Multivariatna analiza –

3. domača naloga

Slavko Žitnik, 63060254

Fakulteta za računalništvo in informatiko, Ljubljana

7.4.2011

# Problem in cilj naloge

V nalogi izvedemo faktorsko analizo nad izbranimi podatki, ki so opisani v drugem razdelku. Uporabimo metodo glavnih osi (PAF) in metodo največjega verjetja (ML) ter primerjamo komunalitete.

# Opis podatkov

Podatke smo pridobili na portalu ESS (*European Social Survey*). Podatki se zbirajo za akademske namene v več kot 30 državah. Cilji anketiranj so spremljanje, interpretacija javnega mnenja in raziskovanje odnosa do spreminjajočih se evropskih institucij. Želijo izboljšati metode meddržavnega anketiranja v Evropi. Navsezadnje žeijo razviti tudi seznam socialnih vedenjskih vzorcev. V prvih 4 sklopih anketiranj je sodelovalo 34 držav.

Posamezna enota anketiranja predstavlja posamezno osebo. Vsi sodelujoči so stari nad 15 let, prebivalci, ki živijo v zasebnih gospodinjstvih, ne glede na njihovo nacionalnost, jezik ali pravni status.

Uporabili bomo rezultate anketiranj iz **Avstrije**, **Nemčije**, **Slovenije** v letu **2006/2007**. V Sloveniji je bila za anketiranje odgovorna Brina Malnar iz Univerze v Ljubljani in skupno 33 ljudi je izvajalo ankete. Več si lahko o teh podatkih preberemo v [3].

## Neodvisne spremenljivke

* **EDULVLA** (*Highest level of education*)

(\* Opomba: Mislim, da je slovenska verzija vprašanja neprimerno zastavljena!)

Vprašanje: Katera je zadnja šola, ki ste jo končali, redno ali izredno?

Možni odgovori:

|  |  |
| --- | --- |
| 0 | nedokončana osnovna šola |
| 1 | dokončana osnovna šola |
| 2 | 2-3 letna poklicna šola |
| 3 | splošna gimnazija, poklicna gimnazija, štiriletna strokovna šola |
| 4 | 2-letna višja (strokovna) šola |
| 5 | visoka šola, fakulteta, akademija |
| 6 | magisterij, doktorat |

* **PDJOBEV** (*Ever had a paid job*)

Vprašanje: Ste imeli kdaj prej plačano delo?

Možni odgovori:

|  |  |
| --- | --- |
| 1 | da |
| 2 | ne |

* **MARITALA** (*Legal marital status*) (\* Enako vprašanje kot MARITAL, vendar se je uporabljal le v ESS1, ESS2)

Vprašanje: Povejte kakšen je vaš sedanji uradni zakonski stan? kateri od naslednjih opisov najbolj ustreza vaši sedanji situaciji.

Možni odgovori:

|  |  |
| --- | --- |
| 01 | poročen |
| 02 | izvenzakonska skupnost |
| 03 | živi ločeno (a legalno še poročen) |
| 04 | živi ločeno (a legalno še vedno v izvenzakonski skupnosti) |
| 05 | ločen |
| 06 | vdovec, vdova |
| 07 | nekoč v izvenzakonski skupnosti, zdaj razpadla |
| 08 | nekoč v izvenzakonski skupnosti, partner umrl |
| 09 | nikoli ni bil poročen in nikoli živel v izvenzakonski skupnosti |

* **AGE** (*Age of respondent, calculated*)

Vprašanje: (\*Avtomatsko pridobljeno iz katerega leta ste bili rojeni / je bila rojena / je bil rojen?)

Možni odgovori:

[starost]

* **GNDR** (*Gender*)

Vprašanje: Spol

Možni odgovori:

|  |  |
| --- | --- |
| 1 | Moški |
| 2 | Ženski |

* **CTZCNTR** (*Citizen of country*)

Vprašanje: Ali ste državljan(ka) Slovenije?

