Multivariatna analiza –

4. domača naloga

Slavko Žitnik, 63060254

Fakulteta za računalništvo in informatiko, Ljubljana

14.4.2011

# Problem in cilj naloge

V nalogi izvedemo regresijsko analizo. Odvisna spremenljivka nam bo predstavljala prvi faktor, ki smo ga dobili v prejšnji domači nalogi. To spremenljivko bomo opisali in pojasnili vpliv na izbrano odvisno spremenljivko.

# 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

## Regresija

Regresija je postopek, da z eno ali več spremenljivkami modeliramo neko novo spremenljivko, za katero skleplamo, da je od teh odvisna. Multipla regresijska analiza proučuje odvisnost ene spremenljivke od več neodvisnih. Iščemo torej naslednji model:

$$Y= β\_{0}+ β\_{1}X\_{1}+ β\_{2}X\_{2}+ β\_{3}X\_{3}+…+β\_{n}X\_{n}+E$$

E predstavlja napako modela oz. vpliv ostalih neodvisnih spremenljivk. Naš cilj je, da poiščemo čimboljše koeficiente $β\_{i}$, ki najbolj določajo spremenljivko *Y* glede na podane spremenljivke $X\_{i}$. Eden izmed postopkov, ki ga lahko za to uporabimo, je metoda najmanjših kvadratov.

## Pomembni termini iz statističnega slovarja

* ANOVA: (angl. Analysis of Variance), analiza variance
* regresija: odnos med pojasnjevano spremenljivko in eno pojasnjevalno spremenljivko ali več pojasnjevalnimi spremenljivkami
* multipla regresija: regresija, s katero se proučuje odvisnost kake spremenljivke od več neodvisnih
* F preizkus: preizkus, ki temelji na F porazdelitvi ustrezne statistike, npr. preizkus enakosti varianc v dveh populacijah ob predpostavki, da se spremenljivka v populacijah porazdeljuje normalno

# Predstavitev rezultatov

## Uporabljene neodvisne spremenljivke

Izvedli bomo univariatno analizo neodvisnih spremenljivk. Izbrali smo naslednje spremenljivke: E**DULVLA** (*Highest level of education*), **PDJOBEV** (*Ever had a paid job*), **AGE** (*Age of respondent, calculated*), **GNDR** (*Gender*), **CTZCNTR** (*Citizen of country*), **HINCTNT** (*Household's total net income, all sources*).

V spodnji tabeli predstavljamo opisno statistiko neodvisnih spremenljivk za vse enote, ki jih imamo v podatkih. Osnovne podatke o nominalnih spremenljivkah sledijo pod njo. Grafični prikaz frekvenčnih porazdelitev vseh spremenljivk se nahaja v prilogi.

| **Descriptive Statistics** |
| --- |
|  | Statistic | Std. Error | Bootstrapa |
| Bias | Std. Error | 95% Confidence Interval |
| Lower | Upper |
| Highest level of education | N | 4805 |  | 0 | 0 | 4805 | 4805 |
| Minimum | 1 |  |  |  |  |  |
| Maximum | 5 |  |  |  |  |  |
| Mean | 3,23 |  | ,00 | ,01 | 3,20 | 3,26 |
| Std. Deviation | ,995 |  | -,001 | ,009 | ,976 | 1,013 |
| Variance | ,991 |  | -,001 | ,019 | ,953 | 1,027 |
| Skewness | ,549 | ,035 | ,001 | ,025 | ,505 | ,599 |
| Kurtosis | -,247 | ,071 | ,006 | ,064 | -,375 | -,115 |
| Age of respondent, calculated | N | 4805 |  | 0 | 0 | 4805 | 4805 |
| Minimum | 14,83 |  |  |  |  |  |
| Maximum | 96,42 |  |  |  |  |  |
| Mean | 47,7408 |  | -,0043 | ,2580 | 47,2559 | 48,2447 |
| Std. Deviation | 17,75838 |  | -,00573 | ,14304 | 17,47094 | 18,02475 |
| Variance | 315,360 |  | -,183 | 5,078 | 305,234 | 324,892 |
| Skewness | ,142 | ,035 | ,001 | ,021 | ,101 | ,183 |
| Kurtosis | -,773 | ,071 | ,000 | ,029 | -,829 | -,713 |
| Household's total net income, all sources | N | 4805 |  | 0 | 0 | 4805 | 4805 |
| Minimum | 1 |  |  |  |  |  |
| Maximum | 12 |  |  |  |  |  |
| Mean | 6,20 |  | ,00 | ,03 | 6,14 | 6,25 |
| Std. Deviation | 2,041 |  | -,001 | ,019 | 2,002 | 2,075 |
| Variance | 4,164 |  | -,003 | ,078 | 4,007 | 4,307 |
| Skewness | ,168 | ,035 | ,000 | ,029 | ,109 | ,226 |
| Kurtosis | -,323 | ,071 | ,001 | ,046 | -,413 | -,232 |
| Valid N (listwise) | N | 4805 |  | 0 | 0 | 4805 | 4805 |
| a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples |

