Kanonične korelacije in diskriminantna analiza

Domača naloga 5 - Multivariatna analiza

Urh Peček in Alen Kahteran

24. 5. 2021

Kazalo

Slike

Tabele

Cilji naloge

Kot je bilo pojasnjeno že v prejšnjih nalogah, bomo analizirali podatke, ki so bili zbrani v sklopu Evropske družboslovne raziskave (ESS) v 9. valu leta 2018. Naši podatki se nanašajo na Dansko prebivalstvo, kjer je sodelovalo 1572 prebivalcev. Analizirali bomo različne vrednote, merjenje s tako imenovanimi Schwartzovimi spremenljivkami, ki smo jih razdelili v sklop odprtosti in sklop konservativnosti. V sklopu odprtosti imamo spremenljivke drugačnost, užitek, vznemerljivost, zabava, kreativnost in svoboda. V sklopu spremenljivk konservativnosti pa imamo spremenljivke varnost, ponižnost, obramba, tradicija, ubogljivost in sprejemljivost. Vse spremenljivke so bile podrobneje opisane že v prejšnjih nalogah.

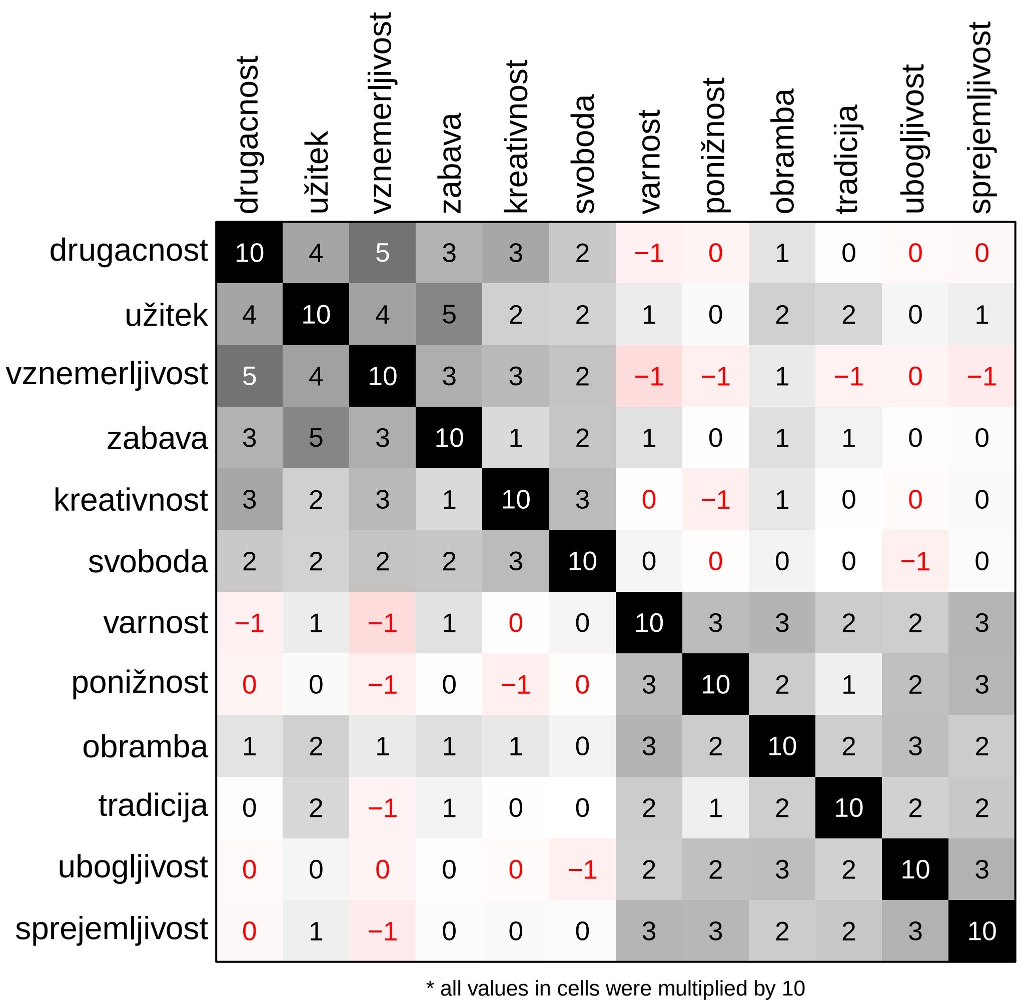
V prvem delu naloge (kanonične korelacije) želimo izračunati korelacijo med dvema sklopoma spremenljivk, to sta sklop spremenljivk odprtosti in sklop spremenljivk konservativnosti. Cilj je narediti dve (ali več) linearni kombinaciji spremenljivk odprtosti in spremenljivk konservativnosti, da bo korelacija med sklopoma spremenljivk čim večja. Ob tem mora veljati, da je linearna kombinacija spremenljivk taka, da bo korelacija med linearnima kombinacijama čim večja. Ker s prvo linearno kombinacijo ne uspemo pojasniti celotne variabilnosti v spremenljivkah, za vsak sklop spremenljivk ustvarimo več tovrstnih linearnih kombinacij, ki pa morajo biti med seboj neodvisne.

V drugem delu naloge (diskriminantna analiza) pa je cilj poiskati takšno kombinacijo merjenih spremenljivk, ki bo maksimalno ločila vnaprej določene skupine in bo napaka pri uvrščanju enot v skupine čim manjša. Želimo torej poiskati takšne razsežnosti, ki najbolje prirejajo enote vnaprej danim skupinam. S spremenljivkami odprtosti in konservativnosti bomo poskušali napovedati kakšno je delovno razmerje (nedoločen čas, določen čas, brez pogodbe) posameznika. V ta namen bomo uporabili spremenljivko Razmerje, opisano v prvi nalogi.

Kanonične korelacije

Korelacijska matrika vsebinskih spremenljivk

Najprej si oglejmo korelacijsko matriko vseh izbranih Schwartzovih spremenljivk odprtosti in konservativnosti. Pojasniti želimo predvsem povezanost med spremenljivkami različnega sklopa, torej med diagonalnima navideznima kvadratoma.



