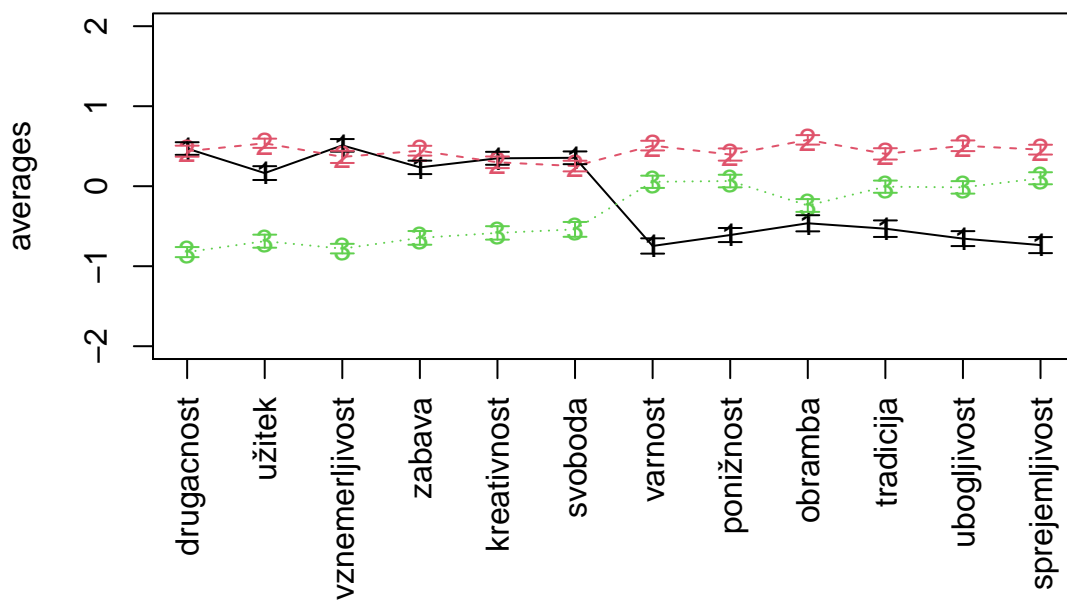
Povezanost faktorjev s prebivališčem

Nazadnje pa preverimo kako je s povezanostjo obeh faktorjev in kraja bivanja v populaciji. Na našem vzorcu se ne kaže, da bi bilo povprečje skupin spremenljivke Prebivališče različno pri nobenem izmed faktorjev odprtosti ali konservativnosti. Spearman koeficient korelacije kaže, da pri stopnji značilnosti 0.05 ne moremo zavrniti ničelne domneve, da je v populaciji vrednost korelacijskega koeficienta med Prebivališčem in odprtostjo ali pa Prebivališčem in konservativnostjo različna od 0, torej da sta spremenljivki povezanosti. V obeh primerih je na vzorcu vrednost Spearmanovega korelacijskega koeficienta pozitivna in zanemarljiva.



