

Dr. sc. Nataša Kurnoga Živadinović

MULTIVARIJATNA KLASIFIKACIJA ŽUPANIJA HRVATSKE

MULTIVARIATE CLASSIFICATION OF CROATIAN COUNTIES

SAŽETAK: Cilj rada je klasifikacija županija Hrvatske u veće grupe sličnih socioekonomskih obilježja. Primijenjene su metode multivarijatne analize: klaster analiza, faktorska analiza i diskriminacijska analiza. Provedene su hijerarhijska i nehijerarhijska klaster analiza. Primijenjene su različite metode hijerarhijske klaster analize, a najbolje interpretabilno rješenje dobiveno je Wardovom metodom s kvadriranim euklidskim udaljenostima. Odabrano je rješenje s četiri klastera, te je provedena nehijerarhijska klaster analiza, K-means metoda. Provedena je faktorska analiza glavnih komponenata s varimaks rotacijom faktora. Na osnovi deset socio-ekonomskih pokazatelja faktorskom su analizom izlučena tri faktora te su izračunani faktorski bodovi. Faktorski bodovi korišteni su u diskriminacijskoj analizi. Diskriminacijska analiza je provedena za tri grupe klaster analize s tri izlučena faktora kao nezavisnim varijablama. Za razvijanje diskriminacijske funkcije primijenjena je Wilksova lambda.

KLJUČNE RIJEČI: klasifikacija, multivarijatna analiza, faktorska analiza, klaster analiza, diskriminacijska analiza.

ABSTRACT: The aim of this paper is to classify Croatian counties in larger groups with similar socio-economic characteristics. Multivariate methods were used for the classification: cluster analysis, factor analysis and discriminant analysis. The hierarchical and non-hierarchical cluster analysis was run. Various methods of hierarchical cluster analysis were employed and the best interpretative solution was provided by Ward's method with squared euclidean distances. Solution with four clusters was chosen and the K-means method of the non-hierarchical cluster analysis was run. Principal component analysis with varimax rotation was performed. Three factors were extracted in the factor analysis of ten socio-economic indicators and the factor scores were calculated. The factor scores were used in discriminant analyses. Discriminant analysis was conducted for the three clusters with the three extracted factors as independent variables. The Wilks lambda was used to develop the discriminant function.

KEYWORDS: classification, multivariate analysis, factor analysis, cluster analysis, discriminant analysis

1. UVOD

Hrvatska je kao zemlja kandidat za pridruživanje Europskoj uniji trebala uspostaviti Nomenklaturu prostornih jedinica za statistiku (NUTS – *Nomenclature des Unités Territoriales Statistiques*) sa svrhom jedinstvenog praćenja regionalne statistike Europske unije kako bi se stvorio sustav regionalne statistike. NUTS klasifikacija omogućuje prikupljanje usporedivih podataka, praćenje, analizu i utvrđivanje stupnja razvoja regija, te praćenje i ocjenu učinaka definiranih mjera regionalne politike. Prostor svake zemlje članice Europske unije se, na osnovi NUTS klasifikacije, dijeli na tri NUTS razine čija je osnovna karakteristika određen broj stanovnika za svaku razinu. Također su definirane i dvije razine lokalnih administrativnih jedinica (LAU). NUTS klasifikacija je razvijena s obzirom na sljedeće kriterije: poštovanje institucionalne podjele zemlje, poštovanje općih geografskih jedinica i hijerarhijska klasifikacija.¹

Klasifikacija prostora na NUTS razine započinje na prvoj NUTS razini gdje to može biti sama zemlja ili više regija. Primjerice², od članica EU-a Cipar, Češka, Danska, Estonija, Irska, Latvija, Litva, Luksemburg, Malta, Slovačka, Slovenija i Švedska imaju po jednu regiju na prvoj NUTS razini koja čini samu zemlju, dok Finska i Bugarska imaju dvije uspostavljene regije na prvoj NUTS razini, zatim Austrija, Belgija, Mađarska i Portugal po tri regije, a preostale zemlje više od četiri regije gdje Njemačka ima najviše regija na prvoj razini, šesnaest. Svaka od regija prve NUTS razine, ako ih ima više, dijeli se na regije druge NUTS razine. Međutim, neke zemlje nemaju uspostavljene regije ni za drugu NUTS razinu: Cipar, Danska, Estonija, Latvija, Litva, Luksemburg, Malta i Slovenija. Nadalje, pojedine članice EU-a, poput Francuske, imaju i regije koje se zemljopisno nalaze izvan europskog kontinenta, u prekomorskim departmanima, ali koje su sastavni dijelovi EU-a, te čine regije druge NUTS razine: Guadalupe, Martinik, Gvajana i Reunion. Svaka se regija druge NUTS razine dijeli na regije treće NUTS razine i tim slijedom dalje na LAU razine. Prema tome, regije niže razine su dijelovi regija više razine.

NUTS razine određuju se prema broju stanovnika. Pri tome treba naglasiti da je prema uputi EUROSTAT-a potrebno uvažiti i druge kriterije, kao što su homogenost regija, struktura gospodarstva, stupanj razvijenosti, prirodno-geografske karakteristike, geopolitički položaj, prostorna veličina, povijesna tradicija i slično. Primjerice, Bratislava ima status regije druge NUTS razine (Bratislavsky kraj), a ima manje od 800.000 stanovnika kao donje granice. S obzirom na broj stanovnika, prva NUTS razina odnosi se na prostor s brojem stanovnika od 3.000.000 do 7.000.000, druga NUTS razina na prostor s brojem stanovnika od 800.000 do 3.000.000, a treća NUTS razina na prostor s brojem stanovnika od 150.000 do 800.000.³

Prema NUTS klasifikaciji u Hrvatskoj je početkom 2007. godine uspostavljena Nacionalna klasifikacija prostornih jedinica za statistiku (NKPJS). Prostorna jedinica za statistiku prve NKPJS razine je Hrvatska, prostorne jedinice za statistiku druge NKPJS razine su Sjeverozapadna Hrvatska, Središnja i Istočna (Panonska) Hrvatska te Jadranska Hrvatska, a prostorne jedinice za statistiku treće NKPJS razine su 20 županija i Grad Zagreb.