V spodnjih tabelah predstavimo še osnovne statistike dihotomnih spremenljivk **PDJOBEV** (*Ever had a paid job*), **GNDR** (*Gender*) in Ali ste državljan Slovenije/Avstrije/Nemčije - **CTZCNTR** (*Citizen of country*).

| **Ever had a paid job** |
| --- |
|  | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Yes | 2449 | 36,0 | 80,6 | 80,6 |
| No | 590 | 8,7 | 19,4 | 100,0 |
| Total | 3039 | 44,7 | 100,0 |  |
| Missing | Not applicable | 3705 | 54,5 |  |  |
| Refusal | 19 | ,3 |  |  |
| No answer | 34 | ,5 |  |  |
| Total | 3758 | 55,3 |  |  |
| Total | 6797 | 100,0 |  |  |

| **Citizen of country** |
| --- |
|  | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Yes | 6610 | 97,2 | 97,3 | 97,3 |
| No | 185 | 2,7 | 2,7 | 100,0 |
| Total | 6795 | 100,0 | 100,0 |  |
| Missing | Refusal | 2 | ,0 |  |  |
| Total | 6797 | 100,0 |  |  |

| **Gender** |
| --- |
|  | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Male | 3222 | 47,4 | 47,4 | 47,4 |
| Female | 3575 | 52,6 | 52,6 | 100,0 |
| Total | 6797 | 100,0 | 100,0 |  |

V naslednji tabeli prikazujemo Pearsonovo korelacijsko matriko odvisnih spremenljivk, ki smo jo opisali že v 1. domači nalogi.

| **Correlationsa** |
| --- |
|  | How interested in politics | Trust in country's parliament | Trust in the legal system | Member of political party | Trust in the European Parliament | How satisfied with life as a whole | State of health services in country nowadays |
| How interested in politics | 1 | -,159\*\* | -,105\*\* | ,186\*\* | -,084\*\* | -,090\*\* | ,004 |
|  | ,000 | ,000 | ,000 | ,000 | ,000 | ,761 |
| Trust in country's parliament | -,159\*\* | 1 | ,569\*\* | -,097\*\* | ,574\*\* | ,244\*\* | ,362\*\* |
| ,000 |  | ,000 | ,000 | ,000 | ,000 | ,000 |
| Trust in the legal system | -,105\*\* | ,569\*\* | 1 | -,063\*\* | ,392\*\* | ,246\*\* | ,351\*\* |
| ,000 | ,000 |  | ,000 | ,000 | ,000 | ,000 |
| Member of political party | ,186\*\* | -,097\*\* | -,063\*\* | 1 | -,019 | -,073\*\* | -,094\*\* |
| ,000 | ,000 | ,000 |  | ,139 | ,000 | ,000 |
| Trust in the European Parliament | -,084\*\* | ,574\*\* | ,392\*\* | -,019 | 1 | ,171\*\* | ,237\*\* |
| ,000 | ,000 | ,000 | ,139 |  | ,000 | ,000 |
| How satisfied with life as a whole | -,090\*\* | ,244\*\* | ,246\*\* | -,073\*\* | ,171\*\* | 1 | ,291\*\* |
| ,000 | ,000 | ,000 | ,000 | ,000 |  | ,000 |
| State of health services in country nowadays | ,004 | ,362\*\* | ,351\*\* | -,094\*\* | ,237\*\* | ,291\*\* | 1 |
| ,761 | ,000 | ,000 | ,000 | ,000 | ,000 |  |
| \*\*. Correlation is significant at the 0.01 level (2-tailed).a. Listwise N=5960 |

## Odvisna spremenljivka Zaupanje

V nadaljevanju bomo modelirali spremenljivko Zaupanje, ki smo jo dobili kot prvi faktor pri faktorski analizi v prejšnji domači nalogi. Spremenljivka najbolj očitno meri zaupanje ljudi v politične ustanove v državi in Evropski uniji.

### Univariatna analiza

Najprej izvedimo univariatno analizo spremenljivke Zaupanje. V spodnji tabeli navajamo osnovne statistike. Ugotovimo, da ima povprečje enako 0, ker je normalizirana. Spremenljivka je tudi asimetrična v levo in nekoliko bolj sploščena. Pod tabelo predstavimo tudi frekvenčni histogram spremenljivke Zaupanje.

| **Descriptive Statistics** |
| --- |
|  | N | Minimum | Maximum | Mean | Std. Deviation | Variance | Skewness | Kurtosis |
| Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | S. E. | Statistic | S. E. |
| Zaupanje | 6027 | -2,04773 | 2,38929 | ,0000000 | ,91635481 | ,840 | -,088 | ,032 | -,480 | ,063 |
| Valid N (listwise) | 6027 |  |  |  |  |  |  |  |  |  |



### Faktorska analiza

V spodnji tabeli smo ponovili faktorsko analizo po metodi največjega verjetja z rotacijo quartimax.