Možni odgovori:

|  |  |
| --- | --- |
| 1 | da |
| 2 | ne |

* **HINCTNT** (*Household's total net income, all sources*)

Vprašanje: S pomočjo te kartice povejte, katera črka najbolje označuje skupni mesečni neto dohodek vašega gospodinjstva (neto – brez davkov in prispevkov). Seštejte dohodke iz vseh virov. Če ne veste natančno, povejte približno.

Možni odgovori:

|  |  |  |
| --- | --- | --- |
| 01 | J | manj kot 650€ |
| 02 | R | od 650€ do 800€ |
| 03 | C | od 800€ do 950€ |
| 04 | M | od 950€ do 1100€ |
| 05 | F | od 1100€ do 1300€ |
| 06 | S | od 1300€ do 1600€ |
| 07 | K | od 1600€ do 1900€ |
| 08 | P | od 1900€ do 2200€ |
| 09 | D | od 2200€ do 3000€ |
| 10 | H | 3000€ in več |

## Odvisne spremenljivke

1. **POLINTR** How interested in politics

Vprašanje: Kaj bi rekli, koliko se zanimate za politiko?

Možni odgovori:

|  |  |
| --- | --- |
| 1 | zelo se zanimam |
| 2 | še kar se zanimam |
| 3 | komaj kaj se zanimam |
| 4 | sploh se ne zanimam |

1. **TRSTPRL** Trust in country's parliament

Vprašanje: Prosim, da s pomočjo lestvice od 0 do 10 ocenite, koliko vi osebno zaupate vsaki od naslednjih ustanov. 0 pomeni, da ustanovi sploh ne zaupate, 10 pa, da ji povsem zaupate. Najprej... (državni zbor)

Možni odgovori:

|  |  |
| --- | --- |
| 00 | sploh ne zaupam |
| 01 |  |
| 02 |  |
| 03 |  |
| 04 |  |
| 05 |  |
| 06 |  |
| 07 |  |
| 08 |  |
| 09 |  |
| 10 | povsem zaupam |

1. **TRSTLGL** Trust in the legal system

Vprašanje: Prosim, da s pomočjo lestvice od 0 do 10 ocenite, koliko vi osebno zaupate vsaki od naslednjih ustanov. 0 pomeni, da ustanovi sploh ne zaupate, 10 pa, da ji povsem zaupate. Najprej... (pravni sistem)

Možni odgovori:

|  |  |
| --- | --- |
| 00 | sploh ne zaupam |
| 01 |  |
| 02 |  |
| 03 |  |
| 04 |  |
| 05 |  |
| 06 |  |
| 07 |  |
| 08 |  |
| 09 |  |
| 10 | povsem zaupam |

1. **MMBPRTY** Member of political party

Vprašanje: Ali ste član kakšne politične stranke?

Možni odgovori:

|  |  |
| --- | --- |
| 1 | da |
| 2 | ne |

1. **TRSTEP** Trust in the European Parliament

Vprašanje: Prosim, da s pomočjo lestvice od 0 do 10 ocenite, koliko vi osebno zaupate vsaki od naslednjih ustanov. 0 pomeni, da ustanovi sploh ne zaupate, 10 pa, da ji povsem zaupate. Najprej... (evropski parlament)

Možni odgovori:

|  |  |
| --- | --- |
| 00 | sploh ne zaupam |
| 01 |  |
| 02 |  |
| 03 |  |
| 04 |  |
| 05 |  |
| 06 |  |
| 07 |  |
| 08 |  |
| 09 |  |
| 10 | povsem zaupam |

1. **STFLIFE** How satisfied with life as a whole

Vprašanje: V celoti gledno, kako zadovoljni ste sedaj s svojim življenjem? Izberite ustrezno vrednost na lestvici od 0 do 10, kjer pomeni 0 levo in 10 desno?