¹ Prema: http://ec.europa.eu/comm/eurostat/ramon/nuts/basicnuts_regions_en.html

² Prema: http://ec.europa.eu/comm/eurostat/ramon/nuts/codelist_en.cfm?list=nuts

³ Prema: http://ec.europa.eu/comm/eurostat/ramon/nuts/basicnuts_regions_en.html

U radu je na osnovi odabranih socioekonomskih pokazatelja, primjenom metoda multivarijatne analize, provedena klasifikacija županija Hrvatske u veće i homogene prostorne jedinice sličnih socioekonomskih obilježja. Od metoda multivarijatne analize primijenjene su klaster, faktorska i diskriminacijska analiza.

2. METODOLOGIJA ISTRAŽIVANJA

Za potrebe klasifikacije županija Hrvatske odabrano je deset socioekonomskih pokazatelja jer se zbog ograničenja o broju izvornih varijabli u faktorskoj analizi nije mogao koristiti veći broj pokazatelja. S obzirom na dostupnost podataka na razini županija, odabrani su sljedeći pokazatelji socioekonomske strukture županija:⁴ bruto domaći proizvod po stanovniku u kunama (bdpst), stopa registrirane nezaposlenosti (stnezap), zaposleni u primarnom sektoru na 1.000 stanovnika (zappst), zaposleni u sekundarnom sektoru na 1.000 stanovnika (zapsst), zaposleni u tercijarnom sektoru na 1.000 stanovnika (zaptst), koeficijent pokrivenosti uvoza izvozom⁵ (koefui), promet s PDV-om trgovina na malo po stanovniku (prometst), doktori medicine na 10.000 stanovnika (drmedst), zubni terapeuti na 10.000 stanovnika (zubterst) i vrijednost prodaje industrijskih proizvoda (PRODCOM) u industrijskim lokalnim jedinicama u tisućama kuna (prodip).

Dio podataka za navedene pokazatelje po županijama preuzet je iz publikacije Statistički ljetopis Republike Hrvatske 2006. godine Državnog zavoda za statistiku (zappst, zapsst, zaptst, koefui, prometst, drmedst i zubterst). Podatci su također preuzeti iz publikacije Statističke informacije 2006. godine (stnezap i prodip) i Priopćenja Državnog zavoda za statistiku (bdpst). Svi korišteni podatci ne odnose se na istu godinu što ovisi o dostupnosti podataka, primjerice korišten je BDP po stanovniku za 2004. godinu jer noviji podatci nisu dostupni.

3. KLASIFIKACIJA ŽUPANIJA HRVATSKE

Klasifikacija županija Hrvatske provedena je primjenom metoda multivarijatne analize na osnovi prethodno navedenih pokazatelja. Primijenjene su klaster, faktorska i diskriminacijska analiza.

Klaster analiza omogućuje razvrstavanje objekata (opažanja ili varijabli) u klaster s obzirom na njihovu sličnost prema mjerenim obilježjima. Svaki klaster bi trebao biti homogen i svi klasteri između sebe heterogeni, odnosno objekti klastera trebali bi nalikovati jedni drugima i biti različiti od objekata u drugim klasterima. Klaster analizom nastoji se utvrditi broj klastera i sastav klastera. Osnovna podjela metoda klaster analize je na hijerarhijske metode koje karakterizira razvijanje hijerarhije i nepoznat broj klastera te nehijerarhijske metode kojima se objekti grupiraju u unaprijed određen broj klastera. Od hijerarhijskih metoda primjenjuju se metoda najbližeg susjeda (*single linkage method*), metoda najudaljenijeg susjeda (*complete linkage method*), metoda prosječne povezanosti (*average lin-*

⁴ U zagradama su navedeni skraćeni nazivi varijabli koje se koriste dalje u radu.

⁵ Koeficijent pokrivenosti uvoza izvozom je relativni broj koordinacije i izračunava se kao omjer izvoza i uvoza.

kage method), Wardova metoda i centroidna metoda, a od nehijerarhijskih K-means metoda. U svrhu utvrđivanja sličnosti objekata koji se nastoje grupirati koriste se mjere udaljenosti. Od mjera udaljenosti najčešće se koriste euklidska udaljenost i kvadrirana euklidska udaljenost.

Osnovni cilj faktorske analize je sažimanje informacija velikog broja izvornih varijabli u manji broj zajedničkih faktora uz minimalan gubitak informacija. Iz osnovnog cilja proizlaze različiti zadatci faktorske analize, kao što su određivanje faktora i faktorskih bodova sa svrhom njihova korištenja u daljnjim analizama, identificiranje dimenzija koje se ne mogu odmah uočiti, istraživanje novog područja – eksploratorna faktorska analiza, testiranje definiranih hipoteza – konfirmatorna faktorska analiza, te zahtjev parsimonije ili štednje, odnosno objašnjavanje većeg broja varijabli na osnovi manjeg broja faktora što dovodi do smanjenja troškova istraživanja.

Diskriminacijska analiza je metoda u kojoj su varijable klasificirane kao zavisne i nezavisne varijable pri čemu bi zavisna varijabla trebala biti mjerena na nominalnoj skali i imati grupa koliko i kategorija nominalne varijable, a nezavisne varijable bi trebale biti mjerene na intervalnoj ili omjernoj skali i može ih biti bilo koji broj, ali za barem dva manje od ukupnog broja opažanja. Diskriminacijskom analizom je moguće u dvije ili više definiranih grupa zavisne varijable klasificirati opažanja na osnovi odabranih nezavisnih varijabli. Diskriminacijska se analiza može provesti na manjem broju faktora izlučenih faktorskom analizom umjesto na velikom broju izvornih varijabli. Primjenom nekoreliranih faktora i njihovih faktorskih bodova može se ukloniti ili barem smanjiti problem multikolinearnosti koja može utjecati na rezultate.

