**TOMŠIČ GAŠPER**

**21150317**

**REDNI ŠTUDIJ**

**DI-UNI-1**

**2P**

Univerza

*v Ljubljani*

Fakulteta

*za družbene vede*



1. **STATISTIČNA NALOGA**

**MULTIVARIATNA ANALIZA**

**NEMČIJA**

**Študijsko leto 2016/2017**

**Nosilec: izr. prof. dr. Aleš Žiberna**

**Tomšič Gašper, 21150317, gasper.sola@gmail.com**

**Družboslovna Informatika**

**Redni študij, Družboslovna Informatika UNI, 2P**

**OCENA 4. NALOGE: \_\_\_\_\_\_\_\_**

**Osnovne informacije o podatkovni datoteki in izbranih spremenljivkah**

**Podatkovna datoteka:**

World values survey – Nemčija

**Izbrane spremenljivke:**

Neodvisne: Justifiable v198-v210 (brez v201 in v207A)

Skupine po : Izobrazba v248

**Odločitev o verjetnostih na populaciji in utemeljitev**

Analizo bom izvajal po predpostavki, da skupine v populaciji niso enako velike, ker bom izvajal linearno diskriminantno analizo po skupinah glede na izobrazbo in imajo nekatere skupine v populaciji izrazito majhno število enot (primer: brez formalne izobrazbe, razne skupine z nedokončano stopnjo izobrazbe)

**Linearna diskriminantna analiza**

**Predpostavke**

Predpostavka o enakosti kovariančnih matrik:

Iz tabele 2 vidimo, da je natančna stopnja značilnosti p manjša od 0.0005, ničelno domnevo pri 0,05% stopnji značilnosti zavrnemo. Predpostavka o enakosti kovariančnih matrik je kršena. V tem primeru bi lahko razmislili o uporabi kvadratne diskriminantne analize, vendar je mi ne bomo.

**Tabela 1: Log determinante**

|  |  |  |
| --- | --- | --- |
| **Log Determinants** | | |
| Highest educational level attained | Rank | Log Determinant |
| No formal education | 12 | .a |
| Incomplete primary school | 13 | 12,274 |
| Complete primary school | 13 | 10,765 |
| Incomplete secondary school: technical/ vocational type | 13 | 7,782 |
| Complete secondary school: technical/ vocational type | 13 | 12,586 |
| Incomplete secondary school: university-preparatory type | 13 | 8,772 |
| Complete secondary school: university-preparatory type | 13 | 11,342 |
| Some university-level education, without degree | 13 | 4,478 |
| University - level education, with degree | 13 | 11,905 |
| Pooled within-groups | 13 | 12,425 |
| The ranks and natural logarithms of determinants printed are those of the group covariance matrices. | | |
| a. Singular | | |

**Tabela 2: Box-ov M test**

|  |  |  |
| --- | --- | --- |
| **Test Resultsa** | | |
| Box's M | | 1602,202 |
| F | Approx. | 2,312 |
| df1 | 637 |
| df2 | 60054,779 |
| Sig. | ,000 |
| Tests null hypothesis of equal population covariance matrices. | | |
| a. Some covariance matrices are singular and the usual procedure will not work. The non-singular groups will be tested against their own pooled within-groups covariance matrix. The log of its determinant is 12,391. | | |

**Diskriminatne funckije**

Iz tabele 4 lahko vidimo, da se le prva diskirminatna funkcija statistično značilno razlikuje med skupinami. Kanonična korelacija za prvo diskriminatno funkcijo je 0,357. V nadaljevanju bomo interpretirali le prvo funkcijo.