V levem delu tabele prikazujemo rezultate iz prejšnje domače naloge, v desnem delu pa rezultate ob ponovnem zagonu faktorske analize in uporabi le tistih spremenljivk, ki so pojasnile dober delež variance (t.j. so imeli komunaliteto nad 0,2 in so v tabeli označeni z odebeljeno pisavo).

|  |  |  |
| --- | --- | --- |
|  | ML – QUARTIMAX - \*rezultati iz prejšnje naloge | ML – QUARTIMAX |
|  | F1 | h2 | F1 | h2 |
| Trust in country's parliament | **,901** | **,828** | **,885** | **,783** |
| Trust in the European Parliament | **,622** | **,398** | **,639** | **,409** |
| Trust in the legal system | **,645** | **,423** | **,648** | **,420** |
| How interested in politics | -,159 | ,037 | 0 | 0 |
| How satisfied with life as a whole | ,314 | ,160 | 0 | 0 |
| State of health services in country nowadays | **,479** | **,551** | **,422** | **,178** |
| lastna vrednost *i* | 2,45 |  | 2,27 |  |
| %poj. var. | 40,85 |  | 56,7 |  |

### Regresijska analiza

Pri regresijski analizi smo uporabili ordinalne in dihotomne spremenljivke, ki smo jih našteli v razdelki 4.1. V programskem paketu SPSS smo uporabili linearne regresijo z metodo enter. Metoda enter pri analizi vstavi vse spremenljivke naenkrat za razliko od ostalih »step-wise« metod.

Metoda pri analizi izvede tudi statistični test ANOVA s pomočjo F-testa. Test testira hipotezo, da so vse uteži $β\_{i}$ enake 0. Iz spodnje tabele lahko vidimo, da je vrednost F višja od kritične vrednosti in lahko test zavrnemo.

| **ANOVAb** |
| --- |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 49,573 | 6 | 8,262 | 9,646 | ,000a |
| Residual | 1582,030 | 1847 | ,857 |  |  |
| Total | 1631,603 | 1853 |  |  |  |
| a. Predictors: (Constant), Household's total net income, all sources, Ever had a paid job, Gender, Citizen of country, Highest level of education, Age of respondent, calculatedb. Dependent Variable: Zaupanje |

V spodnjih dveh tabelah razlagamo zgrajen regresijski model. V prvi tabeli opazimo popravljeni determinacijski koeficient z vrednostjo 0,027, ki nam pove, da z neodvisnimi spremenljivkami pojasnimo le 2,7% variance odvisne spremenljivke zaupanje. V naslednji tabeli pa lahko razberemo izračunane uteži $β\_{i}$ in statistično značilnost t vpliva posameznih neodvisnih spremenljivk.

| **Model Summary** |
| --- |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | ,174a | ,030 | ,027 | ,92549450 |
| a. Predictors: (Constant), Household's total net income, all sources, Ever had a paid job, Gender, Citizen of country, Highest level of education, Age of respondent, calculated |

| **Coefficientsa** |
| --- |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | -1,111 | ,240 |  | -4,629 | ,000 |
| Highest level of education | ,016 | ,023 | ,017 | ,702 | ,482 |
| Ever had a paid job | ,104 | ,067 | ,040 | 1,553 | ,121 |
| Age of respondent, calculated | ,001 | ,001 | ,015 | ,575 | ,565 |
| Gender | -,116 | ,044 | -,062 | -2,673 | ,008 |
| Citizen of country | ,680 | ,158 | ,100 | 4,318 | ,000 |
| Household's total net income, all sources | ,061 | ,011 | ,126 | 5,390 | ,000 |
| a. Dependent Variable: Zaupanje |

Iz zgornje tabele lahko razberemo, da na odvisno spremenljivko zaupanje najbolj vpliva skupen dohodek gospodinjstva in to, da je oseba državljan določene države. Mejni vpliv imata še spol in dejstvo, da je oseba že imela plačano delo, zanemarljiv vpliv pa nosi najvišja dosežena stopnja izobrazbe in starost. Regresijski model v naslednjem razdelku pri interpretaciji prikažemo tudi na sliki.

# Interpretacija in evaluacija rezultatov

V nalogi smo analizirali izbrane statistične podatke iz Evropskih Družboslovnih raziskav (*ESS*). V tej domači nalogi smo s pomočjo multiple regresijske analize zgradili model za odvisno spremenljivko Zaupanje, ki smo jo dobili kot prvi faktor faktorske analize, ki smo jo izvedli v prejšnji domači nalogi.

Ugotovili smo, da na stopnjo zaupanja v poltične institucije in stanje zdravstva v državi najbolj vpliva skupen dohodek gospodinjstva in to, da je oseba državljan določene države. Mejni vpliv imata še spol in dejstvo, da je oseba že imela plačano delo, zanemarljiv vpliv pa nosi najvišja dosežena stopnja izobrazbe in starost.

Rezultat analize prikazujemo tudi na spodnji sliki.

0,100\*

-0,044\*

0,126

0,015\*\*

0,017\*\*

0,040\*\*

EDULVLA

PDJOBEV

AGE

GNDR

CTZCNTR

HINCTNT

E

Legenda:

\*\*\*: < 0,01

\*\*: 0,01 – 0,05

\*: 0,05 – 0,10

F = 9,646

Popravljen R2 = 0,027

# Priloge

## Frekvenčne porazdelitve

### Neodvisne spremenljivke