3.1. Klasifikacija županija Hrvatske primjenom klaster analize

Klaster analizom se objekti (opažanja ili varijable) grupiraju u manje grupe ili klasterne. S obzirom da je cilj klaster analize grupirati slične objekte zajedno, potrebno je utvrditi koliko su ti objekti slični odnosno različiti. Stoga se u klaster analizi primjenjuju različite mjere udaljenosti. Upravo zbog mjera udaljenosti potrebno je voditi računa o mjernim jedinicama odabranih varijabli jer je većina mjera udaljenosti prilično osjetljiva na različite mjerne jedinice. Naime, mjere udaljenosti djeluju prvenstveno na doprinos varijabli mjernih većim mjernim jedinicama. S obzirom da su odabrane varijable izražene u različitim mjernim jedinicama, provedena je transformacija originalnih podataka. Da bi se uklonio utjecaj različitih mjernih jedinica varijabli, primijenjene su standardizirane vrijednosti.

U radu je prvo provedena hijerarhijska klaster analiza, i to različite metode, a zatim je na dobivenim rezultatima provedena nehijerarhijska klaster analiza, odnosno K-means metoda. Na taj je način hijerarhijska metoda klaster analize upotpunjena mogućnošću nehijerarhijske metode klaster analize da se promijeni pripadnost klasteru.

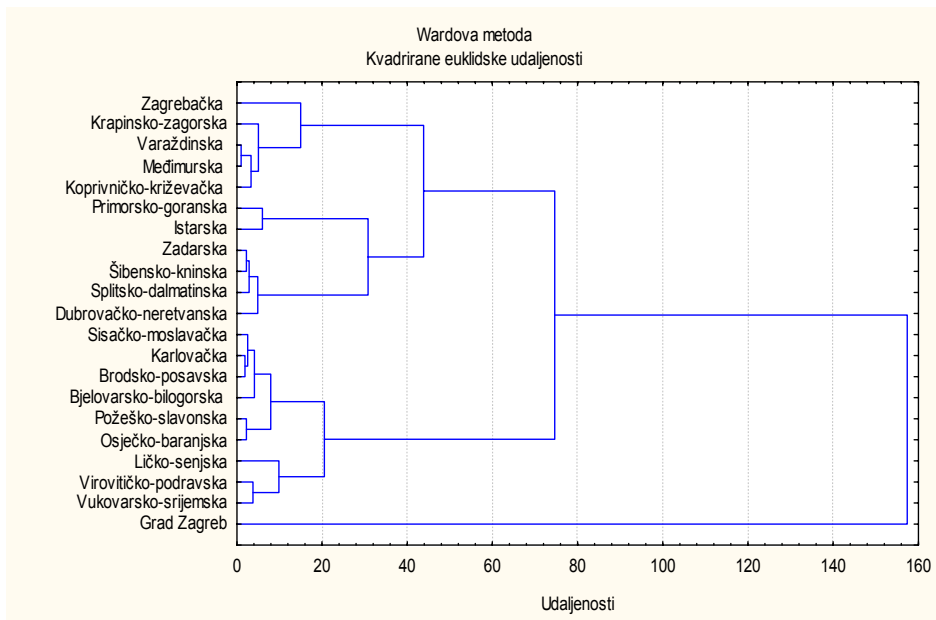
Provedene su različite metode hijerarhijske klaster analize, ali je najbolje interpretabilno rješenje dobiveno Wardovom metodom s kvadriranim euklidskim udaljenostima. Na slici 1 prikazan je dendrogram⁶ provedene klaster analize primjenom Wardove metode s

⁶ Dendrogram je grafički prikaz postupnog kombiniranja objekata u klasterne i udaljenosti između pojedinih razina. Čita se s lijeva na desno gdje su opažanja prikazana na vertikalnoj liniji, a udaljenosti između klastera na kojoj su spojeni na horizontalnoj liniji.

kvadriranim euklidskim udaljenosti. Iz dendrograma je vidljivo da je na osnovi Wardove metode s kvadriranim euklidskim udaljenostima prihvatljivo rješenje s tri i četiri klastera.

U rješenju s tri klastera prvi klaster⁷ čini jedanaest županija: Varaždinska, Međimurska, Koprivničko-križevačka, Krapinsko-zagorska, Zagrebačka, Zadarska, Šibensko-kninska, Splitsko-dalmatinska, Dubrovačko-neretvanska, Primorsko-goranska i Istarska županija. Drugi klaster čini devet županija: Karlovačka, Brodsko-posavska, Požeško-slavonska, Osječko-baranjska, Sisačko-moslavačka, Virovitičko-podravska, Vukovarsko-srijemska, Bjelovarsko-bilogorska i Ličko-senjska županija. Posljednji klaster čini Grad Zagreb.

Slika 1. Dendrogram klaster analize za 20 županija i Grad Zagreb



U rješenju s četiri klastera prvi klaster⁸ čini pet županija: Varaždinska, Međimurska, Koprivničko-križevačka, Krapinsko-zagorska i Zagrebačka županija. Taj se klaster može nazvati *Sjeverozapadna Hrvatska*. Drugi klaster čini šest županija: Zadarska, Šibensko-kninska, Splitsko-dalmatinska, Dubrovačko-neretvanska, Primorsko-goranska i Istarska županija. Taj se klaster može nazvati *Jadranska Hrvatska*. Treći klaster čini devet županija: Karlovačka, Brodsko-posavska, Požeško-slavonska, Osječko-baranjska, Sisačko-moslavačka, Virovitičko-podravska, Vukovarsko-srijemska, Bjelovarsko-bilogorska i Ličko-senjska županija. Taj se klaster može nazvati *Središnja i Istočna Hrvatska*. Posljednji klaster čini Grad Zagreb.

⁷ Županije su unutar klastera poredane prema redosljedju spajanja u klaster. Raspored spajanja u klaster vidljiv je u dendrogramu.

⁸ Županije su unutar klastera poredane prema redosljedju spajanja u klaster. Raspored spajanja u klaster vidljiv je u dendrogramu.