Slika 1: Korelacijska matrika vseh spremenljivk izbranih sklopov

Kanonične rešitve

Izračunajmo kanonične rešitve, torej linearne kombinacije spremenljivk znotraj sklopa in korelacije med sklopoma spremenljivk. Velja, da je kanoničnih rešitev enako število spremenljivk v sklopu z najmanj spremenljivkami. V našem primeru oba sklopa vsebujeta šest spremenljivk, kar pomeni, da imamo šest kanoničnih rešitev. Kanonične rešitve so izračunane na standariziranih podatkih.

V spodnjem tabelaričnem prikazu so Eigenvalues lastne vrednosti, % je delež pojasnjene variabilnosti vseh spremenljivk s kanoničnimi spremenljivkami, Cum % so kumulativne vrednosti deleža pojasnjene variabilnosti, Cor so kanonične korelacije in Sq. Cor so vrednosti mer prekrivanja sklopov.

Tabela 1: Tabelarični prikaz kanoničnih rešitev

Eigenvalues	%	Cum %	Cor	Sq. Cor
0.072	51.045	51.045	0.260	0.067
0.046	32.154	83.199	0.209	0.044
0.015	10.588	93.787	0.122	0.015
0.005	3.766	97.553	0.073	0.005
0.003	2.446	99.999	0.059	0.003
0.000	0.001	100.000	0.001	0.000

Vidimo, da z vsemi kanoničnimi spremenljivkami pojasnimo celotno variabilnost spremenljivk. Prva kanonična korelacija je enaka 0.26, torej gre za pozitivno in šibko povezanost. Z vsako naslednjo kanonično rešitvijo korelacije padajo, posamezni kombinaciji spremenljivk obeh sklopov pa sta pozitivni in zanemarljivo povezani.

Statistična značilnost kanoničnih korelacij

Poglejmo si še, katere kanonične korelacije so statistično značilno različne od nič. Ničelna domneva je, da je opazovana kanonična korelacija in vse naslednje enaka nič. Domneve preverimo z Wilksovim lambda testom.

Tabela 2: Tabelarični prikaz statistične značilnosti kanoničnih korelacij

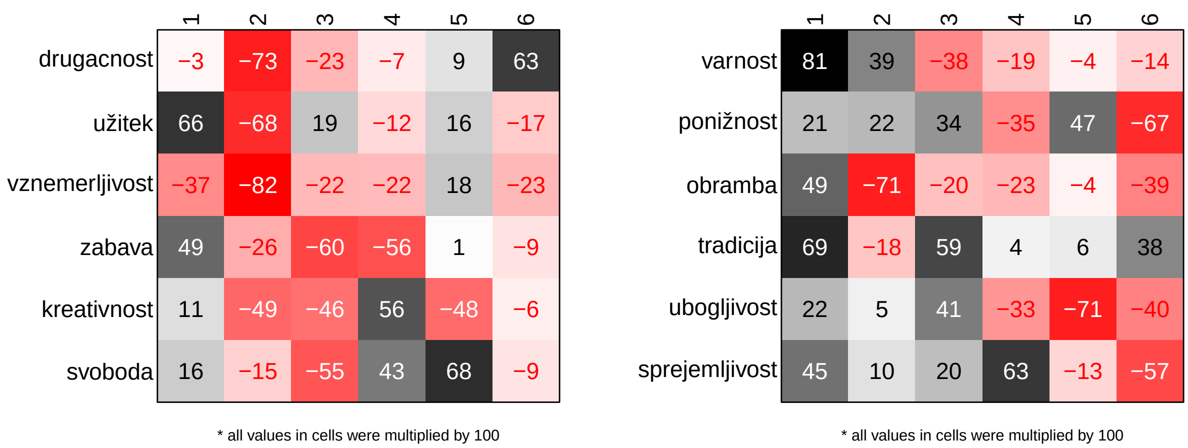
	WilksL	F	df1	df2	p
1 to 6	0.871	5.786	36	6519.461	0.000
2 to 6	0.934	4.093	25	5518.032	0.000
3 to 6	0.977	2.209	16	4540.443	0.004
4 to 6	0.991	1.455	9	3619.118	0.159
5 to 6	0.997	1.290	4	2976.000	0.272
6 to 6	1.000	0.002	1	1489.000	0.963

Na podlagi vrednosti p pri stopnji statistične značilnosti 0.05, zavrnemo prve tri ničelne domneve in sprejmemo sklep, da so prva, druga in tretja kanonična korelacija statistično značilno različne od 0. Za četrto (posledično tudi peto in šesto) kanonično korelacijo pa ne moremo zavrniti domneve, da je ta in vse naslednje enaka nič.

Statistično značilne (α=0.05) so torej prva, druga in tretja kanonična korelacija. Prva je kot rečeno enaka 0.26, druga je enaka 0.21 in tretja je enaka 0.12.

Interpretacija kanoničnih korelacij

Za pomoč pri interpretaciji kanoničnih korelacij si oglejmo pripadajoče uteži, torej regresijske koeficiente kanoničnih spremenljivk odprtosti in konservativnosti. Trojček, ki ga sestavljajo kanonična spremenljivka prve množice, kanonična spremenljivka druge množice in kanonični korelacijski koeficient imenujemo kanonična rešitev. Teh je kot rečeno največ toliko, kolikor je spremenljivk v manjši od množic oz. sklopu, v našem primeru šest.



Slika 2: Korelacije med vrednostmi kanoničnih spremenljivk in spremenljivk v množici odprtosti (levo) in konservativnosti (desno)

Prva kanonična korelacija je najmočnejša (največja) in kot rečeno enaka 0.26. Pri spremenljivkah konservativnosti vidimo, da imata varnost in tradicija močno pozitivno korelacijo, obramba in sprejemljivost srednje močno pozitivno korelacijo, korelacija ponižnosti in ubogljivosti pa je pozitivna in šibka. Pri spremenljivkah odprtosti imata močno pozitivno korelacijo spremenljivki užitka in zabava, korelacija vznemerljivosti je srednje močna in negativna. Šibka pozitivna korelacija je še pri kreativnosti in svobodi, korelacija drugačnosti pa je negativna in zanemarljiva.