Možni odgovori:

|  |  |
| --- | --- |
| 00 | izjemno nezadovoljen |
| 01 |  |
| 02 |  |
| 03 |  |
| 04 |  |
| 05 |  |
| 06 |  |
| 07 |  |
| 08 |  |
| 09 |  |
| 10 | izredno zadovoljen |

1. **STFHLTH** State of health services in country nowadays

Vprašanje: In kakšno je po vašem mnenju splošno stanje zdravstva danes v Sloveniji?

Možni odgovori:

|  |  |
| --- | --- |
| 00 | izjemno slabo |
| 01 |  |
| 02 |  |
| 03 |  |
| 04 |  |
| 05 |  |
| 06 |  |
| 07 |  |
| 08 |  |
| 09 |  |
| 10 | izredno dobro |

# Opis izbrane metode

## Faktorska analiza

Faktorska analiza je podobna kot metoda glavnih komponent. Pri metodi glavnih komponent iščemo nekaj prvih komponent, ki so linearne kombinacije izmerjenih spremenljivk in pojasnjujejo večji del variabilnosti. Pri faktorski analizi poskušamo najti manjšo množico spremenljivk, ki izražajo to, kar izražajo merjene spremenljivke.

Faktorsko analizo uporabimo, ko želimo meriti nekaj, kar v okolju ne moremo neposredno izmeriti. Tako si izberemo nekaj posrednih spremenljivk in skušamo pojasniti nemerljivo.

Delež variance, ki jo pojasnimo s skupnimi faktorji, imenujemo komunaliteta. Spodobi se, da je večja ali enaka vsaj 20%.

V naši nalogi uporabimo metodo glavnih osi (angl. *Principal axis factoring*) in metodo največjega verjetja (angl. *Maximum likelihood*). Slednjo je priporočljivo uporabiti, če imamo vzorce vsaj 400 enot, kar v našem primeru velja.

Pri faktorski analizi tudi rotiramo prostor z namenom, da se koordinatne osi kar se da prilegajo točkam v prostoru. Rotacije delimo na pravokotne – faktorji so med seboj neodvisni, in poševne – faktorji so med seboj odvisni. Mi si bomo za pravokotno rotacijo izbrali QUARTIMAX rotacijo in OBLIMIN za poševno.

## Pomembni termini iz statističnega slovarja

* faktorska analiza: analiza, s katero se ugotavlja pojasnljivost opazovanih spremenljivk z manjšim številom faktorjev (2).
* metoda glavnih osi: metoda za izračunavanje faktorskih uteži.
* metoda največjega verjetja: metoda določanja cenilk, pri kateri je cenilka določena tako, da je za dani vzorec verjetje največje.
* komunaliteta: delež variance spremenljivke, ki je pojasnjen s skupnimi faktorji.

# Predstavitev rezultatov

V prejšnji nalogi smo pri metodi glavnih komponent ugotovili, da prva komponenta pojasni 35% variance, druga pa 16%. Na podlagi tega se pri izvedbi faktorske analize odločimo, da jo bomo izvajali z dvema faktorjema.

## Faktorska analiza brez rotacij

Najprej bomo uporabili metodi PAF in ML ter primerjali njune rezultate.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PAF | | | ML | | |
|  | F1 | F2 | h2 | F1 | F2 | h2 |
| Trust in country's parliament | **,871** | -,258 | **,825** | **,901** | -,127 | **,828** |
| Trust in the European Parliament | **,596** | -,188 | **,391** | **,622** | -,105 | **,398** |
| Trust in the legal system | **,653** | -,023 | **,427** | **,645** | ,083 | **,423** |
| How interested in politics | -,146 | ,114 | ,034 | -,159 | ,107 | ,037 |
| How satisfied with life as a whole | ,351 | ,165 | ,151 | ,314 | ,248 | ,160 |
| State of health services in country nowadays | **,567** | ,548 | **,621** | ,479 | **,567** | **,551** |
| lastna vrednost *i* | 2,45 | 1,0 |  | 2,45 | 1,0 |  |
| %poj. var. | 40,85 | 16,8 |  | 40,85 | 16,8 |  |