**Tabela 3: Lastne vrednosti, pojasnene variance in koeficient kanonične korelacije**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Eigenvalues** | | | | |
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | ,146a | 72,8 | 72,8 | ,357 |
| 2 | ,021a | 10,3 | 83,1 | ,142 |
| 3 | ,012a | 6,1 | 89,2 | ,110 |
| 4 | ,007a | 3,4 | 92,6 | ,082 |
| 5 | ,006a | 3,2 | 95,7 | ,079 |
| 6 | ,005a | 2,3 | 98,0 | ,068 |
| 7 | ,003a | 1,6 | 99,6 | ,057 |
| 8 | ,001a | ,4 | 100,0 | ,027 |
| a. First 8 canonical discriminant functions were used in the analysis. | | | | |

**Tabela 4: Wilksova lambda**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wilks' Lambda** | | | | |
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 through 8 | ,827 | 346,859 | 104 | ,000 |
| 2 through 8 | ,947 | 98,741 | 84 | ,130 |
| 3 through 8 | ,967 | 61,470 | 66 | ,635 |
| 4 through 8 | ,979 | 39,388 | 50 | ,860 |
| 5 through 8 | ,985 | 27,066 | 36 | ,859 |
| 6 through 8 | ,991 | 15,589 | 24 | ,902 |
| 7 through 8 | ,996 | 7,191 | 14 | ,927 |
| 8 | ,999 | 1,306 | 6 | ,971 |

Iz tabele 5 lahko razvidimo, da funkcija 1 meri, koliko skupine opravičujejo homoseksualnost, splav, samomor, izogibanje plačevanja javnega prevoza, spolnosti pred poroko in krajo premoženja. Meri pa tudi, koliko skupine ne opravičujejo prostitucije in sprejemanje podkupnin.

Iz tabele koeficientov ni razvidno, da bi bile kakšne tematske povezave med spremenljivkami z visokimi koeficienti.

**Tabela 5: Standardizirani koeficienti**

|  |  |
| --- | --- |
| Standardized Canonical Discriminant Function Coefficients | |
|  | Function |
|  | 1 |
| Justifiable: Claiming government benefits to which you are not entitled | -0,06 |
| Justifiable: Avoiding a fare on public transport | 0,324 |
| Justifiable: Stealing property | 0,197 |
| Justifiable: Someone accepting a bribe in the course of their duties | -0,236 |
| Justifiable: Homosexuality | 0,629 |
| Justifiable: Prostitution | -0,395 |
| Justifiable: Abortion | 0,337 |
| Justifiable: Divorce | -0,124 |
| Justifiable: Sex before marriage | 0,201 |
| Justifiable: Suicide | 0,327 |
| Justifiable: For a man to beat his wife | -0,126 |
| Justifiable: Parents beating children | -0,054 |
| Justifiable: Violence against other people | 0,139 |

Pri diskriminatni funkciji najbolj pozitivno korelirajo spremenljivke sklopa liberalnih dejanj (z izjemo prostitucije, ki korelira nekoliko manj). Poleg teh spremenljivk korelira tudi izogibanje plačevanja javnega prevoza, kraja premoženja in nasilje nad ostalimi ljudmi. Negativno korelirajo le opravičevanje, da mož tepe ženo in prejemanje podkupnine, vendar sta obe precej šibke korelacije.

**Tabela 6: Strukturna matrika**

|  |  |
| --- | --- |
| **Structure Matrix** | |
|  | Function |
| 1 |
| Justifiable: Homosexuality | ,734 |
| Justifiable: Abortion | ,632 |
| Justifiable: Suicide | ,583 |
| Justifiable: Divorce | ,509 |
| Justifiable: Prostitution | ,257 |
| Justifiable: Sex before marriage | ,483 |
| Justifiable: Stealing property | ,241 |
| Justifiable: Violence against other people | ,115 |
| Justifiable: Avoiding a fare on public transport | ,402 |
| Justifiable: For a man to beat his wife | -,046 |
| Justifiable: Someone accepting a bribe in the course of their duties | -,039 |
| Justifiable: Parents beating children | ,027 |
| Justifiable: Claiming government benefits to which you are not entitled | ,081 |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  Variables ordered by absolute size of correlation within function. | |

Diskriminatna funkcija loči med tistimi z nizko izobrazbo od tistih z visoko izobrazbo.

