

Faktorinės analizės taikymo regionų lyginamajai analizei galimybės

Bronislava Kaminskienė

Matematikos mokslų daktarė
Lietuvos mokslų akademijos
Matematikos ir informatikos instituto
vyr. mokslinė bendradarbė
Akademijos g. 4, Vilnius
Tel. (370 2) 72 96 04
El. paštas: bronek@mail.std.lt

Algimantas Misiūnas

Docentas socialinių mokslų daktaras
Vilniaus universiteto Ekonomikos fakultetas
Sisteminės analizės katedra
Saulėtekio al. 9, 2040 Vilnius
Tel. (370 2) 36 61 41
El. paštas: algism@mail.std.lt

Straipsnyje nagrinėjamos komponentinės ir faktorinės analizės metodų taikymo regionų ekonominiams tyrimams galimybės. Taikant šiuos metodus gauti bendri rodikliai padeda įvertinti socialinį ir ekonominį apskričių išsivystymo lygį. Ekonominio išsivystymo rodiklis palygintas su bendruoju vidaus produktu, tenkančiu vienam gyventojui. Gauti rezultatai rodo, kad taikyti matematiniai statistikos modeliai yra tinkami naudoti regionų lyginamajai analizei.

Įvadas

Darbo tikslas – suformuoti statistinių rodiklių rinkinį, kuris suteiktų galimybę įvertinti susidariusius atskirų teritorinių administracinių vienetų socialinio ir ekonominio išsivystymo skirtumus, nustatyti labiausiai atsilikusias apskritis.

Lietuva, vykdydama keliamus narystės Europos Sąjungoje (ES) reikalavimus, regioninę politiką išskyrė kaip prioritetinę veiklos sritį. Regioninės politikos priemonėmis reguliuojami susidarę atskirų teritorinių vienetų (administraciniai teritoriniai vienetai – apskritys, kurių dabar yra 10) ekonominiai ir socialiniai skirtumai. Formuluojami tokie regioninės politikos tikslai:

1) remti rinkos ekonomikos plėtrą kiekviename šalie regione, sukurti reikiamas ilgalaikės plėtros sąlygas;

2) formuoti socialinius-ekonominius ir aplinkosaugos reikalavimus atitinkančią infrastruktūrą;

3) mažinti regionų gyventojų gyvenimo lygio, kultūros, išsimokslinimo sąlygų skirtumus.

Regioninės politikos tikslams įgyvendinti numatyta keletas priemonių. Viena iš jų – parengti statistinių rodiklių sistemą, charakterizuojančią socialinį-ekonominį regionų išsivystymo lygį. Rodiklių sistema padėtų nustatyti labiausiai atsilikusias apskritis, sudaryti regioninės plėtros programas bei įvertinti pasiektus rezultatus.

Apskričių išsivystymo lygio susidariusiems netolygumams įvertinti naudojami statistiniai rodikliai, kurie apima įvairias gyvenimo sritis – demografiją, sveikatos rizikos veiksnius, užimtumą, laisvalaikio leidimo būdus, aplinkos apsaugą, nusikalstamumą, žemės ūkį, pramonę,

statybą, prekybą, paslaugas, investicijas (iš viso 44 rodikliai). Iš šio didelio kiekio statistinių rodiklių svarbu atrinkti informatyvius, geriausiai apibūdinančius tiriamą sritį. Kai kurioms veiklos rūšims apibūdinti parinkti požymiai yra nesudėtinga. Pavyzdžiui, pramonės efektyvumui nusakyti tinka rodiklis – parduota pramonės produkcija vienam darbuotojui. Tačiau yra veiklos rūšių, kurias apibūdina keletas požymių, ir sunku spręsti, kuris jų yra reikšmingiausias. Pavyzdžiui, nėra požymio, nusakančio visą gautą derlių, tačiau yra apskaitomi linų, cukrinių runkelių, grūdų, bulvių ir daržovių derliai. Tokiu atveju tikslinga tirti rodiklių tarpusavio ryšius ir sudaryti naują bendrą kintamąjį, į kurį pradiniai požymiai būtų įtraukti su tam tikrais svoriais.

