

Milan Papić*
Davor Žmegač**
Igor Zovak***

NOVA METODA KLASTER ANALIZE U MS EXCELU NA PRIMJERU ISTRAŽIVANJA MODELA REGIONALIZACIJE REPUBLIKE HRVATSKE

Sažetak

Klaster analiza služi kao statistička tehnika za utvrđivanje relativno homogenih grupa objekata, odnosno za kategorizaciju pojedinih jedinica analize uzimajući u obzir njihovu sličnost, odnosno različitost prema mjerenim obilježjima. Klaster analiza relativno je složena tehnika zbog čega ne nailazi na širu primjenu u akademskoj zajednici. U ovome radu prikazan je relativno jednostavan način na koji se primjenom MS Excela može napraviti analiza i bez detaljnog poznavanja matematike, odnosno korištenja složenijih statističkih programa. Jednostavnost i primjenjivost analize te dostupnost programa za izradu trebao bi pridonijeti široj upotrebi tog alata u akademskoj zajednici.

Ključne riječi: klaster analiza, MS Excel, regionalizacija, gospodarska specijalizacija

1. Uvod

Klaster analiza predstavlja statističku tehniku za utvrđivanje relativno homogenih grupa objekata (Everitt et al., 2011). Koristi se u različitim granama znanosti za kategorizaciju odnosno klasifikaciju pojedinih jedinica analize (ispitanika, proizvoda, poduzeća, geografskih jedinica) s obzirom na njihovu sličnost odnosno različitost prema nekim njihovim mjerenim obilježjima. Klaster analiza je multivarijantna tehnika kojom se sustavno grupiraju objekti u skupine koje su određene njihovim osobinama (atributima). Posebnost klaster analize u odnosu na ostale slične tehnike u tome je da ulazne varijable nisu plod empirijskog istraživanja nego se koriste varijable koje zadaje sam istraživač (Žmegač, 2020).

* mr. sc. Milan Papić, Libertas međunarodno sveučilište, mpapic@libertas.hr

** dr. sc. Davor Žmegač, Petrokemija d.d., Kutina, dzmegac@libertas.hr

*** Igor Zovak, mag. oec., Libertas međunarodno sveučilište, izovak@libertas.hr

Uporaba klaster analize ima široku primjenu u medicini, biologiji, prostornom planiranju i planiranju općenito, prometnim studijama, ekologiji, istraživanjima tržišta itd. Uz sve prednosti metode, klaster analiza ima i određene nedostatke koji se prije svega ogledaju u ulaznim varijablama koje su korištene kao podloga za utvrđivanje sličnosti (zahtijeva pažnju istraživača) te posebno što konačan broj klastera znatno ovisi o subjektivnim odukama istraživača (Žmegač, 2020).

Cilj je klaster analize unutar određenog skupa objekata formirati grupe tako da su, promatrajući određena obilježja, istovremeno sličnosti unutar grupa i razlike između grupa što veće (Hair et al., 2010). Termin klaster dolazi od engleske riječi *cluster* što znači 'skupina istovrsnih stvari, grozd, skupiti u hrpu'. Primjena klaster analize, a prije svega njezino razumijevanje, zahtijeva detaljnije poznavanje određenih područja matematike (vektorska analiza, euklidska geometrija, metoda najmanjih kvadrata), kao i poznavanje nekih specifičnih računalnih programa (SPSS, Statistica, SAS, Past, Cluster 3.0).

U ovome radu objašnjen je postupak klaster analize u kojem se primjenom MS Excela na relativno jednostavan način i bez detaljnog poznavanja matematičke pozadine može doći do njezinog osnovnog cilja (Papić, 2018).

2. Razrada problema

Metoda je primijenjena u sklopu šireg istraživanja čiji je cilj bio novi regionalni ustroj Republike Hrvatske. Konkretni primjer pokazuje na koji se način mogu grupirati općine unutar jedne regije¹ u klastere, odnosno podregije, sa zajedničkim gospodarskim specijalizacijama² (Žmegač, 2020):

1. vlastiti prihodi jedinice lokalne samouprave po glavi stanovnika
2. udio zaposlenih u djelatnosti poljoprivrede i djelatnosti ribarstva u ukupnom broju zaposlenih
3. udio zaposlenih u djelatnosti prerađivačke industrije u ukupnom broju zaposlenih
4. udio zaposlenih u djelatnosti hoteli i restorani te djelatnosti poslovanje nekretninama, iznajmljivanje i poslovne usluge u ukupnom broju zaposlenih
5. stopa nezaposlenosti
6. koeficijent starenja
7. vitalni indeks
8. udio stanovništva bez škole u ukupnom stanovništvu
9. broj noćenja turista

¹ Za naš primjer koristili smo podatke za Istočnu Hrvatsku (ukupno 153 općine i gradova).

