



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Tarptautinės migracijos srautams įtaką darančių veiksnių analizė

Baigiamasis magistro projektas

Gintarė Pamarnackaitė
Projekto autorė

prof. dr. Vilmantė Kumpikaitė-Valiūnienė
Vadovė
doc. dr. Tomas Ruzgas
Vadovas

Kaunas, 2018



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Tarptautinės migracijos srautams įtaką darančių veiksnių analizė

Baigiamasis magistro projektas

Didžiųjų verslo duomenų analitika (621G12002)

Gintarė Pamarnackaitė

Projekto autorė

prof. dr. Vilmantė Kumpikaitė-Valiūnienė

Vadovė

doc. dr. Tomas Ruzgas

Vadovas

doc. dr. Vytautas Janilionis

Recenzentas

dr. Ineta Žičkutė

Recenzentas

Kaunas, 2018



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Gintarė Pamarnackaitė

Tarptautinės migracijos srautams įtaką darančių veiksnių analizė

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Gintarės Pamarnackaitės, baigiamasis projektas tema „Tarptautinės migracijos srautams įtaką darančių veiksnių analizė“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Turinys

Ižanga.....	10
1. Tarptautinei migracijai įtaką darančių veiksnių teorinė analizė.....	12
1.1. Migracijos samprata	12
1.2. Migracijos tipai	12
1.3. Tarptautinę migraciją aiškinančių teorijų analizė	13
1.4. Tarptautinei migracijai įtaką darantys veiksniai ir jų klasifikacijos	16
1.5. Migracijai įtaką darančių veiksnių tyrimai ir studijose naudojami metodai	23
1.6. Duomenų analizės programinės įrangos apžvalga	25
1.7. Literatūros apžvalgos apibendrinimas.....	27
2. Tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo metodinė dalis	28
2.1. Duomenų matricos formavimas	28
2.2. Tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo analizė.....	29
2.2.1. Duomenų apžvalga.....	29
2.2.2. Homogeninių grupių išskyrimas	29
2.2.3. Statistiškai reikšmingų veiksnių identifikavimas	33
2.2.4. Kintamųjų dimensijos mažinimas	35
2.2.5. Prognostinio modelio sudarymas remiantis kiekybiniais duomenimis	38
3. Tyrimų rezultatai ir jų aptarimas.....	43
3.1. Duomenų matricos formavimas ir žvalgomoji analizė Europos sąjungos šalių pavyzdžiu	43
3.2. Europos Sąjungos šalių grupavimas į homogenines grupes	44
3.2.1. Grupavimas pagal BVP rodiklį vienam gyventojui	45
3.2.2. Grupavimas pagal migracijos srautus ir jų pokytį.....	45
3.2.3. K-vidurkių klasterizacija	47
3.3. Tarptautinės migracijos srautams įtaką darančių veiksnių koreliacinė analizė	49
3.4. Tarptautinės migracijos srautams įtaką darančių veiksnių pagrindinių komponentų analizė	53

3.5. Regresinė analizė	57
3.5.1. Apibendrinti adityvūs modeliai	57
3.5.2. Branduolinė regresinė analizė	60
3.5.3. Regresinės analizės metodų palyginimas	65
Išvados	66
Literatūra	68
Priedai.....	73

Lentelių sąrašas

1.1 lentelė. Stūmos ir traukos veiksniai	18
1.2 lentelė. Susisteminti ekonominės ir socialinės gerovės veiksniai	22
3.1 lentelė. Migracijos srautų apžvalga.....	44
3.2 lentelė. Skurdžių ir turtingų ES šalių išskyrimas pagal BVP.....	45
3.3 lentelė. Migracijos srautų pokyčių lentelė	46
3.4 lentelė. Optimalus ES šalių grupavimas pagal migracijos srautus ir BVP	49
3.5 lentelė. Tarptautinės migracijos srautams statistiškai reikšmingų veiksmių koreliacijos neturtingų šalių grupei	51
3.6 lentelė. Tarptautinės migracijos srautams statistiškai reikšmingų veiksmių koreliacijos turtingų šalių grupei	52
3.7 lentelė. Kaiserio-Meyerio-Olkino reikšmių lentelė neturtingų šalių grupei	54
3.8 lentelė. Bartleto sferiškumo kriterijus ir KMO matas neturtingų šalių grupei.....	55
3.9 lentelė. Pagrindinių komponentų sklaidos lentelė	55
3.10 lentelė. Pagrindinių komponentų tikrinių reikšmių lentelė	56
3.11 lentelė. Kaiserio-Meyerio-Olkino reikšmių lentelė turtingų šalių grupei	57
3.12 lentelė. Bartleto sferiškumo kriterijus ir KMO matas turtingų šalių grupei	57
3.13 lentelė. Apibendrintų adityvių modelių regresija neturtingų šalių grupei.....	58
3.14 lentelė. Apibendrintų adityvių modelių reikšmingumo testas neturtingų šalių grupei	58
3.15 lentelė. Apibendrintų adityvių modelių paklaidos neturtingų šalių grupei	59
3.16 lentelė. Apibendrintų adityvių modelių regresija turtingų šalių grupei	59
3.17 lentelė. Apibendrintų adityvių modelių reikšmingumo testas turtingų šalių grupei	60
3.18 lentelė. Apibendrintų adityvių modelių paklaidos turtingų šalių grupei.....	60
3.19 lentelė. Branduolinės regresijos reikšmingumo testas neturtingų šalių grupei	61
3.20 lentelė. Branduolinė regresija neturtingų šalių grupei	62
3.21 lentelė. Branduolinės regresijos modelių paklaidos neturtingų šalių grupei.....	62
3.22 lentelė. Branduolinės regresijos reikšmingumo turtingų šalių grupei.....	63
3.23 lentelė. Branduolinės regresija turtingų šalių grupei.....	64
3.24 lentelė. Branduolinės regresijos modelių paklaidos turtingų šalių grupei	64
3.25 lentelė. Regresinės analizės modelių paklaidų ir tikslumo palyginimas.....	65

Paveikslų sąrašas

1.1 pav. Darbo jėgos migracija remiantis migracijos teorijomis (Stulgienė ir Daunorienė, 2009 m., p. 988).....	16
1.2 pav. E. S. Lee migracijos modelis.....	18
1.3 pav. Bendrasis emigracijai įtaką darančių veiksnių modelis su juos apibūdinančiais rodikliais (Kumpikaitė ir Žičkutė, 2012 m., p. 744).....	20
2.1 pav. Klasterinės analizės metodai	32
2.2 pav. K-vidurkių klasterizavimas	32
2.3 pav. Pirmoji ir antroji pagrindinės komponentės	36
3.1 pav. BVP vienam gyventojui ir Y (migracijos neto) pasiskirstymas	47
3.2 pav. Optimalus klasterių skaičius – alkūnės metodas	48
3.3 pav. Optimalus klasterių skaičius – Silueto kriterijus	48

Pamarnackaitė, Gintarė. Tarptautinės migracijos srautams įtaką darančių veiksnių analizė. Magistro baigiamasis projektas / vadovai prof. dr. Vilmantė Kumpikaitė-Valiūnienė, doc. dr. Tomas Ruzgas; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Taikomoji matematika (A02), Matematikos mokslai (A). Reikšminiai žodžiai: *aprašomoji statistika, klasterinė analizė, koreliacinė analizė, pagrindinių komponentų analizė, neparamestriniai metodai, RStudio, tarptautinė migracija, migracijos srautai, traukos – stūmos veiksniai.*

Kaunas, 2018. 91 p.

Santrauka

Baigiamajame magistro darbe sprendžiamas aktualus šiuolaikiniam pasauliui uždavinys – analizuojami veiksniai, turintys įtakos migracijos srautams. Kasmet migracijos apimtys auga, o ši reiškinį ypač didina pabėgėlių srautai iš Artimųjų Rytų ir Afrikos. Vokietijos vidaus reikalų ministro T. de Maiziere teigimu (2018 m.) „migracija ir pabėgėlių srautai bus vienas iš pagrindinių pasaulio bendruomenės iššūkių artimiausiais dešimtmečiais“. Pirmoje darbo dalyje aptariami mokslinėje literatūroje rasti tyrimai migracijos srautams ir veiksniams analizuoti. Aptartos tyrimuose naudotos programinės priemonės ir analizės metodai. Antroje dalyje pateikta metodinė informacija, skirta tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo analizei atlikti. Aprašytas duomenų matricos paruošimas, sukurti valstybių grupavimo į homogenines grupes, statistiškai reikšmingų veiksnių, darančių įtaką migracijai, identifikavimo, neparamestrinės regresinės analizės modeliai. Pateiktas originalus, naujas, literatūros analizės metu nerastas, modelis migracijos srautams įtaką darančių veiksnių identifikavimui, jų analizei ir migracijos srautų prognozei sudaryti. Trečioje dalyje, panaudojus sukurtą modelį, atlikta Europos Sąjungos valstybių migracijos srautų analizė ir pateiktos jos išvalgos. Nustatyta, kad tirtos 28 valstybės pagal migracijos srautus ir bendrą vidaus produktą vienam gyventojui gali būti klasterizuojamos į 3 grupes. Skurdžiausiai kategorijai priklauso daugiau nei pusė šių valstybių. Viena iš visų valstybių išsiskiria ypač dideliu imigrantų skaičiumi ir bendru vidaus produktu vienam gyventojui. Ekonominės gerovės veiksniai, vertinant migracijos srautus, aktualūs tik neturtingų šalių grupei. Sukurti migracijos srautų kiekybinio vertinimo prognostiniai modeliai, neturtingų šalių grupei išskirti 34, turtingų šalių grupei – keturi statistiškai reikšmingi veiksniai. Šiais modeliais, atitinkamai, galima paaiškinti 81,12 proc. ir 65,79 proc. duomenų sklaidos.

Pamarnackaitė, Gintarė. Analysis of Factors Influencing International Migration Flows. Master's Final Degree Project / supervisor prof. dr. V. Kumpikaitė-Valiūnienė, doc. dr. Tomas Ruzgas;

The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics (A02), Mathematical Sciences (A).

Keywords: *descriptive analysis, cluster analysis, correlation analysis, principal component analysis, nonparametric regression, RStudio, international migration, migration flows, "push" and "pull" factors.*

Kaunas, 2018. 91 pages.

Summary

The final master thesis deals with an issue that is relevant to a modern society – factors affecting international migration flows. Migration flows are growing every year and this phenomenon is even more magnified by refugee flows from the Middle East and Africa. According to Federal Minister of the Interior T. de Maiziere (2018): “migration and refugee flows will be one of the main challenges for our society in the coming decades”. The first part of this thesis discusses scientific researches on analysis of migration flows and influential factors. Software tools and analytical measures employed for the research are defined. The second part of this thesis contains methodological information for analysing factors affecting international migration flows. Preparation of data matrix is described and models for organizing countries into homogenous groups, identifying statistically important factors affecting migration flows and nonparametric regression analysis are developed. A new and original model for identifying and analysing factors affecting migration flows and forecasting migration flows is created. In the third part of the thesis, using the developed model, the analysis of migration flows in the EU member states is performed and insights are presented. The study revealed that 28 member states, studied in terms of migration flows and GDP per capita, can be clustered into 3 groups. More than half of these countries are classified to the category of “The poorest countries”. One of the member states is distinguished by a particularly high number of immigrants and GDP per capita. Factors of economic welfare are only relevant for the poorest countries. Developed prognostic models for quantitative assessment of migration flows distinguish 34 significant factors for the group of poor countries and 4 statistically significant factors for the group of rich countries. These models explain data dissemination respectively by 81.12% and 65.79%.

Ižanga

Migracija – tai viena iš nedaugelio tarpdisciplininių socialinių procesų sričių, turinti tiek teigiamą, tiek neigiamą įtaką asmenims ir visuomenei. Ji gali būti laikoma demografiniu, ekonominiu, etnologiniu, geografiniu, politiniu, psichologiniu ar socialiniu procesu. Migracijos aktualumą įrodo daugiau nei 50 metų egzistuojantis faktas, kuris dar 2009 metais buvo paskelbtas „Human Development Report“ ataskaitoje: „pasaulio migrantai kasmet sudaro apie 3 proc. visos žmonijos populiacijos.“ Didėjant žmonių skaičiui, proporcingai didėja ir migrantų skaičius. Įprasta manyti, kad daugiausia migruoja skurdžių šalių gyventojai į pasiturinčias valstybes. Tačiau tyrimai atskleidžia, kad tokių migrantų srautai yra gerokai mažesni nei migrantų, keliaujančių iš neturtingų šalių į neturtingas ar iš turtingų šalių į turtingas. Mažiausiai mobilūs ir linkę keisti savo gyvenamąją vietą yra skurdžiausių šalių gyventojai (mažiau nei 1 proc. Afrikos žemyno gyventojų migravo į Europos Sąjungos valstybes) (1).

Baigiamajame magistro darbe nagrinėjama aktuali šių dienų problema – migracija. Migracijos srautai kiekvienais metais auga. Be to, šią problemą didina pabėgėliai iš Artimųjų Rytų ir Afrikos, o ir ateityje daugės migrantų – pabėgėlių ir dėl geografinių priežasčių (tokių kaip klimato kaita, apsemti regionai ir pan.), karų, persekiojimų, geresnės ateities lūkesčių. Vokietijos vidaus reikalų ministro T. de Maiziere teigimu (2018 m.) „migracija ir pabėgėlių srautai bus vienas iš pagrindinių pasaulio bendruomenės iššūkių artimiausiais dešimtmečiais“.

Migracijos priežastys analizuojamos daugelį metų, tačiau tyrimai dažniausiai apsiriboja tik statistinių duomenų analize, pasiremiant keliais rodikliais (Damulienė (8)), o atskirais atvejais – panelinių duomenų analize (Ramirez‘as ir kt. (9), Mayda (29)) ir kiekybiniais tyrimais (Malhotra ir Devi‘s (21)). Paprastai siekiama nustatyti priežastis, kokie veiksniai lemia migraciją, tačiau gautų rezultatų nebandoma pritaikyti prognozei. Tai lemia naujų, išsamesnių tyrimų poreikį, taikant sudėtingesnius analitinius metodus.

Objektas – migracijos srautams įtaką turintys veiksniai.

Tikslas – suformuoti migracijos srautų prognozavimo modelį, remiantis tarptautinės migracijos srautams įtaką darančių veiksnių analize Europos Sąjungos šalyse.

Uždaviniai:

1. Atskleisti migracijos srautams įtaką darančius veiksniai ir jų analizei taikomus tyrimo metodus teoriniame lygmenyje.

2. Išrinkti rodiklius pagal veiksnius, galinčius turėti įtakos migracijos srautams, ir paruošti šių duomenų rinkinį analizei.
3. Sudaryti tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo modelį ir jį realizuoti programiškai.
4. Panaudojant sudarytą modelį ir programines priemones nustatyti statistiškai reikšmingus veiksnius, turinčius įtakos tarptautinės migracijos srautams Europos Sąjungos šalyse.
5. Remiantis tarptautinės migracijos srautams įtaką darančiais veiksniais Europos sąjungos šalyse, sudaryti migracijos srautų prognostinius modelius ir juos palyginti.

Darbe atlikta mokslinės literatūros analizė, išanalizuota programinė įranga ir matematiniai metodai, taikyti panašių uždavinių sprendimui. Iškelto tikslo sprendimui parinkti klasterinės, koreliacinės, pagrindinių komponentų ir nelinearės regresinės analizės metodai, skirti originaliam kiekybinių duomenų modelio kūrimui. Migracijos srautų analizei taikyti metodai realizuoti programiškai. Naudojant sudarytus modelius ir Europos statistikos departamento bei Pasaulio banko (angl. *World bank data*) duomenis, atlikta Europos Sąjungos valstybių migracijos srautų analizė.

1. Tarptautinei migracijai įtaką darančių veiksnių teorinė analizė

1.1. Migracijos samprata

Migracija (lot. *migratio* – kėlimasis, kraustymasis) – ilgalaikis gyvenamosios vietos pakeitimas, kurio rezultatas – gyventojų skaičiaus ir struktūros pasikeitimas išvykimo bei atvykimo vietovėse (Kasnauskienė, 2006). Migracija – gyventojų persikėlimas iš vienos gyvenamosios vietos į kitą visam ar tam tikram laikui, kertant administracines vietovės ribas (1). „Lietuvių kalbos žodynas“ tą patį terminą aiškina kaip žmonių gyvenamosios vietos pasikeitimą dėl besikeičiančių ekonominių, socialinių, politinių, aplinkos (angl. *environmental*) priežasčių (2). Migracija gali būti tiek nuolatinė, tiek laikina. Ekonominę migraciją galima apibūdinti kaip žmonių judėjimą ieškant darbo ar karjeros perspektyvų, socialinę – judėjimą ten, kur geresnė gyvenimo kokybė ar ten, kur gyvena artimieji, politinę – kaip persikėlimą dėl politinio persekiojimo ar karo baimės, aplinkos – išvykimą dėl netenkinamų klimato sąlygų, gamtinių katastrofų ir nelaimių. Gyvenamosios vietos keitimas daro įtaką visai populiacijai, kadangi žmonėms migruojant, gyventojų skaičius arba mažėja, arba didėja. Nuo to, kokios žmonės išvyksta ar įsikelia į šalį, priklauso ir kiti pokyčiai valstybės viduje (specialistų trūkumas ar perviršis ir pan.). Žmonių pasirinkimas keisti gyvenamąją vietą yra skirstomas į dvi kategorijas: vidinę (gyventojų judėjimas valstybės viduje) bei išorinę, kitaip dar tarptautinę (gyventojų judėjimas tarp valstybių), migraciją.

Terminas migracija jungia du glaudžiai susijusius terminus: imigracija (lot. *immigro* – apsigyvenu, įsikuriu) ir emigracija (lot. *emigratio* – išsikėlimas) (3). Valstybės kontrolės institucija imigracijos terminą apibrėžia kaip asmenų, išskyrus asmenis, kurie naudojami laisvo asmenų judėjimo teise, atvykimą gyventi į užsienio valstybę dėl įvairių priežasčių (dirbti, užsiimti verslu, būti su šeima, ieškoti prieglobsčio ir kt.) ir asmenų gyvenimą užsienio valstybės teritorijoje. Šis apibūdinimas tinkamas ir emigracijos apibrėžimui. Imigracija – atvykimas į šalį, emigracija – išvykimas iš šalies (4). Norint vertinti migracijos srautus, imigracija ir emigracija yra sujungiamos ir tiriamos kaip vienas matas – grynoji migracija (migracijos saldo). Remiantis grynąja migracija, galima įvertinti šalies padėtį: jei šalis turi aukštą migracijos normą, ji yra turtinga ir išsivysčiusi, ir priešingai – jei šalis turi žemą normą, joje piliečiai nesijaučia saugūs ir patenkinti.

1.2. Migracijos tipai

Migracija – sudėtingas procesas, apimantis žmones, keičiančius savo gyvenamąją vietą dėl skirtingų priežasčių. Migrantu gali būti žmogus, kuris savo noru persikrausto gyventi į kitą miestą ar šalį, taip pat

migrantu gali būti ir asmuo, kuris perkeliamas gyventi į kitą vietą prieš savo valią. Svarbu atkreipti dėmesį į tai, kad judėjimas toje pačioje srityje yra laikomas mobilumu, o ne migracija, kadangi judėjimas gali trukti visą gyvenimą (kasdienė rutina – darbas, mokymo įstaigos, apsipirkimas parduotuvėse ir pan.) ir nesukelti jokių didesnių nepatogumų. Įvairūs autoriai išskiria skirtingą kiekį migracijos tipų. Vienas plačiausių, paprasčiausių ir labiausiai paplitusių yra migracijos tipų grupavimas į dvi kategorijas (5, 6):

1. Vidinė migracija (gyventojų judėjimas valstybės viduje).
2. Išorinė, kitaip dar tarptautinė migracija (gyventojų judėjimas tarp valstybių).

Tarptautiniai migrantai toliau klasifikuojami:

- į legalius imigrantus, kuriais laikomi asmenys, turintys valstybės, į kurią atvyko, išduotą leidimą dirbti ir gyventi,
- nelegalius imigrantus, kurie persikėlė be leidimo,
- pabėgėlius, kurie kirtę tarptautinę sieną dėl savo kilmės valstybėje patirto persekiojimo ar neramumų.

J. Weinstein'as ir V. Pillai 2001 metais išskyrė dar vieną migracijos grupę – priverstinę migracija. Ji egzistuoja, jei asmuo perkeliamas gyventi prieš savo valią ar yra veikiamas išorinių veiksnių (stichinės nelaimės, karas ir pan.) (5).

Tačiau egzistuoja ir kur kas detalesni migracijos tipų sąrašai. Autorių kolektyvas Jančaitytė, Valavičienė, Augutienė ir Prakapas įvardija 6 tipų grupes, pagal kurias migraciją galima skirstyti:

- **pagal legalumą**: legali, nelegali;
- **pagal trukmę**: ilgalaikė, trumpalaikė;
- **pagal srautų kryptis**: išvyksta, atvyksta;
- **pagal kertamas sienas**: vidinė, išorinė;
- **pagal priežastis**: dėl darbo, išsilavinimo, šeimos aplinkybių;
- **pagal noro išraišką**: savanoriška, priverstinė (1).

Šiame darbe toliau bus analizuojami išorinės (tarptautinės) legalios ilgalaikės migracijos srautai.

1.3. Tarptautinę migraciją aiškinančių teorijų analizė

Šiame poskyryje aptariamos pagrindinės migracijos procesą aiškinančios teorijos ir jų tarpusavio ryšys.

Kadangi migracija yra kompleksinis procesas, todėl vienos teorijos, paaiškinančios šį reiškinį nėra. Stulgienė ir Daunorienė (7), Damulienė (8), Ramirez (9), Kumpikaitė-Valiūnienė ir Žičkutė (10) Čiarnienė, Kumpikaitė ir Taraškevičius (11), Žibas (12) savo darbuose išskiria tokias pagrindines migracijos teorijas:

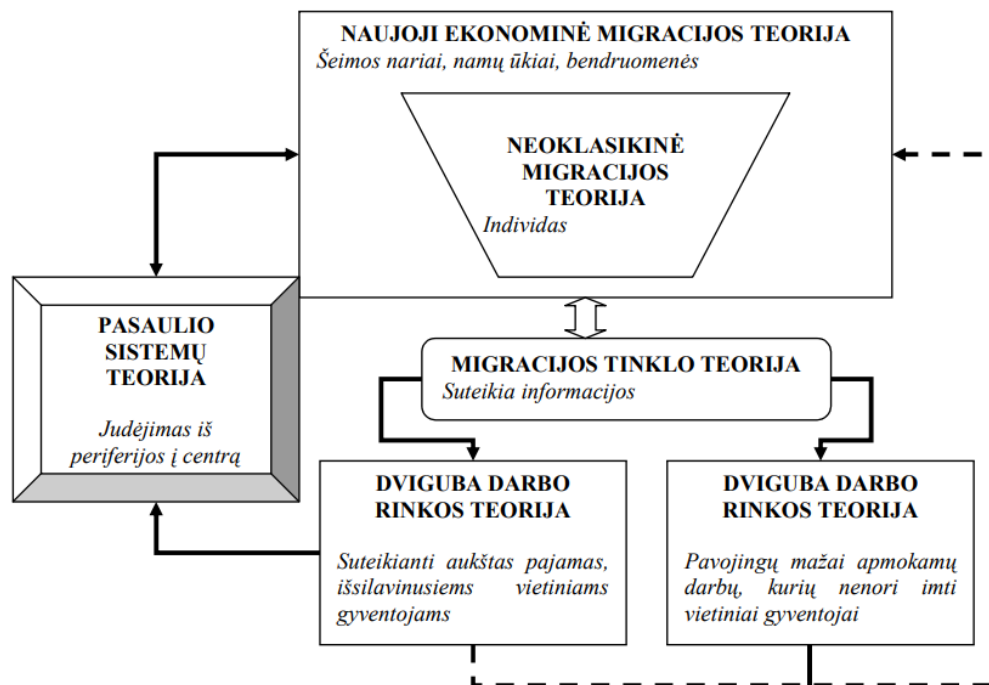
- **neoklasikinė ekonomikos teorija.** Migracija aiškinama kaip dviejų veikiančių jėgų, „stūmimo“ ir „traukimo“, sąveika. Mikrolygiu ši teorija teigia, kad sprendimas dėl migracijos priklauso nuo kiekvieno individualaus apsisprendimo. Gyventojai migruoja atlikę galimos naudos ir kaštų analizę. Makrolygiu tiriami veiksniais, lemiančiais žmogaus pasirinkimą migruoti, įvardijami darbo jėgos paklausos ir pasiūlos skirtumai tarptautinėse rinkose, o pagrindine priežastimi – skirtingas darbo užmokestis skirtingose šalyse;
- **naujoji ekonominė migracijos teorija.** Kaip ir neoklasikinė teorija, nagrinėja migracijos naudą. Tačiau naujosios ekonominės migracijos teorija labiau koncentruojasi į šeimos ir bendruomenės priežastis dėl migracijos, kai skirtumo dėl darbo užmokesčio tarp šalių nėra;
- **dviejų darbo rinkų teorija.** Migracijos procesas aiškinamas ne iš emigranto pusės, o iš šalies, priimančios migrantą. Tigiama, jog valstybėje egzistuoja dvi darbo rinkos: viena – gerai apmokamų ir išsilavinusiems vietiniams gyventojams skirtų darbų rinka, kita – mažai apmokamų, o dažnai ir fiziškai sunkių ir pavojingų, darbų rinka, kuri nėra patraukli vietiniams gyventojams ir dėl šios priežasties dažnai užpildoma imigrantų;
- **migracijos tinklo teorija.** Teorija sako, kad migracija yra pats save skatinantis veiksnys. Emigrantai skleidžia informaciją apie darbo rinkos padėtį šalyje, įsidarbinimo galimybes, darbo užmokesčio normas, gyvenimo galimybes ir pan. ir tuo prisideda prie didėjančio migracijos srauto;
- **migracijos sistemų teorija.** Migracija tai dviejų teorijų mikro- ir makrostruktūrų tarpusavio ryšių rezultatas. Mikrostruktūromis laikomi migrantų įsitikinimai ir patirtys, makrostruktūromis – instituciniai faktoriai;
- **pasaulio sistemų teorija.** Migracija vyksta dėl darbo jėgos judėjimo – iš periferinių regionų į centrinius (būdinga tiek šalies viduje, tiek tarptautiniu mastu) (7, 8, 9, 10, 11 12).

Kiti šaltiniai taip pat išskiria dar vieną – **elgesio teorija**. Ji apibūdina, kaip žmogaus psichologija skatina ekonomiką ir sukelia ekonomines problemas, kurių pačios rinkos negali išspręsti.

1.1 pav. pateiktame A. Stulgienės ir A. Daunorienės (2009 m.) sudarytame paveiksle apibendrintos aptartos migracijos teorijos ir atskleistas jų tarpusavio sąryšis (7). Parodyta, jog naujoji ekonominė migracijos teorija apima ir neoklasikinę migracijos teoriją. Naujoji ekonominė teorija yra platesnė ir migraciją analizuoja vertindama naudą šeimos ir bendruomenės lygiu, tuo tarpu neoklasikinė teorija, naujosios ekonominės migracijos teorijos poaibis, yra paremta pačiu individu ir jo poreikiais. Abi šios teorijos dvipusiu ryšiu yra susijusios su migracijos tinklo teorija, kuri apibūdina migranto santykius su jo šeima ir šalyje likusiais artimaisiais. Migrantas padeda likusiems lengviau įsidarbinti kitoje rinkoje, suteikia daugiau naudingos informacijos. Minėtas dvipusis ryšys atskleidžia, kad potencialus migrantas, bendraudamas su informacijos teikėju, ieško naudingos informacijos, o teikėjas – potencialių migrantų rinkos poreikiams patenkinti.

Naujoji ekonominė, neoklasikinė ir migracijos tinklo teorijos kartu yra susijusios su dviguba darbo rinkos teorija. Vienoje rinkoje stengiamasi surasti aukštą išsilavinimą turinčius specialistus, kurie viliojami aukštomis pajamomis ir geromis perspektyvomis. Taip siekiama užpildyti rinkos trūkumus, kurių neužpildo vietiniai gyventojai. Taip pat aukštų pajamų rinka turi ryšį ir su pasaulio sistemų teorija, kadangi aukšto lygio specialistai turi dideles galimybes kelti savo kvalifikaciją, gerinti materialinę padėtį ir migruoti į kitas aukšto išsilavinimo specialistų norinčias šalis. Antroje rinkoje siekiama rasti individų, užpildysiančių pavojingų ir nepatrauklių darbų rinką, kurioje dažniausiai nenori dirbti vietiniai gyventojai. Ši rinka turi ryšį su naująja ekonomine ir neoklasikine migracijos teorija. Pavojingus ir mažai apmokamus darbus dirbę migrantai grįžta į savo gimtinę patenkinę poreikius ir vėl tampa potencialiais migrantais.