Tirti objektus, kurie turi daug juos apibūdinančių ir tarpusavyje priklausomų rodiklių, taikomi komponentinės ir faktorinės analizės metodai. Komponentinės ir faktorinės analizės metodai tiksliai: sumažinti rodiklių, apibūdinančių tiriamąjį objektą (duomenų redukcija) skaičių, ir nustatyti kintamųjų tarpusavio ryšio struktūrą (klasifikacija) atliekant regionų išsivystymo lygio vertinimą. Taikant komponentinės ir faktorinės analizės metodus įvertinami numanomi veiksniai, atrenkami informatyviausi rodikliai, sudaromos tiesinės jų kombinacijos. Taip nuo pradinių duomenų pereinama prie apibendrintų, tarpusavyje nepriklausomų kintamųjų. Nauji požymiai arba tiesiog atrenkami iš pradinių, arba pradiniai įtraukiami su tam tikrais svoriais. Kintamuosius apibendrinant sumažėja rodiklių skaičius, tačiau informatyvumas nenukenčia. Atsižvelgiant į tai, kad pagrindinių komponentių analizė remiasi visų kintamųjų dispersija, o apibendrintų faktorių analizėje naudojama tik ta kintamųjų dispersijos dalis, kuri yra bendra su kitais kintamaisiais, darbe taikomi abu me-

todai. Pagrindinės komponentės sudaromos apibendrinant augalininkystės produkciją, paslaugų įmonių pajamas, demografijos rodiklius, sveikatos rizikos veiksnius, darbo rinką, laisvalaikio leidimo būdus. Maksimalaus tikėtinumo metodas taikomas nustatant bendrus faktorius, rodančius apskričių socialinį ir ekonominį lygį.

Statistikos duomenys ir tyrimo metodai

Darbe nagrinėjami rodikliai, apibūdinantys Lietuvos apskričių socialinę ir ekonominę būklę. Lyginamajai analizei atlikti sudaryti rodiklių, charakterizuojančių kiekvieną apskritį, rinkiniai. Kadangi apskritys skiriasi gyventojų skaičiumi, teritorijos dydžiu, žemės ūkio naudmenų plotu, lyginti tikslinga naudoti santykinius rodiklius. Pavyzdžiui, BVP, tenkantis vienam gyventojui, tarša viename ploto vienetu, galvijų skaičius viename naudmenų ploto vienetu ir pan.

Tiriant naudoti 1998 metų statistikos duomenys (iš viso 44 rodikliai).

Taigi vertiname daugiamačių atsitiktinį dydį:

$$X = (x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(p)})^T, \quad k = 1, 2, \dots,$$

čia p – rodiklių skaičius, n – stebėjimų skaičius. Tegu X pasiskirstęs pagal p -matį normalųjį dėsnį

$$\varphi(X, M, \Sigma) = \frac{1}{(2\pi)^p |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-M)^T \Sigma^{-1}(X-M)}$$

čia M – vidurkių vektorius, Σ – kovariacinė matrica.

$M = EX = (Ex^{(1)}, Ex^{(2)}, \dots, Ex^{(p)})^T$. Statistinis vidurkių vektorius įvertis

$$\hat{M} = \bar{X} = (\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(p)})^T,$$

$$\bar{x}^{(i)} = \frac{1}{n} \sum_{k=1}^n x_k^{(i)}, \quad k = 1, 2, \dots,$$

Kovariacinės matricos įvertis

$$\hat{\Sigma} = [\hat{\sigma}_{ij}]_{i=1,2,\dots,p}^{j=1,2,\dots,p},$$

$$\hat{\sigma}_{ij} = \frac{1}{n} \sum_{k=1}^n (x_k^{(i)} - \bar{x}^{(i)}) (x_k^{(j)} - \bar{x}^{(j)}),$$