² Ukupno 11 gospodarskih specijalizacija (op. a).

10. bruto domaći proizvod po glavi stanovnika
11. udio zaposlenih u ukupnom stanovništvu.

U prvom koraku postavljen je matematički model kojim se definira početni skup podataka te odgovarajuće funkcije kojima se postiže željeni cilj.³ Nakon toga pristupili smo postupku klaster analize u MS Excelu koji se detaljnije opisuje u daljnjem tekstu.

3. Postupak klaster analize u MS Excelu

Provedeni postupak klasterizacije izveden je u tri koraka koji su nastavno obrazloženi i grafički prikazani.

Prvi korak

Početna baza podataka formirana je tako da su podaci vezani uz gospodarske specijalizacije uneseni u 11 stupaca u koje su upisane odgovarajuće vrijednosti (brojevi od 1 do 11) ukoliko općina ima tu specijalizaciju. Ostale ćelije ostavljene su prazne. Izdvojimo iz ukupne baze podataka dio koji nam je potreban za rješavanje zadanog problema. Na novi radni list kopiramo stupac s imenima gradova i općina te stupce s gospodarskim specijalizacijama (GS). Dio tablice prikazan je na slici 1.

Slika 1. Prikaz početne baze podataka u prvom koraku

	A	B	C	D	E	F	G	H	I	J	K	L
1	Ime grada	Gospodarska specijalizacija 1	Gospodarska specijalizacija 2	Gospodarska specijalizacija 3	Gospodarska specijalizacija 4	Gospodarska specijalizacija 5	Gospodarska specijalizacija 6	Gospodarska specijalizacija 7	Gospodarska specijalizacija 8	Gospodarska specijalizacija 9	Gospodarska specijalizacija 10	Gospodarska specijalizacija 11
2	Daruvar	1	2	3					8		10	
3	Grubišno Polje	1	2	3					8		10	
4	Đulovac	1	2	3					8		10	
5	Veliki Grđevac	1	2	3					8		10	
6	Dežanovac	1	2	3					8		10	
7	Hercegovac	1	2	3					8		10	
8	Končanica	1	2	3					8		10	
9	Sirač	1	2	3					8		10	
10	Velika Trnovitica	1	2	3					8		10	
11	Slavonski Brod	1	2	3	4				8		10	
12	Nova Gradiška	1	2	3	4				8		10	
13	Sibinj	1	2	3	4				8		10	
14	Oriovac	1	2	3	4				8		10	

Izvor: sistematizacija autora.

Drugi korak

U drugom koraku označimo stupac B, ikonu *Sort&Filter* i izaberemo opciju *Sort Smallest to Largest* (slika 2).

³ Detalji matematičkog modela su dio šireg istraživanja i zainteresirani se mogu javiti autorima za dodatne informacije.

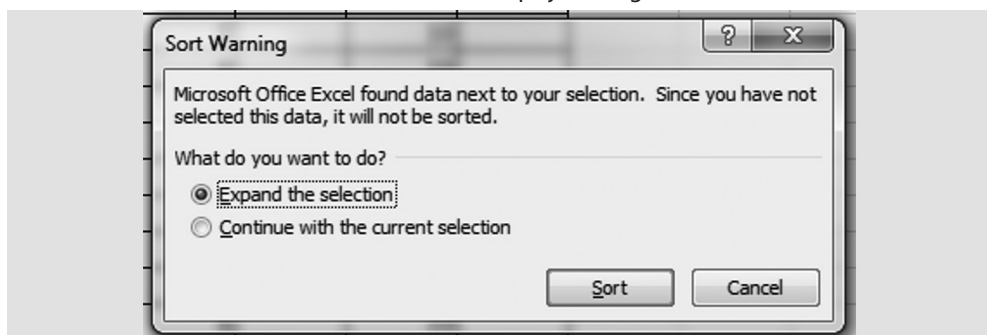
Slika 2. Prikaz baze podataka u drugom koraku

	A	B	C	D	E	F
1	Ime grada	Gospodarska specijalizacija 1	Gospodarska specijalizacija 2	Gospodarska specijalizacija 3	Gospodarska specijalizacija 4	Gospodarska specijalizacija 5
2	Daruvar	1	2	3		
3	Grubišno Polje	1	2	3		
4	Đulovac	1	2	3		
5	Veliki Grđevac	1	2	3		
6	Dežanovac	1	2	3		
7	Hercegovac	1	2	3		

Izvor: sistematizacija autora.

Nastavno, u dijaloškom okviru koji se pojavi izaberemo opciju *Expand the selection* → *Sort* (slika 3).

Slika 3. Prikaz odabrane opcije u drugom koraku



Izvor: sistematizacija autora.