Naujoji ekonominė ir neoklasikinė teorijos taip pat turi ryšį su pasaulio sistemų teorija. Išsivystę šalies centrai bei ekonomiškai išsivysčiusios valstybės yra migrantų traukos centrai, sudarantys galimybes rasti geriau apmokamus darbus, patrauklesnes socialines garantijas ir pan.



1.1 pav. Darbo jėgos migracija remiantis migracijos teorijomis (Stulgienė ir Daunorienė, 2009 m., p. 988)

1.4. Tarptautinei migracijai įtaką darantys veiksniai ir jų klasifikacijos

Žmonių migracija – tai vienas iš nedaugelio tarpdisciplininių socialinių procesų, turinčių didelę įtaką tiek individualiems asmenims, tiek visuomenei. Ji gali būti laikoma demografiniu, politiniu, etnologiniu, geografiniu, psichologiniu ar socialiniu procesu. Kiekviena ši sritis yra tiriama atskirai ir turi savo individualų požiūrį į migracijos procesą. Pavyzdžiui, ekonomika migraciją vertina per jos įtaką šalies, regiono ar individo gerovei, demografija – per gyventojų dinamikos pokyčius. Vienas pirmųjų analitikų, bandžusių formalizuoti migracijos procesų tyrimą, buvo anglų geografas E. G. Ravenstein‘as. Jis 1885 metais paskelbė septynis „migracijos įstatymus“, kurie turėjo paaiškinti ir padėti prognozuoti migracijos modelius tiek tautose, tiek tarp jų (14).

E. G. Ravenstein‘o „migracijos įstatymai“:

1. Dauguma migrantų persikelia tik nedideliu atstumu nuo ankstesnės gyvenamosios vietos.
2. Vyksta absorbcijos procesas – žmonės, gyvenantys šalia greitai augančio miesto, persikelia į jį, o jų paliktas vietas užpildo gyventojai iš tolimesnių vietovių, ir taip toliau, kol veikia traukos (stūmos) dėsnis.
3. Dispersijos procesas yra priešingas absorbcijai ir veikia panašiu principu.

4. Kiekvienas migracijos srautas sukelia kompensuojantį priešpriešinį srautą. Taigi, grynoji migracija iš regiono A į regioną B visada bus mažesnė nei bendroji migracija tarp A ir B regionų.
5. Didesnį atstumą įveikiantys migrantai dažniausiai traukia į didelius komercinius ir pramoninius centrus.
6. Didmiesčių ir miestų gyventojai yra mažiau linkę migruoti nei provincijos gyventojai. Dėl to, vykstant pramonės, prekybos ir transportavimo plėtrai, didėja ir migracijos srautai.
7. Moterys linkusios migruoti dažniau nei vyrai.

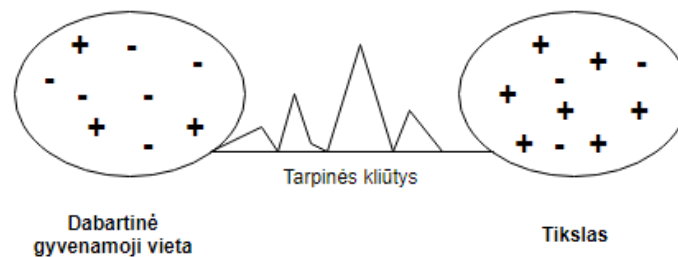
Vėliau šis sąrašas buvo praplėstas ir papildytas dar keliais punktais, kurie teigia, kad ekonominiai faktoriai yra pagrindinė migracijos priežastis.

Žmonės migruoja dėl įvairių priežasčių, tačiau dažniausiai tai lemia noras pagerinti savo kasdienį gyvenimą bei suteikti geresnių galimybių sau ir savo šeimai. Galima sakyti, jog žmogus bus pasirengęs migruoti, jei iš to gaunama nauda bus didesnė, nei patiriami migracijos kaštai (8). Analizuojant migracijos reiškinį, išskiriami įvairūs faktoriai, kurie ją skatina. Veiksnius, lemiančius asmenų judėjimą tarp valstybių, 1966 metais savo darbe „Migracijos teorija“ išskyrė ir sujungė profesorius E. S. Lee. Jis migracijos veiksnius sugrupavo į dvi kategorijas, kurios yra ypač populiarios ir šiomis dienomis: veiksniai, kurie skatina emigraciją („stūmos“ veiksniai) ir veiksniai, kurie skatina imigraciją („traukos“ veiksniai). Nesant papildomų faktorių įtakos, žmogus priima sprendimą migruoti dėl dvejų vietų pranašumų ir trūkumų. Autorius taip pat pastebėjo, kad skirtingi stūmos ir traukos veiksniai gali būti ypač individualūs ir subjektyvūs skirtingiems žmonėms. Pavyzdžiui, geros klimato sąlygos yra patrauklios daugumai gyventojų, blogos – nepatrauklios. Tačiau gera švietimo sistema yra patraukli jaunoms šeimoms, turinčioms vaikų, ir, priešingai, – ji nėra aktuali šeimai, neturinčiai mokyklinio amžiaus vaikų. 1.1 lentelėje pateikti autorės išskirti stūmos ir traukos veiksnių pavyzdžiai, kurie tinkami daugumai gyventojų (1, 12, 13). Visi demografiniai, socialiniai, politiniai, ekonominiai ir aplinkos rodikliai gali būti migracijos priežastimis. Iš pateiktos lentelės nesudėtinga pastebėti, jog vienos šalies trūkumai (stūmos veiksniai) yra kitos šalies privalumai (traukos veiksniai).

1.1 lentelė. Stūmos ir traukos veiksniai

Stūmos veiksniai	Traukos veiksniai
Nedarbas	Didelė darbo rinka
Žemas socialinis ir ekonominis išsivystymo lygis	Aukštas socialinis ir ekonominis išsivystymo lygis
Žemas darbo užmokesčio lygis šalyje	Aukštas darbo užmokesčio lygis šalyje
Neefektyvi valstybės ir vietinės valdžios mokesčių politika	Efektyvi valstybės ir vietinės valdžios mokesčių politika
Mažas teikiamų paslaugų skaičius	Gausus ir kokybiškas įvairių paslaugų kiekis
Didelis nusikalstamumas	Žemas nusikalstamumo lygis
Dažnos stichinės nelaimės (sausra, potvyniai, žemės drebėjimai ir pan.)	Maža stichinių nelaimių rizika
Karas	Politinis saugumas
Religinė diskriminacija	Religinė laisvė
	Patrauklesnis klimatas
	Patrauklesnės gyvenimo sąlygos

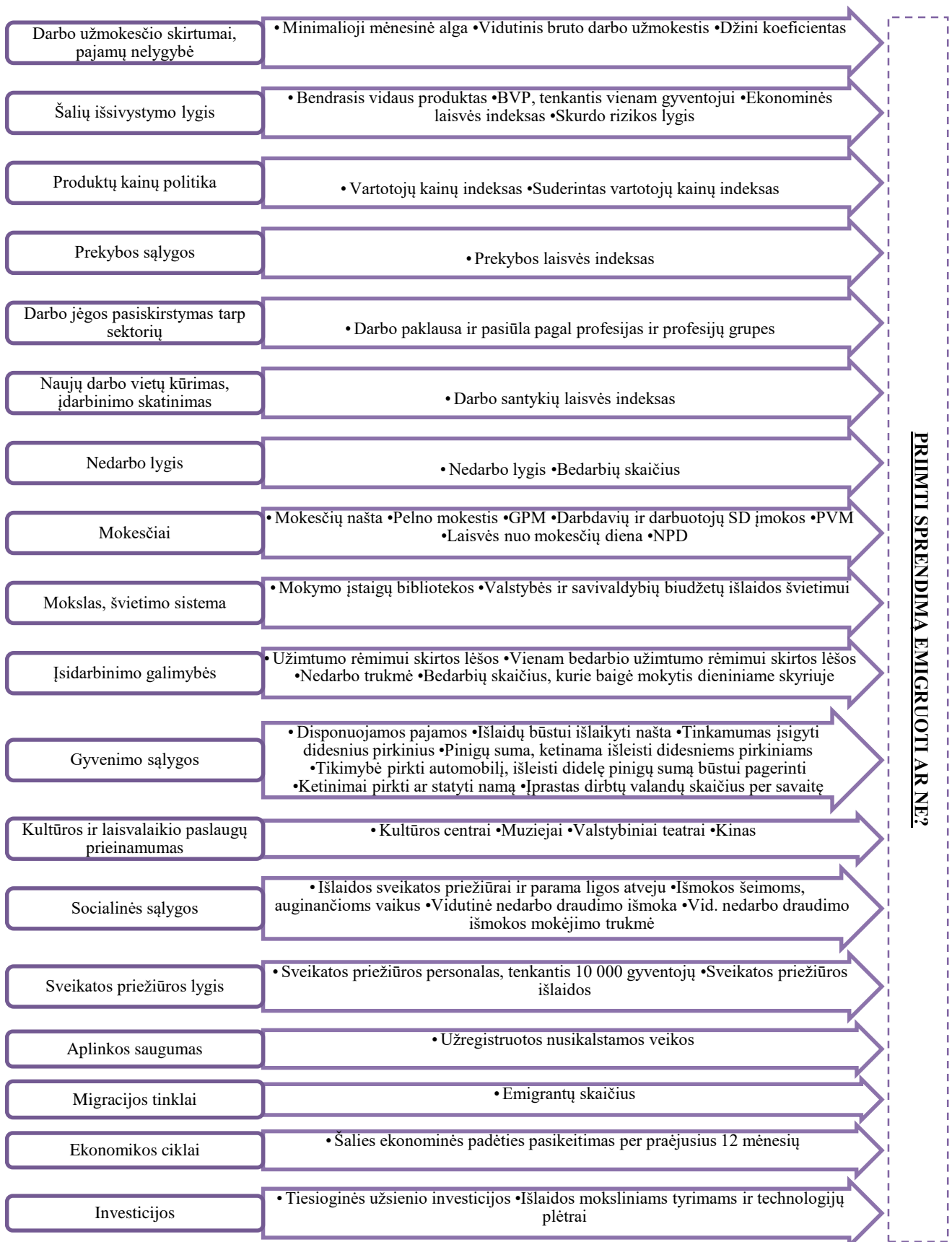
E. S. Lee apibrėžtą migracijos procesą paprasčiausiai atspindi 1.2 paveikslas. Čia atvaizduojami traukos (pliusai) ir stūmos (minusai) veiksniai bei tarpinės kliūtys. Tarpinės kliūtys grafike pateikiamos laužčių pavidalu, tačiau nereikia apsiriboti vien fizinėmis kliūtimis migrantui (sudėtinga, tolima bei varginanti kelionė ir pan.) (13). Pavyzdžiui, griežti imigracijos, įsidarbinimo, nekilnojamo turto įsigijimo įstatymai gali tapti didele kliūtimi potencialiam migrantui. Taip pat svarbu pastebėti, jog plusų ir minusų dažnu atveju turi tiek dabartinė gyvenamoji vieta, tiek vieta, į kurią planuojama persikelti, todėl migrantas, savo subjektyvia nuomone, turi ypač gerai apvarstyti stūmos ir traukos veiksnius, kurie iš jų turi didesnę svorį ir įtaką, nepamirštant įvertinti ir tarpinių kliūčių. Todėl, jei vietovėje, į kurią galvojama persikelti, plusų kiekis atsveria plusus, esančius dabartinėje gyventojų vietovėje, kaip ir matoma aptartame grafike, tikėtina, jog įvyks migracija (6, 15, 16).



1.2 pav. E. S. Lee migracijos modelis

Savo straipsnyje Kumpikaitė ir Žičkutė (17) detalizuoja įvairius migracijos veiksnius ir juos apibūdinančius rodiklius (žr. 1.3 pav.). Tačiau dažniausiai mokslinėje literatūroje migracijai įtaką darantys veiksniai yra skirstomi į socioekonominius arba ekonominius ir neekonominius (socialinius) (Damulienė (8), Kumpikaitė-Valiūnienė ir Žičkutė (10), Žibas (12), Ramirez'as (18)).

Vėlesniame savo tyrime Kumpikaitė-Valiūnienė ir Žičkutė (10), analizuodamos Lietuvos migracijos atvejį pagal 1.3 pav. pateiktą modelį, išskyrė 8 ekonominius ir 11 neekonominių stūmos veiksnių bei 4 ekonominius ir 11 neekonominių migrantams įtaką darančių traukos veiksnių.



PRIMTI SPRENDIMA EMIGRUOTI AR NE?

1.3 pav. Bendrasis emigracijai įtaką darančių veiksnių modelis su juos apibūdinančiais rodikliais (Kumpikaitė ir Žičkutė, 2012 m., p. 744)

Visgi, apibendrinant aptartus migracijai įtaką darančius veiksnius, galima teigti, jog vienu ar kitu būdu jie yra susiję su žmogaus patiriama gerove. Gerovė (angl. *welfare*) užtikrina gerą ir aprūpintą gyvenimą, o ji pati gali būti skirstoma į dvasinę, asmeninę, kultūrinę, socialinę, ekonominę ir pan. Plačiausiai aptariamoms **socialinė ir ekonominė gerovės**, kurios apibrėžia palankią situaciją žmonėms šalyje gyventi, dirbti, kurti šeimą, plėtoti verslą ir kt. – t. y. gyventi visapusiškai pilnavertį gyvenimą (19).

Šalies ekonominė gerovė

Šalies ekonominę gerovę vertina makroekonominiai rodikliai, tokie kaip bendrasis nacionalinis produktas (BNP), grynasis nacionalinis produktas (GNP), nacionalinės pajamos, kurias sudaro grynasis vidaus produktas, grynosios vidaus pajamos, netiesioginiai mokesčiai, asmeninės pajamos, grynosios asmeninės pajamos, tiesioginiai mokesčiai. Dar vienas aktualus ekonominės gerovės veiksnys – nemateriali kaina (angl. *human cost*), sumokama kenksmingomis darbo sąlygomis dirbančio gyventojų. Sudėtingomis sąlygomis dirbančiam žmogui kompensuojama trumpesnėmis darbo valandomis, o tai sąlygoja mažesnę sukuriama produkcijos kiekį, tačiau ilgesnį darbuotojo laisvalaikį, kuris iki tam tikros ribos daro teigiamą įtaką visuomenės gerovei.

Šalies socialinė gerovė

Įprastai šalies socialinę gerovę nusako tokie veiksniai kaip nedarbo lygis, užimtumas, sveikatos apsaugos būklė, realiosios gyventojų pajamos, piliečių raštingumo, kūdikių mirtingumo, gamtos apsaugos lygiai ir pan. Sveikatos apsaugos būklė – rodiklis, parodantis socialinės gerovės lygį, kuris susijęs su gydytojų ir lovų ligoninėse kiekiu, tenkančiu tam tikram gyventojų skaičiui. Ekologinė apsauga taip pat tiesiogiai susijusi su žmogaus sveikata ir jo gyvenimo sąlygomis (13). Valstybės socialinę gerovę galima vertinti ir pagal minimalų pragyvenimo lygį, darbo užmokestį.

1.2 lentelėje pateikiami autorės sugrupuoti migracijai įtaką darantys ekonominės ir socialinės gerovės veiksniai (8, 10, 17, 18, 19).

1.2 lentelė. Susisteminti ekonominės ir socialinės gerovės veiksniai

Ekonominiai gerovės veiksniai	Socialiniai gerovės veiksniai
<ul style="list-style-type: none"> • Su BVP susieti rodikliai – Bendrasis vidaus produktas – BVP, tenkantis vienam gyventojui • Kiti ekonominiai rodikliai – Ekonominės laisvės indeksas – Skurdo rizikos lygis – Tiesioginės užsienio investicijos – Minimalioji mėnesinė alga – Vidutinis darbo užmokestis – Džini koeficientas – Vartotojų kainų indeksas – Suderintas vartotojų kainų indeksas – Disponuojamos pajamos – Išlaidos būstui įsigyti – Galimybė įsigyti didesnius pirkinius – Pinigų suma, ketinama išleisti didesniems pirkiniams – Tikimybė pirkti automobilį, išleisti didelę pinigų sumą būstui gerinti – Ketinimai pirkti ar statyti namą – Įprastas dirbtų valandų skaičius per savaitę 	<ul style="list-style-type: none"> – Švietimas – Mokymo įstaigų bibliotekos – Valstybės ir savivaldybių biudžetų išlaidos švietimui – Išlaidos moksliniams tyrimams ir technologijų plėtrai – Sveikatos apsauga – Sveikatos priežiūros personalas, tenkantis 10 000 gyventojų – Sveikatos priežiūros išlaidos – Išlaidos sveikatos priežiūrai ir parama ligos atveju – Išmokos šeimoms, auginančioms vaikus – Vidutinė nedarbo draudimo išmoka – Vidutinė nedarbo draudimo išmokos mokėjimo trukmė – Aplinka ir gyvenimo sąlygos – Kultūros centrai – Muziejai – Valstybiniai teatrai – Kinas – Užregistruotos nusikalstamos veikos – Demografinė situacija – Emigrantų skaičius – Gimstamumas – Mirtingumas – Darbo rinka – Darbo paklausa ir pasiūla pagal profesijas ir profesijų grupes – Darbo santykių laisvės indeksas – Nedarbo lygis – Bedarbių skaičius – Užimtumo rėmimui skirtos lėšos – Vienam bedarbio užimtumo rėmimui skirtos lėšos – Nedarbo trukmė – Bedarbių skaičius, kurie baigė mokytis dieniniame skyriuje

1.5. Migracijai įtaką darančių veiksnių tyrimai ir studijose naudojami metodai

2013 m. savo darbe „Migracijos problema Lietuvoje ir jos įtaka šalies ekonomikai“ Damulienė (8) tyrė ryšius tarp emigracijos srautų ir devynių pagrindinių makroekonominių rodiklių, tokių kaip BVP vienam gyventojui, privačių asmenų perlaidos, nedarbo lygis, užimtumas, vartotojų kainų indeksas ir pan. Tyrimui naudoti 2001–2010 metų duomenys, taikytas *koreliacinės* analizės metodas.

Skaičiuojant BVP vienam gyventojui ryšį su emigracija nustatyta stipri teigiama priklausomybė ($r = 0,745$). Tai grindžiama tuo, kad didėjant emigracijai šalyje mažėja gyventojų, todėl didėja BVP kiekis, tenkantis vienam žmogui, arba emigrantai atlieka daugiau piniginių perlaidų likusiems šalyje šeimos nariams. Pastarąjį faktą įrodo ir stiprus emigracijos ryšys su privačių asmenų pinigineis perlaidomis ($r = 0,842$). Taip pat stiprus emigracijos koreliacinis ryšys gautas su individualaus vartojimo išlaidų ($r = 0,767$) ir vidutinio darbo užmokesčio ($r = 0,842$) rodikliais. Silpnas teigiamas ryšys ($r = 0,442$) gautas tarp emigracijos ir vartotojų kainų indekso. Labai silpnas neigiamas ryšys gautas vertinant nedarbo ($r = -0,006$) ir užimtumo šalyje ($r = -0,079$) lygį. Autorė teigia, kad emigracijos srautai yra tiek pat dideli, kiek greitas bedarbių skaičiaus augimas. Vidutinis neigiamo stiprumo ryšys gautas tarp emigracijos ir socialinės pašalpos gavėjų skaičiaus ($r = -0,567$).

Dar vieną migracijos srautų tyrimą Lietuvos kontekste 2013 metais publikavo Kumpikaitė ir Žičkutė (20). Darbe siekta atskleisti veiksnius, turinčius įtaką emigracijai iš Lietuvos ir sudaryti emigracijos prognostinį modelį. Tirti tokie kintamieji, kaip nedarbo lygis, nedarbo trukmė, minimalus ir vidutinis darbo užmokestis, skurdo lygis, pajamų nelygybė, mokesčių tarifai. Analizė atlikta naudojant SPSS programinę įrangą, taikytas *tiesinės daugialypės regresinės analizės* metodas, tirti 2001–2011 metų duomenys.

Atlikus duomenų analizę nustatyta, kad emigracijai iš Lietuvos statistiškai reikšmingi trys veiksniai: nedarbo lygis, Džini koeficientas ir laisvės nuo mokesčių dienų skaičius. Modelis, sudarytas iš minėtų veiksnių, paaiškina 70,7 proc. duomenų sklaidos. Svarbiausiu veiksnium lygtyje nustatytas Džini koeficientas. Autorės, besiremiamos sudarytu regresinės analizės modeliu nustatė, kad norint sumažinti emigraciją nedarbo lygis turėtų mažėti mažiausiai 8,5 proc., o Džini koeficientas turėtų nukristi iki 30.

2016 m. lapkričio mėn. Amity universiteto ekonomikos žurnale publikuotame straipsnyje (21) analizuojami svarbiausi faktoriai, darantys įtaką vidinės migracijos srautams Indijoje. Tyrime nustatyti reikšmingi veiksniai, turintys lemiamą vaidmenį priimant sprendimą migruoti. Analizės tikslai – iširti

besikeičiančias vidaus migracijos tendencijas Indijoje, identifikuoti veiksnius, sukeliančius vidaus migraciją, ištirti ryšį tarp migracijos ir socialinių ir ekonominių rodiklių.

Tyrimui naudoti Nacionalinio imties tyrimo biuro (angl. *National Sample Survey Office*, NSSO) 1983–2007 metų duomenys apie 17 Indijos valstijų migraciją ir įvairūs jų socialiniai ir ekonominiai kintamieji. Tiriant migracijos ir socialinių ir ekonominių rodiklių santykį, taikyta **koreliacinė, faktorinė ir tiesinė regresinė** analizės. Parinkti 15 įvairių sričių kintamieji, tokie kaip grynasis vidaus produktas vienam gyventojui, gyventojų skaičius žemiau skurdo ribos (mieste ir kaimo vietovėse), gyventojų augimas, urbanizacija, darbuotojų užimtumo dalis, raštingumo lygis, elektros suvartojimas ir pan. Statistinei duomenų analizei naudota SPSS V.22 programinė įranga.

Atlikus *koreliacinę analizę* nustatyta, kad migracijos srautai yra teigiamai koreliuoti su grynuoju vidaus produktu vienam gyventojui, kreditavimu žemės ūkiui vienam gyventojui ir elektros suvartojimu vienam gyventojui. *Regresinės analizės* rezultatai atskleidžia, kad gyventojų, gyvenančių žemiau skurdo ribos skaičius ir išsilavinimas turi teigiamą ir reikšmingą ryšį, o vienam gyventojui tenkančios pajamos neigiamą ir reikšmingą ryšį su emigracijos apimtimis. Urbanizacija, gyventojų pajamos, raštingumo lygis, gyventojų kreditavimas ir elektros suvartojimas turi teigiamą ir reikšmingą ryšį su imigracijos srautu. *Faktorinės analizės* metu išskirti trys reikšmingi faktoriai. Nustatyta, kad didžiausią įtaką turi vienam gyventojui tenkantis bendrasis vidaus produktas, vieno gyventojų elektros energijos sunaudojimas, vienam gyventojui tenkantis žemės ūkio kreditavimas, gyventojų skaičius žemiau skurdo ribos ir raštingumo lygis.

Kitame tyrime, skelbtame 2017 m. rugpjūčio mėn. (22) analizuojami svarbiausi veiksniai, turintys įtakos jaunimo migracijai iš Kurdistano regiono į Europą. Tyrimo metu autoriai taikė **faktorinės ir klasterinės** analizės metodus, siekdami palyginti gautus rezultatus ir nustatyti pačius svarbiausius kintamuosius, darančius įtaką jaunimo išvykimui iš šalies. Šiam tyrimo tikslui pasiekti buvo naudoti Salahaddin universiteto studentų duomenys, surinkti klausimyno formatu.

Surinktų duomenų tyrimui taikytas *pagrindinių komponentų* (angl. *principal component analysis*, PCA) metodas *faktorinėje* analizėje ir *Ward'o* metodas *klasterinėje* analizėje. Sujungiant *faktorinę* ir *klasterinę* analizę galima sakyti, kad duomenys apžvelgiami iš dviejų viena kitą papildančių perspektyvų. Abiejų šių metodų pagrindinis principas – homogeniškumu paremtas klasifikavimas. Tai reiškia, kad duomenys sujungiami ir grupuojami į tarpusavyje nesusijusias grupes, o grupės viduje esantys kintamieji yra

vienijami kokio nors bendrumo, artumo (latentinio kintamojo). Statistinei šių duomenų analizei naudota SPSS V.22 programinė įranga.

Tyrimo metu pastebėta, jog yra panašumo tarp reikšmingų kintamųjų, gautų tiek vienu, tiek kitu taikytu metodu (faktorių skaičius yra lygus klasterių skaičiui). *Faktorinės* ir *klasterinės* analizės rezultatai atskleidė, kad egzistuoja 5 reikšmingi faktoriai ir 5 klasteriai, iš kurių galima nustatyti kintamuosius, darančius įtaką migracijai. Faktoriai paaiškina 62,29 proc. visos duomenų sklaidos. Abiem metodais kintamųjų paskirstymas atliktas vienodai, išskyrus 4 ir 5 grupes, kuriose, lyginant dvi analizes, vienas kintamasis sukeistas su kitu kintamuoju vietomis. *Klasterinės* analizės metu nustatytas mažiausias atstumas tarp nedarbo ir nepatenkinamo darbo užmokesčio bei didžiausias atstumas tarp studijų ir nepatikimų ar nesąžiningų politinių partijų. *Faktorinės* analizės metu rezultatai nustatyti tokie pat – didžiausia koreliacija tarp nedarbo ir nepatenkinamo darbo užmokesčio, mažiausia – tarp studijų ir nesąžiningos konkurencijos tarp politinių partijų. Tyrimas atskleidė, jog svarbiausi kintamieji, skatinantys jaunimą migruoti iš Kurdistano regiono, yra: ekonominiai – nedarbas, bloga ekonominė situacija, netenkinamas darbo užmokestis ir politiniai – nepasitikėjimas esama vyriausybe.

1.6. Duomenų analizės programinės įrangos apžvalga

Vienas svarbiausių sprendimų atliekant duomenų analizę – tinkamas statistinės analizės programinės įrangos pasirinkimas. Šiame poskyryje aptariamos populiariausios ir dažniausiai duomenų analitikų naudojamos programinės priemonės ir šių priemonių galimybės (23).

Bandant grupuoti statistinės analizės programinę įrangą, jas galima išskirti į 3 kategorijas – specializuoti, universalūs ir profesionalūs įrangos paketai, skirti skirtingoms tikslinėms auditorijoms. Specializuoti paketai pritaikomi konkrečių uždavinių sprendimams ir dažniausiai nereikalauja didelių statistinių žinių. Universaliuose paketuose paprastai būna realizuota šiek tiek daugiau statistinių algoritmų, hipotezių tikrinimų ir pan. Tuo tarpu profesionalūs paketai išsiskiria gebėjimu apdoroti ypač didelius duomenų rinkinius, gausiu ne tik klasikinių, bet ir įvairių specialių statistinių metodų pasirinkimu, galimybėmis kurti ir realizuoti savo paties metodus (24).

MS Excel – vienas populiariausių ir plačiausiai verslo įmonėse naudojamų programinių paketų. Skaičiuoklė nepakeičiama atliekant paprastus skaičiavimus (vidurkiai, dispersijos, standartiniai nuokrypiai ir t. t.), braižant grafikus (histogramos, juostinės diagramos ir t. t.), filtruojant duomenis. Taip pat, įdiegus plėtinį „analizės įrankių paketas“, „Excel“ galimybes galima praplėsti ir tikrinti įvairias hipotezes, atlikti dispersinę, regresinę, koreliacinę, kovariacinę analizę, eksponentinius glodinimus, laiko

eilučių analizę. Plėtinys „PowerPivot“ suteikia galimybę susieti „Excel“ su išorinėmis duomenų bazėmis ir tiesiogiai jų duomenis naudoti analizei (25).

SPSS (angl. *Statistical Package for the Social Sciences*) – statistinės analizės programinis paketas, orientuotas į socialinių mokslų sritį. Ši programinė įranga tinkama pradedančiajam ir patogi pažengusiam vartotojui. Pakete gausu realizuotų statistinių analizės metodų, hipotezių tikrinimų, aukštos kokybės vizualizacijos priemonių. Puikiai tinka dispersinei, regresinei, daugiamatei analizei. Tačiau SPSS sunku apdoroti didelius duomenų kiekius, o pritrūkus funkcionalumo, vartotojui nėra galimybių jį susikurti pačiam. Dažniausiai šią programinę įrangą naudoja medicinos, rinkodaros, kokybės valdymo procesų darbuotojai (26).