Na rezultatima prethodno provedene hijerarhijske klaster analize provedena je K-means metoda nehijerarhijske klaster analize gdje je zadani broj klastera četiri. Inicijalni klasteri centri su automatski određeni. U tabeli 1 je dana klasifikacija županija u četiri klastera K-means metodom te udaljenosti do klasterskog centra za svaku pojedinu županiju unutar pripadajućeg klastera. Što je udaljenost od klasterskog centra veća, to je konkretno opažanje udaljenije od danog klastera. Primjerice, za Primorsko-goransku i Istarsku županiju je udaljenost od klasterskog centra prilično velika u odnosu na ostale županije. To upućuje na moguće odvajanje tih dviju županija u zaseban klaster, a što je vidljivo i na dendrogramu prikazanom na slici 1 ako se promatra rješenje s pet klastera. Prema tome, u rješenju s pet klastera te se dvije županije odvajaju od Zadarske, Šibensko-kninske, Splitsko-dalmatinske i Dubrovačko-neretvanske županije u zaseban klaster.

Tabela 1. Klasifikacija K-means metodom i udaljenosti od klasterskog centra

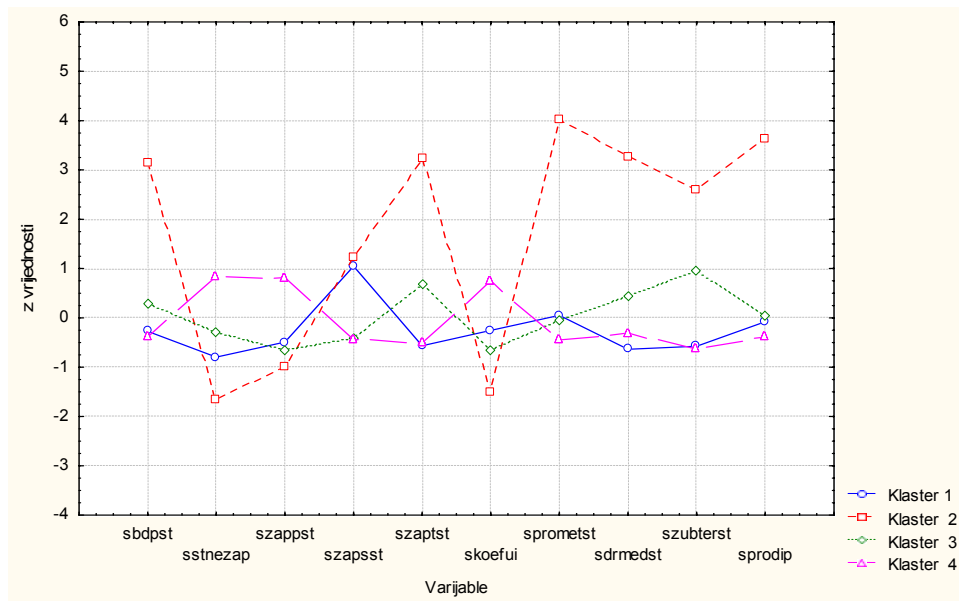
Županija	Pripadnost klasteru	Udaljenosti
Zagrebačka županija	1	0,77
Krapinsko-zagorska županija	1	0,47
Sisačko-moslavačka županija	4	0,54
Karlovačka županija	4	0,46
Varaždinska županija	1	0,39
Koprivničko-križevačka županija	1	0,39
Bjelovarsko-bilogorska županija	4	0,49
Primorsko-goranska županija	3	0,82
Ličko-senjska županija	4	0,83
Virovitičko-podravska županija	4	0,55
Požeško-slavonska županija	4	0,44
Brodsko-posavska županija	4	0,49
Zadarska županija	3	0,50
Osječko-baranjska županija	4	0,32
Šibensko-kninska županija	3	0,65
Vukovarsko-srijemska županija	4	0,60
Splitsko-dalmatinska županija	3	0,30
Istarska županija	3	0,81
Dubrovačko-neretvanska županija	3	0,50
Međimurska županija	1	0,31
Grad Zagreb	2	0,00

Iz tabele 1 je vidljivo da je prvi klaster dobiven K-means metodom potpuno identičan prvom klasteru dobivenom Wardovom metodom koji je nazvan *Sjeverozapadna Hrvatska*. Drugi klaster dobiven K-means metodom čini Grad Zagreb koji je i Wardovom metodom izdvojen u zaseban klaster. Treći klaster dobiven K-means metodom je potpuno identičan drugom klasteru dobivenom Wardovom metodom koji je nazvan *Jadranska Hrvatska*. Četvrti klaster dobiven K-means metodom je identičan trećem klasteru dobivenom Wardovom metodom koji je nazvan *Središnja i Istočna Hrvatska*. Iz rezultata prikazanih u tabeli vidljivo je da je sastav županija po klasterima jednak sastavu klastera dobivenih Wardovom

metodom te se potvrđuju rezultati hijerarhijske klaster analize o četiri klastera: *Sjeverozapadna Hrvatska, Jadranska Hrvatska, Središnja i Istočna Hrvatska* te *Grad Zagreb*.

Svaki klaster moguće je opisati na osnovi korištenih varijabli u klaster analizi. Pri tome je moguće koristiti grafički prikaz sredina za svaki formirani klaster. Slika 2 prikazuje grafički prikaz sredina za četiri formirana klastera.

Slika 2. Grafički prikaz sredina svakog klastera za četiri odabrana klastera



Iz grafičkog prikaza je vidljivo da se drugi klaster, kojeg čini Grad Zagreb, izrazito odvaja od ostalih klastera po odabranim varijablama. Naime, taj klaster ima više vrijednosti za varijable BDP po stanovniku, zaposlenima u tercijarnim djelatnostima, prometu po stanovniku i vrijednosti prodaje industrijskih proizvoda od ostalih klastera, te niže vrijednosti za varijable stopu nezaposlenosti i koeficijentu pokrivenosti uvoza izvozom. Ta sva obilježja odvajaju Grad Zagreb kao zaseban klaster.