Treći korak

Potrebno je označiti stupac C i ponoviti postupak iz prethodnog koraka. Zatim ponoviti isti postupak sa stupcima D, E i sve do L. Na kraju postupka dobivamo tablicu s grupiranim općinama i gradovima s jednakim skupom gospodarskih specijalizacija. Dio dobivene tablice (većina redaka je sakrivena, cijela tablica zauzela bi previše prostora) izgleda prema prikazu na slici 4.

Slika 4. Prikaz baze podataka u trećem koraku

	A	B	C	D	E	F	G	H	I	J	K	L
	Općine	Gospodarska specijalizacija 1	Gospodarska specijalizacija 2	Gospodarska specijalizacija 3	Gospodarska specijalizacija 4	Gospodarska specijalizacija 5	Gospodarska specijalizacija 6	Gospodarska specijalizacija 7	Gospodarska specijalizacija 8	Gospodarska specijalizacija 9	Gospodarska specijalizacija 10	Gospodarska specijalizacija 11
1												
2	Slavonski Brod	1	2	3	4				8			10
3	Nova Gradiška	1	2	3	4				8			10
58	Jasenovac	1	2	3	4				8			10
59												
60	Vinkovci	1			4				8			10
61	Vukovar	1			4				8			10
90	Negoslavci	1			4				8			10
91												
92	Daruvar	1	2	3					8			10
93	Grubišno Polje	1	2	3					8			10
114	Bjelovar	1	2	3					8			10
115												
116	Osijek	1	2	3	4							10
117	Đakovo	1	2	3	4							10
157	Levanjska Varoš	1	2	3	4							10

Izvor: sistematizacija autora.

Rješenje problema je rezultat s četirima podregijama (tablica 1).

Tablica 1. Tablični prikaz rješenja problema

	ZAJEDNIČKE GOSPODARSKE SPECIJALIZACIJE	N
podregija 1	1,2,3,4,8,10	57
podregija 2	1,4,8,10	31
podregija 3	1,2,3,8,10	23
podregija 4	1,2,3,4,10	42
ukupno		153

Izvor: sistematizacija autora.

4. Zaključak

Moglo bi se reći da su autori do ideje o ovoj metodi došli slučajno tijekom šireg istraživanja. Naime, početna baza podataka šireg istraživanja na kojem je baziran ovaj rad, formirana je na način koji ne bi omogućavao primjenu poznate metode klaster analize (bez obzira na to koji računalni program bi se koristio). Nakon formiranja matematičkog modela, pokazalo se da je krajnji cilj klaster analize (unutar određenog skupa objekata formirati grupe tako da su, promatrajući određena obilježja, istovremeno sličnosti unutar grupa i razlike između grupa što veće) moguće ostvariti vrlo jednostavnim postupkom u Excelu (s određenim brojem ponavljanja funkcije sortiranja). Pregledom literature iz tog područja na internetu, pokazalo se da postoje primjeri

metode klaster analize u Excelu, ali se oni baziraju na znatno složenijim matematičkim postupcima. Autori smatraju da metoda objašnjena u ovom radu može utjecati na znatno širu primjenu klaster analize među znanstvenicima, dio kojih ju je dosad često izbjegavao zbog njezine relativne složenosti.

Literatura

1. Everitt, B. S., Landau, S., Leese, M., Stahl, D., Shewhart, W. A. i Wilks, S. S. 2011. *Cluster Analysis*. 5th Edition. West Sussex: John Wiley and Sons, Ltd.
2. Hair, J. F., Black, W. C., Babin, B. J. i Anderson, R. E. 2010. *Multivariate Data Analysis - A Global Perspective*. New York: Pearson Education.
3. Papić, M. 2018. *Primijenjena statistika u MS Excelu*. Zagreb: Likarija d.o.o.
4. Sokal, R. R. i Sneath, P. H. A. 1963. *Principles of numerical taxonomy*. San Francisco: W. H. Freeman.
5. Žmegač, D. 2020. *Unapređenje sustava teritorijalne i funkcionalne organizacije područne (regionalne) samouprave kao okosnica gospodarskoga razvoja Republike Hrvatske*. Doktorski rad. Ekonomski fakultet Sveučilišta u Rijeci.



A new method of cluster analysis in MS Excel on the example of research of the regionalization model of the Republic of Croatia

Abstract

Cluster analysis is used as a statistical technique for determining relatively homogeneous groups of objects, with the aim of categorizing individual units of analysis considering their similarity or differences according to measured characteristics. Cluster analysis is a relatively complex technique, which is probably reason why it does not find a wider practical application in the academic community. This paper presents a relatively simple way with using MS Excel is possible to make an analysis without detailed knowledge of mathematics theory, or the use of complex statistical programs. The simplicity and applicability of presented method analysis in paper and the availability of required program, should contribute to the widespread use of this tool in the academic community.

Key words: Cluster analysis, MS Excel, regionalization, economic specialization