SAS (angl. *Statistical Analysis System*) – tai viena statistinių analizės programinių priemonių lyderių. Programinė įranga suderinama su visomis operacinėmis sistemomis. Ja galima atlikti pilną analizę – nuo duomenų paruošimo, iki ataskaitų pateikimo. Programinė įranga tinkama tiek pradedančiajam (su SAS „Enterprise Guide“ programa paprastą duomenų analizę galima atlikti meniu pagalba), tiek ekspertui. SAS geba apdoroti ypač didelius duomenų kiekius. Tai profesionalus analizės įrankis, siūlantis plačiausią statistinių algoritmų pasirinkimą profesionalių sistemų rinkoje. Norint sistema atlikti neparametrinių metodų analizę, taikoma GAM procedūra. Vartotojui pasigedus tam tikrų jau realizuotų funkcionalumų, jis turi galimybę analizės įrankius susikurti pats (27).

„RStudio“ – atviro kodo programinis paketas, skirtas duomenų manipuliavimui, skaičiavimui, grafiniam atvaizdavimui. Programinė įranga suderinama su visomis operacinėmis sistemomis. Ja galima atlikti pilną analizę – nuo duomenų paruošimo, iki ataskaitų pateikimo. Norint efektyviai išnaudoti šios programos galimybes, reikia nemažai programavimo žinių. „RStudio“ programinė įranga teikia platų statistinių (tiesinių ir netiesinių modelių, klasikinių statistinių bandymų, laiko eilučių analizės, klasifikavimo, klasterizavimo, grupavimo ir t. t.) ir grafinių metodų įvairovę. Faktorinės, pagrindinių komponentų analizės metodai realizuoti *FactoMineR*, *factoextra*, *psych* plėtiniuose. Neparametriniai analizės metodai realizuoti *mgcv* (apibendrinti adityvūs metodai), *np* (branduolinė regresinė analizė) plėtiniuose. Didžiausias sistemos privalumas – ji nuolat atnaujinama, o jei vartotojui nepakanka esamo funkcionalumo, kiekvienas gali kurti savo algoritmus ir jais dalintis su kitais vartotojais. Norint analizei taikyti iš aplinkos sukurtus metodus, tereikia į savo programinę įrangą parsisiųsti ir įdiegti plėtinius (28).

1.7. Literatūros apžvalgos apibendrinimas

Atlikus literatūros analizę pastebima, jog migracijos procesas ir su migracija susiję klausimai nėra nauja problema, tačiau ypač aktuali ir svarbi įvairioms mokslo sritims, tokioms kaip ekonomika, politika, politinė geografija ar sociologija. Kasmet didėjantis gyventojų skaičius lemia didesnius migracijos srautus, o tai skatina gilintis į problemos esmę. Nustatyta, kad nėra vienos bendros teorijos, kuri gebėtų detaliai paaiškinti migracijos priežastis ir identifikuoti jai įtaką darančius veiksnius. Atlikus mokslinės literatūros analizę, konstatuota, kad dažniausiai migracijos srautams įtaką darantys veiksniai gali būti klasifikuojami į ekonominius ir socialinius. Remiantis atlikta analize, šiame darbe siūloma veiksnius grupuoti į ekonominės ir socialinės gerovės veiksnių grupes, susidedančias iš skirtingų rodiklių, kuriuos galima pamatuoti matematiškai. Be to pastebėtina, kad ankstesniuose tyrimuose daugiausia remiamasi apklausų duomenimis (Abubaker'is ir kiti (22)), arba analizuojami maži statistinių duomenų ir veiksnių kiekiai (Damulienė (8), Ramirez ir kt. (9), Mayda (29)). Dažniausiai tyrimuose siekiama nustatyti priežastis, dėl ko vyksta migracija, koks veiksnys turi didžiausią įtaką, tačiau nebandoma rezultatų pritaikyti prognozei. Todėl toliau darbe bus atliekama Europos Sąjungos šalių tarptautinei migracijai įtaką darančių veiksnių analizė ir sudaromas migracijos srautus leidžiantis prognozuoti modelis.

2. Tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo metodinė dalis

Baigiamojo projekto tyrimo metu bus tiriami veiksniai ir jų tarpusavio sąryšis, darantys įtaką Europos Sąjungos valstybių migracijos srautams, pritaikytas tinkamas analizės metodas ir sukurtas originalus modelis prognozei.

Atlikus literatūros analizę, baigiamojo magistro projekto tyrimui nuspręsta pasitelkti keletą programinių įrankių: „MS Excel“ ir statistinės analizės atviro kodo programą „RStudio“. Darbo temos „Tarptautinės migracijos srautams įtaką darančių veiksnių analizė“ įgyvendinimui duomenys surinkti iš Europos statistikos departamento (angl. *Eurostat*) ir Pasaulio duomenų banko (angl. *World bank data*) duomenų bazių. Visi tyrime naudoti duomenys yra prieinami laisvai be papildomų teisinių reikalavimų. Analizuoti 2000–2016 metais sukaupti stebiniai. Iškelto tikslo, suformuoti prognostinį modelį, remiantis tarptautinės migracijos srautams įtaką darančių veiksnių analize Europos Sąjungos kontekste, įgyvendinimui parinkti aprašomosios, klasterinės, koreliacinės, pagrindinių komponentų ir neparаметrinės regresinės analizės metodai.

2.1. Duomenų matricos formavimas

Iš dviejų skirtingų duomenų bazių, atsižvelgus į literatūros analizės metu rastus galimai reikšmingus kintamuosius, migracijos srautų analizei atrinkti Europos šalių demografiniai rodikliai ir įvairūs ekonominę ir socialinę gerovę įvertinantys kiekybiniai kintamieji (sugrupuoti ir detalizuoti priede nr. 5).

Migracijos srautų tyrimui pasirinktas Y kintamasis – migracijos saldo (migracijos neto) tūkstančiui gyventojų, kurio apskaičiavimui naudoti Eurostato pateikiami demografiniai Europos Sąjungos šalių rodikliai. Skaičiavimai atlikti kiekvienai šaliai visiems tyrime analizuojamiems metams atskirai, taikyta 11 formulė. Imigracija – atvykusių į šalį gyventojų skaičius, emigracija – išvykusių iš šalies gyventojų skaičius. Teigiamas Y reikšmė nurodo didesnę imigraciją, neigiamas – emigraciją.

$$Y = \frac{\text{imigracija} - \text{emigracija}}{\text{gyventojų skaičius šalyje}} * 1\,000 \quad (1)$$

Iš skirtingų duomenų bazių surinkti duomenys sujungti „MS Excel“ programa ir paruoštas vienas CSV formato dokumentas tolesnei analizei.

2.2. Tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo analizė

Išanalizavus literatūrą tarptautinės migracijos srautams įtaką darančių veiksnių tyrimo analizei nuspręsta pasirinkti „RStudio“ programinę įrangą ir statistinės analizės metodus – aprašomoji statistika, klasterinė, koreliacinė, pagrindinių komponentų, regresinės analizės. Skyreliuose pateikiama metodinė informacija apie tyrimo atlikimą ir trumpas „RStudio“ komandų aprašymas tyrimo realizacijai.

2.2.1. Duomenų apžvalga

Tyrimo pradžioje apžvelgiami visi atrinkti kintamieji. Tam taikomi aprašomosios statistikos metodai. Aprašomoji statistika apima duomenų žvalgymą iš įvairių pusių ir įvairiais pjūviais. Tai vienas pirmųjų metodų, atliekamų norint susipažinti su duomenų rinkiniu. Būtent žvalgomosios analizės metu pastebimi duomenų suvedimo netikslumai, bendrosios savybės, skirstiniai. Žvalgomąją analizę galima skirti į dvi grupes: skaitinę ir grafinę.

Pagrindinės aprašomosios statistikos rūšys:

- dažnių lentelė. Ji padeda lengvai pastebėti kiek skirtingų reikšmių yra duomenyse, nustatyti mažiausią, didžiausią bei dažniausiai pasikartojančią reikšmę ir pan.;
- skaitinės charakteristikos. Tai duomenų padėtis (charakteristikos – vidurkis, moda, mediana, kvantiliai), jų sklaida (charakteristikos – vidurkis, moda, mediana, kvantiliai) ir dažnių skirstiniai (asimetrijos ir eksceso koeficientai);
- grafiniai duomenų analizės metodai. Iš jų matoma skirstinio forma bei, kaip ir dažnių lentelė, leidžia nustatyti mažiausią bei didžiausią reikšmę, modą ir pan. Kokybiniais ir kiekybiniais kintamiesiems, įgyjantiems kelias reikšmes, braižomos stulpelinės diagramos (24).

2.2.2. Homogeninių grupių išskyrimas

Kadangi kiekvienai Europos Sąjungos šaliai turima maksimaliai 17 stebinių, o modelius norima sudaryti bendresnius, atliekamas šalių grupavimas pagal bendrinius bruožus. Sujungimui parinkti pagrindiniai ir dažniausiai literatūros apžvalgoje minėti veiksniai – migracijos srautas ir bendrasis vidaus produktas (BVP) vienam gyventojui.

Remiantis autorių kolektyvo A. Mihi-Ramirez'o, V. Kumpikaitės-Valiūnienės ir E. Guenca-Garcia straipsniu (9), šalys skirstomos į skurdžias ir turtingas. Šiai analizei naudotasi BVP rodiklio vienam gyventojui penkerių paskutinių stebėtų metų vidurkiu. Šalys padalinamos į dvi lygias grupes.

Dar vienas šalių skirstymas į homogenines grupes atliktas vertinant migracijos srauto rodiklį į stumiančias ir traukiančias migrantus. Analizei naudoti penkerių paskutinių stebėtų metų duomenys. Šalys skirstomos į tris grupes: visais stebėtais metais turinčias neigiamą, teigiamą arba pakitusį (iš teigiamo į neigiamą, arba iš neigiamo į teigiamą) migracijos srautus.

Trečiam grupavimui taikytas metodas – *klasterinė analizė*.

Klasterinė analizė – tai matematinių metodų visuma, skirta objektų ar reikšmių aibėms suskaidyti į tam tikra prasme vienalytes, homogenines grupes (klases, klasterius) taip, kad tos pačios grupės (klasterio) elementai būtų „artimi“ vienas kitam, o elementai iš skirtingų grupių (klasterių) – „tolimi“ vienas kitam (30). Norėdami suskirstyti objektus į klasterius, dažnai nežinome, ar klasteriai iš viso egzistuoja, o jei egzistuoja, tai koks tai yra skaičius. Dėl šios priežasties klasterinė analizė yra egzistuojančių struktūrų paieška, o analizės metodo parinkimas ir rezultatų interpretacija priklauso tik nuo pačio tyrėjo (31).

Pagrindiniai klasterizavimo etapai:

- uždavinio formulavimas;
- požymių parinkimas;
- artumo (panašumo) mato parinkimas;
- atstumo tarp klasterių parinkimas;
- klasterizavimo metodo parinkimas;
- klasterių interpretavimas ir kokybės vertinimas.

Požymių parinkimas. Parenkami požymiai, kurie yra esminiai atskiriant objektus, turintys tik konkrečią informaciją, reikalingą uždaviniui spręsti. Šalių grupavimas atliekamas analizei naudojant du anksčiau aptartus rodiklius ir taikant k-vidurkių metodą. Kiekvienai šaliai paskaičiuojamas penkerių metų migracijos neto ir BVP vienam gyventojui vidurkis. Kadangi kintamieji apskaičiuoti skirtingose skalėse, jie yra standartizuojami.

Standartizavimas:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

čia z – standartizuota reikšmė, x – stebėta reikšmė, μ – populiacijos vidurkis, σ – populiacijos standartinis nuokrypis.

Artumo (panašumo) mato parinkimas. Tai kiekybinis matas, kuris įvertina, kiek du požymių vektoriai yra „panašūs“ ar „nepanašūs“ tarpusavyje. Reikia įvertinti, ar visi požymiai lygiaverčiai ir ar tarp jų nėra dominuojančių. Dažniausiai naudojami metriniai artumo matai (pvz., Euklido), koreliacijos, asociatyvumo koeficientai. Jei požymiai matuojami intervalų arba santykių skalėje, taikomi metriniai artumo matai, jei kintamieji kokybiniai – koreliacijos ir asociatyvumo koeficientai.

Euklido atstumas:

$$d(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}, \quad (3)$$

čia m – požymių skaičius.

Tiesinės koreliacijos koeficientas:

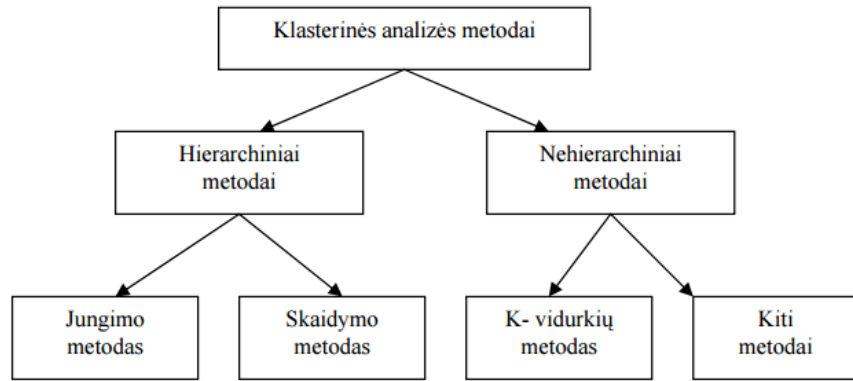
$$r_{XY} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}. \quad (4)$$

Asociatyvumo koeficientas:

$$d(X, Y) = \#(x_i \neq y_j) / m, \quad (5)$$

čia $\#(x_i \neq y_j)$ – nesutampančių požymių reikšmių skaičius.

Atstumo tarp klasterių parinkimas. Klasterinėje analizėje objektų kintamieji gali būti matuojami nominaliojoje, ranginėje, intervalinėje arba santykių skalėse. Taikomi įvairūs metodai, kurie išskiriami į dvi pagrindines klases – hierarchinius ir nehierarchinius metodus. Hierarchiniu metodu klasteriai sudaromi naudojant hierarchinį principą: daroma prielaida, jog visi stebėjimai yra vienas klasteris, kurį sudaro mažesni klasteriai, įtraukiantys dar mažesnius ir t. t., kol kiekvienas atskiras stebėjimas tampa atskiru klasteriu. Taip pat hierarchinė klasterizacija gali būti atlikta ir iš kitos pusės, t. y. skaidymo metodu, kai kiekvienas stebėjimas yra laikomas atskiru klasteriu ir kiekvienoje sekančioje iteracijoje šie klasteriai yra jungiami į didesnius, kol gaunamas vienintelis klasteris. Hierarchiniai metodai efektyvūs tik nedideliems duomenų rinkiniams (laikoma, jog iki 250 stebėjimų), todėl, norint klasterizuoti didelius duomenų kiekius, taikomi nehierarchiniai metodai.

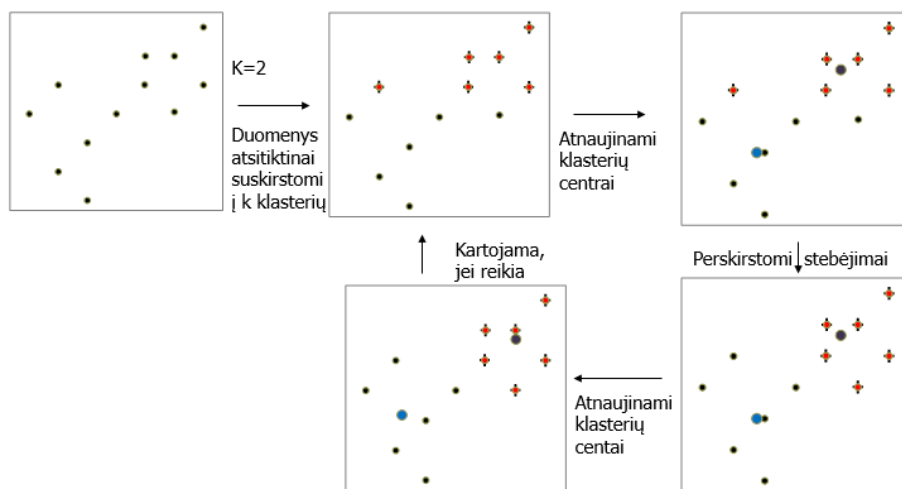


2.1 pav. Klasterinės analizės metodai

K-vidurkių metodas. Šis metodas leidžia lengvai analizuoti didelį objektų skaičių, be to, pateikiami galutiniai klasterių centrai, kurie iškart leidžia stebėti, kaip skiriasi radinių grupių santykiniai tankiai konkrečiuose klasteriuose. Nehierarchinis k-vidurkių metodas skirtas kiekybiniais (intervalų ir santykių skalės) duomenims klasterizuoti.

Klasterizavimo procedūrą sudaro trys žingsniai:

- objektai skirstomi į k pradinių klasterių;
- iš eilės skaičiuojamas kiekvieno objekto atstumas iki klasterių centro. Objektas priskiriamas artimiausiam klasteriui. Klasterių centrai perskaičiuojami;
- 2 žingsnis kartojamas tol, kol perskirstymų daugiau nėra (32).



2.2 pav. K-vidurkių klasterizavimas

Norint efektyviausiai taikyti k -vidurkių metodą, naudinga prieš tai atlikti keletą tikrinimų, skirtų optimaliam klasterių skaičiui imtyje iširti. Šių kriterijų aibė yra labai didelė. Vienas dažniausiai taikomų – alkūnės metodas (angl. *elbow method*), kurio idėja – vykdyti k -vidurkių klasterizavimą visam rinkiniui k (pvz., nuo 1 iki 10) ir kiekvienai skirtingai k reikšmei apskaičiuoti kvadratų sumos klaidas (angl. *sum of squared*, SSE). Dažnai braižomas šios klaidos ir k grafikas. Optimaliu klasterių skaičiumi laikoma ta vertė, kai, didinant skaičių k , klaida mažėja minimaliai. Dar vienas populiarus metodas – kubinio klasterizavimo koeficientas CCC (angl. *Cubic Clustering Criterion*). Optimaliu klasteriu skaičiumi laikoma ta vertė, kur CCC yra didžiausias.

Klasterių interpretavimas ir kokybės vertinimas. Atlikus klasterizavimą, svarbu peržiūrėti rezultatus ir juos įvertinti. Reikalingos ne tik techninės žinios metodo pritaikymui, tačiau ir ekspertinė įžvalga. Tiems patiems duomenims taikant skirtingus klasterinės analizės metodus, galima gauti skirtingus rezultatus. Tuo tarpu, jei duomenys yra linę klasterizuotis, taikydami skirtingus metodus gausime panašius rezultatus, o tai leis matyti, kad skaidymas atliktas tinkamai (31, 33).

Klasterinės analizės realizavimui „RStudio“ programinėje įrangoje naudotas *NbClust* plėtinys ir to paties pavadinimo procedūra (34). Su šia procedūra galima atlikti pilną klasterinę analizę – nuo optimalaus homogeninių grupių skaičiaus parinkimo, iki optimalaus stebinių apjungimo į homogenines grupes.

2.2.3. Statistiškai reikšmingų veiksnių identifikavimas

Išskyrę valstybes į homogenines grupes, kiekvieną grupę tiriamo atskirai. Nustatomi būtent tos grupės priklausomam kintamajam (migracijos neto) reikšmingi veiksniai. Reikšmingų veiksnių išskyrimui taikoma koreliacinė analizė.

Tiriant kiekybinių kintamųjų X ir Y reikšmių kitimą ir jų ryšį, siekiama nustatyti jų ryšio stiprumą. Šiam tikslui vertinti naudojami įvairūs koreliacijos koeficientai. Koreliacijos koeficiento dydis nurodo ryšio stiprumą (skirtingi koeficientai turi skirtingus režius, dažniausiai koeficientas kinta nuo -1 iki +1, arba nuo 0 iki +1). Siekiant duomenims parinkti tinkamą koreliacijos koeficientą, svarbu atsižvelgti į kintamųjų skirtingą, vertinamą ryšio formą ir t. t.

Tikrindami hipotezę apie populiacijos koreliacijos koeficiento lygybę nuliui, atsakome į klausimą apie kintamųjų priklausomybės populiacijoje statistinį reikšmingumą, tačiau iš koreliacijos koeficiento negalima nustatyti koreliacijos priežasties. Du kintamieji X ir Y gali stipriai koreliuoti dėl trijų priežasčių: kintamasis X daro poveikį kintamajam Y , kintamasis Y daro poveikį kintamajam X , abu kintamieji X ir

Y yra veikiami trečio kintamojo. Todėl koreliacinės analizės metu nustatytas ryšys interpretuojamas tik kaip ryšio matas (35).

Pirsono koreliacijos koeficientas. Šis koeficientas įvertina tiesinio ryšio stiprumą ir yra naudojamas, kai stebimų atsitiktinių dydžių X ir Y reikšmės yra išmatuotos intervalų arba santykių skalėje, o jų dvimatis skirstinys yra normalusis. Populiacijos Pirsono koreliacijos koeficiento taškinio įverčio apskaičiavimo formulė:

$$\hat{\rho} = r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \sqrt{\overline{y^2} - (\bar{y})^2}}. \quad (6)$$

Kuo $|r|$ arčiau 1, tuo tiesinis ryšys yra stipresnis. Kai $r > 0$, tai didėjant vieno atsitiktinio dydžio reikšmėms, kito reikšmės didėja tiesiškai. Kai $r < 0$, tai didėjant vieno atsitiktinio dydžio reikšmėms, kito reikšmės mažėja tiesiškai, o kai $r = 0$ – tiesinio ryšio nėra.

Nustatant dydžių X ir Y tiesinę priklausomybę, tikrinama hipotezė apie populiacijos Pirsono koeficiento lygybę nuliui:

$$H_0: \rho = 0$$

$$H_a: \rho \neq 0$$

Hipotezė H_0 tikrinama naudojant Stjudento statistiką:

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}, t \sim St(n-2) \quad (7)$$

čia r – imties Pirsono koreliacijos koef., n – imties didumas.

Kai hipotezė H_0 atmetama, tai tarp dydžių X ir Y statistiškai reikšmingas tiesinis ryšys.

Spirmeno ranginės koreliacijos koeficientas. Koreliacijos koeficientas apibūdina ryšio tarp X ir Y stiprumą monotoniškumo prasme, t. y., kai $\rho_s > 0$, didėjant dydžiui X , dydis Y monotoniškai didėja (nebūtinai tiesiškai), arba mažėja, kai $\rho_s < 0$. Hipotezės:

$$H_0: \rho_s = 0$$

$$H_a: \rho_s \neq 0.$$

Hipotezė H_0 tikrinama naudojant Stjudento statistiką:

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \quad (8)$$

(35).

Prieš analizę normuojami visi nepriklausomi kintamieji ir kiekvienam klasteriui pašalinami tie veiksniai, kurių trūkstatų reikšmių skaičius yra didesnis nei 50 proc.

Koreliacijos skaičiavimui „RStudio“ programiniame pakete naudojamas *stats* plėtinys ir jo procedūra *cor* (36). Kadangi duomenyse esama trūkstatų reikšmių, koreliacijų skaičiavimui nustatoma parinktis *pairwise.complete.obs*, kuri koreliacijas skaičiuoja tik užpildytoms duomenų reikšmių poroms. Nuspręsta, kad koreliacijos laikomos reikšmingomis, kai koreliacijos koeficiento reikšmė $|cor| > 0,30$, o reikšmingumo lygmuo $p < \alpha = 0,05$.

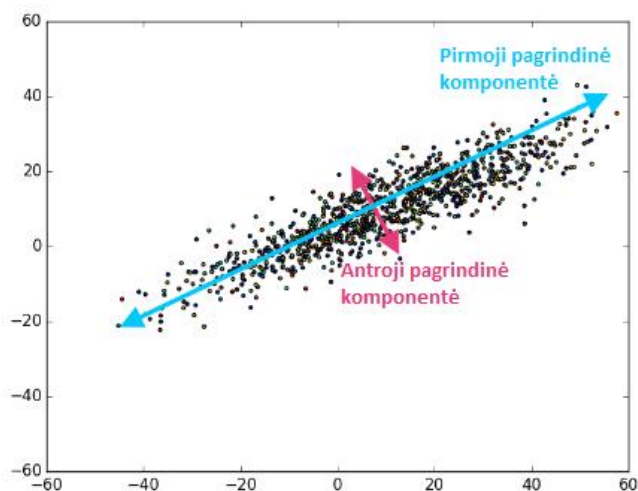
2.2.4. Kintamųjų dimensijos mažinimas

Siekdami sumažinti į modelius įtraukiamų kintamųjų kiekį, iš atrinktų reikšmingų veiksnių kiekvienai tiriamai grupei išskirsime pagrindines komponentes.

Pagrindinių komponentių analizės uždavinys — maksimizuoti kintamųjų tiesinių kombinacijų dispersijas. Metodas dažnai naudojamas duomenų dimensijai sumažinti, kai reikalaujama, kad suprojektuoti į mažesnės dimensijos poerdvį duomenys savyje turėtų išlaikyti kiek galima daugiau pradinių duomenų išsibarstymo (37). Tai matematinė procedūra, kuri koreliuojančius kintamuosius transformuoja į mažesnę skaičių nekoreliuojančių kintamųjų, vadinamų pagrindinėmis komponentėmis. Pagrindinių komponentių analizės tikslas – surasti kryptį, kuria duomenų dispersija yra pati didžiausia. Šis metodas padeda vizualizuoti duomenis, palengvina duomenų suvokimą ir interpretavimą.

Kiekvienai homogeninei grupei nustačius reikšmingus veiksniai, prarasdami kuo mažiau kintamųjų sklaidos, koreliuojančius kintamuosius transformuojame į mažesnę skaičių nekoreliuojančių kintamųjų (pagrindinių komponentių). Pirmąją pagrindinę komponentę vadinama tokia kryptis, kuria duomenų sklaida yra didžiausia. Antroji komponentė parenkama taip, kad jos kryptimi sklaida būtų didžiausia, tačiau kartu jos kryptis taip pat turi būtų statmena pirmajai pagrindinei komponentei (žr. 2.3 pav.). Trečioji parenkama, kad jos sklaida būtų didžiausia ir ji pati būtų ortogonalios tiek pirmajai, tiek antrajai komponentėms ir t. t. Duomenų pasukimas pagrindinių komponentių kryptimi panaikina koreliacijas, todėl transformuoti kintamieji tampa nekoreliuoti. Pirmoji komponentė savyje turi didžiausią sklaidą,

antroji didžiausią sklaidą, kurios nepaaiškina pirmoji ir yra ortogonali pirmajai ir t. t. Pagrindinių komponentių skaičius yra mažesnis arba lygus visų kintamųjų skaičiui (37).



2.3 pav. Pirmoji ir antroji pagrindinės komponentės

Prieš taikant pagrindinių komponentių metodą svarbu patikrinti ar duomenys yra tinkami tokio tipo analizei. Kadangi analizė paremta tarpusavyje koreliuotų kintamųjų apjungimu, todėl aišku, kad nekoreliuotiems duomenims pagrindinių komponentių metodas nėra tinkamas.

Pirmas tikrinamas kriterijus – stebinių koreliacijų (kovariacijų) matrica. Ji leidžia matyti, kurie kintamieji turi ryšį tarpusavyje, o kurie nėra grupuojami ir patys sudaro atskirą grupę.

Vienas iš kriterijų, leidžiančių nustatyti, kad pagrindinių komponentių analizė yra tinkama – Bartleto sferiškumo kriterijus (angl. *Bartlett's test of sphericity*), kuris tikrina, ar tarp stebimų kintamųjų egzistuoja statistiškai reikšminga koreliacija. Tyrimo metu tikrinama hipotezė, kad koreliacijų matrica yra vienetinė (visi stebiniai nekoreliuoti). Jei hipotezė priimama ($p \geq \alpha$, čia α pasirinktas reikšmingumo lygmuo) – analizė neturi prasmės (33, 38).

Kintamųjų tikimą pagrindinių komponentių analizei vertina Kaizerio-Mejerio-Olkinio (*KMO*) (angl. *Keiser–Meyer–Olkin measure*) matas, kuris nustato, ar tarp kintamųjų porų koreliacijos yra paaiškinamos kitais kintamaisiais.

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}}{\sum \sum_{i \neq j} r_{ij} + \sum \sum_{i \neq j} \tilde{r}_{ij}}, \quad (9)$$

čia r_{ij} – kintamųjų X_i ir X_j koreliacijos koeficientas, \tilde{r}_{ij} – X_i ir X_j dalinės koreliacijos koeficientas.

Maža KMO reikšmė rodo, jog kintamųjų porų koreliacijos nėra paaiškinamos kitais kintamaisiais. Literatūroje vadovaujamosi tokiomis KMO normomis: $0,9 < KMO$ – analizė puikiai tinka, $0,8 < KMO \leq 0,9$ – tinka gerai, $0,7 < KMO \leq 0,8$ – tinka patenkinamai, $0,6 < KMO \leq 0,7$ – tinka pakenčiamai, $0,5 < KMO \leq 0,6$ – tinka blogai, $KMO < 0,5$ – analizė netinka (33, 38).

