

PËRDORIMI I ANALIZËS SË KOMPONENTËVE KRYESORË DHE ANALIZËS SË GRUPEVE PËR ZBULIMIN E PABARAZIVE RAJONALE TË VARFËRISË, NJË RAST STUDIMI I PREFEKTURAVE TË SHQIPËRISË

XHAFAJ E.¹, NURJA I.², KOSOVA R.¹

¹Universiteti “Aleksandër Moisiu” Durrës, Fakulteti i Teknologjisë së Informacionit
Departamenti i Matematikës

²Universiteti i New–York–ut, Tiranë

e-mail: genaxhafa@yahoo.com

Përmbledhje

Ky studim ka si qëllim të grupojë prefekturat e Shqipërisë në bazë të disa treguesve të ndryshme të varfërisë. Si fillim paraprakisht përdoret analiza e komponentëve kryesorë (PCA) për të përcaktuar se cilat ndryshore janë më të rëndësishme në shpjegimin e komponentit të varfërisë. Më pas përdoret analiza e grupeve për grupimin e prefekturave. Një nga synimet e këtij punimi është që të tregojë se si harta e varfërisë duke përdorur analizën e grupeve na paraqet grupe homogjene të prefekturave. Në punimin tonë rezultojnë që prefekturat përfshihen në 4 grupe. Ku në një grup futen prefekturat e varfëra, Elbasan, Fier, Dibër, Shkodër, Lezhë një tjetër grup përbëhet nga prefekturat që kanë një intesitet të moderuar të varfërisë të cilat janë Durrës, Vlorë, Gjirokastrë, Korçë, Berat. Një grup të veçantë formon vetëm prefektura e Kukësit, e cila interpretohet si një prefekturë shumë e varfër dhe grupi i fundit përbëhet vetëm nga prefektura e Tiranës e cila interpretohet si prefektura më e zhvilluar.

Abstract

This study aims to group counties of Albania on the basis of several indicators that describe different intensity of poverty. At the beginning, we start by applying the principal component analyses (PCA). After we have defined the principal components and by that, we have reduced the variables, we continue by applying the cluster analyses. The purpose is to detect which variables are responsible for the most of the variance of the component. One aim of this paper is to show how poverty map, using cluster analysis, presents homogeneous groups of counties. The poverty map plays an important role in the visual appearance of the incidence of poverty on the basis of geographic regions. In our paper, it turns out that all counties are divided into 4 groups. The first group consists of the counties that have a moderate intensity of poverty which are Durrës, Vlora, Gjirokastra, Korça, Berat. The second group includes poor counties such as Elbasan, Fier, Dibra, Shkodra, Lezha; a special group forms the only region of Kukës, which is interpreted as a very poor counties and last group consists only of the counties of Tirana, which is interpreted as the most developed counties.

Fjalëkyçe: Analiza e grupeve (cluster analysis), metoda e komponentëve kryesorë, distanca Euklidiane, metoda e Ward-it, prefekturat e Shqipërisë

Hyrje

Qëllimi i këtij punimi është që me anë të analizës së komponentëve kryesorë të përcaktojmë ndryshoret më të rëndësishme në shpjegimin e komponentit të varfërisë dhe në bazë të këtyre ndryshoreve të grupojmë prefekturat e Shqipërisë. Të dhënat janë marrë nga INSTAT-i për vitin 2012.

Një studim i këtij lloji është shkruajtur nga Vázquez & Sumner (2012). Është përdorur analiza e grupeve për të identifikuar dhe klasifikuar 139 shtete të botës duke përdorur disa tregues të mirëqenies. Ka rezultuar që analiza e grupeve prodhon 5 grupe shtetesh për periudhën 2005-2010.

Një punim tjetër është shkruar nga Anderson, (2007). Ky studim shikon një zbatim të analizës së grupeve. Objektivi kryesor i analizës në këtë rast është të identifikojë dallimet e nivelit të zhvillimit midis shteteve në mënyrë që të krijojë grupe homogjene me karakteristika të ngjashme.

Materiali dhe metodat

Si fillim përdoret metoda PCA për të na treguar se sa komponentë janë të rëndësishëm për modelin. Peshat që ka përkatësisht secila ndryshore në secilin komponentë na ndihmojnë në interpretimin e këtyre të fundit. Peshat të cilat kanë vlerë absolute afër 1 na tregojnë që ndryshorja është e rëndësishme në shpjegimin e variancës së të dhënave. Sipas Henry *et.al* (2003) duhen marrë në konsideratë vetëm ato ndryshore me pesha më të mëdha se 0.3.

