Multivariatna analiza –

5. domača naloga

Slavko Žitnik, 63060254

Fakulteta za računalništvo in informatiko, Ljubljana

21.4.2011

# Problem in cilj naloge

V tej nalogi bomo uporabili metodo diskriminantne analize. Z njeno pomočjo bomo poiskali diskriminantne spremenljivke, ki jih dobimo s pomočjo obstoječih, z namenom da bodo kar se da ločevale skupine enot, ki so v našem primeru države.

# 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

## Diskriminantna analiza

Z diskriminantno analizo iščemo tako linearno kombinacijo merjenih spremenljivk – diskriminantna spremenljivka, da bo maksimalno ločila vnaprej določene skupine in bo napaka pri uvrščanju enot v skupine najmanjša. Diskriminantno spremenljivko lahko uporabimo za pojasnjevanje razlik med skupinami ali klasifikacijo novih enot v skupine.

## Pomembni termini iz statističnega slovarja

* diskriminantna analiza: analiza, s katero se poiščejo linearne kombinacije opazovanih spremenljivk in ločijo vnaprej določene skupine tako, da je verjetnosti napakepri uvrščanju enot v skupine najmanjša
* diskriminantna spremenljivka: funkcija, ki je linearna kombinacija opazovanih spremenljivk v diskriminantni analizi in ki se uporablja za uvrstitev opazovanih enot v eno od vnaprej določenih skupin

# Predstavitev rezultatov

| Country | How interested in politics | Trust in the legal system | Trust in country's parliament | Trust in the European Parliament | Member of political party | How satisfied with life as a whole | State of health services in country nowadays |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Austria | 2,47 | **5,99** | **4,89** | 3,95 | 1,85 | **7,53** | **6,66** |
| Germany | 2,40 | 5,46 | 4,09 | 4,00 | 1,96 | 6,71 | 4,39 |
| Slovenia | **2,71** | 4,17 | 4,22 | **4,98** | 1,95 | 6,97 | 5,17 |
| Total | 2,49 | 5,37 | 4,40 | 4,19 | 1,92 | 7,06 | 5,36 |

Naši podatki so zbrani iz treh držav – Avstrije, Nemčije in Slovenije. Te države bomo poskušali ločiti glede na diskriminantni spremenljivki, ki jih bomo dobili iz odvisnih spremenljivk, ki so s področja politike.

Tabela : Povprečja odvisnih spremenljivk

Najprej moramo preveriti, če med skupinami obstajajo kakšne statistično značilne razlike. V tabeli 1 vidimo povprečja posameznih odvisnih spremenljivk po skupinah in skupno. Z odebeljeno smo označili povprečja, ki bolj odstopajo od ostalih. Tako na primer opazimo, da so Avstrijci najbolj zadovoljni z življenjem.

V tabeli 2 z F-testom testiramo ničelno hipotezo enakosti povprečij za posamezno spremenljivko med skupinami. Kot vidimo, lahko hipotezo zavrnemo in nadaljujemo z diskriminantno analizo.

|  | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | Wilks' Lambda | F | df1 | df2 | Sig. |
| How interested in politics | ,982 | 53,256 | 2 | 5957 | ,000 |
| Trust in country's parliament | ,980 | 61,077 | 2 | 5957 | ,000 |
| Trust in the legal system | ,931 | 221,801 | 2 | 5957 | ,000 |
| Trust in the European Parliament | ,971 | 89,145 | 2 | 5957 | ,000 |
| Member of political party | ,964 | 110,350 | 2 | 5957 | ,000 |
| How satisfied with life as a whole | ,973 | 81,767 | 2 | 5957 | ,000 |
| State of health services in country nowadays | ,840 | 568,031 | 2 | 5957 | ,000 |

Tabela : Test enakosti povprečij med skupinami

V tabeli 3 prikazujemo lastni vrednosti obeh diskriminantnih spremenljivk. Opazimo, da prva pojasni 63% skupne variance in ima tudi večji kanonični korelacijski koeficient. Velik del razlik lahko torej pojasnimo že s prvo spremenljivko. V tabeli 4 pokažemo še, da s testom zavrnemo hipotezo, da so aritmetične sredine diskriminantnih spremenljivk enake, torej vemo, da obe prispevata k razlikovanju med državami.