Prva kanonična rešitev torej meri korelacijo med varnostjo in tradicijo ter užitkarstvom in zabavo. Korelacija je močna in pozitivna, Torej bolj kot sta nekomu pomembna varnost in tradicija bolj mu je pomembno užitkarstvo in zabava.

Druga kanonična korelacija je enaka 0.21. Pri spremenljivkah konservativnosti ima obramba močno negativno korelacijo, korelacija varnosti je pozitivna in srednje močna, korelacija ponižnosti je pozitivna in šibka, tradicija ima negativno šibko korelacijo, korelaciji ubogjivosti in sprejemljivosti pa sta pozitivni in zanemarljivi. Pri spremenljivkah odprtosti imajo vse spremenljivke negativno korelacijo. Korelacija spremenljivk drugačnosti, užitka, vznemerljivosti in kreativnosti je močna, korelacija spremenljivk zabave in svobode pa je šibka.

Druga kanonična rešitev torej meri korelacijo med obrambo in vznemerljivosti in drugačnostjo. Korelacija je močna in negativna, torej bolj kot je nekomu pomembna obramba, manj mu je pomembna vznemerljivost oziroma drugačnost.

Podobno interpretiramo še tretjo kanonično rešitev, katere korelacija je enaka 0.12. Velja, da tretja kanonična rešitev meri korelacijo med tradicijo in ubogljivostjo ter zabavo in svobodo. Korelacija je močna in negativna, torej bolj kot smo naklonjeni tradiciji ter smo ubogljivi, manj smo svobodnega duha in prisegamo zabavi.

Delež pojasnjene variabilnosti spremenljivk

Iz spodnjih tabelaričnih prikazov vidimo, kolikšen je povprečen delež pojasnjene variabilnosti vseh spremenljivk istega sklopa s posamezno kanonično spremenljivko odprtosti oziroma konservativnosti, istega sklopa. Velja, da z vsemi kanoničnimi spremenljivkami istega sklopa pojasnimo vso variabilnost vsebinskih spremenljivk istega sklopa.

Tabela 3: Delež pojasnjene variabilnosti s kanoničnimi spr. odprtosti

	R2	cum R2
Kanonična spremenljivka 1:	0.141	0.141
Kanonična spremenljivka 2:	0.335	0.476
Kanonična spremenljivka 3:	0.169	0.645
Kanonična spremenljivka 4:	0.147	0.792
Kanonična spremenljivka 5:	0.125	0.917
Kanonična spremenljivka 6:	0.083	1.000

Največ variabilnosti spremenljivk odprtosti v povprečju pojasnimo z drugo kanonično spremenljivko in sicer 34%, sledita tretja in četrta z nekaj več oziroma manj kot 15%, prva, peta in šesta kanonična spremenljivka pa pojasnijo nekaj več oziroma manj kot 10% variabilnosti spremenljivk odprtosti.

Tabela 4: Delež pojasnjene variabilnosti s kanoničnimi spr. konservativnosti

	R2	cum R2
Kanonična spremenljivka 1:	0.277	0.277
Kanonična spremenljivka 2:	0.126	0.403
Kanonična spremenljivka 3:	0.142	0.545
Kanonična spremenljivka 4:	0.121	0.666
Kanonična spremenljivka 5:	0.125	0.791
Kanonična spremenljivka 6:	0.209	1.000

Pri spremenljivkah konservativnosti v povprečju največ variabilnosti pojasnimo s prvo kanonično spremenljivko in sicer 28%, sledi šesta kanonična spremenljivka z 21%, ostale kanonične spremenljivke pa posamično pojasnijo nekaj več kot 12% variabilnosti spremenljivk konservativnosti.

Pogledamo pa lahko tudi kolikšen je povprečen delež pojasnjene variabilnosti sklopa vsebinskih spremenljivk s kanoničnimi spremenljivkami iz drugega sklopa. Velja, da vsaka nadaljna kanonična spremenljivka pojasni manj variabilnosti vsebinskih spremenljivk drugega sklopa.

Tabela 5: Delež pojasnjene variabilnosti spr. odprtosti s kanoničnimi spr. konservativnosti

	R2	cum R2
Kanonicna spremenljivka 1:	0.019	0.019
Kanonicna spremenljivka 2:	0.005	0.024
Kanonicna spremenljivka 3:	0.002	0.026
Kanonicna spremenljivka 4:	0.001	0.027
Kanonicna spremenljivka 5:	0.000	0.027
Kanonicna spremenljivka 6:	0.000	0.027

S prvo kanonično spremenljivko konservativnosti pojasnimo 2% variabilnosti spremenljivk odprtosti, z drugo kanonično spremenljivko 0.5% in s tretjo 0.2%. Delež pojasnjene variabilnosti spremenljivk odprtosti s četrto, peto in šesto kanonično spremenljivko konservativnosti je zanemarljiv.

Tabela 6: Delež pojasnjene variabilnosti spr. konservativnosti s kanoničnimi spr. odprtosti

	R2	cum R2
Kanonicna spremenljivka 1:	0.010	0.010
Kanonicna spremenljivka 2:	0.015	0.025
Kanonicna spremenljivka 3:	0.002	0.027
Kanonicna spremenljivka 4:	0.001	0.028
Kanonicna spremenljivka 5:	0.000	0.028
Kanonicna spremenljivka 6:	0.000	0.028

S prvo kanonično spremenljivko odprtosti pojasnimo 1% variabilnosti spremenljivk odprtosti, z drugo kanonično spremenljivko 1.5% in s tretjo 0.2%. Delež pojasnjene variabilnosti spremenljivk konservativnosti s četrto, peto in šesto kanonično spremenljivko odprtosti je zanemarljiv.

Diskriminantna analiza

Predpostavke diskriminantne analize

1. Število skupin je vsaj dve

Število skupin v našem primeru določa število kategorij spremenljivke Razmerje. Ta ima 3 kategorije, to so Brez pogodbe, Določen čas in Nedoločen čas. Število skupin je tako 3 in prva predpostavka je izpolnjena.