3.2. Faktorska analiza varijabli korištenih u klasifikaciji

U radu je primijenjen model faktorske analize glavnih komponenta

$$x_i = \lambda_{i1}c_1 + \lambda_{i2}c_2 + \dots + \lambda_{iv}c_v, \quad i = 1, 2, \dots, v \quad (1)$$

gdje je svaka varijabla x_i opisana kao linearna kombinacija novih nekoreliranih glavnih komponenta c_j , $j = 1, 2, \dots, v$. Vrijednosti λ_{ij} su komponentni koeficijenti koji ukazuju na

važnost j -te glavne komponente c_j u kompoziciji i -te varijable x_i . Sve utvrđene glavne komponente ne uzimaju se u obzir nego se, prema određenim kriterijima, izdvaja nekoliko njih, posebno ako je njihov udio u ukupnoj varijanci visok. /3, str. 15/

Prije primjene faktorske analize potrebno je ispitati jesu li ispunjeni potrebni preduvjeti za njezino provođenje. Varijable koje se žele analizirati primjenom faktorske analize trebale bi biti kvantitativne. S obzirom da su varijable numeričke, taj je uvjet ispunjen. Nadalje, faktorska analiza polazi od korelacijske matrice svih izvornih varijabli gdje je korelacijska matrica dobar indikator mogućih faktora i praktički prikazuje moguću faktorsku strukturu. Ukoliko su korelacije između varijabli slabe, faktorska analiza bi mogla biti neprikladna. Korelacijska matrica za deset odabranih izvornih varijabli ukazuje na korelacije između nekih varijabli, odnosno na njihovu značajnu statističku povezanost.⁹ Za ispitivanje prikladnosti podataka za primjenu faktorske analize primjenjuje se i Kaiser-Meyer-Olkinova mjera, za sve varijable zajedno, ali i za svaku varijablu pojedinačno.¹⁰ Budući da vrijednost Kaiser-Meyer-Olkinove mjere iznosi 0,725, te da su pojedinačne vrijednosti također zadovoljavajuće, može se reći da su podaci odabranih varijabli prikladni za provođenje faktorske analize.

Faktorska analiza glavnih komponenti provodi se na nereduciranoj korelacijskoj matrici, odnosno na glavnoj dijagonali se nalaze jedinice čime se uključuje ukupna varijanca. Prema tome, u faktorskoj analizi glavnih komponenti model ne diferencira zajedničku i specifičnu varijancu, te se faktori zasnivaju na ukupnoj varijanci, a izlučeni faktori sadrže i specifičnu varijancu. Budući da se faktorska analiza glavnih komponenti provodi na nereduciranoj korelacijskoj matrici gdje se na glavnoj dijagonali nalaze jedinice i inicijalni komunaliteti su jednaki jedinicama te je njihov zbroj jednak broju izvornih varijabli.

Nakon što je provedena transformacija varijabli u ortogonalne osi izlučuju se faktori. Pri izlučivanju faktora polazi se od svojstvenih vrijednosti. Za faktorsku analizu glavnih komponenti karakteristično je da se faktori izlučuju na osnovu Kaiserova kriterija, odnosno izlučuju se oni faktori čija je svojstvena vrijednost veća od jedan. Svojstvena vrijednost prvog faktora iznosi 6,079, drugog faktora 1,352, trećeg faktora 1,124, a četvrtog faktora 0,61. Prema Kaiserovom kriteriju, izlučena su tri faktora jer je svojstvena vrijednost četvrtog faktora manja od jedan.

Pri određivanju broja faktora može se primijeniti i kriterij postotka objašnjene varijance prema kojem se faktori izlučuju sve dok se više ne mogu odrediti faktori koji značajno smanjuju neobjašnjenu varijancu. U tabeli 2 su, između ostalog, prikazane svojstvene vrijednosti i postotak objašnjene varijance. Vidljivo je da kumulativni postotak varijance objašnjene s tri izlučena faktora iznosi 85,55%. U prirodnim znanostima izlučivanje faktora ne bi trebalo prestati sve dok izlučeni faktori ne objašnjavaju najmanje 95% ukupne varijance. Međutim, u društvenim znanostima istraživači se nerijetko odlučuju za rješenje kojim se objašnjava 60% od ukupne varijance. /4, str. 120/ S obzirom na visok postotak objašnjene varijance, tri izlučena faktora ne zadovoljavaju samo Kaiserov kriterij već i kriterij postotka objašnjene varijance.

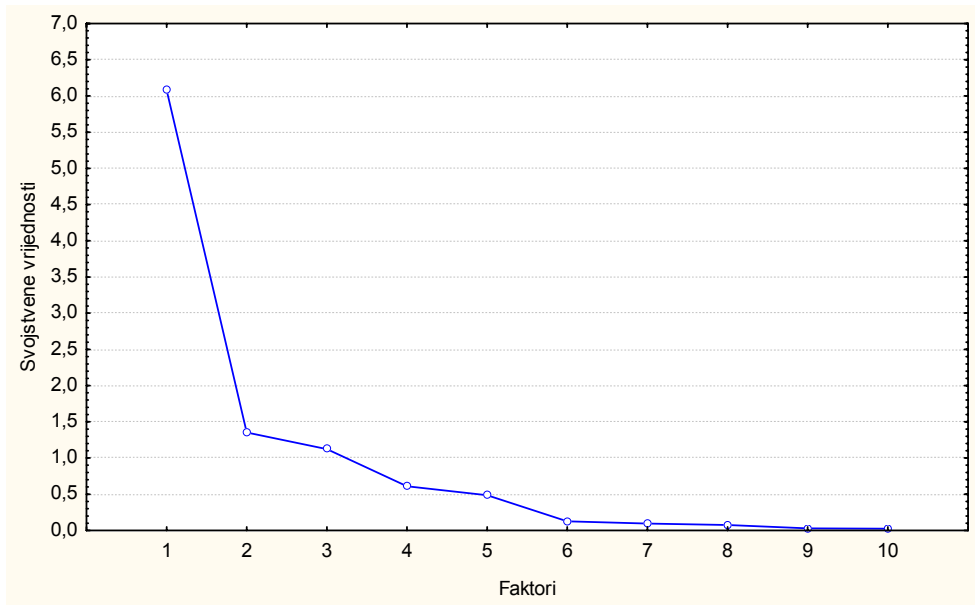
Broj faktora može se odrediti i grafički primjenom Cattelijeva dijagrama (scree plot) gdje oblik dijagrama određuje broj faktora, odnosno vidljiv je prekid između faktora s viso-

⁹ Korelacijska matrica nije prikazana u radu zbog opsežnosti, no može se dobiti od autora.