|  | | | | |
| --- | --- | --- | --- | --- |
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | ,281a | 63,2 | 63,2 | ,468 |
| 2 | ,164a | 36,8 | 100,0 | ,375 |
|  | | | | |

Tabela : Lastne vrednosti, delež variance in kanonični korelacijski koeficient diskriminantnihtnih spremenljivk

|  | | | | |
| --- | --- | --- | --- | --- |
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 through 2 | ,671 | 2376,902 | 14 | ,000 |
| 2 | ,859 | 902,611 | 6 | ,000 |

Tabela : test

V spodnjih dveh tabelah prikazujemo uteži za diskriminantni spremenljivki. V tabeli 5 so standardizirane uteži, ki jih ponavadi uporabljamo za napovedovanje novih primerov. Mi bomo uporabili strukturne uteži, ki so boljše za pojasnjevanje.

|  | | |
| --- | --- | --- |
|  | Function | |
| 1 | 2 |
| How interested in politics | -,022 | ,371 |
| Trust in country's parliament | ,189 | ,107 |
| Trust in the legal system | ,302 | -,951 |
| Trust in the European Parliament | -,666 | ,597 |
| Member of political party | -,325 | -,170 |
| How satisfied with life as a whole | ,087 | ,166 |
| State of health services in country nowadays | ,749 | ,493 |

Tabela : Standardizirane uteži diskriminantnih spremenljivk

V tabeli 5 smo pri obeh diskriminatnih spremenljivkah odebeljeno označili uteži, s katerimi potem na sliki razlagamo razlike med skupinami – državami. Prva diskriminantna spremenljivka nam tako meri odnos državljanov posamezne države do zdravstva in zaupanja v pravni sistem ter članstvo v političnih strankah. Pri drugi spremenljivki pa smmo se odločili, da nam bo poleg stanja v zdravstvu in zaupanja v pravni sistem merila še stopnjo zaupanja v evropski parlament in zainteresiranost v politiko.

|  | | |
| --- | --- | --- |
|  | Function | |
| 1 | 2 |
| State of health services in country nowadays | **,761\*** | **,414** |
| Member of political party | **-,353\*** | -,111 |
| How satisfied with life as a whole | ,286\* | ,166 |
| Trust in country's parliament | ,264\* | ,077 |
| Trust in the legal system | **,395** | **-,432\*** |
| Trust in the European Parliament | -,179 | **,358\*** |
| How interested in politics | -,075 | **,316\*** |

Tabela : Strukturne uteži diskriminantnih spremenljivk

|  | | | |
| --- | --- | --- | --- |
| Drzave - numeric form, 0-AT, 1-DE, 2-SI | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| Avstrija | **,355** | 2116 | 2116,000 |
| Nemcija | **,433** | 2579 | 2579,000 |
| Slovenija | **,212** | 1265 | 1265,000 |
| Total | 1,000 | 5960 | 5960,000 |

Tabela : Priorne verjetnosti za posamezno državo

V tabeli 7 smo z odebeljeno označili priorne verjetnosti za posamezne države, t.j. če bi brez kakršnekoli analize hoteli klasificirati nek primer bi z uporabo večinskega klasifikatorja klasificirali vse v Nemčijo, ki ima priorno verjetnost 43%. Rezultate naše metode lahko smatramo za uspešne, če je predikcija za posamezno državo bolj uspešna od njene priorne verjetnosti. Če primerjamo odebeljene vrednosti med tabelo 7 in 8, ugotovimo, da so za vsako državo verjetnosti pravilnih napovedi večje od priornih.