Dihet që analiza PCA është një teknikë e reduktimit të përmasave brenda disa llojesh të tjera analizash. Komponentët kryesorë janë nxjerrë (ekstratuar) në mënyrë të tillë që të zvogëlojnë përmasat para përdorimit të teknikave të ndryshme shumëpërmasore, më qartë në rastin tonë kjo teknikë është analiza e grupeve (Jolliffe, 2002).

Në punimin tonë, rezultatet (score) e komponentit të parë kryesorë përdoren si inpute (të dhëna) për analizën e grupeve. Këto rezultate kanë prirje matematike 0 dhe shmangie mesatare katrore 1.

Më pas në bazë të rezultateve (score) të komponentit të parë kryesorë realizohet grupimi i prefekturave. Këto rezultate (score) janë vlerësuar me anë të metodës së regresit.

Analiza e grupeve përdoret për të klasifikuar rastet në grupe që janë brenda homogjenë dhe heterogjenë midis njëri-tjetrit në bazë të një numri të caktuar ndryshoresh.

Metodat hierarkike të grupimit

Teknikat hierarkike të grupimit (të analizës së grupeve) rrjedhin nga një seri bashkimesh të njëpasnjshme të rasteve ose nga një seri ndarjesh të njëpasnjëshme. Këto teknika janë më të mirat për databaza të vogla sepse për to llogariten të gjitha distancat për secilin çift rastesh.

Një rëndësi të madhe për ndarjen grupeve ka ngjashmëria midis tyre. Dy raste quhen të ngjashëm kur ndryshimi apo distanca midis tyre është e vogël (Everitt *et.al* 2011). Lind nevoja për të përcaktuar matësit e ngjashmërisë. Konsiderata të rëndësishme përfshijnë natyrën e variablave (të vazhdueshëm, diskret, binar) dhe shkallët e matjes (emërore, rendore, intervalore dhe në formë raporti). Kur grupojmë rastet afërsia (ngjashmëria) tregohet zakonisht nga një distancë, kurse ndryshoret grupohen në bazë të koeficientëve të korrelacionit (Johnson & Wichern, 1982) Kur të gjithë variablat janë të vazhdueshëm për të grupuar rastet e ngjashme përdoren lloje të ndryshme distancash ku një ndër to është distanca Euklidiane.

Kur distanca euklidiane është e vogël, atëherë rastet janë më të ngjashme me njëri-tjetrin. Në analizën e grupeve distancat janë llogariur midis çifteve të rasteve të çdo ndryshore. Kjo distancë varet nga njësitë e matjes së ndryshoreve. Nëse ndryshoret janë matur në njësi të ndryshme ndryshoret me vlera të mëdha do të kenë ndikim (peshë) më të madhe tek distanca se ndryshoret me vlera të vogla. Prandaj është e rëndësishme që nëse ndryshoret janë matur në njësi të ndryshme të standardizohen para se të zbatojmë analizën e grupeve.

Në studimin tonë rastet janë 12 prefekturat e Shqipërisë për të cilat kemi marrë në konsideratë 9 ndryshore që përshkruajnë aspekte të ndryshme të varfërisë: niveli i varfërisë për secilën prefekturë (i shprehur në përqindje), përqindja e familjeve pa ujë dhe higjienë të përshtatëshme, strehim i papërshtatshëm (i shprehur në përqindje; strehim i papërshtatshëm quhet kush e vetëdeklaroi që nuk është banesë e përshtatëshme dhe kush banon në një ndërtesë që është akoma në ndërtim), mbipopullimi (përqindja e familjeve me më shumë se tre persona në një dhomë), shpenzimet e konsumit mujor (shpenzimet e konsumit mujor për familje), përqindja e familjeve që nuk kanë akses për ujë, produkti i përgjithshëm bruto, përqindja e familjeve pa asnjë akses në shërbimet bazë, numri i mjekëve (për 1 000 banorë).

Të dhënat i përkasin vitit 2012 dhe janë marrë nga INSTAT-i. Të dhënat janë përpunuar me anë të programit statistikor SPSS. Për shkak të faktit se madhësia e databazës së marrë në studim është e vogël është përdorur metoda hierarkike e analizës së grupeve. Në mënyrë që të përcaktonim metodën më të përshtatshme për këtë studim i kemi zbatuar të 7 teknikat hierarkike të disponueshme në SPSS. Rezultatet e përfuara nga dendrogrami tregojnë që metoda e Wardit i ka diferencuar më qartë dhe ka gjetur grupet më kompakte.