2. V vsaki skupini sta vsaj dve enoti

Druga predpostavka linearne diskriminantne analize je, da sta v vsaki od skupin za katere želimo napovedati pripadnost enot (nedoločen čas, določen čas, brez pogodbe), vsaj dve enoti.

Če pogledamo frekvenčno porazdelitev spremenljivke Razmerje vidimo, da je Brez pogodbe 98 respondentov, 145 respondentov ima pogodbo za določen čas in največ, 1087 respondentov ima pogodbo za nedoločen čas. Prva predpostavka je torej izpolnjena.

Tabela 7: Frekvenčna porazdelitev spremenljivke Razmerje

Brez pogodbe	Določen čas	Nedoločen čas
98	145	1087

3. Število spremenljivk je manjše od števila enot -1

Tretja predpostavka linearne diskriminantne analize je, da je število spremenljivk manjše od števila enot -1.

Število vsebinskih oziroma Schwartzovih spremenljivk je 12 (6 za odprtost in 6 za konservativnost), število enot pa je 1330. Tudi druga predpostavka je tako izpolnjena.

Za utemeljitev kasneje predpostavljenih verjetnostih v populaciji si oglejmo še deleže po kategorijah spremenljivke Razmerje.

Tabela 8: Deleži po kategorijah spremenljivke Razmerje

Brez pogodbe	Določen čas	Nedoločen čas
0.074	0.109	0.817

Kot smo videli že prej, skupine niso enako velike in posledično pri razvrščanju v skupine na populaciji ne bomo predpostavili enako velikih skupin oziroma verjetnosti.

4. Odsotnost multikolinearnosti

Četrta predpostavka linearne diskriminantne analize je odsotnost multikolinearnosti, kar pomeni, da nobena od vsebinskih spremenljivk ni popolna linearna kombinacija ostalih. Težava z multikolinearnostjo je tako v primeru, ko je katera od korelacij med spremenljivkama blizu 1. V ta namen si oglejmo korelacijsko matriko.

	drugacnost	užitek	vznemerljivost	zabava	kreativnost	svoboda	varnost	ponižnost	obramba	tradicija	ubogljivost	sprejemljivost
drugacnost	10	4	6	3	3	2	0	0	1	0	0	0
užitek	4	10	4	5	2	2	1	0	2	1	0	1
vznemerljivost	6	4	10	3	3	2	−1	0	1	0	0	−1
zabava	3	5	3	10	1	2	1	0	1	0	0	0
kreativnost	3	2	3	1	10	2	0	−1	1	0	0	0
svoboda	2	2	2	2	2	10	1	0	1	0	−1	0
varnost	0	1	−1	1	0	1	10	3	3	2	2	3
ponižnost	0	0	0	0	−1	0	3	10	2	0	2	3
obramba	1	2	1	1	1	1	3	2	10	2	2	2
tradicija	0	1	0	0	0	0	2	0	2	10	2	2
ubogljivost	0	0	0	0	0	−1	2	2	2	2	10	3
sprejemljivost	0	1	−1	0	0	0	3	3	2	2	3	10

* all values in cells were multiplied by 10

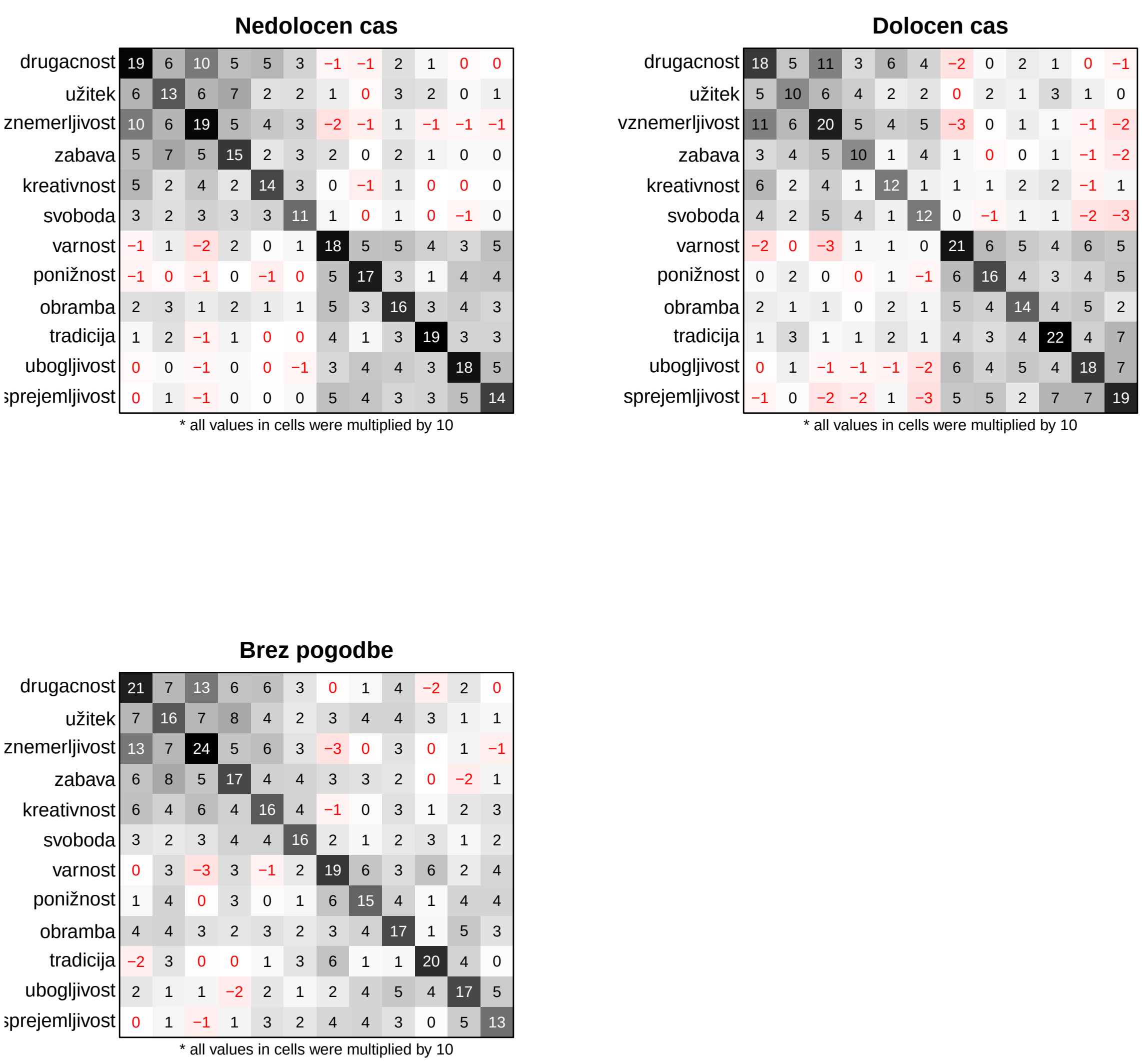
Slika 3: Korelacijska matrika Schwartzovih spremenljivk