|  | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Drzave - numeric form, 0-AT, 1-DE, 2-SI | Predicted Group Membership | | | Total |
|  |  | Avstrija | Nemcija | Slovenija |
| Original | Count | Avstrija | 1354 | 623 | 139 | 2116 |
| Nemcija | 529 | 1837 | 213 | 2579 |
| Slovenija | 288 | 557 | 420 | 1265 |
| % | Avstrija | **64,0** | 29,4 | 6,6 | 100,0 |
| Nemcija | 20,5 | **71,2** | 8,3 | 100,0 |
| Slovenija | 22,8 | 44,0 | **33,2** | 100,0 |
|  | | | | | | |

Tabela : Klasifikacijska matrika

V tabeli 9 so prikazani centroidi izbranih državo za obe diskriminantni spremenljivki. Ti centroidi so narisani na sliki 1, ob kateri bomo tudi razložili rezultate.

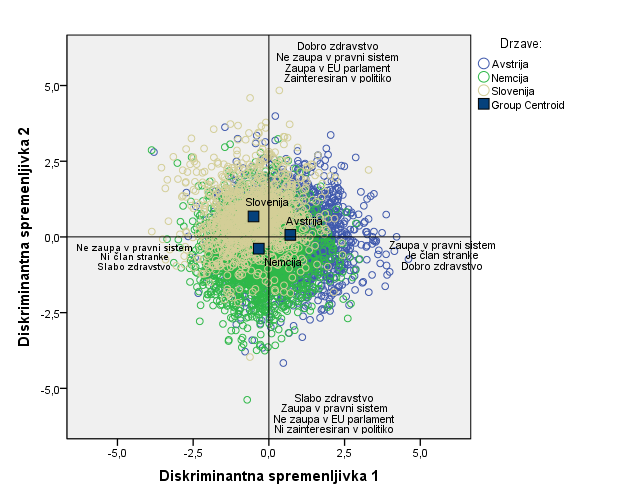
Iz grafa in centroidov na sliki 1 lahko ugotovimo naslednje: Med državami so opazne razlike. Kljub tem razlikam, se del množic enot posameznih držav prekriva med seboj.

Ugotovimo lahko naslednje:

* Za Slovence lahko trdimo, da ne zaupajo v pravni sistem, vendar zaupajo v evropski parlament in so zainteresirani v politiko. Praviloma niso člani političnih strank. Glede stanja v zdravstvu se ne moremo preveč opredeljevati, ker so premajne razlike in lahko rečemo, da ostanemo glede tega nevtralni.
* Za Nemce trdimo, da ne zaupajo v evropski parlament, niso zainteresirani v politiko in so mnenja, da imajo slab zdravstveni sistem. Praviloma niso člani političnih strank. O zaupanju v pravni sistem lahko ostanemo nevtralnega mnenja.
* Pri Avstrijcih lahko upoštevamo le prvo diskriminantno spremenljivko, saj jih druga le malo premakne v smeri ordinatne osi. Zanje lahko rečemo, da zaupajo v svoj pravni sistem, so člani strank in pravijo, da imajo dobro urejeno zdravstvo.

|  | | |
| --- | --- | --- |
| Drzave - numeric form, 0-AT, 1-DE, 2-SI | Function | |
| 1 | 2 |
| Avstrija | ,709 | ,067 |
| Nemcija | -,333 | -,387 |
| Slovenija | -,506 | ,677 |
|  | | |

Tabela : Centroidi za posamezne države



Slika : Projekcija podatkov, določenih z diskriminantnima spremenljivkama

# Interpretacija in evaluacija rezultatov

V nalogi smo analizirali podatke iz sklopa politike raziskav ESS. Ukvarjali smo se z diskriminantno analizo in skušali najti diskriminantni funkciji, ki med seboj ločujeta izbrane države – Slovenijo, Avstrijo in Nemčijo.

Celotno analizo smo opisali v četrtem razdelku. Med državami smo ugotovili razlike. Na kratko lahko rečemo, da je za Slovence značilno, da zaupajo v evropski parlament, Nemci trdijo, da imajo slabo zdravstvo in Avstrijci zaupajo v svoj pravni sistem ter pravijo, da imajo dobro zdravstvo.