$$k = 1, 2, \dots,$$

Pagrindinių komponentių metodas yra taikomas spręsti daugiamatį rodiklių dimensijos mažinimo uždavinius. Pagrindinės komponentės gaunamos kaip pradinių požymių tiesinės kombinacijos ir tenkina šiuos optimizavimo uždavinio reikalavimus: maksimalaus informatyvumo (išsaugoti kuo daugiau informacijos apie pradinis duomenis), tarpusavio nepriklausomumo bei turi kuo mažiau iškraipyti geometrinę pradinių duomenų struktūrą. Komponentinė analizė pradinių rodiklių koreliacinę matricą padalija į ortogonalų vektorių aibę. Šie vektoriai atitinka koreliacinės matricos nuosavus vektorius. Nuosavos reikšmės išskiriamos mažėjančia tvarka, kas yra svarbu, jei norime sumažinti rodiklių skaičių. Kadangi vektoriai ortogonalūs, gautos iš jų komponentės yra tarpusavyje nepriklausomos. Nors kelios komponentės gali paaiškinti didelę bendros kintamųjų dispersijos dalį, tačiau, norint tiksliai atgaminti koreliaciją, reikia visų komponentių. Pradiniai kintamieji yra centruoti, todėl vidurkių vektorius $EX = 0$, o kovariacinė matrica $R = E(XX^T)$. Naudodami pagrindinių komponentių metodą pereisime prie tarpusavyje nekoreliuotų ir normuotų kintamųjų $Z = (z_k^{(1)}, z_k^{(2)}, \dots, z_k^{(p)})$. Pagrindinių komponentių Z ir pradinių rodiklių X ryšys užrašomas

$$Z = L^T \cdot X, \quad X = L \cdot Z,$$

čia $L = [l_{ij}]_{i,j=1,2,\dots,p}$ – ortogonalų pertvarkymų matrica.

Tegu l_j – j -asis matricos L stulpelis. Tuomet l_j parenkamas taip, kad pirmosios komponentės

dispersija $Dz^{(1)}$ būtų didžiausia. Po to l_1 parenkamas taip, kad antroji komponentė $z^{(2)}$ turėtų maksimalią dispersiją ir būtų nekoreliuota su $z^{(1)}$. Analogiškai skaičiuojamos likusios komponentės $z^{(j)}$, $j = 1, 2, \dots, n$: maksimizuojama dispersija su sąlyga, kad $z^{(j)}$ yra nepriklausoma su $z^{(1)}, z^{(2)}, \dots, z^{(j-1)}$. Grafiškai pagrindinių komponentių metodas reikštų, kad koordinatinių ašys perkeliama maksimalaus kintamųjų išsibarsavimo kryptimi.

Taigi pirmai komponentei $z^{(1)}$ rasti sprendžiamas tokio pavidalo optimizavimo uždavinys

$$\max_{l_1} D(l_1 X) = \max_{l_1} E(l_1 X)^2 = \max_{l_1} (l_1 \Sigma l_1^T) = \lambda_1,$$

čia λ_1 yra kovariacinės matricos R didžiausia nuosava reikšmė, o l_1 – ją atitinkantis nuosavas vektorius. Vektorius l_1 yra pirmas matricos L stulpelis, taigi jį suradus sudaroma pirmoji komponentė $z^{(1)} = l_1^T X$. Analogiškai randamos kitų komponentių dispersijos $Dz^{(j)} = \lambda_j$, jas atitinkantys vektoriai l_j sudaromos komponentės $z^{(j)} = l_j^T X$, $j = 2, 3, \dots, p$. Matriciniu pavidalu išraiška rašoma $DZ = L^T R L = \Lambda$.

Nuosavos reikšmės yra randamos iš sąlygos: $|R - \Lambda I| = 0$, čia Λ – diagonali matrica, o elementai ant įstrižainės išdėstyti $\lambda_1 = \lambda_2 = \dots = \lambda_p$. Bendra pradinių rodiklių dispersija yra lygi bendrai komponentių dispersijai ir pradinių rodiklių skaičiumi

$$\sum_{i=1}^p Dz^{(i)} = \sum_{i=1}^p Dx^{(i)} = p$$

Rodiklių skaičiaus mažinimo uždaviniuose paliekamos pirmos komponentės ($p' < p$), kurios paaiškina didelę bendros kintamųjų dispersijos dalį.