Metoda hierarkike sugjeron vetëm udhëzime shumë të kufizuara në lidhje me numrin e grupeve. Një rrugë e mundshme për të zgjidhur këtë problem është grupimi i rasteve që na sugjeron dendrogrami (Mooi, *et. al* 2011). Në punimin tonë nuk e përdorim dot metodën e grupimit të K- Mesatareve sepse madhësia e databazës është shumë e vogël. Zakonisht metoda e K-mesatareve përdoret kur numri i të dhënave është i madh, veçanërisht më i madh se 500.

Përkatësinë se cili qark i korrespondon një grupi të caktuar e kemi gjetur me ndihmën e programit SPSS (4 grupe në rastin tonë). Kemi vendosur këtë numër grupesh duke u bazuar tek dendrogrami. Përpara se të fillojmë të zbatojmë analizën e grupeve si fillim shqyrtojmë nëse ndryshoret kanë kolinearitete midis tyre. Praktikisht Everitt, (2011); Mooi & Sarstedt (2011) argumentojnë që vlerat absolute të korrelacioneve të cilat dalin më të mëdha se 0.9 janë problematike. Në studimin tonë nuk rezultoi asnjë korrelacion më i madh se 0.9. Prandaj do t'i marrim në shqyrtim të gjitha ndryshoret dhe nuk do të mënjanohet asnjë.

Rezultatet dhe diskutime

Tabela 1. Peshat e komponentit të parë kryesorë

	Komponenti 1
Pa asnjë akses	0.817
Nuk kanë akses për ujë	0.394
Produkti i përgjithshëm bruto	-0.575
Konsumi mujor	-0.731
Numri i mjekëve	-0.713
Niveli i varfërisë	0.641
Mbipopullim	0.528
Strehimi i papërshtatshëm	0.791
Ujë dhe higjienë e papërshtatshme	0.796

Tabela 1 na paraqet se si janë të korreluara ndryshoret me komponentin e parë kryesorë. Në fillim kemi bërë një analizë paraprake të ndryshoreve për të vrojtuar nëse ndryshoret i kanë peshat përkatësisht më të mëdha se 0.3. Vërejmë nga tabela 1 që të gjitha peshat e plotësojnë kriterin prandaj ato janë të rëndësishme në shpjegimin e komponentit të parë.

Ky i fundit interpretohet si ``komponenti i nivelit të varfërisë`` (kur themi komponenti i nivelit të varfërisë e nënkuptojmë si në nivelin e të ardhurave dhe të nevojave bazë të papërmbushura; si psh mbipopullimi etj). Shenjat e koeficientëve lidhen me karakteristikat e varfërisë.

Koeficientët pozitivë tregojnë një lidhje të drejtëpërdrejtë midis ndryshores dhe komponentit. Me rritjen e vlerave të ndryshores (kur shenja është pozitive) rriten dhe vlerat komponentit të varfërisë. Koeficientët negativë tregojnë lidhjen e kundërt (Henry *et.al* 2003).

Vlera e statistikës KMO(Kaiser-Meyer-Olkin) është 0.7 gjë që na tregon që ka korrelacion midis të dhënave.

Kriteri Bartlett kontrollon nëse matrica e korrelacioneve është matricë njësi. Niveli i rëndësisë për këtë kriter është 0.01 më i vogël se 0.05 prandaj në këtë rast e hedhim poshtë hipotezën e mësipërme.

Në fakt nuk ka një procedurë të saktë për përcaktimin e numrit të grupeve. Një mënyrë për të përcaktuar numrin e grupeve është që të shohim dendrogramin. (Timm, 2002) .

Dendrogrami tregon grafikisht distancat në të cilat qarqet janë bashkuar. Dendrogrami lexohet nga e majta në të djathtë; vijat vertikale janë prefekturat të cilat janë grupuar. Ky grafik na tregon në mënyrë të qartë që të dhënat tona mund të ndahen në 4 grupe.