Vidimo, da vrednosti korelacijskih koeficientov med nobenim parom spremenljivk niso blizu 1 in predpostavka o odsotnosti multikolinearnosti je izpolnjena.

5. Var-cov matrika je v vsaki populacijski skupini enaka

Peta predpostavka je enakost variančno-kovariančne matrike numeričnih (vsebinskih) spremenljivk v populacijskih skupinah. V primeru kršene predpostavke lahko uporabimo kvadratno diskriminantno analizo.

Na spodnjih slikah vidimo, da se variančno kovariančne matrike po delovnih razmerjih nekoliko razlikujejo, v vseh skupinah pa bolj ali manj lahko prepoznamo sklopa spremenljivk, še najmanj v skupini brez pogodbe, kjer je tudi najmanjše število enot. Tudi zadnja predpostavka je tako na prvi pogled izpolnjena.



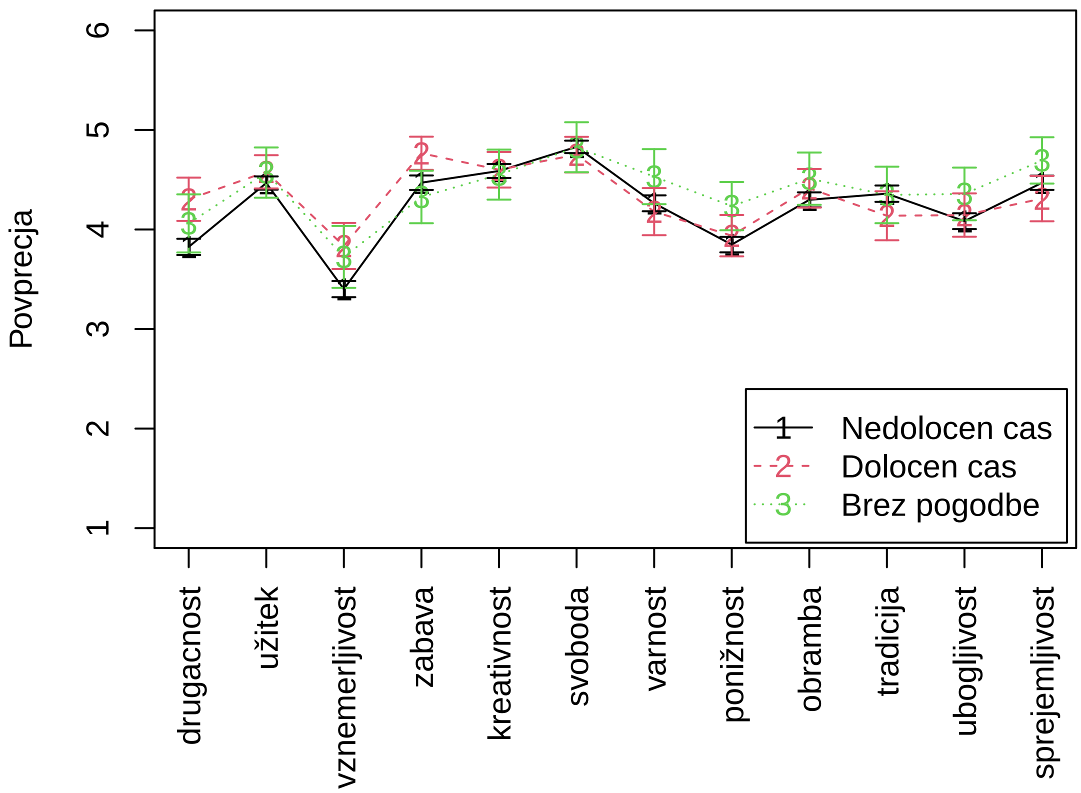
Slika 4: Var-cov matrike populacijksih skupin

Podrobneje in objektivno predpostavko o enakosti variančno-kovariančnih matrik lahko preverimo z Box-ovim M testom. Ničelna domneva je, da so variančno-kovariančne matrike po spremenljivkah enake. Slaba lastnost testa je, da ni robusten na odstopanja od normalnosti. Če je predpostavka o normalnosti spremenljivk kršena, bo večja verjetnost, da bodo enote prej razvrščene v skupino z večjimi kovariancami.

Pri stopnji statistične značilnosti 0.05, na podlagi vrednosti p (p = 0.6153) predpostavke o enakosti variančno-kovariančnih matriko ne moremo zavrniti in tako je tudi z objektivnega stališča predpostavka izpolnjena. χ^2 testna statistika zavzame vrednost 150 pri 156 stopinjah prostosti.

6. Razlikovanje povprečij vsebinskih spremenljivk

Naslednja predpostavka, s katero v bistvu preverimo oziroma utemeljimo smiselnost uporabe metode diskriminantne analize je, da pogledamo kako se razlikujejo povprečja spremenljivk na podlagi katerih razvrščamo v skupine glede na delovno razmerje. Če so razlike v povprečjih majhne, lahko pričakujemo, da bomo z diskriminatnimi funkcijami slabo ločevali med skupinami.