Pagrindinių komponentių analizėje labai svarbi svorių matrica $W = [w_{ij}]_{i,j=1,2,\dots,p}$. Elementas w_{ij} – komponentės $z^{(j)}$ svoris į rodiklį $x^{(i)}$, $i, j = 1, 2, \dots, p$. Svoriai yra labai svarbūs nagri-

nejant pradinių rodiklių ir komponentų tarpusavio ryšius ir norint jas interpretuoti, suteikti prasmę. Svorii matrica W gaunama sunormavus pagrindines komponentes $Z = \Lambda^{1/2} \cdot L^T X = W^T X$, $W = \Lambda^{1/2} \cdot L^T$. Atgaminata kovariacinė matrica gaunama iš lygybės $WW^T = R_{p'}$.

Darbe skirtumai tarp matricų laikomi nedideli, jei absoliučiu dydžiu yra mažesni už konstantą 0,2.

Faktorinė analizė remiasi prielaida, kad stebimų rodiklių $x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(p)}$, $k = 1, 2, \dots, n$ tarpusavio ryšių struktūrą galima paaiškinti šių požymių priklausomybe (tiesine ar kitokia) nuo mažesnio skaičiaus $p' < p$, neišmatuojamų faktorių $f_k^{(1)}, f_k^{(2)}, \dots, f_k^{(p')}$. Šie faktoriai vadinami apibendrintais ir sudaromi taip, kad būtų tarpusavyje nekoreliuoti. Tačiau aprašyti pradinius kintamuosius neužtenka vien apibendrintų faktorių, kiekvienas $x^{(i)}$, $i = 1, 2, \dots, p$, kartu priklauso ir nuo tam tikros savo liekanos $u^{(i)}$. Taigi faktorinės analizės metodų tikslas – išskirti ir interpretuoti bendrus faktorius, minimizuoti jų skaičių bei ryšio tarp požymio $x^{(i)}$ ir jo liekanos $u^{(i)}$ stiprumą.

Faktoriai turi maksimaliai paaiškinti pradinių kintamųjų koreliaciją. Pirmasis bendras faktorius $f^{(1)}$ randamas iš sąlygos, kad dviejų pradinių požymių koreliacija lygi nuliui, jei $f^{(1)}$ poveikis įvertintas. Jei tai nepaaiškina koreliacinės matricos, tuomet ieškomi tokie du atsitiktiniai dydžiai $f^{(1)}$ ir $f^{(2)}$, kad, įvertinus jų poveikį, dviejų pradinių kintamųjų koreliacijos būtų lygios nuliui, ir t. t.

Tegu atsitiktinis dydis $X = (x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(p)})^T$, $k = 1, 2, \dots, n$, turi p -matį normalųjį pasiskirstymą ir yra centruotas. Tiesinis faktorinės analizės modelis užrašomas

$$X = QF + U,$$

čia $Q = [q_{ij}]_{i=1,2,\dots,p}^{j=1,2,\dots,p'}$ – faktorinių svorių matrica, kuri nusako tiesinį ryšį tarp pradinių

kintamųjų ir apibendrintų faktorių $F = (f_k^{(1)}, f_k^{(2)}, \dots, f_k^{(p')})^T$, $U = (u_k^{(1)}, u_k^{(2)}, \dots, u_k^{(p)})^T$ – liekanų (pradinių požymių dalis, kurios nepaaiškina apibendrinti faktoriai) vektorius.