Taip pat vertinamas ir kintamųjų stebinių tikimas analizei pagal MSA_i (angl. *Kaiser measure of sampling adequacy*) reikšmes.

$$MSA_i = \frac{\sum \sum_{j \neq i} r_{ij}}{\sum \sum_{j \neq i} r_{ij} + \sum \sum_{j \neq i} \tilde{r}_{ij}} \quad (10)$$

Reikšmingais laikomi tie kintamieji, kurių $MSA_i > 0,5$. Jei reikšmė mažesnė – kintamuosius verta pašalinti.

Kai aptarti kriterijai yra patikrinti ir tenkinami, galima ieškoti pagrindinių komponentų. Iš pradžių apskaičiuojamos tikrinės reikšmės ir tikrinių vektorių kovariacijų matrica. Svarbu, kad kintamieji būtų matuojami tais pačiais vienetais. Todėl, jei jie matuojami skirtingais vienetais, kintamuosius reikia centruoti ir normuoti.

Tarkime turime aibę d -mačių stebinių $\{z_i, i = 1, \dots, n\}$ imtį, kur kiekvienas stebiny z_i yra d -matis vektorius. Iš kiekvienos koordinatės atimamas atitinkamos koordinatės vidurkis:

$$\bar{z}^j = \frac{1}{n} \sum_{i=1}^n z_i^j, \quad (11)$$

čia $j = \overline{1, d}$.

Tuomet gauname pataisytus duomenis $v_i = z_i - (\bar{z}^1, \bar{z}^2, \dots, \bar{z}^d)$. Apskaičiuojame kovariacinę matricą:

$$cov(v^k, v^j) = \sum_{i=1}^n \frac{(v_i^k - \bar{v}^k)(v_i^j - \bar{v}^j)}{n - 1}, \quad (12)$$

čia $k, j = \overline{1, d}$, o dispersija apskaičiuojama:

$$DV^k = cov(v^k, v^k). \quad (13)$$

Sudaroma kovariacijų matrica $A_{d \times d}$ ir matricai randamos tikrinės reikšmės λ :

$$\det(A - \lambda I) x = 0. \quad (14)$$

Visoms tikrinėms reikšmėms λ randami tikriniai vektoriai x ir jie normuojami:

$$(A - \lambda I)x = 0. \quad (15)$$

Gauti tikriniai vektoriai ir yra pagrindinės komponentės, kurios tarpusavyje yra statmenos (37, 39).

Atlikus analizę ir radus pagrindines komponentes, kyla klausimas, kiek komponentių tikslinga pasilikti tolesnei analizei. Šio skaičiaus parinkimui taikomos kelios taisyklės. Viena iš jų – tikrinių reikšmių tikrinimas. Jei reikšmė didesnė už 1, tuomet komponentė laikoma reikšminga (komponentės indėlis didesnis nei vidutinis). Kita – komponentių bendrai sukauptos duomenų sklaidos apie vidurkį kiekis.

Norint įvykdyti pagrindinių komponentių analizę, kai duomenyse turima trūkstamų reikšmių, naudojamas *FactoMineR* papildinys (40). Jame realizuota procedūra *imputePCA*, kuri užpildo trūkstamas reikšmes pradiniam duomenų rinkinyje ir gražina naują rinkinį su visomis reikšmėmis. Iteratyvus pagrindinių komponentių analizės algoritmas užpildo trūkstamas reikšmes taip, kad naujosios reikšmės neturėtų jokios reikšminės įtakos tolesnių analizių rezultatui.

Pagrindinių komponentių analizės prielaidų tikrinimui „RStudio“ programiniame pakete naudojamas *stats* papildinys ir procedūra *bartlett.test*, MSA_i reikšmės apskaičiuojamos papildinio *psych* procedūra *KMO*, o *KMO* matas suskaičiuojamas R kalba parašyta funkcija, kuri pateikta su visu programos kodu priede nr. 1.

Pagrindinėms komponentėms išskirti naudojama *FactoMineR* papildinio procedūra *PCA*. Ji taip pat apskaičiuoja tikrines reikšmes ir kiekvienos komponentės sukauptą duomenų sklaidą.

2.2.5. Prognostinio modelio sudarymas remiantis kiekybiniais duomenimis

Tarptautinės migracijos srautų prognostinio modelio sudarymui bus taikomi regresinės analizės metodai.

Pagrindiniai regresinės analizės etapai:

- formuluojamas uždavinys;
- prognozei parenkami kintamieji;
- patikrinama, kaip susijęs priklausomas ir nepriklausomi kintamieji, parenkamas modelis;
- tikrinama, ar hipotetinis modelis yra tinkamas;
- tinkamas modelis naudojamas prognozei.

Tyrimo pradžioje išbandyta paprasta daugialypė tiesinė regresija.

Tiesinės daugialypės regresijos modelis:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (16)$$

Čia ε yra nepriklausomai vienodai pasiskirstę atsitiktiniai dydžiai, turintys normalųjį skirstinį su nuliniu vidurkiu ir dispersija σ^2 , $\beta_0, \beta_1, \dots, \beta_n$ – modelio koeficientai. Tiesinė regresija reiškia, kad kintamasis Y nuo kintamųjų X_1, X_2, \dots, X_n priklauso tiesiškai (33). Tačiau kadangi liekanų normalumo hipotezė nebuvo tenkinama, parinkti neparamestriniai regresinės analizės metodai (apibendrintas adityvus ir branduolinis modeliai).

Apibendrintas adityvus metodas

Tiriant neparamestrinius metodus, apibendrintų adityvių modelių metodas yra vienas universaliausių ir taikomų dažniausiai. Didelė dalis neparamestrinių metodų nėra efektyvūs dideliems kintamųjų kiekiams, kadangi duomenų sklaida didina įverčių poslinkį. Apibendrintų adityvių modelių metodas išsprendžia šią problemą. Metodas pasižymi tuo, jog nereikalauja įprastų parametrinų prielaidų, bet atskleidžia priklausomo ir nepriklausomų kintamųjų struktūrą.

Pagrindinės metodo savybės:

- apibendrintų adityvių modelių metodas tinka neparamestriniams ar pusiau parametriniam modeliams sudaryti;
- tinkamas daugiamačiams duomenims;
- vertina laisvės laipsnių skaičių ar suglodinimo parametą.

Apibendrintų adityvių modelių metode galima taikyti Gauso, Puasono, binominį ar Gama skirstinius.

Tarkime, Y yra priklausomas kintamasis, o X_1, X_2, \dots, X_p – nepriklausomų kintamųjų aibė. Standartinės tiesinės regresinės analizės modelis aprašytas 7 formule, kur $E(\varepsilon) = 0$ ir $D(\varepsilon) = \sigma^2$. Tuo tarpu apibendrintas adityvus modelis priklausomą kintamąjį aprašo:

$$Y = s_0 + s_1(X_1) + s_2(X_2) + \dots + s_p(X_p) + \varepsilon \quad (17)$$

kur $s_j(X), j = 1, 2, \dots, p$ yra glodinimo (angl. *smooth*) funkcija, $E(\varepsilon) = 0$ ir $D(\varepsilon) = \sigma^2$.

Kad būtų galima įvertinti, glodinimo funkcijos s_j turi tenkinti standartizuotas sąlygas, tokias kaip $E(s_j(X_j)) = 0$. Šios funkcijos vertinamos neparametriniu būdu (41).

Norint „RStudio“ pakete atlikti apibendrintą adityvią regresinę analizę, reikalingas *mgcv* papildinys ir jame realizuota procedūra *gam* (42). Procedūroje nurodoma kuriamo modelio lygtis. Kadangi metodas pusiau neparametrinis, galima daliai nepriklausomų kintamųjų nurodyti glodinimo funkciją, o dalį palikti tiesinius. Taip pat netiesiniams kintamiesiems galima nurodyti mazgų skaičių k , kuris nustato viršutinę laisvės laipsnių ribą, susijusią su glodiniu. Parenkamas *GCV.Cp* glodinimo parametro metodas, šios procedūros metu kartu vykdomas ir apibendrintas kryžminis patvirtinimas bei, modelyje papildomai nenurodžius glodumo parametru, juos bando parinkti optimaliai geriausius. Taigi, tikslus k skaičiaus parinkimas nėra kritiškai svarbus: turi būti pakankamai didelis, kad būtų užtikrintas pakankamas laisvės laipsnių skaičius, kad pakankamai gerai atvaizduotų „slepimą“ informaciją, bet pakankamai mažas, kad išlaikytų pagrįstą skaičiavimo efektyvumą. Akivaizdu, kad „didelio“ ir „mažo“ k parametro parinkimas priklauso nuo konkretaus sprendžiamo uždavinio. Vertinant sudaryto modelio korektiškumą, vykdoma procedūra *gam.check*. Šia procedūra patikrinama, ar modelyje optimaliai ir tinkamai parinkta k reikšmė. Jei kintamajam pastebima maža p reikšmė ($p < \alpha$), k dydis yra artimas *edf* (apskaičiuotas laisvės laipsnių skaičius), tai signalizuoja, jog k parinktas per mažas ir jį reikėtų didinti, priešingai – modelis tinkamas.

Branduolinė regresija

Branduolinė regresija – dar vienas neparametrinis metodas, skirtas netiesinio ryšio radimui tarp atsitiktinių dydžių X ir Y . Šio metodo tikslas – apskaičiuoti ir panaudoti tinkamus svorius $w_{ij(\ker)}$:

$$Y_{i(\ker)} = \sum_{j=1}^n w_{ij(\ker)} Y_j . \quad (18)$$

Kiekvienam duomenų taškui n priskiriamas individualus svoris $w_{ij(\ker)}, j = 1, 2, \dots, n$, pritaikant jį kiekvienam taškui X_i (arba pritaikant individualius svorius $w_{0j(\ker)}$ taškui X_0).

Svorių nustatymui dažniausiai taikoma tokia formulė:

$$w_{ij(\ker)} = \frac{K \left[\frac{x_i - x_j}{h} \right]}{\sum_{j=1}^n K \left[\frac{x_i - x_j}{h} \right]} , \quad (19)$$

čia $u = \frac{x_i - x_j}{h}$, tada $K(u)$ yra mažėjanti u funkcija, o $h > 0$ vadinama branduolio pločio glodinimo parametru. Glodinimo plotis nustato, kaip greitai mažėja svoriai, kai atstumas nuo taško X_0 didėja (43). Branduolio pločio glodinimo parametrai gali būti parenkami keliais metodais: fiksuotas, vienodas h parametras visiems stebiniams, ir kintamo pločio, kiekvienam stebiniui priskiriant po branduolį (priešingai nei fiksuoto pločio atveju, plotis kiekvienam stebiniui nėra pastovus).

Vykdamt branduolinę regresinę analizę „RStudio“ pakete reikalingas np papildinys ir jame realizuota procedūra $npreg$ (44). Procedūroje nurodoma kuriamo modelio lygtis. Nustatoma glodinimo parametro parinktis „ $cv.aic$ “. Vertinant modelio korektiškumą, vykdoma procedūra $npsigtest$, kurios pagalba nustatomi reikšmingi modelio kintamieji.

Modelių vertinimas

Vertinant sudarytų regresinės analizės metodų tikslumą, duomenys padalinami į apmokymo ir tikrinimo imtis (80 proc. apmokymui, 20 proc. tikrinimui). Kiekvienam modeliui apskaičiuojamos paklaidų metrikos (MSE , $RMSE$, MAE) ir determinacijos koeficientai (R^2).

Vidutinė kvadratinė paklaida MSE (angl. *mean square error*) parodo, kiek vidutiniškai modelio įverčiai skiriasi nuo tikrųjų reikšmių:

$$MSE = \frac{1}{n} \sum_{t=1}^n (x_s - x_p)^2. \quad (20)$$

Kvadratinė šaknis iš vidutinės kvadratinės paklaidos $RMSE$ (angl. *root mean square error*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_s - x_p)^2}. \quad (21)$$

Vidutinė absoliutinė paklaida MAE (angl. *mean absolut error*) parodo, kiek vidutiniškai modelio įvertis yra artimas tikrajai reikšmei:

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_s - x_p|. \quad (22)$$

Determinacijos koeficientas R^2 nusako priklausomo ir nepriklausomų kintamųjų priklausomybės stiprumą:

$$R^2 = 1 - \frac{\sum (x - x_p)^2}{\sum (x - \bar{x})^2}, \quad (23)$$

čia x_s – stebėta reikšmė, x_p – prognozuota reikšmė, \bar{x} – vidutinė visų stebinių reikšmė.

Paklaidų vertinimui „RStudio“ programinėje įrangoje taikomas plėtinys *caret* ir procedūra *postResample*, kurioje nurodomos suprognozuotos ir tikrosios reikšmės.

3. Tyrimų rezultatai ir jų aptarimas

Šiame skyriuje aprašyta Europos Sąjungos šalių tarptautinei migracijai įtaką darančių veiksnių tyrimo eiga ir rezultatai.

3.1. Duomenų matricos formavimas ir žvalgomoji analizė Europos sąjungos šalių pavyzdžiu

Baigiamojo magistro projekto darbui duomenys surinkti iš Europos statistikos departamento (Eurostato) ir Pasaulio banko (angl. *World bank data*) duomenų bazių. Analizuoti 2000–2016 metais sukaupti duomenys. Duomenų rinkinyje surinkta informacija apie 28 Europos Sąjungos valstybių kiekvienų metų imigrantų, emigrantų ir gyventojų skaičių (tolesniame tyrime naudota kitų kintamųjų perskaičiavimui) ir ekonominės ir socialinės gerovės veiksnius (pagal išskirtus 1.2 lentelėje). Visgi duomenų analizei iš viso atrinkti 147 įvairūs veiksniai, kurie gali turėti įtakos migracijos srautams. Tai yra daugiau veiksnių, nei jų buvo identifikuota analizuojant mokslinę literatūrą. Detalus veiksnių sąrašas naudotas tyrime pateiktas priede nr. 5.

Iš duomenų bazių surinkus ir sujungus duomenis, „RStudio“ programiniu paketu atlikta šių duomenų žvalgomoji analizė, siekiant išsiaiškinti kintamųjų savybes.

Visiems kintamiesiems paskaičiuotos skaitinės charakteristikos – mažiausia, vidutinė, didžiausia reikšmė, kvartilai (priedas nr. 1). Nustatyta, kad didžiosios dalies kintamųjų sklaida nėra didelė (kintamieji apskaičiuoti procentine išraiška), tačiau keletas jų (mokslinių tyrimų ir taikomosios technikos specialistai, suaugę žmonės, naujai užsikrėtę ŽIV) pateikti vienetais. Pastarieji veiksniai, naudojantis gyventojų skaičiumi atitinkamoje šalyje atitinkamu metu, perskaičiuoti į procentinius dydžius. Žvalgomosios analizės metu taip pat pastebėta, kad duomenyse yra trūkstamų reikšmių, o peržiūrėjus kiekvieno kintamojo histogramas (kintamieji pasiskirstę ne pagal normalųjį dėsnį), pateiktas priede nr. 2, nustatyta, kad klasikiniai mažiausių kvadratų ar didžiausio tikėtimumo metodai nebus tinkami.

Detaliau apžvelgiamas priklausomas kintamasis – migracijos neto. 3.1 lentelėje pateikiamos kiekvienos tiriamos šalies 17 stebinių (po vieną per metus) migracijos srautų minimalios, maksimalios, vidutinės reikšmės ir standartiniai nuokrypiai. Galima pastebėti, jog šešios valstybės (Lietuva, Latvija, Rumunija, Estija, Lenkija ir Bulgarija) pasižymi neigiamu migracijos srauto vidurkiu. Keturios iš jų per visą laikotarpį nėra turėjusios didesnės imigracijos, nei emigracijos. Likę 22 šalys išsiskiria teigiamu vidurkiu. Dešimt iš jų (Kroatija, Graikija, Portugalija, Čekija, Slovėnija, Anglija, Vokietija, Airija, Ispanija, Kipras) per laikotarpį yra turėjusios tiek didelę emigraciją, tiek imigraciją. Tuo tarpu kitos 12 valstybių

visais 17 metų išsiskyrė savybe pritraukti migrantų srautus ir jų emigracija neviršydavo imigracijos. Atsižvelgiant į standartinį nuokrypį – didžiausias fiksuotas Kiprui. Šalis turėjusi tiek ypač didelę emigraciją, tiek labai didelę imigraciją.

3.1 lentelė. Migracijos srautų apžvalga

Šalis	Mažiausia reikšmė	Didžiausia reikšmė	Vidutinė reikšmė	Standartinis nuokrypis
Lietuva	-24,81	-3,36	-8,73	4,97
Latvija	-16,81	-3,60	-7,33	3,79
Rumunija	-7,94	0,00	-1,60	2,25
Estija	-2,78	1,83	-1,14	1,13
Lenkija	-2,86	0,40	-0,88	0,72
Bulgarija	-1,30	0,00	-0,17	0,33
Kroatija	-5,36	3,94	0,00	2,40
Graikija	-6,00	5,78	0,41	3,50
Slovakija	0,17	1,31	0,59	0,32
Prancūzija	0,00	1,77	0,62	0,56
Portugalija	-3,54	6,55	0,80	2,84
Čekija	-2,75	8,19	1,46	2,60
Vengrija	0,43	2,79	1,68	0,53
Slovėnija	-0,25	9,24	2,15	2,65
Suomija	0,50	3,33	2,20	0,89
Olandija	0,55	4,58	2,26	1,33
Danija	0,90	5,98	2,57	1,32
Anglija	-5,46	5,12	2,59	2,25
Vokietija	-0,68	14,74	3,23	3,55
Belgija	0,00	6,39	3,32	1,80
Italija	2,19	8,18	4,57	1,95
Austrija	1,93	12,77	4,67	2,68
Airija	-5,66	22,58	5,48	8,32
Švedija	2,77	11,89	5,48	2,33
Ispanija	-5,38	16,33	5,85	7,24
Malta	0,00	22,38	7,19	7,66
Kipras	-17,28	27,13	10,24	12,66
Liuksemburgas	5,97	21,50	14,21	4,61

3.2. Europos Sąjungos šalių grupavimas į homogenines grupes

Siekdami bendresnių modelių, tiriamas šalis pagal bendrinius bruožus sujungiame į keletą grupių, kurios vėliau bus tiriamos atskirai. Šalių grupavimas vykdomas keliais būdais.

3.2.1. Grupavimas pagal BVP rodiklį vienam gyventojui

Bendrasis vidaus produktas (BVP) – vienas pagrindinių rodiklių, parodančių už kokią vertę per tam tikrą laikotarpį šalyje buvo sukurta prekių ir paslaugų (45). BVP išraiška vienam gyventojui parodo šalies ekonominę išsivystymą: kuo šalies ekonomika stipresnė, tuo šis rodiklis didesnis.

Remiantis autorių kolektyvo Ramirez'o, Kumpikaitės-Valiūnienės ir Guenca-Garcia straipsniu (9), šalys pagal BVP rodiklį vienam gyventojui padalintos į 2 lygias grupes. Taip išskirtos skurdžių ir turtingų šalių kategorijos. Analizei naudoti paskutinių penkerių turimų (2012–2016) metų BVP vienam gyventojui vidurkiai.

3.2 lentelė. Skurdžių ir turtingų ES šalių išskyrimas pagal BVP

Skurdžios šalys		Turtingos šalys	
Šalis	BVP vienam gyventojui, Eur	Šalis	BVP vienam gyventojui, Eur
Bulgarija	5 600	Kipras	21 020
Rumunija	7 040	Ispanija	22 660
Latvija	10 340	Italija	25 660
Kroatija	10 500	Anglija	30 780
Lenkija	10 560	Prancūzija	31 360
Vengrija	10 640	Belgija	33 940
Lietuva	11 200	Vokietija	33 980
Estija	13 120	Suomija	34 480
Slovakija	13 720	Austrija	36 260
Čekija	15 640	Olandija	38 400
Portugalija	16 380	Švedija	40 900
Graikija	17 040	Airija	44 380
Slovėnija	17 640	Danija	44 920
Malta	18 020	Liuksemburgas	79 760

Į skurdžių šalių kategoriją patekusių valstybių BVP rodiklio vidurkis – 12 674 eurų, standartinis nuokrypis – 3 768 eurų, į turtingųjų šalių kategoriją patekusių valstybių BVP rodiklio vidurkis – 37 036 eurų, standartinis nuokrypis didesnis, nei pirmosios grupės – 13 753 eurų.

3.2.2. Grupavimas pagal migracijos srautus ir jų pokytį

Migracijos srautų grupavimas pagal BVP rodiklį yra gana subjektyvus, kadangi slenkstis, nuo kada šalis laikoma turtinga, parenkamas kiekvieno individualia ekspertine įžvalga. Dėl šios priežasties šalys taip pat suskirstytos ir pagal migracijos neto matą – į stumiančias ir traukiančias migrantus. Šio metodo

pagalba Europos Sąjungos valstybės padalintos į 3 grupes: šalis, turinčias neigiamą migracijos neto, šalis turinčias teigiamą migracijos neto, ir šalis, kurioms migracijos neto pasikeitė iš teigiamo į neigiamą ar atvirkščiai. Grupavimas atliktas vertinant paskutiniųjų penkerių turimų metų duomenis (2012 – 2016 metai). Rezultatai pateikti 3.3 lentelėje. Neigiamas migracijos srautas nustatytas šešioms, pasikeitęs – aštuonioms, teigiamas – 14 valstybių.

3.3 lentelė. Migracijos srautų pokyčių lentelė

Šalis	Metai					T - teigiama N - neigiama P - pokytis
	2012	2013	2014	2015	2016	
Lietuva	-7,08	-5,66	-4,19	-7,67	-10,45	N
Latvija	-5,80	-7,05	-4,32	-5,36	-6,21	N
Kroatija	-0,92	-1,15	-2,41	-4,25	-5,36	N
Rumunija	-0,15	-0,41	-1,85	-3,12	-3,55	N
Bulgarija	-0,34	-0,15	-0,29	-0,59	-1,30	N
Lenkija	-1,53	-1,47	-1,21	-1,07	-0,74	N
Portugalija	-3,54	-3,45	-2,88	-1,01	-0,81	N
Slovėnija	0,31	0,24	-0,24	0,25	0,51	P
Estija	-2,78	-1,99	-0,56	1,83	0,78	P
Graikija	-6,00	-5,38	-4,37	-4,14	0,96	P
Čekija	-1,12	0,40	0,14	0,37	2,39	P
Kipras	-0,73	-13,95	-17,28	-2,36	2,95	P
Airija	-4,46	-2,39	0,52	2,91	4,89	P
Ispanija	-3,04	-5,38	-2,04	-0,04	1,88	P
Slovakija	0,63	0,44	0,32	0,58	0,72	T
Prancūzija	1,10	1,51	0,48	1,03	1,02	T
Vengrija	1,09	0,43	1,25	1,53	1,40	T
Italija	4,12	3,04	2,32	2,19	2,37	T
Belgija	3,24	1,56	2,56	5,06	2,76	T
Suomija	3,23	3,33	2,94	2,27	3,07	T
Anglija	2,78	3,27	4,86	5,12	3,80	T
Danija	1,93	3,03	4,26	5,98	3,81	T
Olandija	0,84	1,00	1,93	3,23	4,58	T
Vokietija	4,38	5,38	6,94	14,74	6,04	T
Austrija	4,73	5,65	7,38	12,77	7,48	T
Švedija	5,41	6,82	7,85	8,04	11,89	T
Liuksemburgas	19,12	19,27	20,10	19,82	16,39	T
Malta	10,18	14,48	21,76	22,38	19,42	T

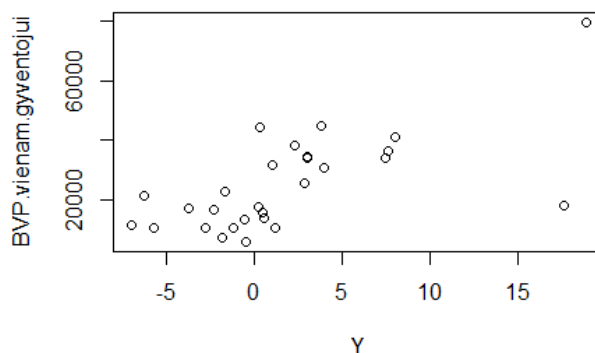
Apjungiant 3.2 ir 3.3 lentelių rezultatus pastebima, kad griežtai neigiamu migracijos srautu pasižyminčios šalys yra priskiriamos skurdžiųjų grupei ir turi vienus žemiausių BVP vienam gyventojui rodiklių visoje Europos Sąjungoje. Šalys, kurių migracijos srautų rodiklis kintantis, yra iš dvejų grupių – tiek turtingųjų, tiek skurdžiųjų kategorijų. Tuo tarpu 11 šalių iš 14, turinčių griežtai teigiamą migracijos

srautą, yra priskiriamos turtingų valstybių grupei. Šie rezultatai leidžia patvirtinti dažniausiai literatūroje aptariamus ir minimus vienus svarbiausių migracijos motyvų – ekonominius veiksnius. Žmonės yra linkę emigruoti iš mažiau turtingų šalių į pasižyminčias geresne ekonomine situacija.

3.2.3. K-vidurkių klasterizacija

Norėdami šalis sujungti į kuo labiau homogeniškesnes grupes, jas klasterizuojame pagal migracijos srautų (Y) ir BVP vienam gyventojui penkerių metų vidurkius (2012–2016).

Pirmasis tyrimo etapas – grafinis klasterizuojamų duomenų atvaizdavimas.



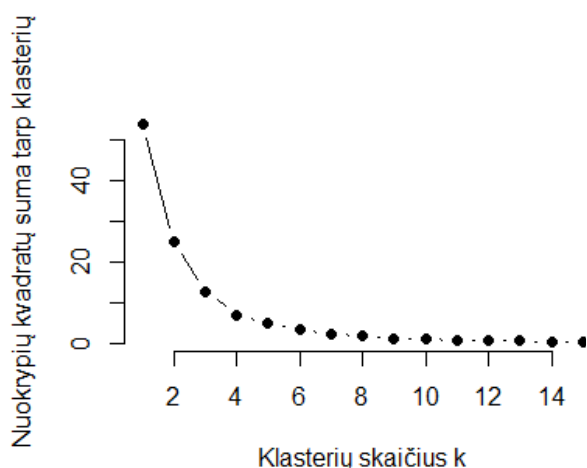
3.1 pav. BVP vienam gyventojui ir Y (migracijos neto) pasiskirstymas

Iš 3.1 paveikslo galima pastebėti, jog duomenys yra linkę jungtis į homogeniškas grupes ir galima išvelgti nuo dviejų didelių iki šešių ar net daugiau mažesnių klasterių. Taip pat pastebima, jog kintamųjų Y ir BVP vienam gyventojui mastelis yra skirtingas. Dėl šios priežasties toliau, parenkant optimalų klasterių skaičių, bus naudojami standartizuoti duomenys.

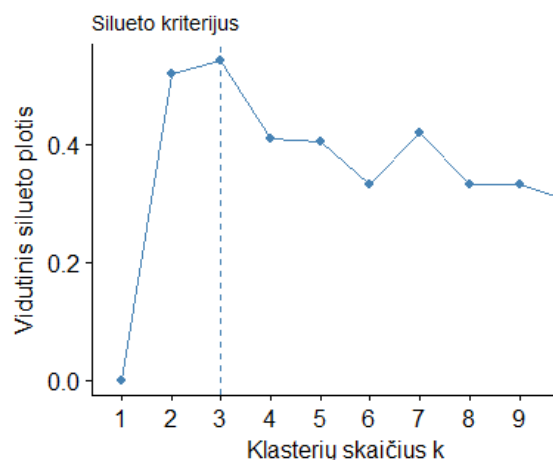
Prieš klasterių skaičiaus parinkimą numatyta, kad klasterizavimui taikysime k-vidurkių metodą. Taip pat sieksime sudaryti ne daugiau nei aštuonis skirtingus klasterius.

Pirmasis metodas, kuriuo tikrinamas optimalus klasterių skaičius skirstant 28 Europos Sąjungos valstybes į homogenines grupes, realizuotas taikant alkūnės metodą (angl. *Elbow Method*). Šiuo metodu ieškoma nuokrypių kvadratų sumos tarp klasterių parametro alkūnė (lūžio taškas), kuri ir identifikuoja optimalų klasterių skaičių. Pastaruoju atveju šis skaičius yra 4 (žr. 3.2 pav.). Antras metodas, kuriuo tikrintas duomenų klasterių skaičius, Silueto kriterijus. Iš 3.3 paveikslo galima matyti, kad duomenys

tinkami klasterizavimui. Didžiausia gauta Silueto kriterijaus reikšmė, didesnė nei 0,4, gauta ties 3 klasteriais, kurie ir yra optimalus skaičius taikant šį metodą.