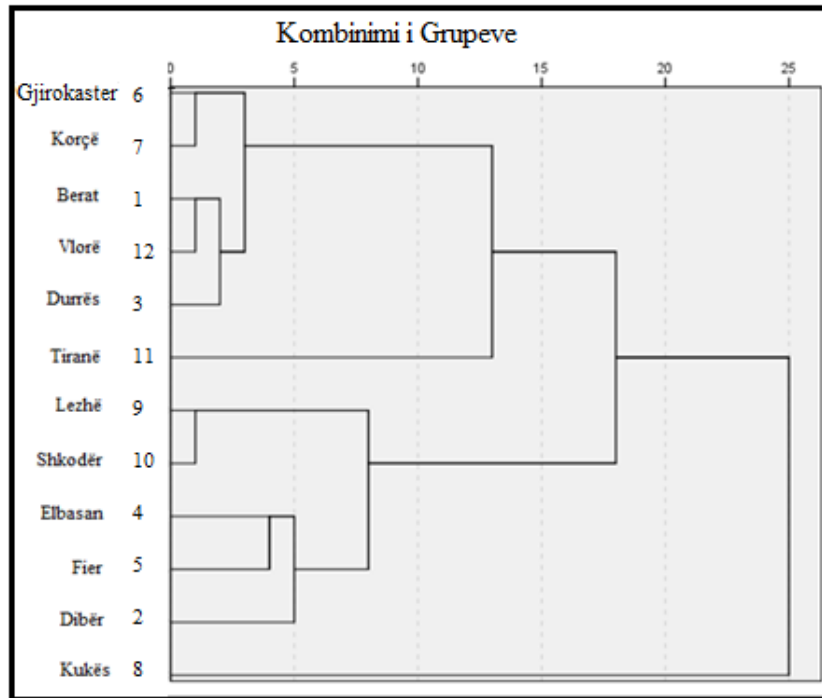


Figura 1. Dendrogram-i

Tabela2. Prefekturat që i përkasin çdo grupi

Prefekturat	Grupi	Prefekturat	Grupi
Berat	1	Korçë	1
Dibër	2	Kukës	3
Durrës	1	Lezhë	2
Elbasan	2	Shkodër	2
Fier	2	Tiranë	4
Gjirokastrër	1	Vlorë	1

Pasi kemi përfutur grupet përgjithsisht është e dobishme të përshkruajmë secilin grup duke përdorur disa mjete përshkruese për të kuptuar më mirë diferencat që ekzistojnë midis tyre. Në mënyrë që të karakterizojmë grupet është llogaritur mesatarja për secilin grup. Është paraqitur në tabelën 3.

Tabela 3. Mesatarja për secilin grup

Grupet	1	2	3	4
Ujë dhe higjienë e papërshtatshme	4.04	5.3	14.3	2
Strehimi i papërshtatshëm	7.82	7.84	14	3.3
Mbipopullim	5.4	6.9	8.4	4.7
Niveli i varfërisë	12.7	14.8	21.8	14.2
Numri i mjekëve	14	10.2	8.9	13.3
Konsumi mujor	47 029	46 571	41 000	53 349
Produkti përgjithshëm bruto	73 128 200 000	86 371 400 000	27 898 000 000	510 093 000 000
Pa akses për ujë	8.6	21.04	16	11.2
Pa asnjë lloj aksesi	1.01	1.58	5.3	0.55

Grupi shumë i varfër

Niveli më i lartë i varfërisë është regjistruar në grupin e tretë (në tabelë) që është krijuar nga prefektura e Kukësit e cila ka shkallën më të lartë të varfërisë nga të katërta grupet 21.8%, konsumin mujor më të vogël, përqindjen më të madhe të familjeve pa asnjë lloj aksesi, numrin më të vogël të mjekëve, produktin e përgjithshëm bruto më të vogël, përqindjen më të madhe të familjeve të mbipopulluara, përqindjen më të madhe të familjeve me strehim të papërshtatshëm. Prefektura e Kukësit është e prekur më shumë nga varfëria në të gjitha dimensionet e këtij fenomeni. Këta faktorë e klasifikojnë prefekturën e Kukësit si më të varfëren nga të gjitha prefekturat e Shqipërisë.

Grupi më i zhvilluar.

Grupi i katërt i cili përbëhet vetëm nga prefektura e Tiranës paraqet prefekturën me nivelin më të lartë të standardit të jetesës nga të gjitha prefekturat. Ka konsumin mujor (për familje) më të lartë nga të gjitha prefekturat. Përqindjen

më të ulët të familjeve me ujë dhe higjienë të papërshtatshme, përqindjen më të ulët të familjeve me strehim të papërshtatshëm, përqindjen më të ulët të familjeve të mbipopulluara, produktin e përgjithshëm bruto më të lartë dhe përqindjen më të ulët të familjeve pa asnjë lloj akses.

Grupi i varfër

Ky grup formohet nga prefekturat Elbasan, Fier, Dibër, Shkodër, Lezhë, të cilat konsiderohen jo shumë të zhvilluara (grupi 2 në tabelë).