¹⁰ Korelacijska matrica nije prikladna za primjenu faktorske analize ako je vrijednost Kaiser-Meyer-Olkinove mjere manja od 0.5.

kim svojstvenim vrijednostima i faktora s niskim svojstvenim vrijednostima. Na slici 3 je prikazan Cattellijev dijagram gdje grafički prikaz potvrđuje izbor triju izlučenih faktora.

Slika 3. Cattellijev dijagram (*scree plot*)



U tabeli 2 je prikazana inicijalna matrica faktorske strukture te pripadajući komunaliteti za svaku pojedinu varijablu.¹¹ Niske vrijednosti komunaliteta ukazuju na varijable koje bi se mogle izostaviti iz analize. Iz tabele je vidljivo da su za sve varijable vrijednosti komunaliteta visoke. Ispitivanjem inicijalne matrice faktorske strukture uočava se da pojedine varijable koreliraju s nekoliko faktora te da je prvi faktor definiran visokim faktorskim opterećenjima većeg broja varijabli. Budući da inicijalna matrica nije interpretabilna, provodi se rotacija faktora kojom se mijenja odnos između varijabli i faktora.

¹¹ Matrica faktorske strukture prikazuje korelacije između pojedinih izvornih varijabli i pojedinih faktora. Komunalitet varijable je zbroj kvadrata faktorskih opterećenja, a tumači se kao doprinos varijanci varijable od zajedničkih faktora, odnosno kao proporcija varijance varijable objašnjene izlučenim faktorima.

Tabela 2. Inicijalna matrica faktorske strukture, komunaliteti, svojstvene vrijednosti i objašnjena varijanca

Varijable	Faktori			Komunaliteti
	1.	2.	3.	
bdpst	0,834	0,263	0,293	0,851
stnezap	-0,733	0,425	-0,203	0,759
zappst	-0,576	0,579	0,450	0,870
zapsst	0,445	-0,665	0,545	0,937
zaptst	0,924	0,314	-0,066	0,956
koefui	-0,638	0,266	0,598	0,835
prometst	0,884	0,092	0,146	0,811
drmedst	0,817	0,306	0,003	0,761
zubterst	0,895	0,217	-0,272	0,922
prodip	0,898	0,085	0,198	0,852
Svojstvene vrijednosti	6,079	1,352	1,124	
% objašnjene varijance	60,792	13,523	11,235	
Kumulativni % objašnjene varijance	60,792	74,314	85,550	

S obzirom da su izlučeni faktori korišteni u daljnjim analizama, provedena je ortogonalna rotacija faktora koja rezultira nekoreliranim faktorima. Odabrana je metoda varimaks rotacije jer rezultira pojednostavljenjem stupaca u matrici faktorske strukture, odnosno pojednostavljenjem faktora. Naime, kao što je već rečeno, prvi faktor ima velik broj visokih faktorskih opterećenja. Nakon provedene rotacije faktora, također je potrebno odrediti matricu faktorske strukture¹² na kojoj se zasniva interpretacija faktora. Tabela 3 prikazuje matricu faktorske strukture nakon provedene varimaks rotacije faktora, te pripadajuće komunalitete, svojstvene vrijednosti i objašnjenu varijancu. Iz matrice faktorske strukture nakon provedene rotacije faktora vidljiva je izmijenjena struktura faktorskih opterećenja, što ukazuje da rotirana struktura faktorskih opterećenja omogućuje bolju interpretaciju faktora u odnosu na inicijalnu nerotiranu faktorsku matricu. Također je vidljivo da su komunaliteti nakon rotacije jednaki komunalitetima prije rotacije. Nadalje, vidljivo je da je postotak objašnjene varijance i dalje ostao jednak 85,55%, ali se preraspodijelio po faktorima. Nakon provedene rotacije prvi faktor objašnjava 46,84% varijance dok je prije rotacije objašnjavao 60,79%, drugi faktor objašnjava 20,62% varijance dok je prije rotacije objašnjavao 13,52%, a treći faktor objašnjava 18,09% varijance dok je prije rotacije objašnjavao 11,24%.

¹² Uz matricu faktorske strukture, rezultat rotacije faktora je i matrica faktorskog sklopa. Međutim, te su dvije matrice jednake ako su faktori ortogonalni, odnosno nakon ortogonalne rotacije. Te dvije matrice su različite ako su faktori korelirani, odnosno nakon kosokutne rotacije. Budući da je provedena ortogonalna rotacija, dovoljno je prikazati matricu faktorske strukture.

Tabela 3. Matrica faktorske strukture nakon varimaks rotacije faktora, komunaliteti, svojstvene vrijednosti i objašnjena varijanca

Varijable	Faktori			Komunaliteti
	1.	2.	3.	
bdpst	0,879	-0,002	0,280	0,851
stnezap	-0,421	0,356	-0,675	0,759
zappst	-0,118	0,867	-0,324	0,870
zapsst	0,100	-0,085	0,959	0,937
zaptst	0,932	-0,292	0,051	0,956
koefui	-0,314	0,857	-0,032	0,835
prometst	0,811	-0,217	0,326	0,811
drmedst	0,847	-0,198	0,064	0,761
zubterst	0,828	-0,485	-0,018	0,922
prodip	0,826	-0,186	0,368	0,852
Svojstvene vrijednosti	4,684	2,062	1,809	
% objašnjene varijance	46,841	20,618	18,090	
Kumulativni % objašnjene varijance	46,841	67,460	85,550	

Iz matrice faktorske strukture nakon varimaks rotacije vidljivo je da prvi faktor ima značajna faktorska opterećenja pozitivnog predznaka za šest varijabli: BDP po stanovniku, zaposleni u tercijarnom sektoru na 1.000 stanovnika, promet s PDV-om trgovina na malo po stanovniku, doktori medicine na 10.000 stanovnika, zubni terapeuti na 10.000 stanovnika i vrijednost prodaje industrijskih proizvoda (PRODCOM) u industrijskim lokalnim jedinicama. Drugi faktor ima značajno faktorsko opterećenje pozitivnog predznaka za dvije varijable: koeficijent pokrivenosti uvoza izvozom i zaposleni u primarnom sektoru na 1.000 stanovnika. Treći faktor ima značajno faktorsko opterećenje pozitivnog predznaka za varijablu zaposleni u sekundarnom sektoru na 1.000 stanovnika, a negativnog predznaka za varijablu stopa registrirane nezaposlenosti.