Slika 5: Razlikovanje povprečij vsebinskih spremenljivk

Glede na povprečja spremenljivk odprtosti lahko pričakujemo, da bomo dobro ločevali med respondenti s pogodbo za določen čas in tistimi s pogodbo za nedoločen čas, saj se povprečja med tema dvema skupinama najbolj razlikujejo. Pri spremenljivkah konservativnosti pa pričakujemo, da bomo dobro ločevali predvsem respondente brez pogodbe in ostali dve skupini.

Neke razlike v povprečjih torej so, so pa razlike zelo majhne in posledično lahko pričakujemo, da metoda ne bo delovala najbolje.

S testom Manova lahko testiramo ničelno domnevo, da so povprečja intervalnih spremenljivk med skupinami enaka. Na podlagi vrednosti p ($p < 0.001$) lahko pri stopnji statistične značilnosti 0.05 zavrnemo ničelno domnevo, da so povprečja med skupinami glede na vsebinske spremenljivke enaka in sprejmemo sklep, da se vsaj eno povprečje pri skupinah spremenljivke Razmerje razlikuje od ostalih dveh.

Preverimo lahko, med katerimi kategorijami spremenljivke Razmerje obstajajo razlike v povprečjih.

Tabela 9: Dodatek k testu Manova

drugačnost	$F(2; 1327) = 8.588$	$p < 0.01$
užitek	$F(2; 1327) = 0.952$	$p = 0.39$
vznemerljivost	$F(2; 1327) = 8.053$	$p < 0.01$
zabava	$F(2; 1327) = 4.81$	$p < 0.01$
kreativnost	$F(2; 1327) = 0.055$	$p = 0.95$
svoboda	$F(2; 1327) = 0.341$	$p = 0.71$
varnost	$F(2; 1327) = 2.148$	$p = 0.12$
ponižnost	$F(2; 1327) = 4.127$	$p < 0.05$
obramba	$F(2; 1327) = 1.685$	$p = 0.19$
tradicija	$F(2; 1327) = 1.616$	$p = 0.2$
ubogljivost	$F(2; 1327) = 1.943$	$p = 0.14$
sprejemljivost	$F(2; 1327) = 2.889$	$p < 0.10$

Vidimo, da pri stopnji statistične značilnosti 0.05 lahko zavrnemo domnevo o enakosti povprečij med skupinami spremenljivke nadzor pri vsebinskih spremenljivkah drugačnosti, vznemerljivosti, zabave in ponižnosti. Spremenljivka sprejemljivost je mejno statistično značilna, ostale spremenljivke pa ne razlikujejo med vsaj dvema povprečjema skupin.

Zaključimo lahko, da so predpostavke linearne diskriminatorne analize izpolnjene in lahko nadaljujemo z analizo. Zaradi majhnega razlikovanja med povprečji pa lahko pričakujemo, da metoda ne bo delovala najbolje.

Diskriminantne funkcije

Pomembnost in število diskriminantnih funkcij

Izvedimo sedaj linearno diskriminantno analizo na standardiziranih podatkih. Verjetnosti pripadnosti enot posamezni skupini določimo glede na zgoraj navedene deleže po kategorijah spremenljivke Razmerje iz podatkov. Za oceno vrednosti kanoničnih korelacij uporabimo Jack-knife metodo, kjer model ocenimo na vseh enotah razen ene, na kateri pa model evalviramo. Tako dobimo nepristransko oceno modela.

V spodnjem izpisu prvo množico spremenljivk predstavljajo vse Schwartzove spremenljivke, druga množica spremenljivk pa določa skupine. V izpisu sta dve kanonični korelaciji, ki sta narejeni na podlagi dummy spremenljivk iz spremenljivke Razmerje, ki sta dve.

Tabela 10: Vrednosti kanoničnih korelacij

Eigenvalues	%	Cum %	Cor	Sq. Cor
0.028	64.04	64.04	0.166	0.028
0.016	35.96	100.00	0.125	0.016

S prvo diskriminantno funkcijo pojasnimo 64% razlik med povprečji spremenljivk, z drugo kanonično korelacijo pa 36% razlik med povprečji spremenljivk. Korelacija med dummy spremenljivko, ki določa funkcijo in prvo diskriminantno funkcijo je enaka 0.028 in z drugo diskriminantno funkcijo je enaka 0.016. Pomen diskriminantnih funkcij torej pada.

Poglejmo še ali sta diskriminantni funkciji statistično značilni. Preverjamo ničelno domnevo, da so povprečja diskriminantnih funkcij po skupinah enaka.

Tabela 11: Statistična značilnost diskriminantnih funckij

	WilksL	F	df1	df2	p
1 to 2	0.957	2.433	24	2632	0.000
2 to 2	0.984	1.912	11	1317	0.034

Pri stopnji značilnosti 0.05 lahko na podlagi vrednosti p za obe diskriminantni funkciji zavrnemo domnevo o enakosti povprečij diskriminantnih funkcij po skupinah in sprejmemo sklep, da so povprečja pri obeh diskriminantih funkcijah po skupinah različna. Tako bomo v nadaljevanju interpretirali in napovedovali za obe diskriminantni funkciji.

Pomen diskriminantnih funkcij

Nadalje nas zanima, med katerimi skupinami ločuje posamezna diskriminantna funkcija. Eden izmed načinov, da to pogledamo so centroidi torej povprečja po skupinah.

Tabela 12: Centro

	LD1	LD2
Nedoločen čas	-0.078	-0.012
Določen čas	0.416	-0.182
Brez pogodbe	0.249	0.407

Prva linearna diskriminantna funkcija najbolj ločuje med osebami s pogodbo za nedoločen čas in s pogodbo za določen čas. Ločuje tudi med ostalima kombinacijama zaposlitvenega razmerja. Druga linearna diskriminantna funkcija pa najbolj ločuje med osebami brez pogodbe in osebami s pogodbo za določen čas. Ločuje pa tudi med osebami s pogodbo za nedoločen čas in brez pogodbe.