Faktorja, ki jih razkrijeta metodi PAF in ML sta precej podobna. Vrednosti gledamo le, kjer je komunaliteta večja od 20%. Prvi faktor ima v obeh primerih močne uteži pri spremenljivkah zaupanja in mnenju o zdravstvenih storitvah. Drugi faktor se izkaže nekoliko boljši pri metodi ML, kjer bolje določa zadovoljstvo z življenjem. Ker ni tako zelo očitnih razlik, bi lahko 2. faktor tudi zanemarili.

Za nadaljevanje faktorske analize z rotacijami si izberimo metodo ML.

## Faktorska analiza z rotacijami.

Za pravokotno rotacijo uporabimo QUARTIMAX in OBLIMIN za poševno.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ML - QUARTIMAX | | | ML – OBLIMIN - Structure | | | ML – OBLIMIN - Pattern | | |
|  | F1 | F2 | h2 | F1 | F2 | h2 | F1 | F2 | h2 |
| Trust in country's parliament | **,901** | -,127 | **,828** | **,891** | ,533 | **,828** | **,809** | ,201 | **,828** |
| Trust in the European Parliament | **,622** | -,105 | **,398** | **,622** | ,355 | **,398** | **,572** | ,120 | **,398** |
| Trust in the legal system | **,645** | ,083 | **,423** | **,580** | ,507 | **,423** | **,447** | ,323 | **,423** |
| How interested in politics | -,159 | ,107 | ,037 | -,186 | -,033 | ,037 | -,208 | ,052 | ,037 |
| How satisfied with life as a whole | ,314 | ,248 | ,160 | ,212 | ,396 | ,160 | ,059 | ,372 | ,160 |
| State of health services in country nowadays | ,479 | **,567** | **,551** | ,260 | **,741** | **,551** | -,053 | **,762** | **,551** |
| lastna vrednost *i* | 2,45 | 1,0 |  | 2,45 | 1,0 |  | 2,45 | 1,0 |  |
| %poj. var. | 40,85 | 16,8 |  | 40,85 | 16,8 |  | 40,85 | 16,8 |  |

V spodnji tabeli je prikazana korelacija med obema faktorjema pri poševni metodi OBLIMIN. Iz tega sklepamo, da faktorja nista čisto neodvisna.

| **Factor Correlation Matrix** | | |
| --- | --- | --- |
| Factor | 1 | 2 |
| 1 | 1,000 | ,410 |
| 2 | ,410 | 1,000 |
| Extraction Method: Maximum Likelihood.  Rotation Method: Oblimin with Kaiser Normalization. | | |

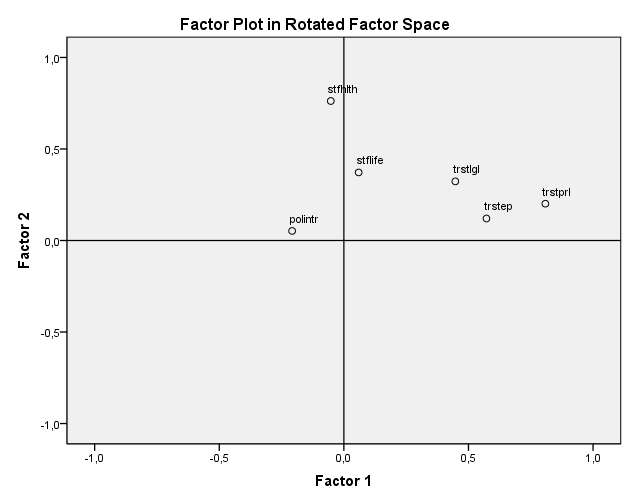
Metoda QUARTIMAX je pravokotna rotacija. Glede na rezultate v tabeli lahko ugotovimo, da se metoda OBLIMIN s pravokotnimi projekcijami (»structure« uteži) odreže bolje. To lahko razberemo tudi iz spodnjih dveh slik, kjer je očitno, da Faktor 2 pri OBLIMIN metodi bolje ločuje spremenljivko za stanje zdravstva.