**Tabela 7: Povprečja spremenljivk po skupinah**

|  |  |
| --- | --- |
| **Functions at Group Centroids** | |
| Highest educational level attained | Function |
| 1 |
| No formal education | -,660 |
| Incomplete primary school | -,537 |
| Complete primary school | -,450 |
| Incomplete secondary school: technical/ vocational type | -,140 |
| Complete secondary school: technical/ vocational type | -,006 |
| Incomplete secondary school: university-preparatory type | ,132 |
| Complete secondary school: university-preparatory type | ,436 |
| Some university-level education, without degree | ,632 |
| University - level education, with degree | ,513 |
| Unstandardized canonical discriminant functions evaluated at group means | |

**Tabela 8: Začetne vrednosti skupin izračunane iz velikosti skupin**

|  |  |  |  |
| --- | --- | --- | --- |
| **Prior Probabilities for Groups** | | | |
| Highest educational level attained | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| No formal education | ,010 | 18 | 18,000 |
| Incomplete primary school | ,066 | 122 | 122,000 |
| Complete primary school | ,229 | 421 | 421,000 |
| Incomplete secondary school: technical/ vocational type | ,041 | 75 | 75,000 |
| Complete secondary school: technical/ vocational type | ,323 | 593 | 593,000 |
| Incomplete secondary school: university-preparatory type | ,023 | 43 | 43,000 |
| Complete secondary school: university-preparatory type | ,112 | 205 | 205,000 |
| Some university-level education, without degree | ,011 | 20 | 20,000 |
| University - level education, with degree | ,184 | 338 | 338,000 |
| Total | 1,000 | 1835 | 1835,000 |