Grupi me një intensitet të moderuar të varfërisë

Ky grup është i përbërë nga prefektura Durrës, Berat, Gjirokastrë, Korçë, Vlorë. Më poshtë po paraqesim hartën e varfërisë e ndërtuar në bazë të grupeve

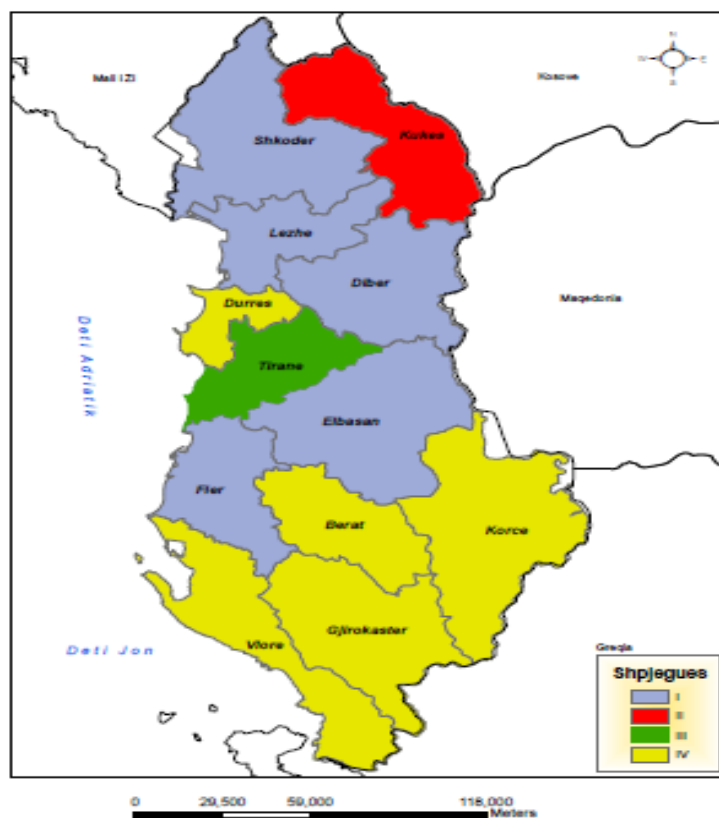


Figura 1. Harta e varfërisë në Shqipëri

Përfundime

Në këtë punim treguam se si metoda PCA përdoret në analizën e grupeve. Analiza e grupeve përdoret për të klasifikuar rastet në grupe që janë brenda homogjenë dhe heterogjenë midis njëri tjetrit në bazë të një grupi të caktuar ndryshorësh. Në artikullin tonë është përdorur metoda hierarkike e cila është e përshtatshme kur vëllimi i zgjedhjes është më i vogël se 500. Rezultoi që prefekturat ndahen në 4 grupe ku në njërin nga grupet futen prefekturat e varfëra Elbasan, Fier, Dibër, Shkodër, Lezhë, një tjetër grup përcaktohet nga prefekturat me një intesitet të moderuar të varfërisë të cilat janë Durrës, Gjirokastrë, Vlorë, Korçë, Berat një grup të veçantë formon vetëm prefektura e Kukësit e cila interpretohet si një prefekturë shumë e varfër dhe grupi i fundit përbëhet vetëm nga prefektura e Tiranës e cila interpretohet si prefektura më e zhvilluar.

Literatura

Instituti i Statistikave të Shqipërisë: www.instat.gov.al

Mooi E, Sarstedt M. (2011): A Concise Guide to Market Research ,pp 241-242; 254-255 ISBN 978-3-642-12540-9, Springer Heidelberg Dordrecht London New York

Johnson R, Wichern D. (1982): Applied Multivariate Statistical Analysis; 345-355 ISBN 0-13-187715-1

Everitt B, Landau S, Leese M, Stahl D.(2011): Cluster Analysis 4-9 ISBN 978-0-470-74991

Rencher A. (2002): Methods of Multivariate Analysis 467 -468 ISBN 0-471-41889-7

Landau S, Brian E. (2004): A Handbook of Statistical Analyses using SPSS; 323-327 ISBN 1-58488-369-3

Vázquez S, Sumner A. (2012): Beyond Low and Middle Income Countries: What if There Were Five Clusters of Developing Countries? Ids Working paper Volume 2012 No 404

Anderson E. (2007): Identifying chronically deprived countries: results from cluster analysis, CPRC Working Paper 70, ISBN 1-904049-69-9

Timm N. (2002): Applied multivariate analysis; 531, ISBN 0-387-95347-7, Springer-Verlag New York

Jolliffe. T (2002): Principal Components Analysis; 230-240, Second Edition, ISBN 0-387-95442-2, Springer-Verlag New York

Vyas S, Kumanarayake L. (2006): Constructing social-economic status indices, How to use PCA. Health Policy and Planning, ISSN 0268-1080, Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine.

Henry C, Sharma M, Lapenu C, Zeller M. (2003): The International Bank for Reconstruction and Development, ISBN 0-8213—5588-0; 150-160