Poglejmo še katere spremenljivke imajo največji vpliv na vrednosti diskriminantne funkcije. Torej spremenljivke, ki najbolj ločujejo med skupinami. V spodnjem grafičnem prikazu so podane standarizirane vrednosti diskriminantne funkcije.

Slika 6: Standarizirane vrednosti diskriminantne funkcije

Visoke pozitivne uteži pri prvi diskriminantni funkciji imata predvsem drugačnost in vznemerljivost, ki sta močno korelirani s prvo diskriminantno funkcijo. Šibko ali srednje pozitivno korelirane s prvo diskriminantno funkcijo so še zabava, ponižnost, obramba, ubogljivost in užitek. Šibko negativno korelirana je tradicija. Ostale spremenljivke so zanemarljivo korelirane s prvo diskriminantno funkcijo.

Sklepamo lahko, da sta ljudem s pogodbo za določen čas pomembna predvsem drugačnost in vznemerljivost, ljudem s pogodbo za nedoločen čas pa tradicija.

Z drugo diskriminantno funkcijo je izrazito najmočneje negativno korelirana spremenljivka zabava, korelacija je močna. Pozitivno korelirane so spremenljivke sprejemljivost, varnost, in ponižnost, ki so srednje močno korelirane z drugo diskriminantno funkcijo. Ostale korelacije so manjše.

Sklepamo lahko, da so ljudem brez pogodbe najbolj pomembne varnost, sprejemljivost in ponižnost, ljudem s pogodbo za določen čas pa predvsem zabava.

Kvaliteta ocenjenega modela

Na podlagi Scwhartzovih spremenljivk lahko napovemo delovno razmerje posameznika, kako uspešna pa ta napoved bo, si bomo ogledali sedaj. Preverili bomo torej kako kvaliteten je ocenjen diskriminantni model. Posamezno enoto uvrstimo v skupino na podlagi ocenjenih diskriminantnih funkcij in sicer v tisto skupino katere centroid v prostoru diskriminantnih funkcij ji je najbližje.

Kot rečeno smo za oceno modela vzeli vse enote razen ene in na tej eni enoti model evalvirali, torej po principu Jack-knife moetode. Model je tudi upošteval, da skupine niso enako velike.

V spodnji tabeli je prva spremenljivka originalna spremenljivka Nadzor, druga spremenljivka pa je napovedana pripadnost Nadzoru na podlagi linearnih diskriminantnih funkcij.

Iz klasifikacijske tabele razberemo, da smo pravilno razvrstili 82% enot. Opomba: Vse enote so bile razvrščene v skupino z največjim deležem oziroma frekvenco, torej skupino z delovnim razmerjem za nedoločen čas. To je tudi vzrok, da je delež pravilno razvrščenih enot zelo visok. Enote z delovnim razmerjem za določen čas so torej 100% pravilno razvrščene, enote iz drugih dveh skupin pa 100% napačno razvrščene.

Za predstavo so velja ogledati še delež pravilno razvrščenih enot, če bi enote v skupine razvrščali povsem slučajno. Ta delež je enak 67%, kar ni tako zelo manjše od deleža naše ocene, ki je obenem delež pravilno razvrščenih enot, ko smo vse enote razvrstili v največjo skupino. Pod črto, naš model bi lahko bil boljši, ni pa popolnoma slab.

Tabela 13: Klasifikacijska tabela

	Nedoločen čas	Določen čas	Brez pogodbe	Sum
Nedoločen čas	1087	0	0	1087
Določen čas	145	0	0	145
Brez pogodbe	98	0	0	98
Sum	1330	0	0	1330

Kaj pa bi se zgodilo, če bi predpostavili, da so skupine v populaciji enako velike? Torej je verjetnost pripadnosti enote posamezni skupini $\frac{1}{3}$.

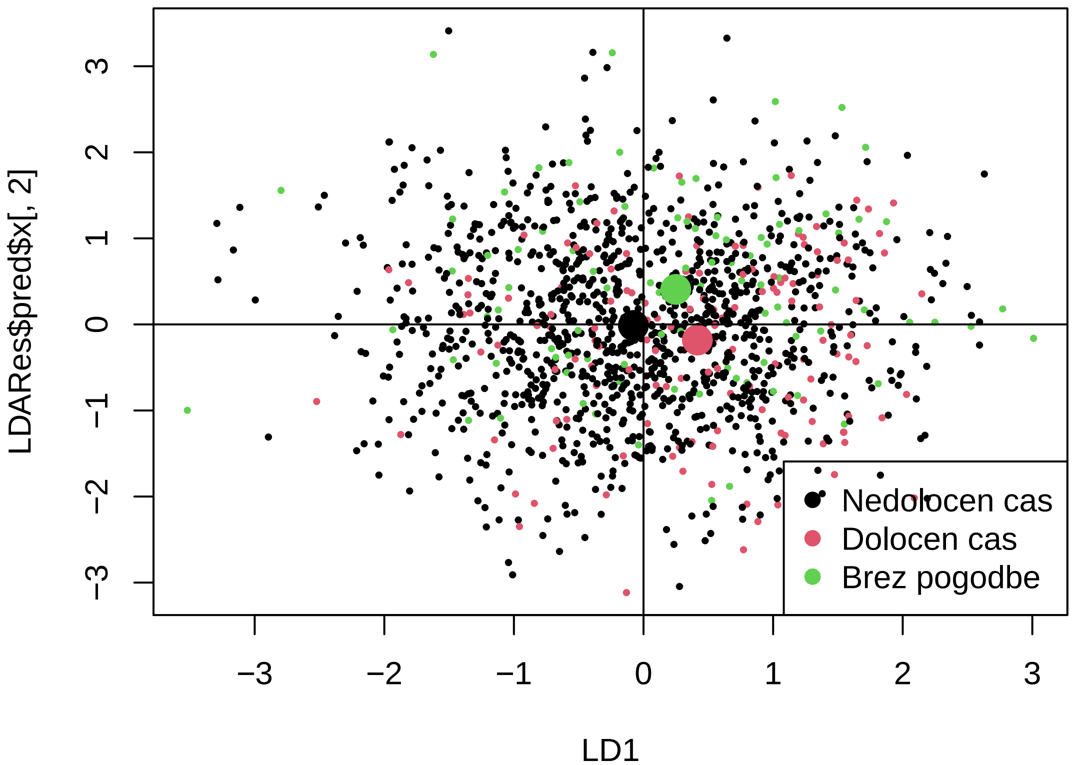
Tabela 14: Klasifikacijska tabela z enako velikimi skupinami