Za potrebe diskriminacijske analize, nakon izlučivanja faktora, su izračunani faktorski bodovi.¹³ Osnova za izračunavanje faktorskih bodova je matrica faktorske strukture, a faktorski se bodovi izračunavaju za svako opažanje na osnovi izlučenih faktora. Prema tome, svaki faktor ima faktorskih bodova koliko i opažanja.

3.3. Provođenje diskriminacijske analize na osnovi rezultata klaster i faktorske analize

U radu je diskriminacijska analiza primijenjena za provjeru rješenja klaster analize jer je na osnovi diskriminacijske analize moguće otkriti koje županije nisu spojene u odgovarajuću grupu, te specificirati najprikladniju grupu za te županije.

¹³ Faktorski bodovi nisu prikazani u radu zbog opsežnosti ali se mogu dobiti od autora.

Jedan je od preduvjeta provođenja diskriminacijske analize da u svakoj grupi ima barem dva opažanja što nije slučaj rješenja prethodno opisane klaster analize s četiri klastera. Naime, četvrti klaster ima samo jedno opažanje, Grad Zagreb. S obzirom da se Grad Zagreb izdvaja i po drugim klasterskim rješenjima (K-means metoda) kao zaseban klaster, diskriminacijska analiza je provedena bez četvrtog klastera. Prema tome, diskriminacijska analiza je provedena na 20 županija, za tri grupe definirane klaster analizom uz tri izlučena faktora i faktorskim bodovima kao nezavisnim varijablama. Za razvijanje diskriminacijske funkcije primijenjena je Wilksova lambda.

Za predviđanje kojoj grupi pripadaju opažanja, odnosno u ovom slučaju kojem klasteru pripadaju županije, primijenjene su klasifikacijske funkcije. Klasifikacijske funkcije se računaju za svaku grupu i mogu se izravno koristiti za klasifikaciju opažanja. U tabeli 4 su prikazani koeficijenti klasifikacije funkcije gdje je vidljivo da za svaki od triju klastera postoji funkcija. Primjerice, funkcija za prvi klaster je sljedeća:

$$-6,99404 - 7,85108 \cdot faktor1 - 0,66732 \cdot faktor2 + 4,43311 \cdot faktor3 \quad (2)$$

Tabela 4. Koeficijenti klasifikacijske funkcije

	klaster1	klaster2	klaster3
faktor1	-7,85108	6,47960	-3,63782
faktor2	-0,66732	3,24346	-2,95027
faktor3	4,43311	-1,71752	-1,00563
konstantni član	-6,99404	-4,25711	-2,60052

Klasifikacijska matrica prikazuje rezultate korištenja funkcija u klasifikaciji opažanja. U tabeli 5 dana je klasifikacijska matrica koja prikazuje broj korektno i pogrešno klasificiranih opažanja, u ovom primjeru županija. Za *a-priori* klasifikacijske vjerojatnosti određeno je da su *proporcionalne veličinama grupa*. Stoga su *a-priori* klasifikacijske vjerojatnosti određene s obzirom na broj županija u svakom klasteru. Prvi klaster ima pet županija pa je njegova *a-priori* vjerojatnost jednaka 0,25, za drugi klaster koji ima šest županije jednaka je 0,3, a za treći klaster koji ima devet županija jednaka je 0,45. Na dijagonali klasifikacijske matrice nalaze se korektno klasificirana opažanja, dok izvandijagonalni elementi prikazuju pogrešno klasificirana opažanja s obzirom na predviđene i stvarne grupe. Omjer pogrešno klasificiranih opažanja i ukupnog broja opažanja daje proporciju korektno klasificiranih opažanja. Iz tabele je vidljivo da su sve županije korektno klasificirane.

Tabela 5. Klasifikacijska matrica

	klaster1 p=0,25	klaster2 p=0,30	klaster3 p=0,45
klaster1	5	0	0
klaster2	0	6	0
klaster3	0	0	9
Ukupno	5	6	9

3.4. Uspostavljena klasifikacija prostornih jedinica za statistiku Hrvatske

Na početku 2007. godine u Hrvatskoj je uspostavljena Nacionalna klasifikacija prostornih jedinica za statistiku (NKPJS). U skladu s NUTS klasifikacijom uspostavljene su tri NKPJS razine. Prostorna jedinica za statistiku prve NKPJS razine je Hrvatska kao administrativna jedinica. Prostorne jedinice za statistiku druge NKPJS razine su tri neadministrativne jedinice: Sjeverozapadna Hrvatska, Središnja i Istočna (Panonska) Hrvatska te Jadranska Hrvatska. Navedene tri prostorne jedinice nastale su grupiranjem županija kao administrativnih jedinica niže razine. Prostorne jedinice za statistiku treće NKPJS razine čine 21 administrativnu jedinicu, odnosno 20 županija i Grad Zagreb koji ima status grada i županije jer kao glavni grad predstavlja jedinstvenu teritorijalnu i samoupravnu jedinicu. U tabeli 6 prikazana je Nacionalna klasifikacija prostornih jedinica za statistiku.