Norint aprašyti koreliuotus rodiklius $x^{(i)}$ ($i = 1, 2, \dots, p$) mažesniu skaičiumi apibendrintų faktorių $f^{(r)}$ ($r = 1, 2, \dots, p'$) reikia įvertinti nežinomus faktorinius svorius \hat{q}_{ir} , liekanų dispersijas \hat{v}_i ir pačius faktorius $\hat{f}^{(r)}$. Tam, kad modelis būtų vienareikšmiškai identifikuojamas, įvedamos papildomos sąlygos. Papildomi q_{ij} ir v_i ryšiai apibrėžiami dviem būdais: matrica $Q^T V Q$ – diagonali, o elementai ant įstrižainės skirtingi ir išdėstyti mažėjančia tvarka. Nežinomi parametrai q_{ir} ir v_i vertinami maksimalaus tikėtino metodo. Maksimalaus tikėtino metodo gauti faktorinių svorių ir liekanų dispersijų įverčiai yra efektyvūs. Nustatčius minimalų faktorių, kurių reikia paaiškinti pradinių kintamųjų ir skaičių koreliaciją bei radus faktorinius svorius, įvertinami patys faktoriai. Jiems rasti taikomas Barleto metodas.

Rezultatai

Nusakyti regionų ekonominę raidą sudarytas *ekonominio lygio indeksas*. Šis rodiklis yra sudėtinis ir turi plačią prasmę, nes jo komponentės apima įvairius šalies ūkio sritis ir veiktas – finansus, investicijas, žemės ūkį, pramonę, statybą, prekybą, paslaugas. Regionų ekonominę padėtis nagrinėjama dviem aspektais, vienas iš jų apibūdina finansinius išteklius, investicijų patrauklumą, kitas – gamybos efektyvumą, pasiektus rezultatus. Norint įvertinti gautų rezultatų patikimumą, regionų ekonominio lygio rodiklis bei jo komponentės lyginamos su BVP, tenkančiu vienam gyventojui. Šis palyginimas padės įvertinti komponentių įtaką, perspektyvias ir efektyvias regionų ekonominio gyvenimo sri-

tis bei veiklos rūšis. Tam naudojami rodikliai ($p = 10$):

Vidutinis mėnesinis apskaičiuotas darbo užmokestis;

Pelnas;

Materialinės investicijos;

Tiesioginės užsienio investicijos;

Derlius;

Gyvulininkystės produkcija, perskaičiuota į pieno rodiklius;

Parduota pramonės produkcija;

Statybos įmonių ir bendrovių atliktų darbų apimtis;

Prekybos, restoranų ir kitų maitinimo įstaigų mažmeninė prekių apyvarta;

Paslaugos.

Rodikliai nėra vienodo pobūdžio. Visi požymiai, išskyrus *derlių* ir *paslaugas*, teikia išsamią informaciją apie konkrečią sritį, veiklos rūšį, yra matuojami piniginiiais ar kiekio vienetais. Augalininkystės produkciją, paslaugų sritį apibūdina keletas rodiklių ir sunku spręsti, kurie jų tinkamesni analizuoti. Šioms veiklos rūšims nusakyti sudaryti bendri kintamieji, kurie neturi matavimo vienetų. Norint išvengti apskričių dydžio įtakos, imami santykiniai dydžiai, t. y. kiek vienam gyventojui tenka 1000-iais litų.

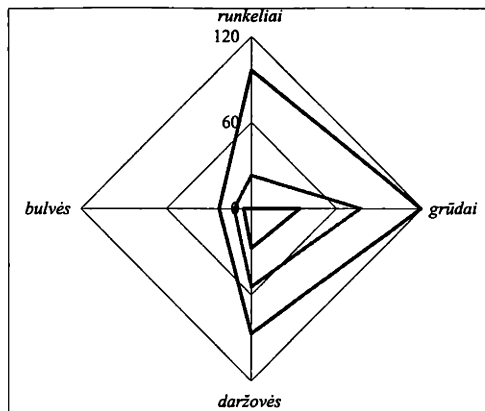
Rodiklių *pelnas*, *materialinės investicijos*, *užsienio investicijos*, *prekyba* skirtumai labai dideli. Ypač bloga padėtis yra Tauragės apskrityje – čia rodiklių *užmokestis*, *pelnas*, *materialinės investicijos*, *užsienio investicijos*, *statyba* reikšmės yra mažiausios. Vilniaus apskrityje geriausi rezultatai darbo užmokesčio, užsienio investicijų, statybos ir prekybos srityse. Daugiausia materialinių investicijų vienam gyventojui tenka Klaipėdos apskrityje, Telšių apskritis pirmauja pramonės ir statybos veikla. Taigi yra apskričių ekonominio pobūdžio skirtumų, išsiskiria prioritetinės sritys, veiklos rūšys. Apibendrinti

visą rodiklių teikiamą informaciją, nusakyti ir palyginti esamą regionų ekonominį lygį taikomi komponentinės ir faktorinės analizės metodai. Tačiau prieš tai įvertinsime žemės ūkio ir paslaugų įmonių darbo rodiklius, sudarysime apibendrintus kintamuosius.