3.2 pav. Optimalus klasterių skaičius – alkūnės metodas



3.3 pav. Optimalus klasterių skaičius – Silueto kriterijus

Kadangi skirtingais kriterijais gaunami skirtingi rezultatai, optimalaus klasterių skaičiaus parinkimui iš viso patikrinti 26 kriterijai: *Duda, Beale, Ratkowsky, Dunn, Scott, Marriot, TraceW, DB, Siluteto, Ball, PtBiserial, Gap, McClain, Tau, SDindex, Rubin, KL, CH, CCC, TrCovW, Gamma, Gplus, Hartigan, Friedman, Cindex, SDbw* (29). Iš jų:

- 3 kriterijai optimaliu skaidymu parinko 2 klasterius;
- 12 kriterijų optimaliu skaidymu parinko 3 klasterius;
- 6 kriterijai optimaliu skaidymu parinko 5 klasterius;
- 5 kriterijai optimaliu skaidymu parinko 8 klasterius.

Kiekvienu kriterijumi gauto optimalaus klasterių skaičiaus rezultatai pateikti priede nr. 4. Taigi, pagal daugumos taisyklę, įvertinus visus tirtus kriterijus, optimalus grupių skaičius 28 valstybėms yra 3. „RStudio“ programinėje įrangoje tyrimui naudotas *NbClust* plėtinys, kuriame parinkus klasterizavimą pagal k-vidurkių metodą, valstybės buvo sugrupuotos į homogenines grupes. Grupavimas paaiškina 76,1 proc. duomenų sklaidos. Valstybių jungimo į homogenines grupes rezultatai pateikti 3.4 lentelėje.

3.4 lentelė. Optimalus ES šalių grupavimas pagal migracijos srautus ir BVP

Europos Sąjungos šalių grupės					
1 grupė			2 grupė		3 grupė
Bulgarija	Kipras	Portugalija	Airija	Malta	Liuksemburgas
Čekija	Kroatija	Rumunija	Anglija	Olandija	
Estija	Latvija	Slovakija	Austrija	Prancūzija	
Graikija	Lenkija	Slovėnija	Belgija	Suomija	
Ispanija	Lietuva	Vengrija	Danija	Švedija	
			Italija	Vokietija	

Galima pastebėti, jog didžiausia sudaryta grupė (1 grupė) yra iš 15 valstybių. Peržvelgus šių šalių rodiklius matoma, kad jos išsiskiria žemu BVP vienam gyventojui ir dažniausiai neigiamu (arba labai žemu teigiamu) migracijos srautų vidurkiu. Vidutinis BVP vienam gyventojui šioje grupėje – 13 540 eurų, o vidutinis migracijos srautų vidurkis yra -2,1.

Antrą pagal dydį grupę (2 grupė) sudaro 12 valstybių. Pastarosios išsiskiria labai aukštu BVP vienam gyventojui bei teigiamu migracijos srautų vidurkiu. Vidutinis BVP vienam gyventojui šioje grupėje yra 34 423 eurai (3 kartus didesnis, nei pirmoje), o vidutinis migracijos srautų vidurkis lygus 5,08. Galima būtų išskirti nebent Malta, kurios BVP rodiklis žemiausias iš 12 valstybių, tačiau migracijos srautas yra antras pagal didumą iš visų tiriamų 28 valstybių.

Trečią, pačią mažiausią grupę, sudaro tik viena valstybė – Liuksemburgas. Ši šalis yra lyg išskirtis visoje imtyje. Ji iš visų ES šalių išsiskiria pačiu didžiausiu BVP vienam gyventojui (beveik dvigubai didesniu, nei antroje vietoje esanti Danija) ir didžiausiu migracijos srautų vidurkiu. Liuksemburgas išsiskiria ne tik Europos Sąjungos kontekste, tačiau ir visame pasaulyje (net 40 proc. valstybės pridėtinės vertės sukuria bankai). 2016 metų duomenimis tai antra didžiausią BVP vienam gyventojui turinti valstybė (po Kataro) visame pasaulyje. Dėl šių priežasčių Liuksemburgas baigiamajame magistro darbe bus laikomas išskirtimi ir toliau nebus analizuojamas.

Apibendrinant gautus klasterizavimo rezultatus, pirmai ir antrai šalių grupėms suteikiame pavadinimus atitinkamai – neturtingos ir turtingos valstybės.

3.3. Tarptautinės migracijos srautams įtaką darančių veiksnių koreliacinė analizė

Išskyrę Europos Sąjungos šalis į homogenines grupes, tolesnei analizei pasirenkame 3.2.3 skyrelyje gautus klasterizavimo rezultatus ir kiekvieną grupę, neturtingų ir turtingų šalių, tiriame atskirai.

Prieš analizę normuojame visus nepriklausomus kintamuosius. Kiekvienai grupei nustatome būtent jai reikšmingus veiksnius. Kiekvienam kintamajam patikrinamas trūkstatų reikšmių skaičius ir jei jis didesnis nei 50 proc., tuomet kintamasis šalinamas. Tokiu būdu neturtingų šalių grupei pašalinti 21 kintamasis, o Y koreliacija skaičiuojama su likusiu 126 kintamaisiais. Turtingų šalių grupei pašalinti 22 kintamieji, o Y koreliacija skaičiuojama su likusiais 125 kintamaisiais. Iš žvalgomosios duomenų analizės pastebėta, kad kintamieji nepasižymi normaliuoju skirstiniu, todėl skaičiuojamas Spirmeno ranginės koreliacijos koeficientas. Reikšmingais veiksniais laikomi tuomet, kai koreliacijos koeficiento reikšmė $|cor| > 0,30$, o reikšmingumo lygmuo $p < \alpha = 0,05$. Toliau lentelėse (3.5 ir 3.6) pateikiami statistiškai reikšmingi migracijos srautams įtaką turintys veiksniai, surikiuoti pagal koreliacijos stiprumą: nuo didžiausio neigiamo iki didžiausio teigiamo.

Neturtingų šalių grupei nustatyti reikšmingi kintamieji:

3.5 lentelė. Tarptautinės migracijos srautams statistiškai reikšmingų veiksnių koreliacijos neturtingų šalių grupei

Kintamojo grupė	Kintamasis	Y	p_reikšmė
Socialinė gerovė	x115 Aukštasis išsilavinimas, akademinis personalas	-0,697	0,0000
Socialinė gerovė	x94 Nauji ŽIV atvejai	-0,520	0,0000
Socialinė gerovė	x99 Gyventojai, moterys	-0,520	0,0000
Socialinė gerovė	x34 Mirtingumo lygis	-0,488	0,0000
Socialinė gerovė	x79 Tuberkuliozės paplitimas	-0,476	0,0000
Socialinė gerovė	x147 Gyventojai be dušo/vonios namuose	-0,452	0,0000
Socialinė gerovė	x131 Nedarbo lygis	-0,409	0,0000
Ekonominė gerovė	x6 Koreguoti taupymo būdai: mineralų išekvojimas	-0,364	0,0000
Socialinė gerovė	x146 Gyventojai mažo intensyvumo darbo zonoje	-0,361	0,0000
Socialinė gerovė	x78 ŽIV paplitimas	-0,334	0,0001
Socialinė gerovė	x12 Amžių priklausomybės santykis	-0,332	0,0000
Socialinė gerovė	x143 Skyrybų lygis	-0,332	0,0000
Ekonominė gerovė	x7 Gamtinių išteklių išekvojimas	-0,312	0,0000
Socialinė gerovė	x145 Skyrybų ir santuokų santykis	-0,303	0,0000
Ekonominė gerovė	x90 Gaminių eksportas	0,303	0,0000
Ekonominė gerovė	x66 Valdžios sektoriaus vartojimo išlaidų augimas	0,311	0,0000
Socialinė gerovė	x17 Antiretrovirusinio gydymo aprėptis	0,323	0,0002
Socialinė gerovė	x117 Reikalingas laikas, norint įregistruoti turtą	0,323	0,0000
Socialinė gerovė	x51 Energijos importas	0,325	0,0000
Ekonominė gerovė	x103 Pelno mokestis	0,332	0,0000
Socialinė gerovė	x133 Žmonės su viduriniu arba žemesniu išsilavinimu	0,335	0,0000
Socialinė gerovė	x130 Vidutinės pajamos	0,343	0,0000
Socialinė gerovė	x77 Geresnio vandens šaltinis	0,347	0,0000
Ekonominė gerovė	x91 Gaminių importas	0,361	0,0000
Socialinė gerovė	x122 Tuberkuliozės atvejų aptikimo sparta	0,365	0,0000
Socialinė gerovė	x108 Kaimo vietovės gyventojų augimas	0,385	0,0000
Socialinė gerovė	x46 Darbdaviai	0,449	0,0000
Ekonominė gerovė	x31 Verslo pradžios procedūrų sąnaudos	0,451	0,0000
Ekonominė gerovė	x138 BVP vienam gyventojui	0,458	0,0000
Socialinė gerovė	x95 Nuolatinė dirbama žemė	0,473	0,0000
Socialinė gerovė	x76 Patobulintos sanitarinės priemonės	0,492	0,0000
Socialinė gerovė	x100 Gyventojai, vyrai	0,520	0,0000
Socialinė gerovė	x127 Miesto gyventojų skaičiaus augimas	0,705	0,0000
Socialinė gerovė	x98 Gyventojų skaičiaus augimas	0,735	0,0000

Iš rezultatų matoma, kad neturtingų šalių grupės migracijos srautams nustatyti 34 reikšmingi kintamieji. 14 iš jų turi neigiamą koreliacijos koeficientą ir yra stumiantys išvykti veiksniai, 20 – teigiamą ir yra traukiantys atvykti veiksniai. 26 veiksniai susiję su socialine, o 8 su ekonomine gerove.

Pastebima, kad skurdesnėse šalyse daugiau emigruoja aukštą išsilavinimą turintys gyventojai, akademinis personalas. Taip pat emigracijai aktualūs sveikatos rodikliai – didesnis sergamumas ŽIV, tuberkulioze ir blogos sanitarinės sąlygos, darbo rinkos veiksniai – išvykti skatina mažas užimtumą ir nedarbas.

Pirmos grupės šalims imigraciją didina gyventojų skaičiaus augimas, geriau išvystyta medicininė priežiūra (greitesnis ir tikslesnis ligų aptikimas). Imigruoti skatina ir geresnės verslo steigimo sąlygos, didesnis darbdavių skaičius, aukštesnės vidutinės pajamos. Pastebima, kad skurdesnių šalių grupėje daugiau imigruoja žmonės su viduriniu ar žemesniu išsilavinimu.

Turtingų šalių grupei nustatyti reikšmingi kintamieji:

3.6 lentelė. Tarptautinės migracijos srautams statistiškai reikšmingų veiksnių koreliacijos turtingų šalių grupei

Kintamojo grupė	Kintamasis	Y	p_reikšmė
Socialinė gerovė	x131 Nedarbo lygis	-0,345	0,0000
Socialinė gerovė	x18 Dirbamoji žemė	-0,309	0,0000
Socialinė gerovė	x108 Kaimo vietovių gyventojų skaičiaus augimas	0,303	0,0000
Socialinė gerovė	x114 Mokslinių tyrimų ir taikomosios technikos specialistai	0,323	0,0009
Socialinė gerovė	x11 Nesimokantys paaugliai	0,375	0,0000
Socialinė gerovė	x98 Gyventojų skaičiaus augimas	0,454	0,0000

Iš rezultatų matoma, kad turtingų šalių grupės migracijos srautams nustatyti 6 reikšmingi kintamieji, 2 iš jų turi neigiamą koreliacijos koeficientą ir yra stumiantys išvykti veiksniai, 4 – teigiamą ir yra traukiantys atvykti veiksniai. Taip pat pastebima, kad visi 6 išskirti veiksniai yra susiję su socialine gerove.

Turtingesnių šalių grupėje didėjantis nedarbas ir dirbamos žemės trūkumas neigiamai veikia migracijos srautus ir didina išvykstančiųjų skaičių.

Į pasiturinčias šalis daugiausia imigruoja arba mokslų netęsiantys jaunuoliai, arba aukštą išsilavinimą turintys asmenys. Todėl logiška, kad gyventojų skaičiaus augimas taip pat koreliuoja su imigracija.

Apibendrinant neturtingų ir turtingų šalių grupių rezultatus (3.5 ir 3.6 lenteles), pastebima, kad nedarbo lygis yra emigraciją vienijantis veiksnys, o natūralus gyventojų skaičiaus didėjimas – imigraciją skatinantis faktorius. Taip pat pastebėta, kad vargingesnėse valstybėse geras išsilavinimas yra išstumiantis veiksnys, o turtingesnėse – pritraukiantis. Socialinės gerovės veiksniai aktualūs tiek neturtingoms, tiek turtingoms šalims, tuo tarpu ekonominės gerovės rodikliai aktualūs tik neturtingų valstybių grupei.

3.4. Tarptautinės migracijos srautams įtaką darančių veiksnių pagrindinių komponentių analizė

Siekdami sumažinti duomenų dimensijas, iš atrinktų reikšmingų kintamųjų kiekvienai tiriamai grupei išskirsime pagrindines komponentes.

Prieš analizę tikslinga patikrinti kiekvieno veiksnio tinkamumą metodui. Stebinių tinkamumą galima nustatyti pagal MSA_i (angl. *Kaiser measure of sampling adequacy*) reikšmes. Reikšmingais laikomi tie kintamieji, kurių $MSA_i > 0,5$. Kintamųjų tinkamumą pagrindinių komponentių analizei padeda įvertinti keletas kriterijų. Vienas iš jų – Kaizerio-Mejerio-Olkino (KMO) (angl. *Keiser–Meyer–Olkin measure*) matas, kuris nustato, ar tarp kintamųjų porų koreliacijos yra paaiškinamos kitais kintamaisiais. Kitas – Bartleto sferiškumo kriterijus (angl. *Bartlett's test of sphericity*), kuris leidžia nustatyti, ar tarp stebimų kintamųjų egzistuoja statistiškai reikšminga koreliacija (32).

Neturtingų šalių grupė

Analizuojami neturtingų šalių grupei reikšmingi kintamieji, aprašyti 3.3 skyrelyje. Peržvelgus MSA_i reikšmes (3.7 lentelė) pastebima, jog visi kintamieji yra tinkami tolesnei analizei, nes visų jų reikšmės didesnės nei 0,5.

3.7 lentelė. Kaiserio-Meyerio-Olkino reikšmių lentelė neturtingų šalių grupei

Kaiserio-Meyerio-Olkino kriterijus	MSA
Bendra MSA = 0,83	reikšmės
Kintamieji:	kiekvienam
	kintamajam:
Koreguoti taupymo būdai: mineralų išekvojimas	0,88
Gamtinių išteklių išekvojimas	0,80
Amžių priklausomybės santykis	0,69
Antiretrovirusinio gydymo aprėptis	0,89
Verslo pradžios procedūrų sąnaudos	0,88
Mirtingumo lygis	0,88
Darbdaviai	0,84
Energijos importas	0,82
Valdžios sektoriaus vartojimo išlaidų augimas	0,60
Patobulintos sanitarinės priemonės	0,85
Geresnis vandens šaltinis	0,74
ŽIV paplitimas	0,78
Tuberkuliozės paplitimas	0,82
Eksportas	0,61
Gaminių importas	0,68
Nauji ŽIV atvejai	0,84
Nuolatinė dirbama žemė	0,79
Gyventojų skaičiaus augimas	0,90
Gyventojai, moterys	0,88
Gyventojai, vyrai	0,88
Pelno mokestis	0,95
Kaimo gyventojų augimas	0,78
Aukštasis išsilavinimas, akademinis personalas	0,92
Reikalingas laikas, norint įregistruoti turtą	0,77
Tuberkuliozės atvejų aptikimo sparta	0,85
Miesto gyventojų skaičiaus augimas	0,90
Vidutinės pajamos	0,90
Nedarbo lygis	0,51
Žmonės su viduriniu ir žemesniu išsilavinimu	0,64
BVP vienam gyventojui	0,83
Skyrybų lygis	0,70
Skyrybų ir santuokų santykis	0,66
Gyventojai mažo intensyvumo darbo zonoje	0,70
Gyventojai be dušo/vonios namuose	0,90

Taip pat šiam duomenų rinkiniui paskaičiuotas *KMO* matas ir Bartleto sferiškumo kriterijus. Rezultatai pateikti 3.8 lentelėje.

3.8 lentelė. Bartleto sferiškumo kriterijus ir KMO matas neturtingų šalių grupei

Bartleto sferiškumo kriterijus
Bartleto testas. K -kvadratai = 98322, $df = 33$, p -reikšmė $< 2,2 \cdot 10^{-16}$
df – laisvės laipsniai
KMO = 0,8056427

Bartleto sferiškumo kriterijaus reikšmė p -reikšmė $< 0,05$, todėl hipotezė „visi stebimi kintamieji yra nekoreliuoti“ atmesta. KMO matas lygus 0,806, tai reiškia, jog duomenys gerai tinka analizei. Galime teigti, jog pagrindinių komponentių analizė šiam duomenų rinkiniui yra galima ir turi prasmę.

Toliau analizėje taikomas pagrindinių komponentių išskyrimo metodas. Nustatant komponentių skaičių remiamasi tikrinėmis reikšmėmis ir kiekvienos komponentės sukaupta duomenų sklaidos dalimi.

Iš pradžių nustatome kokią duomenų sklaidos dalį lemia kiekviena iš 34 komponentių (angl. *proportion of variance*) ir kokią sklaidos dalį komponentės lemia kartu (angl. *cumulative proportion*). Rezultatai 3.9 lentelėje.

3.9 lentelė. Pagrindinių komponentių sklaidos lentelė

	Standartinis nuokrypis	Komponentės sukaupta sklaida	Bendrai sukaupta sklaida		Standartinis nuokrypis	Komponentės sukaupta sklaida	Bendrai sukaupta sklaida
PC1	3,5338	0,3673	0,3673	PC18	0,4814	0,0068	0,9671
PC2	2,2375	0,1472	0,5145	PC19	0,4579	0,0062	0,9733
PC3	1,9110	0,1074	0,6220	PC20	0,4230	0,0053	0,9786
PC4	1,3615	0,0545	0,6765	PC21	0,3550	0,0037	0,9823
PC5	1,3247	0,0516	0,7281	PC22	0,3292	0,0032	0,9855
PC6	1,1721	0,0404	0,7685	PC23	0,3044	0,0027	0,9882
PC7	0,9990	0,0294	0,7979	PC24	0,2945	0,0026	0,9907
PC8	0,9741	0,0279	0,8258	PC25	0,2496	0,0018	0,9926
PC9	0,8942	0,0235	0,8493	PC26	0,2373	0,0017	0,9942
PC10	0,8632	0,0219	0,8712	PC27	0,2136	0,0013	0,9956
PC11	0,8050	0,0191	0,8902	PC28	0,1957	0,0011	0,9967
PC12	0,7678	0,0173	0,9076	PC29	0,1659	0,0008	0,9975
PC13	0,6781	0,0135	0,9211	PC30	0,1645	0,0008	0,9983
PC14	0,6247	0,0115	0,9326	PC31	0,1502	0,0007	0,9990
PC15	0,5977	0,0105	0,9431	PC32	0,1454	0,0006	0,9996
PC16	0,5751	0,0097	0,9528	PC33	0,1205	0,0004	1,0000
PC17	0,5048	0,0075	0,9603	PC34	0,0000	0,0000	1,0000

Pastebima, jog pirmosios dvi komponentės kartu lemia 51 proc. dispersijos. Norėdami pasiekti 90 proc. dispersijos ribą, turėtume naudoti 12 komponentių. Tačiau vertinant tai, jog pagrindinių komponentių

analizės tikslas –minimaliai prarandant informacijos pakeisti stebimą reiškinį charakterizuojančių požymių aibę kelių faktorių rinkiniu (46), norime tą kiekį sumažinti. Optimaliam komponentių kiekiui parinkti tikriname tikrines reikšmes.

3.10 lentelė. Pagrindinių komponentių tikrinių reikšmių lentelė

Komponentė	Tikrinė reikšmė	Komponentės sukaupta sklaida	Bendrai sukaupta sklaida
komp. 1	12,4880	36,7294	36,7294
komp. 2	5,0064	14,7246	51,4540
komp. 3	3,6520	10,7412	62,1952
komp. 4	1,8536	5,4519	67,6471
komp. 5	1,7549	5,1615	72,8086
komp. 6	1,3739	4,0409	76,8494
komp. 7	0,9979	2,9351	79,7845
komp. 8	0,9488	2,7907	82,5752
komp. 9	0,7996	2,3519	84,9271
komp. 10	0,7451	2,1915	87,1185
...

3.10 lentelėje pateiktos pirmų 10 komponentių tikrinės reikšmės (angl. *eigenvalue*). Jei reikšmė didesnė už 1, tuomet komponentė reikšminga (komponentės indėlis aiškinant duomenų sklaidą didesnis nei vidutinis). Taikant šią taisyklę analizei reikia naudoti šešias pagrindines komponentes. Žiūrint į sukaupią duomenų sklaidos dalį – jos paaiškina net 76,85 proc. Todėl galime teigti, kad šešių pagrindinių komponentių palikimas tolesnei analizei yra tinkamas ir korektiškas.

Turtingų šalių grupė

Analizuojami turtingų šalių grupei reikšmingi kintamieji, aprašyti 3.3 skyrelyje. Dėl patogesnio modelių užrašymo, kintamųjų pavadinimai keičiami į X1,...,X6.

- X1 Nesimokantys paaugliai
- X2 Dirbamoji žemė
- X3 Gyventojų skaičiaus augimas
- X4 Kaimo vietovių gyventojų skaičiaus augimas
- X5 Mokslinių tyrimų ir taikomosios technikos specialistai
- X6 Nedarbo lygis

Nors kintamųjų kiekis ir nėra didelis (šeši nepriklausomi kintamieji), tačiau verta patikrinti, ar jie tinkami pagrindinių komponentių analizei. Peržvelgus MSA_i reikšmes (3.11 lentelė) pastebima, kad keturiems kintamiesiems jos mažesnės nei 0,5.

3.11 lentelė. Kaiserio-Meyerio-Olkinio reikšmių lentelė turtingų šalių grupei

Kaiserio-Meyerio-Olkinio kriterijus						
Bendra MSA = 0,47						
MSA reikšmės kiekvienam kintamajam:						
X1	X2	X3	X4	X5	X6	
0,62	0,62	0,42	0,49	0,42	0,39	

Taip pat, nors Bartleto sferiškumo kriterijaus reikšmė $p_reikšmė < 0,05$ ir hipotezė „visi stebimi kintamieji yra nekoreliuoti“ atmeta, tačiau KMO matas lygus 0,467, o tai reiškia, jog duomenys pagrindinių komponentių analizei netinka (žr. 3.12 lentelę) ir metodas šiam duomenų rinkiniui neturi prasmės.

3.12 lentelė. Bartleto sferiškumo kriterijus ir KMO matas turtingų šalių grupei

Bartleto sferiškumo kriterijus
Bartleto testas. $K\text{-kvadratai} = 45,018$, $df = 5$, $p_reikšmė < 1,439 \cdot 10^{-10}$
df – laisvės laipsniai
KMO = 0,4671207

3.5. Regresinė analizė

Atlikus šalių klasteriams reikšmingų kintamųjų dimensijų mažinimą, sudaryti regresinės analizės modeliai. Kadangi tiesinės regresinės analizės prielaidos nebuvo tenkinamos (netenkinama liekanų normalumo prielaida), todėl analizei taikyti neparimetriniai metodai: apibendrinti adityvūs modeliai ir branduolinė regresinė analizė.

Regresiniams modeliams sudaryti naudojami skirtingi duomenys – neturtingų šalių grupei pagrindinių komponentių metodu išskirtos komponentės (PCA1, PCA2, PCA3, PCA4, PCA5, PCA6), turtingų grupei – koreliacinės analizės metu (3.6 lentelė) nustatyti reikšmingi kintamieji. Taip pat duomenys išskaidomi į apmokymo ir tikrinimo rinkinius (apmokymui 80 proc., tikrinimui 20 proc.).

3.5.1. Apibendrinti adityvūs modeliai

Neturtingų šalių grupė

Sudarome modelį kiekvienam nepriklausomam kintamajam nurodant glodinimo funkciją, tačiau nenurodant jokių papildomų parametrų reikšmių.

3.13 lentelė. Apibendrintų adityvių modelių regresija neturtingų šalių grupei

Formulė:
 $Y \sim s(\text{PCA1}) + s(\text{PCA2}) + s(\text{PCA3}) + s(\text{PCA4}) + s(\text{PCA5}) + s(\text{PCA6})$

Parametrų koeficientai:

	Stand. koef.	Klaida	t_reikšmė	Pr(> t)
(konstanta)	-0,01789	0,02732	-0,655	0,513

Kintamųjų reikšmingumas:

	edf	ref.df	F	p_reikšmė
s(PCA1)	5,031	9	41,573	$< 2 \cdot 10^{-16}$
s(PCA2)	3,353	9	1,362	0,0030
s(PCA3)	6,381	9	5,094	0,0000
s(PCA4)	2,818	9	6,149	0,0000
s(PCA5)	5,963	9	1,731	0,0081
s(PCA6)	6,713	9	4,011	0,0000

 $R^2 = 0,858$

* Pr(>|t|) – Stjudento kriterijaus p reikšmė
 * edf – apskaičiuotas laisvės laipsnių skaičius
 * ref.df – parinktas laisvės laipsnių skaičius

Iš rezultatų (žr. 3.13 lentelę) pastebima, kad visos šešios pagrindinės komponentės yra reikšmingos modeliui ($p_reikšmė < \alpha = 0,05$). Taip pat nurodoma, kad sudarytas modelis paaiškina 85,8 proc. duomenų sklaidos (R^2).

Atliekame sudaryto modelio diagnostinius patikrinimus.

3.14 lentelė. Apibendrintų adityvių modelių reikšmingumo testas neturtingų šalių grupei

Parametro k tinkamumo tikrinimas.
 Maža p_reikšmė (k-indeksas < 1) signalizuoja, kad k parinktas per mažas, ypač jei edf artimas k'.

	k'	edf	k-indeksas	p_reikšmė
s(PCA1)	9	5,03	0,95	0,25
s(PCA2)	9	3,35	1,04	0,68
s(PCA3)	9	6,38	1,11	0,93
s(PCA4)	9	2,82	1,01	0,61
s(PCA5)	9	5,96	0,99	0,46
s(PCA6)	9	6,71	1,02	0,57

* edf – apskaičiuotas laisvės laipsnių skaičius

Matoma, kad visoms komponentėms $p_reikšmė > \alpha$, edf nėra artimas k' , todėl galime teigti, jog k parinktas optimaliai (visoms komponentėms $k' = k - 1 = 9$).

Taip pat bandyta didinti k reikšmes intervale [10;30], tačiau geresnio modelio, nei su automatiškai parinktais parametrais, sudaryti nepavyko.

Metodo patikimumo vertinimas

Atliekame sudaryto modelio tikrinimą ir įvertiname paklaidas.

3.15 lentelė. Apibendrintų adityvių modelių paklaidos neturtingų šalių grupei

MSE	RMSE	MAE	R ²
0,19685	0,44368	0,32459	0,78651

Iš rezultatų pastebima, kad sudarytas modelis paaiškina net 78,65 proc. duomenų sklaidos.

Turtingų šalių grupė

Sudarome modelį kiekvienam nepriklausomam kintamajam nurodant glodinimo funkciją, tačiau nenurodant jokių papildomų parametrų reikšmių.

3.16 lentelė. Apibendrintų adityvių modelių regresija turtingų šalių grupei

Formulė:				
$Y \sim s(X1) + s(X2) + s(X3) + s(X4) + s(X5) + s(X6)$				
Parametrų koeficientai:				
	Stand. koef.	Klaida	t_reikšmė	Pr(> t)
(laisvasis narys)	3,5609	0,1852	19,23	<2*10 ⁻¹⁶
Kintamųjų reikšmingumas:				
	edf	ref.df	F	p_reikšmė
s(X1)	3,9*10 ⁻⁷	9	0,000	0,5925
s(X2)	0,4245	9	0,131	0,0939
s(X3)	5,6520	9	4,868	0,0000
s(X4)	6,7660	9	4,224	0,0000
s(X5)	1,7500	9	3,864	0,0000
s(X6)	7,2180	9	7,190	0,0000

R ² = 0,665				
* Pr(> t) – Stjudento kriterijaus p reikšmė				
* edf – apskaičiuotas laisvės laipsnių skaičius				
* ref.df – parinktas laisvės laipsnių skaičius				

Iš rezultatų pastebima, kad du kintamieji (X_1 – nesimokantys paaugliai ir X_2 – dirbamoji žemė) yra nereikšmingos modeliui ($p_reikšmė > \alpha = 0,05$), todėl juos šaliname. Atlikę modelio kintamųjų reikšmingumo testą nustatome, kad likę keturi veiksniai yra reikšmingi.