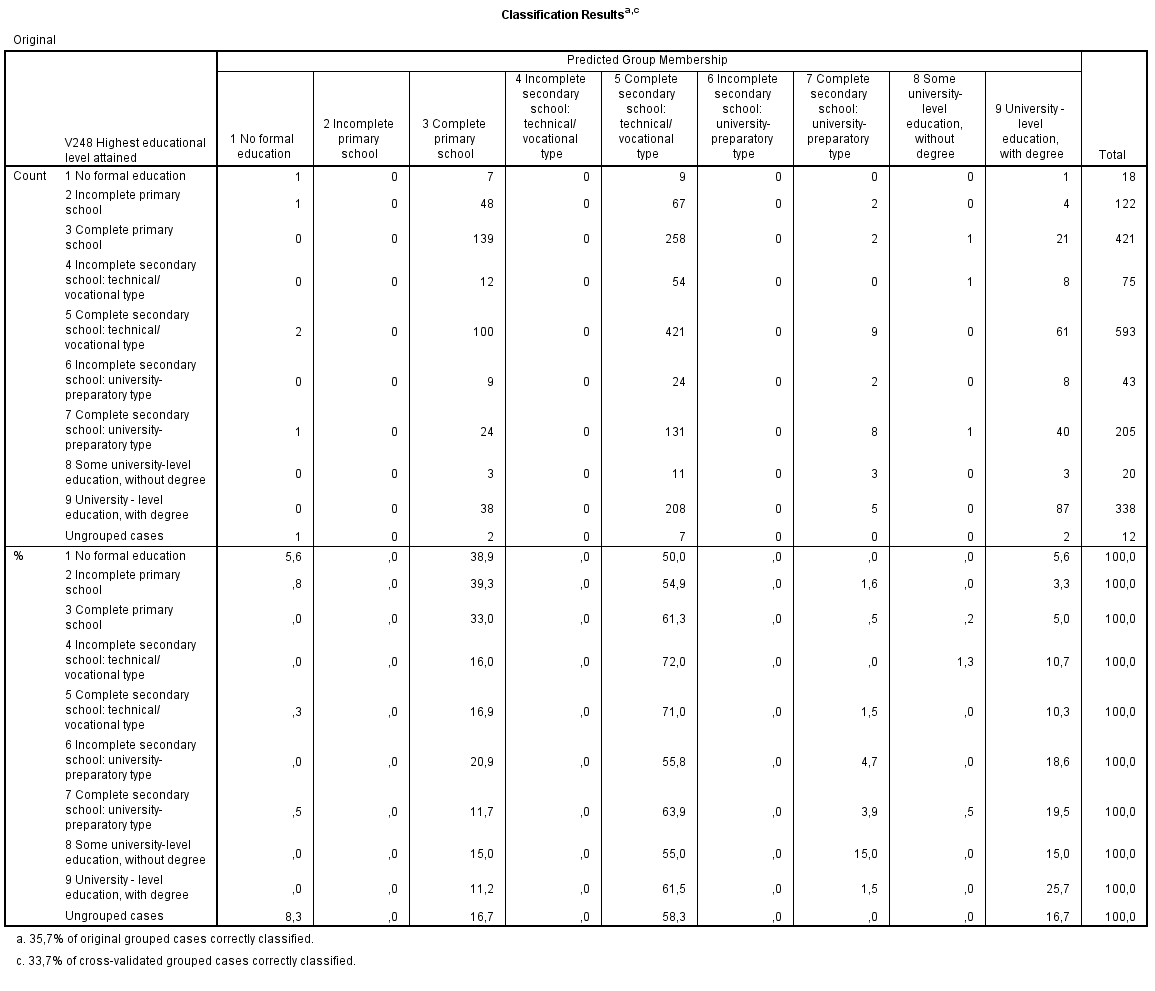
**Ocena kvalitete modela**

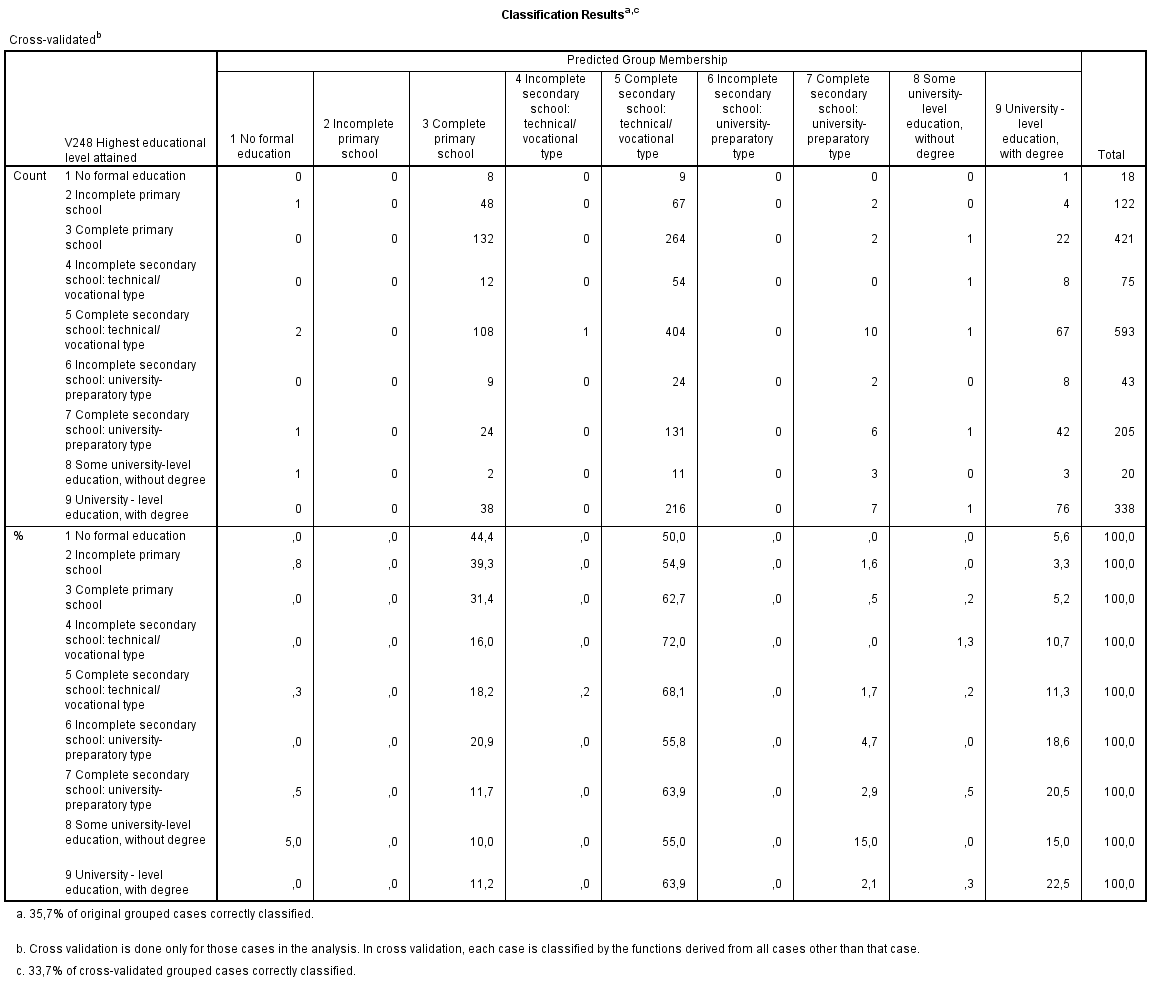
Po originalni razvrstitvi smo vse skupine (razen 5. – končana srednja tehnična šola ) klasificirali narobe. Pri 5. skupini smo razvrstili 71.0% eno pravilno. Med ostale skupine, ki smo jih še uvrstili najbolje je 3. skupina (končana osnovna šola) z 33.0% in 9. skupina (univerzitetna izobrazba) z 25.7% . Skupno smo pravilno razvrstili 35.7% enot.