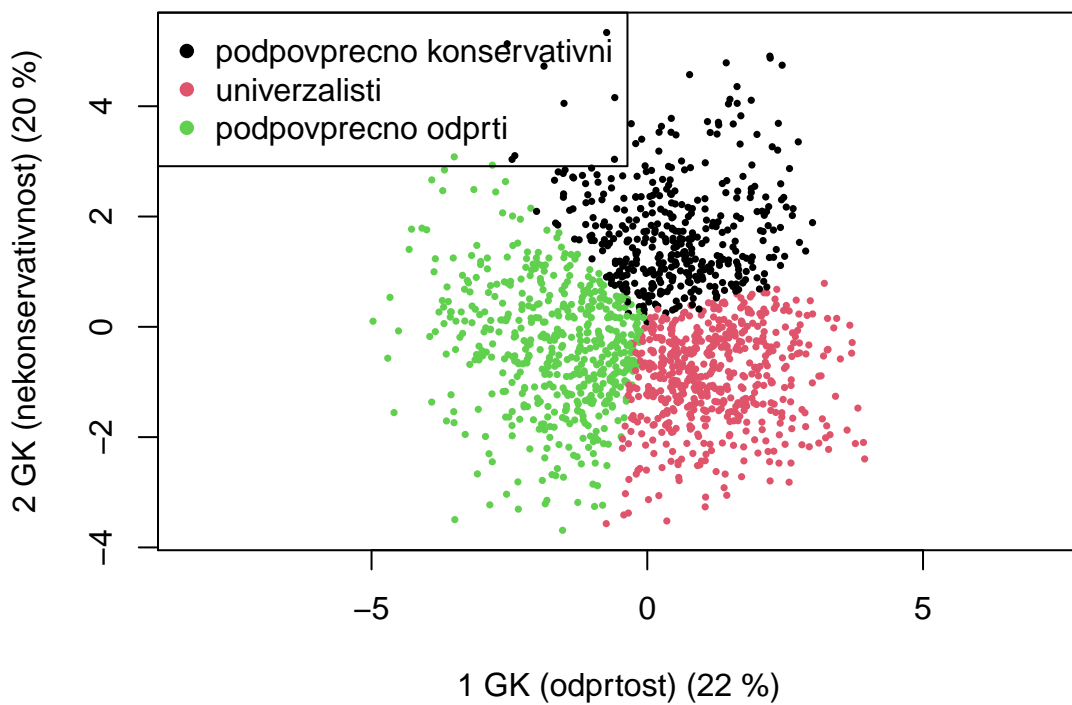
Enote v prostoru

Izrisali bomo enote v prostoru glavnih komponent ter faktorjev in označili skupine, ki smo jih dobili iz razvrščanja. Podatke smo pred razvrščanjem standarizirali, za razvrščanje pa smo uporabili metodo k-means s tremi skupinami. Skupine tudi poimenujemo, skupino ena označimo kot podpovprečno konservativne, skupino 2 kot univerzalistične in skupino 3 kot podpovprečno odprte.



Enote v prostoru glavnih komponent

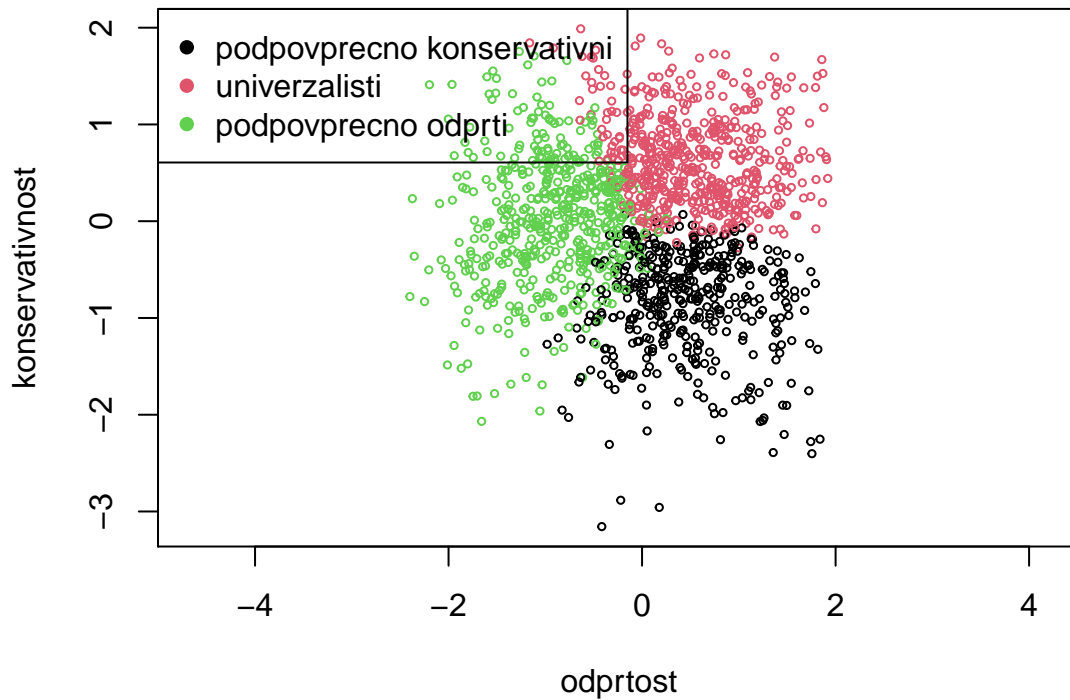
Izrišimo enote v prostoru glavnih komponent in jih pobarvajmo glede na pripadnost skupini iz razvrščanja.



Vidimo, da prva glavna komponenta odprtosti pojasni 22% celotne variabilnosti in druga glavna komponenta nekonservativnosti pojasni 20% celotne variabilnosti.

Enote v prostoru faktorjev

Izrišimo tudi enote v prostoru faktorjev in jih pobarvajmo glede na pripadnost skupini iz razvrščanja. Spodnji graf bomo tudi interpretirali.



Ljudje, ki so v skupini podpovprečno odprti imajo nizke vrednosti faktorja odprtosti in morda malo podpovprečne vrednosti faktorja konservativnosti. Skupina podpovprečno odprti ima nizke vrednosti faktorja konservativnosti in v povprečju pozitivno vrednost faktorja odprtosti. Univerzalisti pa imajo nadpovprečne pozitivne vrednosti tako pri faktorju odprtosti kot konservativnosti.

Zaključek - Alen