Atliekame sudaryto modelio diagnostinius patikrinimus.

3.17 lentelė. Apibendrintų adityvių modelių reikšmingumo testas turtingų šalių grupei

Parametro k tinkamumo tikrinimas.				
Maža p -vertė (k -indeksas < 1) signalizuoja, kad k parinktas per mažas, ypač jei edf artimas k' .				
	k'	edf	k-indeksas	$p_reikšmė$
s(X3)	9	5,73	1,09	0,88
s(X4)	9	6,76	1,04	0,71
s(X5)	9	1,74	0,97	0,3
s(X6)	9	7,36	1,01	0,46

* edf – apskaičiuotas laisvės laipsnių skaičius

Matoma, kad visoms komponentėms $p_reikšmė > \alpha$, edf nėra artimas k' , todėl galime teigti, jog k parinktas optimaliai (visoms komponentėms $k' = k - 1 = 9$).

Metodo patikimumo vertinimas

Atliekame sudaryto modelio tikrinimą ir įvertiname paklaidas.

3.18 lentelė. Apibendrintų adityvių modelių paklaidos turtingų šalių grupei

MSE	RMSE	MAE	R^2
7,99943	2,82833	1,86936	0,63457

Iš rezultatų pastebima, kad sudarytas modelis paaiškina net 63,46 proc. duomenų sklaidos.

3.5.2. Branduolinė regresinė analizė

Neturtingų šalių grupė

Sudarome modelį nurodydami tik modelio lygtį, be jokių papildomų parametru. Atliekame į modelį įtrauktų kintamųjų reikšmingumo testą.

3.19 lentelė. Branduolinės regresijos reikšmingumo testas neturtingų šalių grupei

Branduolinės regresijos reikšmingumo testas	
Reikšmingumo tikrinimas kintamiesiems: PCA1 (1), PCA2 (2), PCA3 (3), PCA4 (4), PCA5 (5), PCA6 (6)	
Individualūs reikšmingumo testai kintamiesiems	
	p_reikšmė:
PCA1	$< 2 * 10^{-16}$
PCA2	0,1078
PCA3	$< 2 * 10^{-16}$
PCA4	0,0150
PCA5	$< 2 * 10^{-16}$
PCA6	$< 2 * 10^{-16}$
<hr/>	
Pašalinama PCA2:	
Branduolinės regresijos reikšmingumo testas	
Reikšmingumo tikrinimas kintamiesiems: PCA1 (1), PCA3 (2), PCA4 (3), PCA5 (4), PCA6 (5)	
Individualūs reikšmingumo testai kintamiesiems	
	p_reikšmė:
PCA1	$< 2 * 10^{-16}$
PCA3	$< 2 * 10^{-16}$
PCA4	0,0100
PCA5	$< 2 * 10^{-16}$
PCA6	$< 2 * 10^{-16}$

Iš rezultatų matoma (žr. 3.19 lentelę), kad modeliui reikšmingos penkios pagrindinės komponentės ($p_reikšmė < \alpha = 0,05$), o PCA2 komponentę iš modelio verta šalinti. Pašalinę PCA2 tyrimą kartojame ir nustatome, kad likusios penkios komponentės yra statistiškai reikšmingos.

3.20 lentelė. Branduolinė regresija neturtingų šalių grupei

	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6
Branduolio pločio parametro reikšmė	2,64835	1,632288	0,904387	1,029141	3,327479	3888753
Branduolinės regresijos prognozė: lokali-tiesinė						
Branduolio pločio parametro tipas: fiksuotas						
Standartizuota liekamoji paklaida: 0,1744214						
R ² : 0,9715135						
Pašalinama PCA2:						
	PCA1	PCA3	PCA4	PCA5	PCA6	
Branduolio pločio parametro reikšmė	1,613945	1,621275	0,736954	2575026	1,397705	
Branduolinės regresijos prognozė: lokali-tiesinė						
Branduolio pločio parametro tipas: fiksuotas						
Standartizuota liekamoji paklaida: 0,202889						
R ² : 0,9615092						

Nustatyta, kad pradinis sudarytas modelis (žr. 3.20 lentelę), su visomis šešiomis komponentėmis, paaiškina net 97,15 proc. duomenų sklaidos (R^2). Modelis su penkiomis komponentėmis paaiškina 96,15 proc. duomenų sklaidos (R^2). Taigi komponentė PCA2 lėmė vos 1 proc. duomenų sklaidos, todėl jos pašalinimas iš modelio yra korektiškas.

Metodo patikimumo vertinimas

Atliekame sudaryto modelio tikrinimą ir įvertiname paklaidas.

3.21 lentelė. Branduolinės regresijos modelių paklaidos neturtingų šalių grupei

MSE	RMSE	MAE	R ²
0,20156	0,44896	0,31299	0,81122

Turtingų šalių grupė

Sudarome modelį nurodydami tik modelio lygtį, be jokių papildomų parametru. Atliekame į modelį įtrauktų kintamųjų reikšmingumo testą.

3.22 lentelė. Branduolinės regresijos reikšmingumo turtingų šalių grupei

Branduolinės regresijos reikšmingumo testas	
Reikšmingumo tikrinimas kintamiesiems: X1 (1), X2 (2), X3 (3), X4 (4), X5 (5), X6 (6)	
Individualūs reikšmingumo testai kintamiesiems	
	p_reikšmė:
X1	0,6867
X2	$< 2,22 \cdot 10^{-16}$
X3	$< 2,22 \cdot 10^{-16}$
X4	$< 2,22 \cdot 10^{-16}$
X5	0,0025
X6	0,0125
<hr/>	
Pašalinamas X1 – nesimokantys paaugliai:	
Reikšmingumo tikrinimas kintamiesiems: X2 (1), X3 (2), X4 (3), X5 (4), X6 (5)	
Individualūs reikšmingumo testai kintamiesiems	
	p_reikšmė:
X2	$< 2 \cdot 10^{-16}$
X3	$< 2 \cdot 10^{-16}$
X4	$< 2 \cdot 10^{-16}$
X5	0,7444
X6	0,0201
<hr/>	
Pašalinamas X5 – mokslinių tyrimų ir taikomosios technikos specialistai	
Reikšmingumo tikrinimas kintamiesiems: X2 (1), X3 (2), X4 (3), X6 (4)	
Individualūs reikšmingumo testai kintamiesiems	
	p_reikšmė:
X2	$< 2 \cdot 10^{-16}$
X3	$< 2 \cdot 10^{-16}$
X4	$< 2 \cdot 10^{-16}$
X6	0,0351

Į modelį įtraukus visus šešis kintamuosius, nustatyta, jog statistiškai reikšmingi penki kintamieji ($p_reikšmė < \alpha = 0,05$), o kintamąjį „X1 – nesimokantys paaugliai“ iš modelio verta šalinti. Pašalinę X1 tyrimą kartojame ir tikriname likusių penkių kintamųjų reikšmingumą dar kartą. Antru žingsniu nustatytas „X5 – mokslinių tyrimų ir taikomosios technikos specialistų“ kintamojo nereikšmingumas. Jį taip pat pašaliname ir patikrinę likusius kintamuosius nustatome, kad jie visi reikšmingi (žr. 3.22 lentelę).

3.23 lentelė. Branduolinės regresija turtingų šalių grupei

	X1	X2	X3	X4	X5	X6
Branduolio pločio parametro reikšmė	1,47*10 ⁸	4,430402	0,185621	0,392235	7,432039	1,360425
Branduolinės regresijos prognozė: lokali-tolydi Branduolio pločio parametro tipas: fiksuotas Standartizuota liekamoji paklaida: 1,212969 R ² : 0,9181616						
Pašalinami X1 – nesimokantys paaugliai ir X5 – mokslinių tyrimų ir taikomios technikos specialistai						
	X2	X3	X4	X6		
Branduolio pločio parametro reikšmė	4,969993	0,203337	0,302856	1,38397		
Branduolinės regresijos prognozė: lokali-tolydi Branduolio pločio parametro tipas: fiksuotas Standartizuota liekamoji paklaida: 1,278014 R ² : 0,9083712						

Pastebima, kad sudarytas pradinis modelis su visais kintamaisiais paaiškina net 91,18 proc. duomenų sklaidos (R²). Modelis su keturiais kintamaisiais paaiškina 90,83 proc. sklaidos (R²). Taigi kintamieji X1 ir X5 lėmė mažiau nei 0,5 proc. duomenų sklaidos.

Metodo patikimumo vertinimas

Atliekame sudaryto modelio tikrinimą ir įvertiname paklaidas.

3.24 lentelė. Branduolinės regresijos modelių paklaidos turtingų šalių grupei

MSE	RMSE	MAE	R ²
8,09640	2,84542	1,79538	0,65800

Nors pats sudarytas modelis deklaravo daugiau nei 90 proc. sklaidos paaiškinimo, tačiau tikrinimas atskleidžia, kad tikrasis sudaryto modelio tikslumas – 65,80 proc.

3.5.3. Regresinės analizės metodų palyginimas

Palyginame skirtingais metodais gautus rezultatus.

3.25 lentelė. Regresinės analizės modelių paklaidų ir tikslumo palyginimas

Duomenys	Metodas	MSE	RMSE	MAE	R ²
Neturtingų šalių grupė	Apibendrinti adityvūs modeliai	0,19685	0,44368	0,32459	0,78651
	Branduolinė regresija	0,20156	0,44896	0,31299	0,81122
Turtingų šalių grupė	Apibendrinti adityvūs modeliai	7,99943	2,82833	1,86936	0,63457
	Branduolinė regresija	8,09640	2,84542	1,79538	0,65800

Tyrimo metu 2 šalių klasteriams sudaryti 6 nparametrinės regresinės analizės modeliai. Galima pastebėti, kad neturtingų šalių grupės, lyginant su turtingų šalių, modeliai pasižymi mažesnėmis paklaidomis ir didesniu tikslumu. Tam įtakos turi modeliui nustatyta aiškesnė duomenų struktūra.

Lyginant apibendrintus adityvius modelius ir branduolinę regresiją, šiek tiek mažesnės MSE ir RMSE klaidos gautos pirmuoju atveju, o MAE klaida – antruoju atveju. Didesne paaiškinamos duomenų sklaidos dalimi išsiskyrė antras, branduolinės regresijos, metodas. Neturtingų šalių grupei, į modelį įtraukus 34 statistiškai reikšmingų veiksnių 5 pagrindines komponentes, sudarytas modelis paaiškina 81,12 proc., tuo tarpu antrai grupei, į modelį įtraukus 4 statistiškai reikšmingus veiksniai, paaiškinama 65,80 proc. duomenų sklaidos.

Išvados

1. Išanalizavus literatūrą nustatyta, kad migracijos problema yra aktuali, o didžiausią įtaką sprendimui migruoti turi ekonominės ir socialinės gerovės veiksniai. Moksliniuose straipsniuose daugiausia dėmesio skiriama teorinių migracijos srautų modelių analizėms, retesniais atvejais taikoma klasterinė ar faktorinė analizės.
2. Tyrimui atrinkti 147 rodikliai, galintys turėti įtaką migracijos srautams. Jie sugrupuoti į ekonominės (su bendru vidaus produktu susieti rodikliai, kiti ekonominiai rodikliai) ir socialinės (švietimas, sveikatos apsauga, aplinka ir gyvenimo sąlygos, demografinė situacija, darbo rinka) gerovės veiksnius.
3. Migracijos srautų analizei ir prognozei, pritaikant klasterinės, koreliacinės, pagrindinių komponenčių ir neparimetrinės regresinės analizės metodus, sukurtas originalus modelis, realizuotas „RStudio“ aplinkoje, pritaikant kituose tyrimuose nenaudotas procedūras.
4. Panaudojus sukurtą modelį ir programines priemones atlikta Europos Sąjungos šalių tarptautinės migracijos srautams įtaką darančių veiksnių analizė. Nustatyta, kad šalys gali būti klasterizuojamos pagal migracijos srautus ir bendrą vidaus produktą vienam gyventojui (neturtingos ir turtingos valstybės). Dviem iš klasterių parinkti reikšmingi veiksniai, turintys įtakos migracijai. Neturtingų šalių grupei nustatyti aštuoni ekonominės ir 26 socialinės gerovės, tuo tarpu turtingoms šalims šeši socialinės gerovės statistiškai reikšmingi migracijos srautams įtaką turintys veiksniai.
5. Statistiškai reikšmingi kintamieji apjungti (išskirtos pagrindinės komponentės) ir sudaryti migracijos srautus prognozuojantys modeliai. Palyginti du neparimetrinės: apibendrintų adityvių modelių ir branduolinės regresijos modeliai.
 - Didesniu tikslumu tiek vienai, tiek kitai šalių grupei išsiskyrė branduolinės regresijos metodas.
 - Skurdesnių šalių grupės migracijos srautų prognostinis modelis (įtraukiant 34 statistiškai reikšmingus veiksnius) pasižymi didesniu tikslumu ir paaiškina 81,12 proc. duomenų sklaidos, o turtingesnių šalių (įtraukiant keturis regresinei analizei statistiškai reikšmingus veiksnius) paaiškina 65,79 proc. duomenų sklaidos.
 - Tyrimo metodai ir sukurtas modelis tinkamas migracijos srautų prognozei Europos Sąjungos šalyse ir leidžia numatyti tendencijas. O tai vyriausybėms ir politikos formuotojams galėtų padėti priimti atitinkamus sprendimus pagal valstybių tikslus ir prioritetus.

Vienareikšmiškai negalima teigti, kad visi atskleisti veiksniai turi įtakos migracijos srautams, tačiau tyrimas pateikia įdomių ir detalesnių tolimesnių tyrimų reikalaujančių įžvalgų apie faktorius, kurie

įprastai mokslinėje literatūroje nėra analizuojami, kaip turintys ryšį su migracijos srautai. Tai galėtų būti atspirties taškas, atliekant tolimesnius tyrimus migracijos srityje. Todėl baigiamajame magistro darbe išskirti statistiškai reikšmingi veiksniai ateityje galėtų būti tikrinami tradiciniais kiekybiniais apklausų metodais, apklausiant migrantus.

Literatūra

1. JANČAITYTĖ, R. ir kt. *Tarpkultūrinės kompetencijos didinimas bei įvairovės valdymo gebėjimų stiprinimas* [interaktyvus]. Vilnius: Mykolo Romerio universitetas, 2009 [žiūrėta 2018-02-27]. ISBN 9789955697220. Prieiga per:
<http://www.iom.lt/images/publikacijos/failai/1427789>
2. *Žodžio migracija reikšmė* [interaktyvus], [žiūrėta 2018-02-20]. Prieiga per:
<http://www.lietuviuzodynas.lt/terminai/Migracija>
3. *Imigracija žodžio reikšmė, emigracija žodžio reikšmė* [interaktyvus]. 1985 [žiūrėta 2018-02-27]. Prieiga per: <http://www.zodziai.lt/>
4. Valstybės kontrolė. *Imigracijos procesų valdymas* [interaktyvus]. 2015 [žiūrėta 2018-02-27]. Prieiga per: <https://www.vkontrole.lt/failas.aspx?id=3327>
5. *Migration - Types Of Migration* [interaktyvus], [žiūrėta 2018-03-02]. Prieiga per:
<http://family.jrank.org/pages/1169/Migration-Types-Migration.html>
6. KRIPAITIS, R. ir B. ROMIKAITYTĖ. *Tarptautinė darbo jėgos migracija: jos esmė, formos ir ją sąlygojantys veiksniai* [interaktyvus]. 2005, Nr. 5, p. 171-179 [žiūrėta 2018-03-04]. ISSN 1648-9098. Prieiga per: <http://etalpykla.lituanistikadb.lt/fedora/objects/LT-LDB-0001:J.04~2005>
7. STULGIENĖ, A. ir A. DAUNORIENĖ. *Migracijos poveikis darbo jėgos rinkos pusiausvyrai* [interaktyvus]. *Ekonomika ir vadyba*: 2009. 14 [žiūrėta 2018-03-20]. ISSN 1822-6515. Prieiga per: <http://internet.ktu.lt/lt/mokslas/zurnalai/ekovad/14/1822-6515-2009-984.pdf>
8. DAMULIENĖ, A. *Migracijos problema Lietuvoje ir jos įtaka šalies ekonomikai* [interaktyvus]. Vilnius, 2013, [žiūrėta 2018-03-02]. ISSN 2029-8234. Prieiga per:
https://www.mruni.eu/upload/iblock/582/009_damuliene.pdf
9. MIHI-RAMÍREZ, A, KUMPIKAITĖ-VALIŪNIENĖ, V. ir E. CUENCA-GARCÍA. *An inclusive analysis of determinants of international migration. The case of European rich and poor countries. Technological and Economic Development of Economy*, 2017, 608-626. [interaktyvus], [žiūrėta 2018-03-28]. Prieiga per:
<https://doi.org/10.3846/20294913.2017.1306726>
10. KUMPIKAITĖ-VALIŪNIENĖ, V. ir I. ŽIČKUTĖ. *Emigration after socialist regime in Lithuania: why the West is still the best?*. *Baltic Journal of Management*, 2017, 12.1: 86-110.
<http://dx.doi.org/10.1108/BJM-02-2016-0053>

11. ČIARNIENĖ, R., KUMPIKAITĖ, V. ir A. TARAŠKEVIČIUS. *Makroekonominių veiksnių poveikis žmonių migracijos procesams: teoriniai ir praktiniai aspektai*. Economics & management: 2009. 14: p. 553-559. ISSN 1822-6515, [interaktyvus], [žiūrėta 2018-04-14].
Prieiga per: <http://ecoman.ktu.lt/index.php/Ekv/article/viewFile/9430/4746>
12. ŽIBAS, K. *Šiuolaikiniai migracijos procesai*. 2011, p. 9-22, [interaktyvus], [žiūrėta 2018-04-24]. ISSN 1822-5152. Prieiga per:
http://eltalpykla.vdu.lt:8080/xmlui/bitstream/handle/1/32869/ISSN2351-6461_2011_N_2_12.PG_9-22.pdf?sequence=1
13. *Post-Apocalyptic Push and Pull Factors* [interaktyvus]. 2013 [žiūrėta 2018-03-02]. Prieiga per:
<http://zombiebased.com/tag/everett-lee/>
14. ALESHKOVSKI, I. ir V. IONTSEV. *Mathematical models of migration. Department of Population, Faculty of Economics, Moscow State 'Lomonosov' University, Russia, 2008* [interaktyvus], [žiūrėta 2018-04-12]. Prieiga per: <http://www.eolss.net/sample-chapters/c15/e1-26-09-06.pdf>
15. *Push and Pull Factors* [interaktyvus]. 2011 [žiūrėta 2018-03-02]. Prieiga per:
http://cgge.aag.org/Migration1e/ConceptualFramework_Jan10/ConceptualFramework_Jan105.html
16. *Migration trends* [interaktyvus], [žiūrėta 2018-03-20]. Prieiga per:
http://www.bbc.co.uk/schools/gcsebitesize/geography/migration/migration_trends_rev3.shtml
17. KUMPIKAITĖ, V. ir ŽIČKUTĖ, I. *Emigracijai įtaką darančių veiksnių analizė*. Economics and management, 2012, 17.2: p.740-746.
18. MIHI RAMÍREZ, A. (2016). *Influence of socio-economic factors on the international migration flow in rich-poor countries: Doctoral dissertation : Social sciences, economics (04S)*. Kaunas. ISBN: 9786090212417.
19. KUMPIKAITĖ-VALIŪNIENĖ, V. *Gyvenimo gerovė*. Paskaitų konspektas.
20. KUMPIKAITĖ, V. ir I. ŽIČKUTĖ. *Regression Analysis of Economic Factors Influencing Emigration Rate in Lithuania*. 2013 [interaktyvus], [žiūrėta 2018-04-28]. Prieiga per:
<https://doi.org/10.1016/j.sbspro.2013.08.701>
21. MALHOTRA, N ir P. DEVI. *Analysis of Factors Affecting Internal Migration in India*. Guru Nanak Dev University, Amritsar, India, 2017 [interaktyvus], [žiūrėta 2018-04-25]. Prieiga per:
<http://amity.edu/UserFiles/admaa/a446aPaper%203.pdf>

22. H. T. A. BLBAS. *A Comparison results of factor analysis and cluster analysis to the migration of young people from the Kurdistan Region to Europe*. The official scientific journal of Salahaddin University-ErbilZJPAS (2017), 29 (4); 44-55. Prieiga per:
DOI:10.21271/ZJPAS.29.4.5
23. OZGUR, C., DOU, M., LI, Y. ir ROGERS, G. (2017). *Selection of statistical software for data scientists and teachers*. Journal of Modern Applied Statistical Methods, 16(1), 753-774.
[interaktyvus], [žiūrėta 2018-03-25]. Prieiga per: doi:10.22237/jmasm/1493599200
24. About SAS, [žiūrėta 2018-05-02]. Prieiga per: http://www.sas.com/en_us/company-information.html
25. „PowerPivot“: *galinga duomenų analizė ir duomenų modeliavimas Excel*“, [žiūrėta 2018-05-02]. Prieiga per: <https://support.office.com/lt-lt/article/%E2%80%9EPowerPivot-galinga-duomen%C5%B3-analiz%C4%97-ir-duomen%C5%B3-modeliavimas-%E2%80%9EExcel-d7b119ed-1b3b-4f23-b634-445ab>
26. PUKĖNAS, K. *Kokybinių duomenų analizė SPSS programa*. [interaktyvus]. Kaunas, 2009, [žiūrėta 2018-05-05]. Prieiga per:
http://www.lsu.lt/sites/default/files/dokumentai/studentams/norminiai_dokumentai/paskaitos/kokybinyr_biniu_duomenu_analize_SPSS_programa.pdf
27. SAS Institute Inc. SAS/STAT SAS/STAT® 13.1 User’s Guide. Cary, NC: SAS Institute Inc. 2013 [žiūrėta 2018-05-10]. Prieiga per:
<https://support.sas.com/documentation/onlinedoc/stat/131/glmselect.pdf>
28. *About RStudio*, [žiūrėta 2018-05-02]. Prieiga per: <https://www.rstudio.com/about/>
29. MAYDA, A., *International Migration: A Panel Data Analysis of Economic and Non-Economic Determinants* (2005). IZA Discussion Paper No. 1590. [žiūrėta 2018-04-01]. Prieiga per:
SSRN: <https://ssrn.com/abstract=725441>
30. RUZGAS, T. *The nonparametric estimation of multivariate distribution density applying clustering procedures*. Physical sciences, mathematics (01 P). Vilnius: Technika, 2007
[interaktyvus], [žiūrėta 2018-03-28]. Prieiga per: http://old.mii.lt/files/imi_dis_07_ruzgas.pdf
31. *Klasterinė analizė* [interaktyvus], [žiūrėta 2018-04-01]. Prieiga per:
<https://auzi22.wordpress.com/category/statistika/page/2/>
32. *K-vidurkių metodas* [interaktyvus], [žiūrėta 2018-04-01]. Prieiga per:
<https://auzi22.wordpress.com/2009/11/21/k-vidurkiu-metodas/>

33. ČEKANA VIČIUS, V. ir G. MURAU SKAS. *Statistika ir jos taikymas II*. Vilnius: TEV, 2002. ISBN 9955491167.
34. *R documentation, Package NbClust*, 2015, [interaktyvus], [žiūrėta 2018-04-18]. Prieiga per: <https://cran.r-project.org/web/packages/NbClust/NbClust.pdf>
35. JANILIONIS, V. *Mokymai apie kiekybinių ir kokybinių HSM tyrimų duomenų analizės metodus. Koreliacinės ir regresinės analizės pagrindai* [interaktyvus]. 2015 [žiūrėta 2018-03-25]. Prieiga per: http://www.lidata.eu/index.php?file=files/mokymai/Janilionis_III/jan_III.html&cour
36. *R documentation, Package stats*, 2012, [interaktyvus], [žiūrėta 2018-04-20]. Prieiga per: <https://www.rdocumentation.org/packages/stats/versions/3.5.0/topics/stats-package>
37. KAVALIAUSKAS, M. Daugiamatė statistinė analizė. *Pagrindinių komponentų analizė, faktorinė analizė*. Paskaitų konspektas, 2017.
38. PAKALNIŠKIENĖ, Vilmantė. *Tyrimo ir įvertinimo priemonių patikimumo ir validumo nustatymas*. Metodinė priemonė [interaktyvus], [žiūrėta 2018-03-25]. ISBN 978-609-459-096-2. Prieiga per: https://www.vu.lt/site_files/LD/Tyrimo_ir_%C4%AFvertinimo_priemoni%C5%B3_patikimumo_ir_valid
39. JAUREGUI, J. *Principal component analysis with linear algebra* [interaktyvus]. 2012 [žiūrėta 2018-04-04]. Prieiga per: <http://www.math.union.edu/~jauregui/PCA.pdf>
40. F. HUSSON, S. LE ir J. PAGES. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*, 2017 [interaktyvus], [žiūrėta 2018-04-25]. Prieiga per: doi:10.1201/b10345-2
41. SAS help. *Additive Models and Generalized Additive Models*. [interaktyvus], [žiūrėta 2018-05-01]. Prieiga per: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_am_sect014.htm
42. S. N. WOOD, N. PYA ir B. SÄFKEN. *Smoothing Parameter and Model Selection for General Smooth Models*, *Journal of the American Statistical Association*, 2017 [interaktyvus], [žiūrėta 2018-05-01]. 111:516, 1548-1563, Prieiga per: DOI: 10.1080/01621459.2016.1180986
43. CAO, Y. and K. .L. POH. *A non-parametric regression approach for missing value imputation in microarray*. 2008, p. 25-34 [interaktyvus], [žiūrėta 2018-05-03]. ISBN 978-83-60434-44-4 Prieiga per: <http://iis.ipipan.waw.pl/2008/proceedings/iis08-02.pdf>

44. *R documentation, Package npreg* [interaktyvus], [žiūrėta 2018-05-01]. Prieiga per:
<https://www.rdocumentation.org/packages/np/versions/0.60-7/topics/npreg>
45. *BVP (Bendrasis vidaus produktas)*, [interaktyvus], [žiūrėta 2018-03-21]. Prieiga per:
<http://www.finansistas.net/bvp.html>
46. *Faktorinė analizė* [interaktyvus], [žiūrėta 2018-03-25]. Prieiga per:
<https://auzi22.wordpress.com/2009/11/21/faktorine-analize/>

Tarptautinei migracijai įtaką darančių veiksmų analizės R programos kodas

Klasterinė analizė

```
#-----  
#Programos autorė: Gintare Pamarnackaite  
#-----  
#Tarptautinės migracijos srautams įtaką darančių veiksmų analizė  
# Kaunas, 2018  
#-----  
#Klasterinė analizė  
  
# optimalaus klasterių skaičiaus parinkimas  
library(NbClust)  
library(mclust)  
library(factoextra)  
library(cluster)  
mydata1 <- read.csv("C:/Users/Gintare/Google Drive/Magistras/clustering.csv", header = TRUE, sep = ";")  
scaled.dat <- as.matrix(scale(mydata1[,2:3]))  
plot(scaled.dat)  
#-----  
# WSS - Alkūnės metodas  
set.seed(123)  
k.max <- 15  
data <- scaled.dat  
wss <- sapply(1:k.max,  
              function(k){kmeans(data, k, nstart=50,iter.max = 15 )$tot.withinss})  
wss  
plot(1:k.max, wss,  
     type="b", pch = 19, frame = FALSE,  
     xlab="Klasterių skaičius k",  
     ylab="Nuokrypių kvadratų suma tarp klasterių")  
#-----  
# Silueto metodas  
fviz_nbclust(scaled.dat, kmeans, method = "silhouette")+  
  labs(subtitle = "Silueto kriterijus")+xlab("Klasterių skaičius k")+ylab("Vidutinis silueto plotis")  
#-----  
# 26 skirtingi kriterijai  
nb<-NbClust(scaled.dat,min.nc=2,max.nc=8,index="all",method="kmeans")  
nb  
summary(nb)  
#-----  
#Klasterizavimas Kmeans  
# parenkamas atsitiktinis klasterių centras  
set.seed(132)  
#k-vidurkių klasterizavimas
```

```

k<- 3
scaled.dat <- scale(mydata1[,2:3])
kmeansResult <- kmeans(scaled.dat, k)
round(kmeansResult$centers, digits=3)
kmeansResult$cluster
#-----
#hierarchinis
library(stats)
clust<-hclust(dist(scaled.dat), method = "ward.D")
plot(clust)
clusterCUT<-cutree(clust,3)
table(clusterCUT,mydata1$Country)
#-----
#-----
# Koreliacinė analizė
library("DataExplorer")
duom<- read.csv("C:/Users/Gintare/Google Drive/Magistras/orig.csv", header = TRUE, sep = ";")
#-----
#Klasterio numeris
cl=1
mydat<-duom[duom$Cluster %in% cl,]
data<-mydat[,c(2:length(mydat))]
data[, c(3:length(data))] <- sapply(data[, c(3:length(data))], as.numeric)
#-----
# ištrinami kintamieji su mažiau nei 50 % užpildytų reikšmių
miss <- c()
for(i in 1:ncol(data)) {
  if(length(which(is.na(data[,i]))) > 0.5*nrow(data)) miss <- append(miss,i)
}
data2 <- data[,-miss]
#-----
# Spirmeno koreliacijos skaičiavimas
data22<-data2[,c(3:length(data2))]
core<-(round(cor(data22, use = "pairwise.complete.obs",method="spearman"),3))
write.csv(core,"C:/Users/Gintare/Google Drive/Magistras/orig_cl_1.csv")
#-----
#-----
# Trūkstatų reikšmių duomenų rinkinyje užpildymas ir pagrindinių komponentių analizė
library("DataExplorer")
library("missMDA")
require("caret")
library("stats")
library("FactoMineR")
library("factoextra")
library("psych")
duom<- read.csv("C:/Users/Gintare/Google Drive/Magistras/orig_cl_1.csv", header = TRUE, sep = ";")
#Klasterio numeris
cl=1