Po navzkrižni klasifikaciji je klasificiranje precej podobno. Ponovno smo edino pravilno klasificirali 5. skupino (končana srednja tehnična šola) sicer malo slabše z 68,1%. Prav tako sta nekoliko slabši 3. skupina (končana osnovna šola) z 31.4% in 9. skupina (univerzitetna izobrazba) z 22.5%.

Skupno je bilo navzkrižno pravilno razvrščenih 33.7% enot.

Nasploh je model precej slab, kar je bilo že razvidno na začetku, ko smo prekršili predpostavko. Prav tako imamo veliko število skupin, za kar je normalno, da bodo enote slabše razvrščene.

**Tabela 9: Klasifacijska tabela – originalna razvrstitev**

**Tabela 10: Klasifacijska tabela – navzkrižna klasifikacija**

**Grafičen prikaz**

Predpostavili smo, da niso vse skupine enako velike, v tabeli 11 so začetne vrednosti skupin.

**Tabela 11: Začetne vrednosti skupin izračunane iz velikosti skupin**

|  |  |  |  |
| --- | --- | --- | --- |
| **Prior Probabilities for Groups** | | | |
| Highest educational level attained | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| No formal education | ,010 | 18 | 18,000 |
| Incomplete primary school | ,066 | 122 | 122,000 |
| Complete primary school | ,229 | 421 | 421,000 |
| Incomplete secondary school: technical/ vocational type | ,041 | 75 | 75,000 |
| Complete secondary school: technical/ vocational type | ,323 | 593 | 593,000 |
| Incomplete secondary school: university-preparatory type | ,023 | 43 | 43,000 |
| Complete secondary school: university-preparatory type | ,112 | 205 | 205,000 |
| Some university-level education, without degree | ,011 | 20 | 20,000 |
| University - level education, with degree | ,184 | 338 | 338,000 |
| Total | 1,000 | 1835 | 1835,000 |

**Slika 1: ''Zemljevid prostora diskriminatnih funckcij**

Territorial Map

(Assuming all functions but the first two are zero)

Canonical Discriminant

Function 2

-4,0 -3,0 -2,0 -1,0 ,0 1,0 2,0 3,0 4,0

+---------+---------+---------+---------+---------+---------+---------+---------+

4,0 + 35 59 +

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

3,0 + + + +35 + 59 + + + +

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

2,0 + + + +35 + 59 + + +

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

I 35 59 I

1,0 + + + 35 + +59 + + +

I 35 59 I

I 35 59 I

I 35 59 I

I 35 \* \* 59 I

I 35 \*\* 59 I

,0 + + + 35 \*\* \* + 59 + + +

I 35 59 I

I 35 \* 59 I

I 35 \* 59 I

I 35 59 I

I 35 59 I

-1,0 + + + 35 + + 59 + + +

I 35 59 I

I 35 59 I

I 35 59 I

I 35 599999 I

I 35 57777799999999 I

-2,0 + + + 35 + + 57 + 7777777799999999 +

I 35 57 77777777999I

I 35 57 777I

I 35 57 I

I 35 57 I

I 35 57 I

-3,0 + + + 35 + 57 + + +

I 35 57 I

I 35 57 I

I 35 57 I

I 35 57 I

I 35 57 I

-4,0 + 35 57 +

+---------+---------+---------+---------+---------+---------+---------+---------+