Žemės ūkio efektyvumo tyrimas

Žemės ūkio vieta šalies ekonominėje ir socialinėje raidoje labai svarbi. 1999 m. žemės ūkyje sukurta 8,8 proc. (išankstiniai duomenys) bendrosios pridėtinės vertės (1998 m. – 10,2 proc.), o užimti gyventojai žemės ūkyje sudarė 20,1 proc. (1998 m. – 21,2%) visų užimtųjų. Todėl tikslinga įvertinti kiekvieno regiono žemės ūkio gamybos efektyvumą. Statistikos departamentas kaupia įvairius žemės ūkio duomenis: pasėlių plotas, derlius, gyvulių, paukščių skaičius, gautos produkcijos kiekis. Gautos augalininkystės ir gyvulininkystės produkcijos rodikliai labiau tinkami charakterizuoti ekonominę regionų raidą. Todėl gyvulininkystei apibūdinti pasirinkta gyvulininkystės produkcija perskaičiuota į pieno rodiklius (toliau – *gyvulininkystė*). Palyginti apskritis imamos santykinės rodiklių reikšmės – kiekis šimtui hektarų žemės ūkio naudmenų šimtais kilogramų. 1998 m. didžiausias produkcijos kiekis gautas Tauragės apskrityje – 1205, mažiausias Utenos apskrityje – 744, šalies vidurkis – 1022.

Augalininkystės gamybos efektyvumą nusako gautas derlius. Kadangi nėra požymio, tinkančio apibūdinti visą augalininkystės produkciją, atrinkti linų, cukrinių runkelių, grūdų, bulvių ir daržovių derlių duomenys. Pagrindinių komponentų metodu nustatomi informatyviausi požymiai, sudaromas bendras kintamasis – *derlius*. Taigi nuo penkių pradinių požymių ($p = 5$) pereisime prie vieno ($p' = 1$). Analizei atlikti naudojamos santykinės rodiklių reikšmės



1 pav. Žemės ūkio rodiklių skirtumai

1 lentelė. Augalininkystės produkcijos rodiklių koreliacijos koeficientai

	<i>Linai</i>	<i>Runkeliai</i>	<i>Grūdai</i>	<i>Bulvės</i>	<i>Daržovės</i>
<i>Linai</i>	1	0,82	0,76	-0,32	-0,06
<i>Runkeliai</i>		1	0,88	0,00	0,25
<i>Grūdai</i>			1	0,24	0,20
<i>Bulvės</i>				1	0,30
<i>Daržovės</i>					1

– derliaus kiekis, tenkantis šimtui hektarų žemės ūkio naudmenų šimtais kilogramų. Apskričių augalininkystės rodiklių skirtumai vaizduojami 1 paveiksle: didžiausia reikšmė, šalies vidurkis ir mažiausia reikšmė.

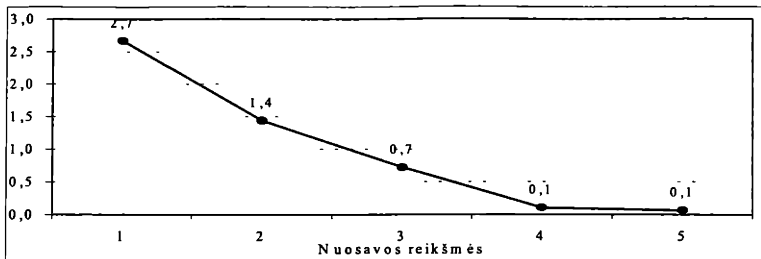
Marijampolės apskrityje gaunami didžiausi linų, runkelių ir grūdų derliai. Utenos apskrityje augalininkystės produktyvumas mažiausias. Tikėtina, kad augalininkystės požymiai tarpusavyje tiesiškai priklausomi. Apskaičiuoti jų koreliacijos koeficientai.