	Nedoločen čas	Določen čas	Brez pogodbe	Sum
Nedoločen čas	430	327	330	1087
Določen čas	37	56	52	145
Brez pogodbe	32	29	37	98
Sum	499	412	419	1330

Če pogledamo rezultate, ponovno po Jack-knife metodi, je delež pravilno razvrščenih enot 39%, kar je bistveno manj od prejšnjega, ki pa ga zaradi že navedenega razloga vseeno velja jemati z rezervo. Slabše smo torej napovedali enote z zaposlitvenim razmerjem za nedoločen čas, z razmerjem za določen čas in brez pogodbe pa smo enote napovedali bolje.

Enote v prostoru ocenjenih diskriminantnih funkcij

Spodnji graf prikazuje enote v prostoru ocenjenih dveh diskriminantnih funkcij, ki sta tudi statistično značilni. Vidimo, da status zaposlitvenega razmerja ni jasno ločen v prostoru glavnih komponent, kar je posledica tega, da diskriminantni funkciji ne ločujeta najbolje, oziroma tega, da je v eni skupini večina enot.



Slika 7: Enote v prostoru ocenjenih diskriminantnih funkcij

Vsebinski povzetek

V nalogi smo se ukvarjali s kanonično korelacijo ter diskriminantno analizo.

Pri kanonični korelaciji smo najprej izračunali koliko kanoničnih korelacij uporabiti. Ugotovili smo da 3 saj so prve tri imele p vrednost 0.004 ali manj ($\alpha = 0.05$), medtem ko je imela četrta 0.159, peta 0.272 ter šesta 0.963. Prve tri korelacije so bile enake 0.26, 0.21 in 0.12. Prva kanonična korelacija se je bila zelo korelirana v sklopu spremenljivk ki so izražale konservativnost in sicer zelo močno pozitivno korelacijo pri varnosti ter najmanj pri ponižnost, kjer je bila šibka. Pri sklopu spremenljivk ki so odražale odprtost sta bili spremenljivki užitek in zabava močno pozitivno korelirani, vznemerljivost je bila srednje močno negativno korelirana, medtem ko so ostale spremenljivke bile bodisi šibko bodisi zanemarljivo korelirane. Pri drugi kanonični korelaciji se je v sklopu konservativnosti najbolj izražala obramba in sicer z močno negativno korelacijo, varnost je bila srednje pozitivno močna, ponižnost pozitivno šibka, tradicija negativno šibka, ubogljivost in sprejemljivost pa zanemarljivi. Pri sklopu odprtostu je opaziti da so vse spremenljivke negativno korelirane in sicer močno korelirane so bile spremenljivke drugačnost, užitek ter vznemerljivost. Zelo blizu jim sledi kreativnost, medtem ko sta zabava in svoboda šibko korelirani. Tretja kanonična korelacija v sklopu konservativnih spremenljivk meri predvsem tradicijo in ubogljivost ki sta bili srednje močno do močno korelirani, medtem ko so v sklopu spremenljivk odprtosti najbolj izražene zabava in svoboda ki sta bili močno negativno korelirani, ter kreativnost ki je bila srednje močno negativno korelirana.

Največ variabilnosti spremenljivk odprtosti smo pojasnili z drugo kanonično spremenljivko in sicer 34%, sledi tretja in četrta z 16.9% in 14.7%. Prva je imela 14.1%, peta in šesta pa 12.5% oz. 8.3%. Pri spremenljivkah konservativnosti največ variabilnosti pojasnimo s prvo kanonično spremenljivko in sicer 28%, sledi šesta z 21%, tretja kanonična spremenljivka pojasni 14.2%, medtem ko druga, peta in četrta pojasnijo 12.6%, 12.5% in 12.1% variabilnosti. Pri preverjanju kolikšen delež variabilnosti odprtosti pojasnimo s kanoničnimi spremenljivkami konservativnosti in obratno, ugotovimo da je pri vseh kanoničnih spremenljivkah bilo to manj od 2%.

Pri diskriminantni analizi smo najprej preverjali predpostavke le-te. Za spremenljivko Razmerje iz prve naloge smo ugotovili da izpolni vse predpostavke in sicer da je bilo število skupin vsaj dve, da so v vsaki skupini bile vsaj dve enoti, da je bilo število spremenljivk manjše od števila enot -1, da nismo imeli opravka z multikolinearnostjo, da je Var-cov matrika v vsaki skupini enaka (Smo potrdili z Box-ovim M testom; p=0.6153) ter da so se povprečja vsebinskih spremenljivk razlikovala. S testom MANOVA, smo zavrnili ničelno domnevo ($p<0.001$), da so povprečja enaka. Po podrobnejšem pregledu smo ugotovili da smo lahko zavrnili domnevo o enakosti povprečij med skupinami spremenljivke nadzor pri vsebinskih spremenljivkah drugačnosti, vznemerljivosti, zabave in ponižnosti.

Nato smo ugotovili da s prvo diskriminantno funkcijo pojasnimo 64% razlik med povprečji spremenljivk, z drugo pa 36%. Obe diskriminantni funkciji sta bili statistično značilni (1 - $p<0.001$, 2 - $p=0.034$) in smo tako lahko sprejeli sklep da so povprečja pri obeh diskriminantnih funkcijah po skupinah različna.

Pri oceni modela, smo ugotovili da smo pravilno razvrstili 82% enot, kar je bila posledica tega da smo imeli eno skupino zelo veliko in je posledično model vse razvrstil v največjo skupino. Ko smo preverjali kolikšen delež enot bi bil pravilno razvrščen če bi enote naključno razvrščali v skupine bi bil ta enak 67%. Ko smo predpostavili da so skupine enako velike, je bil delež pravilno razvrščenih enot enak 39%.