-4,0 -3,0 -2,0 -1,0 ,0 1,0 2,0 3,0 4,0

Canonical Discriminant Function 1

Symbols used in territorial map

Symbol Group Label

------ ----- --------------------

1 1 No formal education

2 2 Incomplete primary s

3 3 Complete primary sch

4 4 Incomplete secondary

5 5 Complete secondary s

6 6 Incomplete secondary

7 7 Complete secondary s

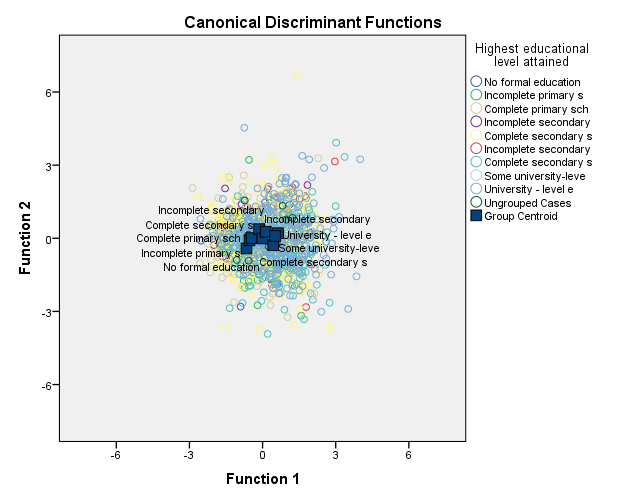
8 8 Some university-leve

9 9 University - level e

\* Indicates a group centroid

Na sliki 2 je grafičen prikaz razporeditve skupin z centroidi. Na grafu je tudi diskriminatna funckija 2, ki je sicer nismo interpretirali, vendar če gledamo samo po prvi diskriminatni funkciji, vidimo, da so centroidi zelo skupaj, na pozitivni strani so pretežno skupine z visoko izobrazbo, ne negativni strani pa skupine z nižjo izobrazbo, kar smo videli že pri tabeli 7.

**Slika 2: Enote v prostoru DF**



**Zaključek**

V tej nalogi smo naredili linearno diskriminatno analizo na neodvisnih spremenljivkah opravičljivosti in odvisno spremenljivko izobrazba. Sprva smo preverili predpostavke, kjer se je izkazalo, da linearna diskriminantna analiza ne bo dobra, ker je bila prekršena predpostavka o enakosti kovariančnih matrik. Analizo smo vseeno opravili. Pri izbiri začetne verjetnosti smo predpostavili, da skupine v populaciji niso enako velike, zato smo naredili diskriminatno analizo z izračunanimi začetnimi vrednostmi. Med vsemi diskriminatnimi funkcijami je bila le ena statistično značilna in sicer je merila, koliko skupine opravičujejo homoseksualnost, splav, samomor, izogibanje plačevanja javnega prevoza, spolnosti pred poroko in krajo premoženja, ter ne opravičujejo prostitucije in sprejemanje podkupnin. Funckija je tudi ločila po izobrazbi in sicer bolj pozitivne vrednosti so imeli tisti z višjo izobrazbo in bolj negativne tisti z nižjo izobrazbo. Na splošno je bil model slab, pravilno je bilo razvrščenih le 35,7% enot. Za boljši model bi morali izbrati kvadratno diskriminatno analizo.

**Programski stavki**

DISCRIMINANT

/GROUPS=V248(1 9)

/VARIABLES=V198 V199 V200 V202 V203 V203A V204 V205 V206 V207 V208 V209 V210

/ANALYSIS ALL

/PRIORS SIZE

/STATISTICS=UNIVF BOXM TABLE CROSSVALID

/PLOT=COMBINED MAP

/CLASSIFY=NONMISSING POOLED.