Iš 1 lentelės matyti, kad požymių *linai*, *runkeliai* ir *grūdai* koreliacijos koeficientai yra didesni už 0,7, taigi rodikliai yra stipriai tarpusavyje priklausomi. Jų ryšį būtų galima paaiškinti tuo, kad

linai, cukriniai runkeliai ir grūdai naudojami kaip žaliava maisto ir tekstilės pramonėje.

Stipri požymių koreliacija rodo, kad tikslinga taikyti pagrindinių komponentų metodą, atrinkti informatyviausius rodiklius bei sudaryti tiesines jų kombinacijas. Pirmą komponentę maksimizuoja pradinių kintamųjų dispersiją. Antra turi didžiausią dispersiją iš likusių ir t. t. Pagal radimo būdą komponentių dispersijos vadinamos nuosavomis reikšmėmis. Reikšmingiausios komponentės atrenkamos pagal nuosavos reikšmės dydį. Komponentių nuosavos reikšmės vaizduojamos 2 paveiksle.

Kaiserio kriterijus yra plačiai taikomas rodiklių mažinimo uždaviniuose. Juo remiantis



2 pav. Komponentių nuosavos reikšmės

siūloma palikti tas komponentes, kurių nuosavos reikšmės didesnės už 1. Aiškinama tuo, kad informatyvios tik tos komponentės, kurių dispersija didesnė už pradinio kintamojo (jų dispersijos lygios 1). Pritaikius minėtą kriterijų atrenkamos dvi pirmosios komponentės. Pirmos komponentės nuosava reikšmė lygi 2,7, ji paaiškina 53,4 proc. bendros kintamųjų dispersijos. Norint išsaugoti kuo daugiau informacijos, labai svarbi antra komponentė, kuri lygi 1,4 ir paaiškina 28,8 proc. Suminė dviejų komponentių dispersija $\sum_{i=1}^2 Dz^{(i)}$ sudaro 82,2 proc. bendros kintamųjų dispersijos $\sum_{i=1}^5 Dx^{(i)}$. Dvi komponentės ($p' = 2$) gali nusakyti pradinis rodiklius ($p = 5$). Atrinkus informatyviausias komponentes, nustatoma duomenų struktūra, apskaičiuojami komponentių ir pradinių kintamųjų koreliacijos koeficientai. Šie koeficientai vadinami svoriais.

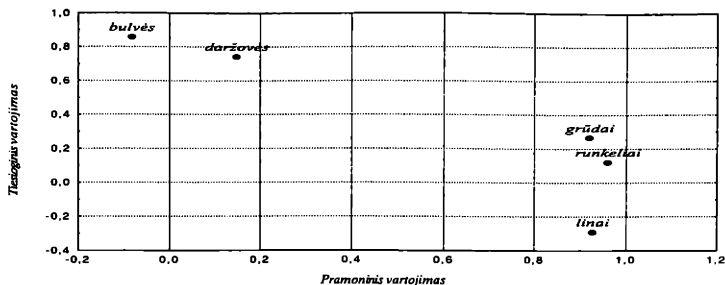
Iš 2 lentelės matyti, kad pirmą komponentę nusako rodiklius *linai*, *runkeliai* ir *grūdai*. Tai, kad šie požymiai susiję, parodė ir pradinės koreliacinės matricos analizė (1 lentelė). Pirmą komponentę interpretuojame kaip *pramoninį vartojimą*. Antra komponentė turi reikšmingą

2 lentelė. Komponentių svoriai pradiniam rodikliams

	Pramoninis vartojimas	Tiesioginis vartojimas
<i>Linai</i>	0,93	-0,29
<i>Runkeliai</i>	0,96	0,12
<i>Grūdai</i>	0,92	0,27
<i>Bulvės</i>	-0,08	0,86
<i>Daržovės</i>	0,15	0,74

ryšį su kintamaisiais *bulvės* ir *daržovės*. Palyginti tik nedidelė bulvių ir įvairių daržovių dalis naudojama kaip žaliava pramonėje, jos vartojamos tiesiogiai. Antrą komponentę galime įvardyti kaip *tiesioginį vartojimą*. Pradinių kintamųjų ir komponentių ryšiai vaizduojami 3 paveiksle. Pradiniai rodikliai yra taškai, o komponentės atidėtos ant ašių.