```

```

mydat<-duom[duom$Cluster %in% cl,]
data<-mydat[,c(2:length(mydat))]
data[, c(3:length(data))] <- sapply(data[, c(3:length(data))], as.numeric)
data22<-data[,c(3:length(data))]
data22 <- scale(data22)
#-----
#duomenų su trūkstamomis reikšmėmis paruosimas
data221<-data22
ncomp <- estim_ncpPCA(data221)
ncomp$ncp
res.imp <- imputePCA(data221, ncp = ncomp$ncp)
#-----
#pagrindinių komponentių analizės prielaidų tikrinimas
library("corpcor")
d<-data.frame(res.imp$completeOb)
d<-d[,c(2:length(d))]
KMO(d) # siuloma naudoti, jei KMO > 0.6 (arba > 0.5)
bartlett.test(d)
kmo.test <- function(d){
  cor.sq = cor(d)^2
  cor.sumsq = (sum(cor.sq)-dim(cor.sq)[1])/2
  library(corpcor)
  pcor.sq = cor2pcor(cor(d))^2
  pcor.sumsq = (sum(pcor.sq)-dim(pcor.sq)[1])/2
  kmo = cor.sumsq/(cor.sumsq+pcor.sumsq)
  return(kmo)
}
kmo.test(d)
#-----
#PKA analizė su trūkstamomis reikšmėmis
res.pca <- prcomp(d, scale = T, rotation= "varimax")
PCA<-PCA(d)
PCA$eig

PCA1 <- (res.pca$x[,1])
PCA2 <- (res.pca$x[,2])
PCA3 <- (res.pca$x[,3])
PCA4 <- (res.pca$x[,4])
PCA5 <- (res.pca$x[,5])
PCA6 <- (res.pca$x[,6])
d$PCA1=PCA1
d$PCA2=PCA2
d$PCA3=PCA3
d$PCA4=PCA4
d$PCA5=PCA5
d$PCA6=PCA6
write.csv(d,"C:/Users/Gintare/Google Drive/Magistras/pcarezstand.csv")
#-----
#-----

```

```

#Regresinė analizė
library(mgcv)
library(np)
library(caret)
dd<- read.csv("C:/Users/Gintare/Google Drive/Magistras/pcarezstand.csv", header = TRUE, sep = ",")
#-----
#apmokymo ir tikrinimo duomenų rinkinių paruošimas
smp_size <- floor(0.8 * nrow(dd))
set.seed(3)
train_ind <- sample(seq_len(nrow(dd)), size = smp_size)
ddtr <- dd[train_ind, ]
ddte <- dd[-train_ind, ]
#-----
#tiesinė regresinė analizė
linearMod <- lm(Y~ PCA1+ PCA2+ PCA3 +PCA4+ PCA5+PCA6, data=ddtr)
#-----
#apibendrintų adityvių modelių regresija
model.g = gam(Y ~s(PCA1)+ s(PCA2)+ s(PCA3)+s(PCA4)+ s(PCA5)+s(PCA6) , data = ddtr, family=gaussian(),
select="TRUE", method="GCV.Cp")
summary(model.g)
plot(model.g, residuals = TRUE)
plot(model.g,seWithMean = TRUE)
#diagnostinis modelio tikrinimas
gam.check((model.g))
rsd <- residuals(model.g)
gam(rsd~s(PCA1,k=20,bs="cs"),gamma=1.4,data=ddtr)
gam(rsd~s(PCA2,k=20,bs="cs"),gamma=1.4,data=ddtr)
gam(rsd~s(PCA3,k=20,bs="cs"),gamma=1.4,data=ddtr)
gam(rsd~s(PCA4,k=20,bs="cs"),gamma=1.4,data=ddtr)
gam(rsd~s(PCA5,k=20,bs="cs"),gamma=1.4,data=ddtr)
gam(rsd~s(PCA6,k=20,bs="cs"),gamma=1.4,data=ddtr)
AIC(model.g)
#sudaryto modelio tikrinimas ir paklaidų vertinimas
p<-as.matrix(predict(model.g, newdata=ddte))
ddte$pred<-p
error<-ddte$Y - ddte$pred
postResample(pred = ddte$pred, obs = ddte$Y)
mse <- (mean(error^2))
rmse <- sqrt(mean(error^2))
mae <- mean(abs(error))
#-----
#branduolinė regresija
freq <- npreg(Y ~ PCA1+PCA3+PCA4+PCA5+PCA6, regtype = "ll",bwmethod = "cv.aic",gradients = TRUE, data = ddtr)
summary(freq)
#diagnostinis modelio tikrinimas
npsigtest(freq)
#sudaryto modelio tikrinimas ir paklaidų vertinimas
p<-as.matrix(predict(freq,newdata=ddte))
ddte$pred<-p

```

```

error<-ddte$Y - ddte$pred
postResample(pred = ddte$pred, obs = ddte$Y)
mse <- (mean(error^2))
rmse <- sqrt(mean(error^2))
mae <- mean(abs(error))
#-----

```

2 priedas

Tarptautinės migracijos srautų tyrimui naudotų kintamųjų skaitinės charakteristikos

##	Y	x1	x2	x3
##	Min. : -24.807	Min. : -16.7256	Min. : -18.2534	Min. : 11.22
##	1st Qu.: -0.147	1st Qu.: 0.1855	1st Qu.: -0.2195	1st Qu.: 14.82
##	Median : 1.490	Median : 2.2944	Median : 1.8223	Median : 16.72
##	Mean : 2.204	Mean : 2.1909	Mean : 1.9911	Mean : 16.85
##	3rd Qu.: 3.808	3rd Qu.: 4.4348	3rd Qu.: 4.2380	3rd Qu.: 18.24
##	Max. : 27.128	Max. : 22.3193	Max. : 20.1067	Max. : 28.35
##		NA's : 52	NA's : 52	NA's : 32
##	x4	x5	x6	x7
##	Min. : 0.00000	Min. : 5.007	Min. : 0.00000	Min. : 0.00002
##	1st Qu.: 0.02827	1st Qu.: 18.427	1st Qu.: 0.00016	1st Qu.: 0.02957
##	Median : 0.10027	Median : 22.267	Median : 0.00893	Median : 0.10212
##	Mean : 0.28379	Mean : 22.070	Mean : 0.08100	Mean : 0.28794
##	3rd Qu.: 0.35962	3rd Qu.: 26.094	3rd Qu.: 0.06359	3rd Qu.: 0.37361
##	Max. : 2.13166	Max. : 39.997	Max. : 1.14631	Max. : 2.14100
##	NA's : 158	NA's : 44	NA's : 204	NA's : 84
##	x8	x9	x10	x11
##	Min. : -13.335	Min. : 2.220	Min. : 3.633	Min. : 0.00685
##	1st Qu.: 1.871	1st Qu.: 4.018	1st Qu.: 8.012	1st Qu.: 0.45537
##	Median : 5.101	Median : 4.805	Median : 11.274	Median : 1.59459
##	Mean : 5.181	Mean : 5.123	Mean : 14.296	Mean : 2.71219
##	3rd Qu.: 8.876	3rd Qu.: 5.450	3rd Qu.: 18.222	3rd Qu.: 3.99541
##	Max. : 21.744	Max. : 18.190	Max. : 43.856	Max. : 23.27368
##	NA's : 44	NA's : 28	NA's : 28	NA's : 167
##	x12	x13	x14	x15
##	Min. : 38.45	Min. : 7.282	Min. : 0.2542	Min. : 40.88
##	1st Qu.: 45.67	1st Qu.: 30.496	1st Qu.: 1.5968	1st Qu.: 63.84
##	Median : 48.47	Median : 47.491	Median : 2.4777	Median : 70.42
##	Mean : 48.51	Mean : 43.603	Mean : 2.9078	Mean : 69.44
##	3rd Qu.: 51.54	3rd Qu.: 56.885	3rd Qu.: 3.7815	3rd Qu.: 73.68
##	Max. : 60.05	Max. : 73.732	Max. : 14.4942	Max. : 91.05
##		NA's : 28		NA's : 133
##	x16	x17	x18	x19
##	Min. : 0.00912	Min. : 1.00	Min. : 6.337	Min. : 0.3158
##	1st Qu.: 2.29942	1st Qu.: 22.00	1st Qu.: 15.805	1st Qu.: 0.6412
##	Median : 9.84773	Median : 41.00	Median : 25.069	Median : 1.0063
##	Mean : 13.25115	Mean : 40.35	Mean : 25.599	Mean : 1.1604
##	3rd Qu.: 20.96105	3rd Qu.: 59.00	3rd Qu.: 33.444	3rd Qu.: 1.4358
##	Max. : 50.73394	Max. : 83.00	Max. : 58.897	Max. : 5.1450
##	NA's : 37	NA's : 245	NA's : 28	NA's : 28
##				

x20	x21	x22	x23
## Min. : 3.219	Min. : 0.2048	Min. : 7.90	Min. : 95.7
## 1st Qu.: 5.485	1st Qu.: 3.7525	1st Qu.: 9.50	1st Qu.: 98.8
## Median : 6.898	Median :11.5670	Median :10.25	Median : 99.6
## Mean : 7.414	Mean :13.9588	Mean :10.56	Mean : 99.3
## 3rd Qu.: 8.890	3rd Qu.:19.5897	3rd Qu.:11.36	3rd Qu.: 99.8
## Max. :14.224	Max. :77.2065	Max. :16.70	Max. :100.0
## NA's :220	NA's :320	NA's :28	NA's :186
x24	x25	x26	x27
## Min. : 49.62	Min. : 2.700	Min. : 3.639	Min. : 100.0
## 1st Qu.:334.00	1st Qu.: 4.175	1st Qu.: 29.872	1st Qu.: 100.0
## Median :454.00	Median : 5.200	Median : 43.498	Median : 100.0
## Mean :417.71	Mean : 5.377	Mean : 49.323	Mean : 233.8
## 3rd Qu.:544.19	3rd Qu.: 6.325	3rd Qu.: 65.429	3rd Qu.: 500.0
## Max. :709.74	Max. :11.200	Max. :132.298	Max. :1000.0
## NA's :392	NA's :364	NA's :348	NA's :340
x28	x29	x30	x31
## Min. : 5.44	Min. : 5.672	Min. : 31.98	Min. : 0.100
## 1st Qu.: 21.30	1st Qu.:10.702	1st Qu.: 85.82	1st Qu.: 1.550
## Median : 30.72	Median :15.417	Median : 97.15	Median : 5.100
## Mean : 36.02	Mean :16.656	Mean : 94.16	Mean : 6.761
## 3rd Qu.: 41.55	3rd Qu.:20.932	3rd Qu.:105.09	3rd Qu.:10.350
## Max. :110.94	Max. :37.755	Max. :114.89	Max. :40.400
## NA's :119	NA's :37		NA's :117
x32	x33	x34	x35
## Min. :-25.549	Min. : 78.16	Min. : 6.10	Min. : 0.010
## 1st Qu.: -5.068	1st Qu.: 89.92	1st Qu.: 9.00	1st Qu.: 2.188
## Median : -1.083	Median : 92.17	Median : 9.80	Median : 3.000
## Mean : -1.316	Mean : 91.84	Mean :10.13	Mean : 3.898
## 3rd Qu.: 2.316	3rd Qu.: 94.36	3rd Qu.:11.20	3rd Qu.: 4.273
## Max. : 11.943	Max. :100.00	Max. :15.30	Max. :33.108
## NA's :12	NA's :131	NA's :28	NA's :309
x36	x37	x38	x39
## Min. :61.80	Min. : 0.1859	Min. : 0.1858	Min. : 2.150
## 1st Qu.:71.94	1st Qu.: 54.5558	1st Qu.: 54.5560	1st Qu.: 9.208
## Median :76.08	Median : 85.5183	Median : 85.3145	Median :10.735
## Mean :75.50	Mean : 89.6109	Mean : 89.7262	Mean :10.896
## 3rd Qu.:79.19	3rd Qu.:115.0881	3rd Qu.:115.1616	3rd Qu.:13.492
## Max. :84.07	Max. :253.2620	Max. :253.1463	Max. :18.744
## NA's :420	NA's :56	NA's :55	NA's :427
x40	x41	x42	x43
## Min. :12.25	Min. :0.0811	Min. : 0.0374	Min. : 0.2234
## 1st Qu.:18.76	1st Qu.:0.5035	1st Qu.: 1.4684	1st Qu.: 5.1757
## Median :22.58	Median :0.6007	Median : 8.1876	Median :15.8108
## Mean :22.75	Mean :0.7181	Mean :15.2706	Mean :22.2954
## 3rd Qu.:27.36	3rd Qu.:0.9003	3rd Qu.:21.8634	3rd Qu.:32.2160
## Max. :33.10	Max. :1.7647	Max. :69.8734	Max. :93.9046
## NA's :424	NA's :425	NA's :65	NA's :65
##			

```

x44          x45          x46          x47
## Min.    : 2.842    Min.    : 0.01622    Min.    : 1.000    Min.    : 1.000
## 1st Qu.:20.826    1st Qu.: 0.70835    1st Qu.: 3.500    1st Qu.: 2.700
## Median :34.715    Median : 1.70904    Median : 4.100    Median : 4.600
## Mean   :36.850    Mean   : 10.37653    Mean   : 4.296    Mean   : 6.824
## 3rd Qu.:50.828    3rd Qu.: 5.86453    3rd Qu.: 5.000    3rd Qu.: 8.900
## Max.   :82.239    Max.   :100.00000    Max.   :12.800    Max.   :45.200
## NA's   :244      NA's   :42
##      x48          x49          x50          x51
## Min.    :10.30    Min.    :29.00    Min.    :37.7    Min.    : -65.69
## 1st Qu.:22.40    1st Qu.:59.67    1st Qu.:48.8    1st Qu.: 32.91
## Median :27.10    Median :66.70    Median :52.4    Median : 56.14
## Mean   :26.79    Mean   :66.39    Mean   :52.9    Mean   : 53.15
## 3rd Qu.:30.70    3rd Qu.:73.70    3rd Qu.:57.0    3rd Qu.: 75.42
## Max.   :40.60    Max.   :88.30    Max.   :75.3    Max.   :100.00
##      NA's   :35
##      x52          x53          x54          x55
## Min.    :26.09    Min.    : 0.111    Min.    : 1.878    Min.    : 18.54
## 1st Qu.:38.30    1st Qu.:19.586    1st Qu.:32.664    1st Qu.: 34.90
## Median :42.16    Median :22.157    Median :38.370    Median : 48.34
## Mean   :42.00    Mean   :22.667    Mean   :36.918    Mean   : 57.96
## 3rd Qu.:45.72    3rd Qu.:26.185    3rd Qu.:42.600    3rd Qu.: 70.37
## Max.   :64.58    Max.   :36.095    Max.   :97.805    Max.   :222.70
## NA's   :145      NA's   :119      NA's   :37
##      x56          x57          x58          x59
## Min.    : -20.308    Min.    : -20.6720    Min.    : 30.29    Min.    :45.32
## 1st Qu.: 2.114    1st Qu.: -2.8168    1st Qu.: 51.37    1st Qu.:72.28
## Median : 5.116    Median : 0.5733    Median : 67.76    Median :76.44
## Mean   : 5.593    Mean   : 1.2820    Mean   : 68.63    Mean   :75.81
## 3rd Qu.: 9.127    3rd Qu.: 4.8445    3rd Qu.: 85.05    3rd Qu.:80.47
## Max.   :38.392    Max.   :35.2348    Max.   :111.08    Max.   :91.67
## NA's   :1      NA's   :442
##      x60          x61          x62          x63
## Min.    : -16.6214    Min.    : 1.642    Min.    : 2.910    Min.    : -58.323
## 1st Qu.: 0.7639    1st Qu.: 5.536    1st Qu.: 7.202    1st Qu.: 1.708
## Median : 1.9741    Median : 8.647    Median : 8.871    Median : 3.609
## Mean   : 2.0998    Mean   :10.311    Mean   : 9.228    Mean   :10.926
## 3rd Qu.: 3.5043    3rd Qu.:12.680    3rd Qu.:11.218    3rd Qu.: 8.143
## Max.   :16.3917    Max.   :46.144    Max.   :20.166    Max.   :451.716
## NA's   :1      NA's   :9      NA's   :9      NA's   :4
##      x64          x65          x66          x67
## Min.    : -89.7089    Min.    : 1.094    Min.    : -10.662    Min.    :12.34
## 1st Qu.: 0.5101    1st Qu.:22.409    1st Qu.: 0.301    1st Qu.:18.12
## Median : 2.2016    Median :32.716    Median : 1.548    Median :19.35
## Mean   : 7.3639    Mean   :33.205    Mean   : 1.616    Mean   :19.77
## 3rd Qu.: 5.4480    3rd Qu.:37.598    3rd Qu.: 3.192    3rd Qu.:21.02
## Max.   :219.8321    Max.   :73.689    Max.   :13.088    Max.   :27.94
## NA's   :3      NA's   :28      NA's   :1
##

```

```

x68          x69          x70          x71
## Min.    :2.326  Min.    : 6.703  Min.    : 8.331  Min.    : 64.77
## 1st Qu.:4.276  1st Qu.:10.143  1st Qu.:19.528  1st Qu.: 95.16
## Median  :5.106  Median  :11.657  Median  :23.560  Median  : 99.43
## Mean    :5.172  Mean    :11.792  Mean    :24.194  Mean    : 98.72
## 3rd Qu.:5.691  3rd Qu.:13.053  3rd Qu.:27.723  3rd Qu.:102.82
## Max.    :8.627  Max.    :18.891  Max.    :54.683  Max.    :120.67
## NA's    :115    NA's    :115
##      x72          x73          x74          x75
## Min.    : 4.871  Min.    : 1.673  Min.    : -16.4292  Min.    : 22.91
## 1st Qu.:17.881  1st Qu.: 7.244  1st Qu.:  0.2368  1st Qu.: 37.05
## Median  :21.819  Median  :11.198  Median  :  1.6587  Median  : 48.30
## Mean    :21.572  Mean    :14.406  Mean    :  2.0233  Mean    : 56.68
## 3rd Qu.:25.731  3rd Qu.:18.552  3rd Qu.:  3.4710  3rd Qu.: 68.71
## Max.    :39.841  Max.    :71.741  Max.    : 20.4931  Max.    :187.47
## NA's    :12     NA's    :9      NA's    :1
##      x76          x77          x78          x79
## Min.    : 73.80  Min.    : 85.30  Min.    :0.01000  Min.    : 2.000
## 1st Qu.: 97.10  1st Qu.: 99.30  1st Qu.:0.01000  1st Qu.: 7.275
## Median  : 98.70  Median  :100.00  Median  :0.01000  Median  : 11.000
## Mean    : 96.14  Mean    : 99.15  Mean    :0.01658  Mean    : 22.220
## 3rd Qu.: 99.50  3rd Qu.:100.00  3rd Qu.:0.02000  3rd Qu.: 24.000
## Max.    :100.00  Max.    :100.00  Max.    :0.05000  Max.    :157.000
## NA's    :28     NA's    :28     NA's    :236
##      x80          x81          x82          x83
## Min.    : 3.614  Min.    :10.69  Min.    : -4.480  Min.    : -9.680
## 1st Qu.:38.930  1st Qu.:22.88  1st Qu.: 1.042  1st Qu.: 1.061
## Median  :61.870  Median  :27.56  Median  : 2.221  Median  : 2.010
## Mean    :57.272  Mean    :27.16  Mean    : 2.692  Mean    : 2.772
## 3rd Qu.:76.486  3rd Qu.:30.88  3rd Qu.: 3.393  3rd Qu.: 3.463
## Max.    :97.494  Max.    :41.51  Max.    :45.667  Max.    :43.070
## NA's    :2      NA's    :2      NA's    :2
##      x84          x85          x86          x87
## Min.    : 1.991  Min.    :68.20  Min.    :47.60  Min.    : 0.70
## 1st Qu.: 3.752  1st Qu.:75.42  1st Qu.:56.80  1st Qu.:19.20
## Median  : 5.178  Median  :78.00  Median  :60.35  Median  :27.70
## Mean    : 5.472  Mean    :78.21  Mean    :61.26  Mean    :28.27
## 3rd Qu.: 6.336  3rd Qu.:81.10  3rd Qu.:64.90  3rd Qu.:36.90
## Max.    :13.156  Max.    :88.00  Max.    :77.90  Max.    :54.70
## NA's    :12     NA's    :2      NA's    :2      NA's    :151
##      x88          x89          x90          x91
## Min.    : 0.500  Min.    :92.36  Min.    :31.34  Min.    :44.22
## 1st Qu.: 5.164  1st Qu.:97.16  1st Qu.:63.58  1st Qu.:64.98
## Median  : 6.831  Median  :98.01  Median  :77.61  Median  :70.75
## Mean    : 8.130  Mean    :97.49  Mean    :73.43  Mean    :69.74
## 3rd Qu.: 9.684  3rd Qu.:98.62  3rd Qu.:83.23  3rd Qu.:75.32
## Max.    :53.850  Max.    :99.90  Max.    :97.44  Max.    :84.78
## NA's    :283    NA's    :440    NA's    :9      NA's    :9
##

```

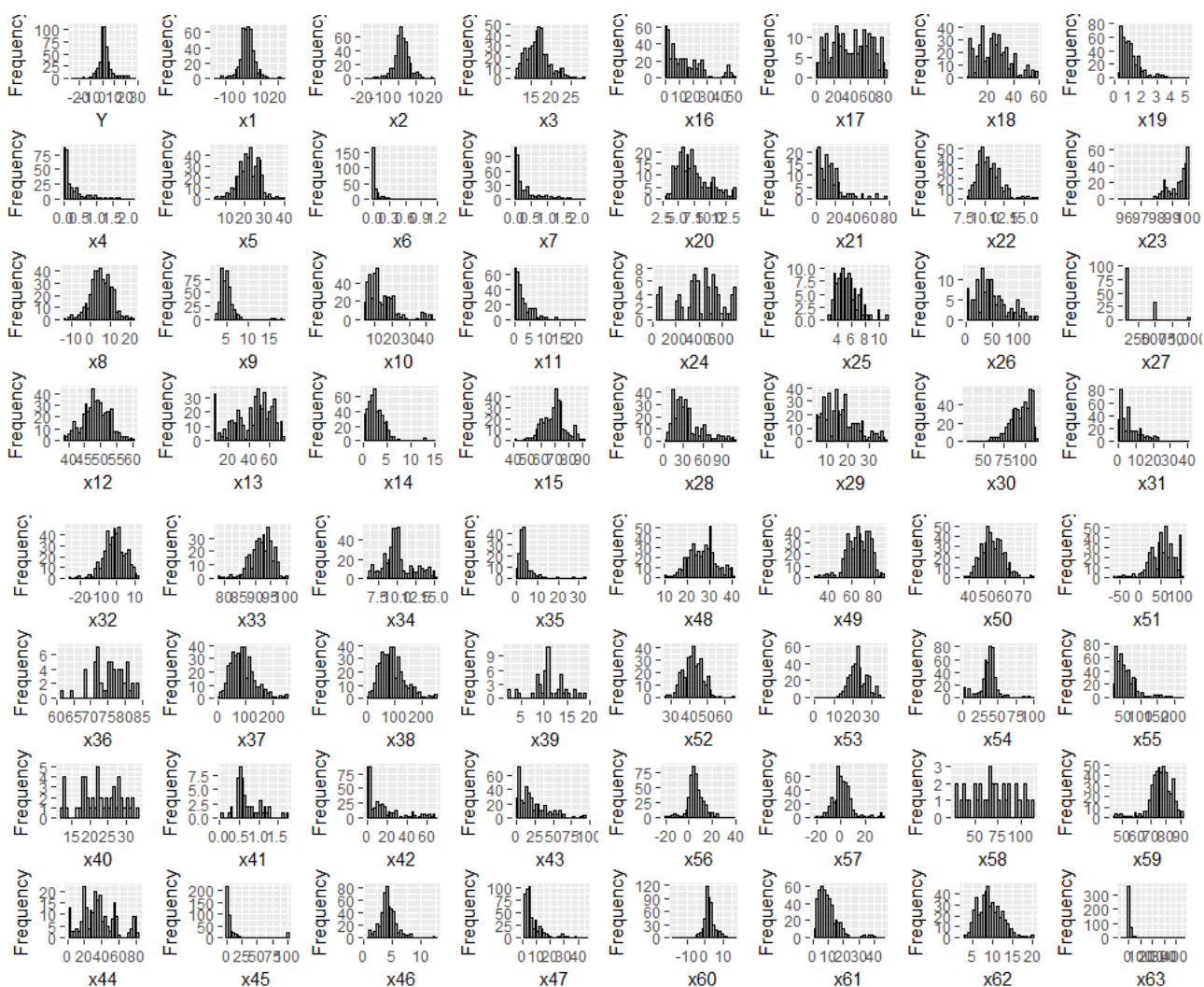

x92	x93	x94	x95
## Min. : 4.322	Min. :0.3279	Min. :0.00097	Min. :0.007311
## 1st Qu.:13.457	1st Qu.:1.1250	1st Qu.:0.00478	1st Qu.:0.386100
## Median :16.484	Median :1.4185	Median :0.00691	Median :0.984074
## Mean :16.830	Mean :1.4852	Mean :0.01051	Mean :2.203371
## 3rd Qu.:21.203	3rd Qu.:1.8421	3rd Qu.:0.01129	3rd Qu.:2.210828
## Max. :36.962	Max. :3.4654	Max. :0.04716	Max. :9.989569
##	NA's :1	NA's :236	NA's :28
x96	x97	x98	x99
## Min. :0.06112	Min. :1.887	Min. :-2.8510	Min. :49.33
## 1st Qu.:0.29514	1st Qu.:2.777	1st Qu.: -0.1989	1st Qu.:50.51
## Median :0.73485	Median :3.284	Median : 0.2655	Median :51.04
## Mean :1.32086	Mean :3.329	Mean : 0.2261	Mean :51.27
## 3rd Qu.:2.02641	3rd Qu.:3.775	3rd Qu.: 0.6539	3rd Qu.:51.52
## Max. :8.15423	Max. :6.255	Max. : 2.8910	Max. :54.21
##	NA's :159	NA's :3	
x100	x101	x102	x103
## Min. :45.79	Min. : 8.60	Min. : 2.500	Min. : -0.40
## 1st Qu.:48.48	1st Qu.:12.78	1st Qu.: 2.500	1st Qu.: 8.40
## Median :48.96	Median :17.60	Median : 2.500	Median :12.70
## Mean :48.73	Mean :17.02	Mean : 2.883	Mean :13.67
## 3rd Qu.:49.49	3rd Qu.:20.90	3rd Qu.: 2.500	3rd Qu.:17.40
## Max. :50.67	Max. :26.40	Max. :10.400	Max. :32.40
##	NA's :372	NA's :28	NA's :163
x104	x105	x106	x107
## Min. : 7.10	Min. :-6.274	Min. :0.2225	Min. : 2.103
## 1st Qu.:15.05	1st Qu.: 1.839	1st Qu.:0.6986	1st Qu.:18.364
## Median :21.00	Median : 3.814	Median :1.2336	Median :29.944
## Mean :23.37	Mean : 3.928	Mean :1.4304	Mean :27.853
## 3rd Qu.:32.20	3rd Qu.: 5.745	3rd Qu.:1.9817	3rd Qu.:35.565
## Max. :47.30	Max. :19.508	Max. :3.9109	Max. :50.373
## NA's :5	NA's :283	NA's :38	
x108	x109	x110	x111
## Min. :-5.9933	Min. :1.041	Min. :13.70	Min. :17.60
## 1st Qu.: -1.3517	1st Qu.:1.052	1st Qu.:21.40	1st Qu.:29.70
## Median :-0.5847	Median :1.058	Median :25.00	Median :35.30
## Mean :-0.7229	Mean :1.057	Mean :25.77	Mean :36.29
## 3rd Qu.: 0.1440	3rd Qu.:1.060	3rd Qu.:29.05	3rd Qu.:41.30
## Max. : 2.1930	Max. :1.070	Max. :42.40	Max. :63.00
##	NA's :196	NA's :341	NA's :341
x112	x113	x114	x115
## Min. : 1.23	Min. :-0.03197	Min. : 0.04917	Min. :22.50
## 1st Qu.:17.19	1st Qu.: 0.00096	1st Qu.: 0.28993	1st Qu.:37.40
## Median :21.03	Median : 0.02755	Median : 1.18122	Median :40.68
## Mean :20.90	Mean : 0.61700	Mean : 4.47443	Mean :41.21
## 3rd Qu.:24.71	3rd Qu.: 0.86446	3rd Qu.: 2.47607	3rd Qu.:44.36
## Max. :62.94	Max. : 6.45959	Max. :84.02565	Max. :61.15
## NA's :29	NA's :246	NA's :148	NA's :146
##			