Tabela 6. Nacionalna klasifikacija prostornih jedinica za statistiku (NKPJS)

NKPJS razine	Regije	Broj stanovnika
Prva NKPJS razina	Hrvatska	4.437.460
Druga NKPJS razina	Sjeverozapadna Hrvatska	1.658.935
Treća NKPJS razina	Grad Zagreb	779.145
	Zagrebačka županija	309.696
	Krapinsko-zagorska županija	142.432
	Varaždinska županija	184.769
	Koprivničko-križevačka županija	124.467
	Međimurska županija	118.426
Druga NKPJS razina	Središnja i Istočna (Panonska) Hrvatska	1.351.517
Treća NKPJS razina	Bjelovarsko-bilogorska županija	133.084
	Virovitičko-podravska županija	93.389
	Požeško-slavonska županija	85.831
	Brodsko-posavska županija	176.765
	Osječko-baranjska županija	330.506
	Vukovarsko-srijemska županija	204.768
	Karlovačka županija	141.787
	Sisačko-moslavačka županija	185.387
Druga NKPJS razina	Jadranska Hrvatska	1.427.008
Treća NKPJS razina	Primorsko-goranska županija	305.505
	Ličko-senjska županija	53.677
	Zadarska županija	162.045
	Šibensko-kninska županija	112.891
	Splitsko-dalmatinska županija	463.676
	Istarska županija	206.344
	Dubrovačko-neretvanska županija	122.870

Izvor: Nacionalna klasifikacija prostornih jedinica za statistiku (NKPJS), NN 35/07. (Podatci o broju stanovnika po županijama preuzeti su iz Statističkog ljetopisa Republike Hrvatske 2006., Državni zavod za statistiku Republike Hrvatske, Zagreb, 2006., str. 630.)

Prethodno provedenim metodama multivarijatne analize dobivena je slična klasifikacija županija. Razlika je u klasifikaciji Ličko-senjske županije koja je multivarijatnim metodama grupirana sa županijama Središnje i Istočne Hrvatske. Također je razlika i u izdvojanju Grada Zagreba u zasebnu grupu na osnovi odabranih socioekonomskih pokazatelja. Međutim, prema uvjetima NUTS klasifikacije Grad Zagreb ne može sam biti regija druge NUTS razine jer ima manje od 800.000 stanovnika.

4. ZAKLJUČAK

Klasifikacija županija Hrvatske provedena je primjenom metoda multivarijatne analize na osnovi deset odabranih socioekonomskih pokazatelja u veće i homogene prostorne jedinice. Provedene su različite metode hijerarhijske klaster analize, ali je najbolje interpretabilno rješenje dobiveno Wardovom metodom s kvadriranim euklidskim udaljenostima. Izdvojila su se četiri klastera: Sjeverozapadna Hrvatska, Jadranska Hrvatska, Središnja i Istočna Hrvatska, te Grad Zagreb. Nakon provedene hijerarhijske klaster analize, provedena je nehijerarhijska klaster analiza, K-means metoda s četiri zadana klastera. Usporedbom rezultata hijerarhijske i nehijerarhijske klaster analize zaključeno je da se rezultati u potpunosti podudaraju.

Na deset odabranih socioekonomskih pokazatelja provedena je faktorska analiza glavnih komponenata. Budući da inicijalna matrica nije bila interpretabilna, provedena je ortogonalna varimaks rotacija. Izlučena su tri faktora te su za izlučene faktore izračunani faktorski bodovi za svaku županiju. Izračunani faktorski bodovi korišteni su u diskriminacijskoj analizi.

Na temelju dobivenih rezultata klaster i faktorske analize provedena je diskriminacijska analiza. Za razvijanje diskriminacijske funkcije primijenjena je Wilksova lambda. Diskriminacijska analiza je provedena na 20 županija, bez Grada Zagreba, jer se diskriminacijska analiza ne može provesti ukoliko jedna od definiranih grupa ima samo jedno opažanje. Kao zavisna varijabla, korištene su tri grupe definirane klaster analizom, a kao nezavisne varijable, tri izlučena faktora i njihovi faktorski bodovi. Rezultati diskriminacijske analize potvrđuju rezultate klaster analize.

LITERATURA

1. Armstrong, H., Jouan de Kervenoael, R. J., Li, X., Read, R., *A Comparison of the Economic Performance of Different Micro-states and Between Micro-states and Larger Countries*, World Development, Vol. 26, No. 4, pp. 639-656, 1998.
2. Armstrong, H., Read, R., *Western European Micro-states and EU Autonomous Regions: The Advantages of Size and Sovereignty*, World Development, Vol. 23, No. 7, pp. 1229-1245, 1995.
3. Harman, H. H., *Modern Factor Analysis*, 3rd Ed., The University of Chicago, Chicago 1976.
4. Hair, J. F., Black, W.C., Babin, B. J., Anderson, R. E., Tatham, R. L., *Multivariate Data Analysis*. 6th Ed., Prentice Hall, Upper Saddle River, 2006.

5. Kurnoga Živadinović, N., *Klasifikacija prostornih jedinica prema stupnju ekonomske razvijenosti*, doktorska disertacija, Ekonomski fakultet – Zagreb, 2007.
6. Liou, F. M., Ding, C. G., *Subgrouping Small States Based on Socioeconomic Characteristics*, *World Development*, Vol. 30, No. 7, pp. 1289-1306, 2002.
7. Lovrinčević, Ž., Marić, Z., Rajh, E., *Kako optimalno regionalizirati Hrvatsku?*, *Ekonomski pregled*, godina 56, broj 12, pp. 1109-1160, 2005.
8. *Nacionalna klasifikacija prostornih jedinica za statistiku (NKPJS)*, NN 35/07.
9. Soares, J. O., Marques, M. M. L., Monteiro, C. M. F., *A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions*, *European Journal of Operational Research*, 145, pp. 121-135, 2003.
10. *Statistički ljetopis Republike Hrvatske 2006.*, Državni zavod za statistiku Republike Hrvatske, Zagreb, 2006.
11. *Statističke informacije 2007.*, Državni zavod za statistiku Republike Hrvatske, Zagreb, 2007.
12. Ulengin, F., Ulengin, B., Onsel S., *A Power-based Measurement Approach to Specify Macroeconomic Competitiveness of Countries*, *Socio-Economic Planning Sciences*, Vol. 36, No. 3, pp. 203-226, 2002.
13. <http://ec.europa.eu/comm/eurostat/ramon/nuts>
14. <http://www.dzs.hr/>