Faktor 1 obsega zaupanje v politične sisteme. Faktor 2 pa malo bolje določa stanje zdravstva v državi.

QUARTIMAX:

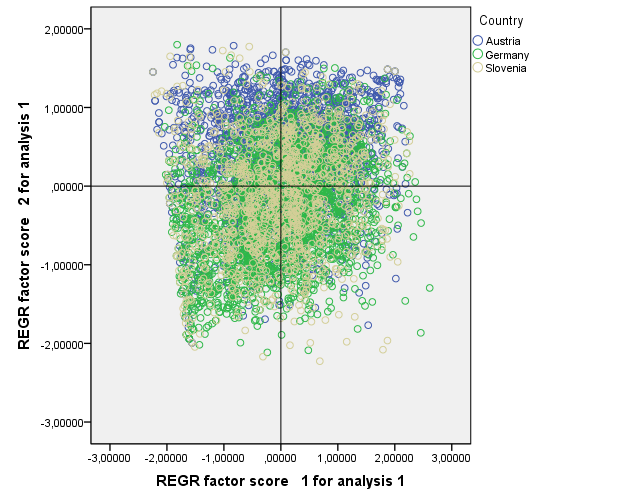


OBLIMIN:

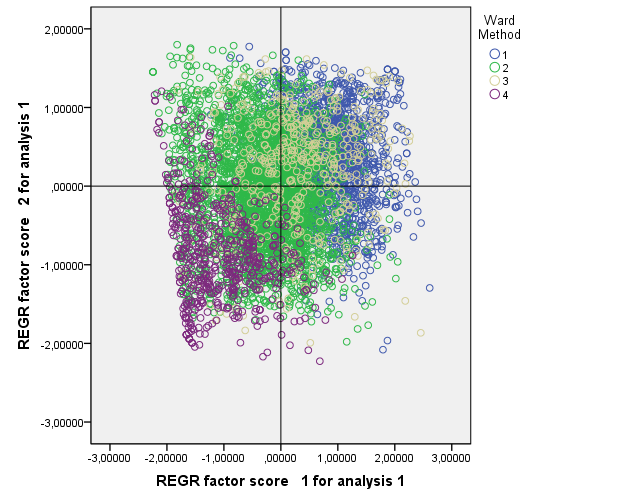


## Povezanost med faktorji in drugimi spremenljivkami

Na spodnji sliki vidimo, kako faktorja ločujeta enote glede na državo. Opazimo, da za ta primer nismo dobili nič pametnega.



Malo bolje se izkaže, če uporabimo 4 skupine, ki smo jih z razvrščanjem izbrali v prejšnji nalogi. Vidimo, da se tri skupine lepo ločijo med sabo. Modra – 1. skupina, ki smo jo poimenovali ZADOVOLJNEŽI se lepo ločijo po tem, da zaupajo v politične zadeve in stanje zdravstva. V 4. skupini so PESIMISTI – ne zaupajo političnim ustanovam in niso zadovoljni z zdravstvom.



# Interpretacija in evaluacija rezultatov

V nalogi smo se ukvarjali s faktorsko analizo. Ugotovili smo, da prvi faktor obsega spremenljivke, ki določajo zaupanje v politične ustanove, drugi faktor pa malo bolje stanje zdravstva v državi.

Pri analizi skupin iz prejšnje naloge smo ugotovili, da zelo lepo identificiramo dve skupini. Pri analizi držav pa rezultati ne pokažejo nič pametnega.