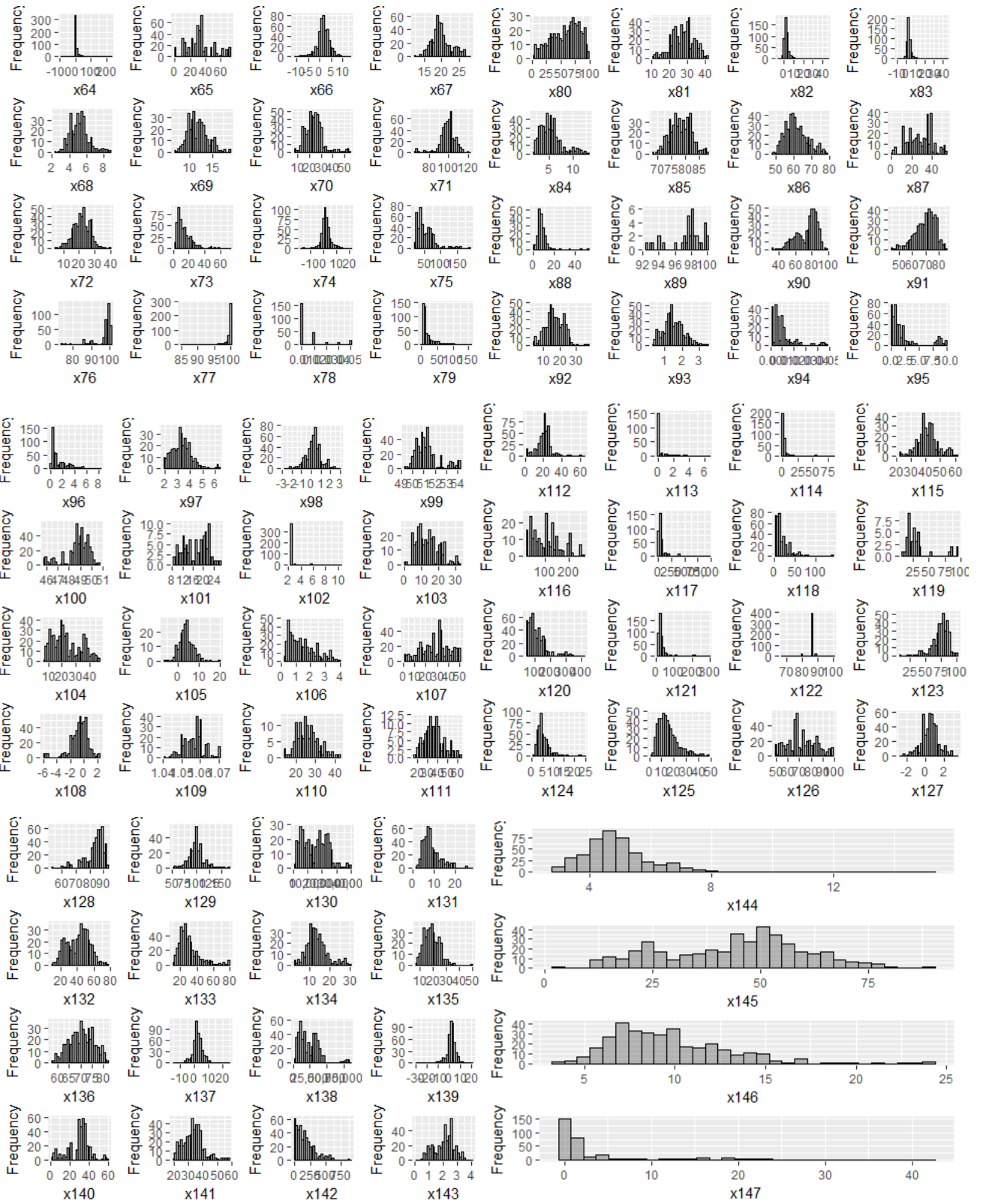
x116	x117	x118	x119
## Min. : 23.0	Min. : 1.00	Min. : 3.50	Min. : 9.995
## 1st Qu.: 55.0	1st Qu.: 15.50	1st Qu.: 7.50	1st Qu.:17.566
## Median :107.5	Median : 25.00	Median : 14.25	Median :25.124
## Mean :108.2	Mean : 51.63	Mean : 19.66	Mean :31.280
## 3rd Qu.:143.8	3rd Qu.: 52.00	3rd Qu.: 25.00	3rd Qu.:32.148
## Max. :263.0	Max. :956.00	Max. :138.00	Max. :95.459
## NA's :254	NA's :125	NA's :100	NA's :433
## x120	x121	x122	x123
## Min. : 45.61	Min. : 8.862	Min. : 67.00	Min. : 9.00
## 1st Qu.: 72.43	1st Qu.: 16.172	1st Qu.: 87.00	1st Qu.: 68.00
## Median : 97.93	Median : 22.314	Median : 87.00	Median : 75.00
## Mean :114.64	Mean : 36.410	Mean : 86.41	Mean : 73.09
## 3rd Qu.:137.50	3rd Qu.: 31.012	3rd Qu.: 87.00	3rd Qu.: 83.00
## Max. :410.17	Max. :294.648	Max. :100.00	Max. :100.00
## NA's :2	NA's :12	NA's :2	NA's :118
## x124	x125	x126	x127
## Min. : 0.700	Min. : 2.200	Min. :49.63	Min. : -2.6979
## 1st Qu.: 3.225	1st Qu.: 9.325	1st Qu.:64.44	1st Qu.: -0.1260
## Median : 4.400	Median :13.400	Median :70.06	Median : 0.4737
## Mean : 5.346	Mean :15.206	Mean :72.15	Mean : 0.4339
## 3rd Qu.: 6.500	3rd Qu.:18.900	3rd Qu.:81.64	3rd Qu.: 0.9533
## Max. :24.500	Max. :48.400	Max. :97.90	Max. : 3.3393
## NA's :2	NA's :2		
## x128	x129	x130	x131
## Min. :53.90	Min. : 51.93	Min. : 1582	Min. : 1.800
## 1st Qu.:81.60	1st Qu.: 89.76	1st Qu.: 7033	1st Qu.: 5.900
## Median :85.65	Median : 99.25	Median :15897	Median : 7.800
## Mean :83.42	Mean : 98.70	Mean :15305	Mean : 8.947
## 3rd Qu.:88.30	3rd Qu.:105.73	3rd Qu.:22471	3rd Qu.:10.700
## Max. :93.50	Max. :162.81	Max. :39707	Max. :27.500
## NA's :2	NA's :192	NA's :103	NA's :2
## x132	x133	x134	x135
## Min. : 8.80	Min. :12.40	Min. : 2.50	Min. : 7.90
## 1st Qu.:29.12	1st Qu.:21.80	1st Qu.:10.20	1st Qu.:14.47
## Median :43.15	Median :27.40	Median :13.00	Median :18.50
## Mean :41.09	Mean :30.97	Mean :13.59	Mean :19.01
## 3rd Qu.:50.70	3rd Qu.:35.67	3rd Qu.:16.10	3rd Qu.:23.02
## Max. :76.30	Max. :79.50	Max. :30.00	Max. :50.20
## NA's :42	NA's :6	NA's :129	NA's :128
## x136	x137	x138	x139
## Min. :57.60	Min. : -14.5599	Min. : 3000	Min. : -28.9569
## 1st Qu.:66.62	1st Qu.: 0.4606	1st Qu.:10825	1st Qu.: 0.4214
## Median :70.40	Median : 1.9993	Median :21250	Median : 2.3580
## Mean :70.33	Mean : 2.1382	Mean :23835	Mean : 2.0985
## 3rd Qu.:74.30	3rd Qu.: 3.8615	3rd Qu.:33575	3rd Qu.: 4.1049
## Max. :82.10	Max. : 24.7648	Max. :84400	Max. : 20.8496
## NA's :2		NA's :2	NA's :30
##			

x140	x141	x142	x143
## Min. : 2.21	## Min. :20.51	## Min. : 0.10	## Min. :0.100
## 1st Qu.:22.03	## 1st Qu.:28.80	## 1st Qu.: 50.35	## 1st Qu.:1.500
## Median :31.98	## Median :33.49	## Median :136.00	## Median :2.200
## Mean :29.59	## Mean :33.71	## Mean :170.86	## Mean :2.073
## 3rd Qu.:35.33	## 3rd Qu.:37.24	## 3rd Qu.:236.70	## 3rd Qu.:2.500
## Max. :59.50	## Max. :58.94	## Max. :898.10	## Max. :4.000
## NA's :52	## NA's :36	## NA's :17	## NA's :59
x144	x145	x146	x147
## Min. : 2.900	## Min. : 1.60	## Min. : 3.700	## Min. : 0.100
## 1st Qu.: 4.300	## 1st Qu.:31.15	## 1st Qu.: 7.225	## 1st Qu.: 0.500
## Median : 4.700	## Median :46.30	## Median : 9.050	## Median : 0.800
## Mean : 4.925	## Mean :43.91	## Mean : 9.669	## Mean : 4.463
## 3rd Qu.: 5.300	## 3rd Qu.:54.80	## 3rd Qu.:11.475	## 3rd Qu.: 3.750
## Max. :15.100	## Max. :87.70	## Max. :24.200	## Max. :42.100
## NA's :39	## NA's :66	## NA's :138	## NA's :149

3 priedas

Tarptautinės migracijos srautų tyrimui naudotų kintamųjų histogramos





Optimalaus klasterių skaičiaus parinkimas skirtingais kriterijais

Kriterijus	Optimalus klasterių skaičius	Kriterijus	Optimalus klasterių skaičius
Duda	2	Tau	3
Beale	2	SDindex	3
Ratkowsky	2	Rubin	5
Dunn	3	KL	5
Scott	3	CH	5
Marriot	3	CCC	5
TraceW	3	TrCovW	5
DB	3	Gamma	5
Silhouette	3	Gplus	8
Ball	3	Hartigan	8
PtBiserial	3	Friedman	8
Gap	3	Cindex	8
McClain	3	SDBw	8

Tarptautinės migracijos srautams įtaką darančių veiksnių analizės tyrimui naudotų kintamųjų sąrašas

EG – ekonominės gerovės veiksniai. SG – socialinės gerovės veiksniai.

Trum-pinys	Grū-pė	Pogrūpis	Detalus pavadinimas
x1	EG	Kiti ekonominiai rodikliai	Koreguotos grynosios nacionalinės pajamos (metinis augimas %)
x2	EG	Kiti ekonominiai rodikliai	Koreguotos grynosios nacionalinės pajamos vienam gyventojui (metinis augimas %)
x3	EG	Su BVP susieti rodikliai	Koreguojami sutaupymai: pagrindinio kapitalo vartojimas (% nuo BNP)
x4	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: energijos išekvojimas (% nuo BNP)
x5	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: bendrosios santaupos (% nuo BNP)
x6	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: mineralų išekvojimas (% nuo BNP)
x7	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: gamtinių išteklių išekvojimas (% nuo BNP)
x8	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: grynieji nacionaliniai sutaupymai (% nuo BNP)
x9	EG	Su BVP susieti rodikliai	Koreguoti taupymo būdai: išlaidos švietimui (% nuo BNP)
x10	SG	Demografinė situacija	Jaunų mamų skaičius (1000 15-19 metų amžiaus moterų)
x11	SG	Švietimas	Nesimokantys paaugliai (% nuo visų žemesniojo vidurinio mokyklinio amžiaus vaikų skaičiaus)
x12	SG	Demografinė situacija	Amžių priklausomybės santykis (% nuo darbingo amžiaus gyventojų)
x13	SG	Aplinka ir gyvenimo sąlygos	Žemės ūkio paskirties žemė (% nuo žemės ploto)
x14	EG	Su BVP susieti rodikliai	Žemės ūkis, pridėtinė vertė (% nuo BVP)
x15	SG	Švietimas	Švietimo darbuotojų kompensavimas, iš viso (% nuo visų išlaidų valstybės institucijose)
x16	SG	Aplinka ir gyvenimo sąlygos	Alternatyvi ir branduolinė energija (% nuo viso energijos naudojimo)
x17	SG	Sveikatos apsauga	Antiretrovirusinio gydymo aprėptis (% nuo žmonių, gyvenančių su ŽIV)
x18	SG	Aplinka ir gyvenimo sąlygos	Ariamoji žemė (% nuo žemės ploto)
x19	SG	Aplinka ir gyvenimo sąlygos	Ginkluotųjų pajėgų personalas (% nuo visos darbo jėgos)
x20	EG	Kiti ekonominiai rodikliai	Banko kapitalo ir turto santykis (%)
x21	EG	Kiti ekonominiai rodikliai	Banko likvidumo rezervai banko turto santykiui (%)

Trum-pinys	Grū-pė	Pogrūpis	Detalus pavadinimas
x22	SG	Demografinė situacija	Gimstamumas (1 000 žmonių)
x23	SG	Sveikatos apsauga	Gimdymai, kuriuose dalyvauja kvalifikuotas sveikatos priežiūros personalas (% nuo visų gimdymų)
x24	EG	Kiti ekonominiai rodikliai	Komercinių bankų paskolos gavėjai (1000 suaugusiųjų)
x25	SG	Aplinka ir gyvenimo sąlygos	Mirtis dėl sužalojimo (% nuo visų mirčių)
x26	EG	Su BVP susieti rodikliai	Centrinės valdžios skola, iš viso (% nuo BVP)
x27	SG	Sveikatos apsauga	Vaikai (0-14 metų) sergantys ŽIV
x28	EG	Kiti ekonominiai rodikliai	Komercinių bankų filialai (100 000 suaugusiųjų)
x29	EG	Kiti ekonominiai rodikliai	Kompensacija darbuotojams (% nuo išlaidų)
x30	EG	Kiti ekonominiai rodikliai	Vartotojų kainų indeksas (2010 m. = 100)
x31	EG	Su BVP susieti rodikliai	Verslo pradžios procedūrų sąnaudos (% nuo BNP vienam gyventojui)
x32	EG	Su BVP susieti rodikliai	Einamosios sąskaitos balansas (% nuo BVP)
x33	SG	Švietimas	Bendros švietimo išlaidos, iš viso (% nuo visų išlaidų valstybės institucijose)
x34	SG	Demografinė situacija	Mirtingumo lygis (1 000 žmonių)
x35	EG	Kiti ekonominiai rodikliai	Indėlio palūkanų norma (%)
x36	SG	Aplinka ir gyvenimo sąlygos	Atstumo iki pasienio balas (0 = mažiausias veikimas, 100 = siena)
x37	EG	Su BVP susieti rodikliai	Vidaus kreditai privačiam sektoriui (% nuo BVP)
x38	EG	Su BVP susieti rodikliai	Bankų vidaus kreditai privačiam sektoriui (% nuo BVP)
x39	SG	Švietimas	Žmonės, su bent magistro arba lygiaverčiu išsilavinimu (% nuo visų gyventojų)
x40	SG	Švietimas	Žmonės, su bent jau bakalauro arba lygiaverčiu išsilavinimu (% nuo visų gyventojų)
x41	SG	Švietimas	Žmonės, su doktorantūros arba lygiaverčiu išsilavinimu (% nuo visų gyventojų)
x42	SG	Aplinka ir gyvenimo sąlygos	Elektros energijos gamyba iš hidroelektrinių šaltinių (% nuo visų)
x43	SG	Aplinka ir gyvenimo sąlygos	Elektros gamyba iš gamtinių dujų šaltinių (iš viso%)
x44	SG	Aplinka ir gyvenimo sąlygos	Elektros energijos gamyba iš branduolinių šaltinių (iš viso %)
x45	SG	Aplinka ir gyvenimo sąlygos	Elektros energijos gamyba iš naftos šaltinių (iš viso %)
x46	SG	Darbo rinka	Darbdaviai, iš viso (% nuo visų darbo vietų)
x47	SG	Darbo rinka	Užimtumas žemės ūkyje (% nuo bendro užimtumo)
x48	SG	Darbo rinka	Užimtumas pramonėje (% nuo bendro užimtumo)

Trum- pinys	Grū- pė	Pogrūpis	Detalus pavadinimas
x49	SG	Darbo rinka	Užimtumas paslaugų sektoriuje (% nuo bendro užimtumo)
x50	SG	Darbo rinka	Užimtumas ir gyventojų santykis, gyventojai 15+ amžiaus (iš viso %)
x51	SG	Aplinka ir gyvenimo sąlygos	Energijos importas (% nuo visos sunaudotos energijos)
x52	SG	Švietimas	Išlaidos viduriniam ugdymui (% nuo švietimo vyriausybės išlaidų)
x53	SG	Švietimas	Išlaidos aukštojo mokslo srityje (% nuo visų vyriausybės išlaidų švietimui)
x54	EG	Su BVP susieti rodikliai	Išlaidos (% nuo BVP)
x55	EG	Su BVP susieti rodikliai	Prekių ir paslaugų eksportas (% nuo BVP)
x56	EG	Kiti ekonominiai rodikliai	Prekių ir paslaugų eksportas (% metinis augimas)
x57	EG	Su BVP susieti rodikliai	Išorinis prekių ir paslaugų balansas (% nuo BVP)
x58	EG	Su BVP susieti rodikliai	Užsienio skola (% nuo BNP)
x59	EG	Su BVP susieti rodikliai	Galutinio vartojimo išlaidos ir kt. (% nuo BVP)
x60	EG	Kiti ekonominiai rodikliai	Galutinio vartojimo išlaidos ir kt. (metinis augimas, %)
x61	EG	Kiti ekonominiai rodikliai	Maisto eksportas (% nuo prekių eksporto)
x62	EG	Kiti ekonominiai rodikliai	Maisto importas (% nuo prekių importo)
x63	EG	Su BVP susieti rodikliai	Tiesioginės užsienio investicijos, grynosios įplaukos (% nuo BVP)
x64	EG	Su BVP susieti rodikliai	Tiesioginės užsienio investicijos, gryniesi pinigų srautai (% nuo BVP)
x65	SG	Aplinka ir gyvenimo sąlygos	Miško plotas (% nuo žemės ploto)
x66	EG	Kiti ekonominiai rodikliai	Valdžios sektoriaus galutinio vartojimo išlaidos (metinis augimas, %)
x67	EG	Su BVP susieti rodikliai	Valdžios sektoriaus galutinio vartojimo išlaidos (% nuo BVP)
x68	SG	Švietimas	Vyriausybės išlaidos švietimui, iš viso (% nuo BVP)
x69	SG	Švietimas	Vyriausybės išlaidos švietimui, iš viso (% nuo valdžios sektoriaus išlaidų)
x70	EG	Su BVP susieti rodikliai	Bendrosios vidaus santaupos (% nuo BVP)
x71	EG	Su BVP susieti rodikliai	Bendrosios nacionalinės išlaidos (% nuo BVP)
x72	EG	Su BVP susieti rodikliai	Bendras taupymas (% nuo BVP)
x73	EG	Kiti ekonominiai rodikliai	Eksportas aukštųjų technologijų srityje (% nuo pagaminto eksporto)
x74	EG	Kiti ekonominiai rodikliai	Namų ūkio vartojimo išlaidų vienam gyventojui augimas (metinis, %)
x75	EG	Su BVP susieti rodikliai	Prekių ir paslaugų importas (% nuo BVP)

Trum-pinys	Grū-pė	Pogrūpis	Detalus pavadinimas
x76	SG	Sveikatos apsauga	Patobulintos sanitarinės priemonės (% nuo gyventojų, turinčių galimybę naudotis)
x77	SG	Sveikatos apsauga	Geresnio vandens šaltinis (% nuo gyventojų, turinčių galimybę naudotis)
x78	SG	Sveikatos apsauga	ŽIV paplitimas (% nuo neinfekuotų 15-49 metų amžiaus gyventojų)
x79	SG	Sveikatos apsauga	Tuberkuliozės paplitimas (100 000 gyventojų)
x80	SG	Aplinka ir gyvenimo sąlygos	Asmenys, naudojantys internetą (% nuo visų gyventojų)
x81	EG	Su BVP susieti rodikliai	Pramonė, pridėtinė vertė (% nuo BVP)
x82	EG	Kiti ekonominiai rodikliai	Vartotojų kainų infliacija (metinė, %)
x83	EG	Su BVP susieti rodikliai	Infliacija, BVP deflatorius (metinis, %)
x84	EG	Kiti ekonominiai rodikliai	Tarptautinio turizmo išlaidos (% nuo viso importo)
x85	SG	Darbo rinka	Darbo jėga su aukštesniu išsilavinimu (% nuo visų darbingo amžiaus žmonių su aukštesniu išsilavinimu)
x86	SG	Darbo rinka	Darbo jėga su tarpiniu išsilavinimu (% nuo visų darbingo amžiaus gyventojų, turinčių tarpinį išsilavinimą)
x87	EG	Kiti ekonominiai rodikliai	Darbo mokestis ir įmokos (% nuo viso komercinio pelno)
x88	EG	Kiti ekonominiai rodikliai	Paskolos palūkanų norma (%)
x89	SG	Švietimas	Raštingumas tarp suaugusiųjų (% nuo 15 metų ir vyresnių žmonių)
x90	EG	Kiti ekonominiai rodikliai	Gaminių eksportas (% nuo viso eksporto)
x91	EG	Kiti ekonominiai rodikliai	Gaminių importas (% nuo viso importo)
x92	EG	Su BVP susieti rodikliai	Gamyba, pridėtinė vertė (% nuo BVP)
x93	SG	Aplinka ir gyvenimo sąlygos	Karinės išlaidos (% nuo BVP)
x94	SG	Sveikatos apsauga	Suaugę žmonės (vyresni nei 15 metų amžiaus), naujai užsikrėtę ŽIV
x95	SG	Aplinka ir gyvenimo sąlygos	Nuolatinė dirbama žemė (% nuo sausumos ploto)
x96	SG	Aplinka ir gyvenimo sąlygos	Asmeniniai pervedimai (% nuo BVP)
x97	SG	Sveikatos apsauga	Gydytojai (1 000 žmonių)
x98	SG	Demografinė situacija	Gyventojų augimas (metinis, %)
x99	SG	Demografinė situacija	Gyventojai, moterys (% nuo visų gyventojų)
x100	SG	Demografinė situacija	Gyventojai, vyrai (% nuo visų gyventojų)
x101	SG	Aplinka ir gyvenimo sąlygos	Skurdo santykis pagal nacionalines skurdo ribas (% nuo visų gyventojų)

Trum-pinys	Grū-pė	Pogrūpis	Detalus pavadinimas
x102	SG	Aplinka ir gyvenimo sąlygos	Prastai besimaitinantys gyventojai (% nuo visų gyventojų)
x103	EG	Kiti ekonominiai rodikliai	Pelno mokestis (% nuo viso komercinio pelno)
x104	SG	Aplinka ir gyvenimo sąlygos	Procentinis moterų kiekis nacionaliniame parlamente
x105	EG	Kiti ekonominiai rodikliai	Tikroji palūkanų norma (%)
x106	SG	Švietimas	Mokslinių tyrimų ir taikomosios veiklos išlaidos (% nuo BVP)
x107	SG	Demografinė situacija	Kaimo gyventojai (% nuo visų gyventojų)
x108	SG	Demografinė situacija	Kaimo vietovių gyventojų augimas (metinis%)
x109	SG	Demografinė situacija	Lyčių santykis gimimo metu (vyrai/moterys)
x110	SG	Sveikatos apsauga	Rūkymo paplitimas, moterys (% nuo visų suaugusiųjų)
x111	SG	Sveikatos apsauga	Rūkymo paplitimas, vyrai (% nuo visų suaugusiųjų)
x112	EG	Su BVP susieti rodikliai	Mokesčių pajamos (% nuo BVP)
x113	EG	Kiti ekonominiai rodikliai	Tarptautinės prekybos mokesčiai (% nuo visų pajamų)
x114	SG	Švietimas	Mokslinių tyrimų ir taikomosios technikos specialistai (milijonui žmonių)
x115	SG	Švietimas	Aukštasis išsilavinimas, akademinis personalas (%)
x116	SG	Aplinka ir gyvenimo sąlygos	Laikas, reikalingas elektros energijai gauti (dienomis)
x117	SG	Aplinka ir gyvenimo sąlygos	Reikalingas laikas, norint įregistruoti turtą (dienomis)
x118	SG	Aplinka ir gyvenimo sąlygos	Laikas, reikalingas verslui pradėti (dienomis)
x119	EG	Kiti ekonominiai rodikliai	Bendras skolos aptarnavimas (% nuo prekių, paslaugų ir pirminių pajamų eksporto)
x120	EG	Su BVP susieti rodikliai	Prekyba (% nuo BVP)
x121	EG	Su BVP susieti rodikliai	Prekyba paslaugomis (% nuo BVP)
x122	SG	Sveikatos apsauga	Tuberkuliozės atvejų aptikimo sparta (% nuo visų tuberkuliozės formų)
x123	SG	Sveikatos apsauga	Sėkmingo tuberkuliozės gydymo rodiklis (% nuo naujų atvejų skaičius)
x124	SG	Darbo rinka	Nedarbas su aukštesniuoju išsilavinimu (% nuo visos darbo jėgos, turinčios išsilavinimą)
x125	SG	Darbo rinka	Nedarbo lygis su pagrindiniu išsilavinimu (% nuo visos darbo jėgos su pagrindiniu išsilavinimu)
x126	SG	Demografinė situacija	Miesto gyventojai (% nuo visų gyventojų)
x127	SG	Demografinė situacija	Miesto gyventojų skaičiaus augimas (metinis, %)
x128	SG	Darbo rinka	Darbo užmokestį gaunantys darbuotojai, iš viso (% nuo visų darbo vietų)

Trum- pinys	Grū- pė	Pogrūpis	Detalus pavadinimas
x129	EG	Kiti ekonominiai rodikliai	Būsto kainų indeksas
x130	SG	Darbo rinka	Vidutinės pajamos
x131	SG	Darbo rinka	Nedarbo lygis (%)
x132	SG	Darbo rinka	Ilgalaikis nedarbas (12 mėn. ir daugiau, % nuo bendro nedarbo)
x133	SG	Švietimas	Žmonės su viduriniu arba žemesniu išsilavinimu (% nuo visų gyventojų)
x134	SG	Aplinka ir gyvenimo sąlygos	Smurtas, terorizmas, vandalizmas rajone (EU-SILC apklausa, %)
x135	SG	Aplinka ir gyvenimo sąlygos	Triukšmas iš kaimynų ar gatvės (%)
x136	SG	Darbo rinka	Užimtumas 16-64m
x137	EG	Su BVP susieti rodikliai	BVP vienam gyventojui augimas (metinis, %)
x138	EG	Su BVP susieti rodikliai	BVP vienam gyventojui
x139	EG	Su BVP susieti rodikliai	BNP augimas (metinis, %)
x140	SG	Aplinka ir gyvenimo sąlygos	Socialinės įmokos (% nuo visų pajamų)
x141	EG	Kiti ekonominiai rodikliai	Prekių ir paslaugų mokesčiai (% nuo visų pajamų)
x142	SG	Demografinė situacija	Pilietybės suteikimai (100 000 gyventojų)
x143	SG	Demografinė situacija	Skyrybų lygis (skyrybų skaičius/vidutinio gyventojų skaičiaus 1000 gyventojų)
x144	SG	Demografinė situacija	Santuokų lygis (skyrybų skaičius/vidutinio gyventojų skaičiaus, 1000 gyventojų)
x145	SG	Demografinė situacija	Skyrybų ir santuokų santykis (100 gyventojų)
x146	SG	Darbo rinka	Gyventojai nuo 0 iki 59 metų, gyvenantys labai mažo darbo intensyvumo namų ūkiuose (% nuo visų namų ūkių)
x147	SG	Aplinka ir gyvenimo sąlygos	Gyventojai be vonios ir dušo nuosavuose namuose (% nuo visų gyventojų)