Iš 3 paveikslo matyti, kad rodiklius *linai*, *runkeliai*, *grūdai*, *bulvės*, *daržovės* ($p = 5$) galima aprašyti dviem komponentėmis ($p' = 2$) – *pramoniniu vartojimu* ir *tiesioginiu vartojimu*, nes pradiniai rodikliai suskirstomi į dvi grupes. Kadangi tiriant apskričių ekonominį išsivystymą svarbesni pramoniniai požymiai, rodiklius *bulvės* ir *daržovės* atmesime. Pakartotinai pagrindinių komponentių metodu tiriami požymių *linai*, *runkeliai*, *grūdai* ($p = 3$) tarpusavio ryšiai. Pagal nuosavos reikšmės dydį atrenkama pirmą kom-



3 pav. Komponentių svoriai pradiniam požymiams

3 lentelė. Komponentės derlius svoriai pradiniam rodikliams

	Linai	Runkeliai	Grūdai
Derlius	0,91	0,96	0,94

4 lentelė. Skirtumai tarp pradinių ir atgamintų koreliacijos koeficientų

	Linai	Runkeliai	Grūdai
Linai	0,17	-0,06	-0,10
Runkeliai		0,08	-0,02
Grūdai			0,12

ponentė. Jos nuosava reikšmė lygi 2,6 ir paaiškina 87,9 proc. bendros kintamųjų dispersijos. Taigi augalininkystės produkciją galima apibūdinti viena komponente, nes kitos sudaro palyginti nedidelę bendros dispersijos dalį (12,1 proc.). Komponentė vadinama *derliumi* ir bus naudojama tolesnei apskričių ekonominio išsivystymo analizei. Toliau pateikiami pagrindinės komponentės svoriai pradiniam kintamiesiems.

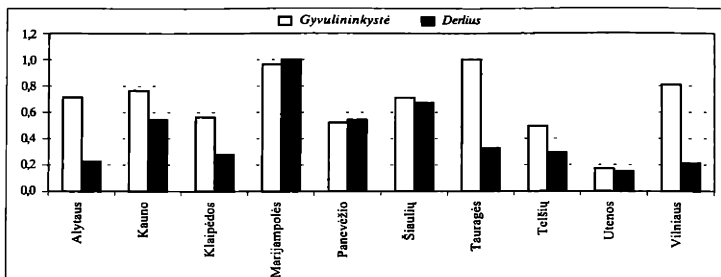
Iš 3 lentelės matyti, kad svoriai yra dideli (reikšmės didesnės už 0,9). Patikrinti, kaip gerai naujas rodiklis aprašo tiriamus, skaičiuojami skirtumai tarp pradinės ir atgamintos koreliacinių matricių. Žemiau pateikiami rezultatai.

Komponentė *derlius* tinka aprašyti rodiklius *linai*, *runkeliai*, *grūdai*, nes skirtumai tarp koreliacijos koeficientų yra maži (mažesni už absoliutų dydį 0,2). Komponentinės analizės metu atrinkti informatyviausi požymiai ir sudaryta

pagrindinė komponentė – *derlius* rodo pramoninį žemės ūkio produkcijos naudojimą.

Įvertinti gyvulininkystės ir augalininkystės efektyvumą apskirtyse rodiklis *gyvulininkystė* ir komponentė *derlius* perskaiciuojami į indeksus pagal JTVP pasiūlytą skaičiavimo metodiką. Gauti dydžiai patenka į intervalą nuo 0 (blogiausias rezultatas) iki 1 (geriausias rezultatas). Indeksų reikšmės vaizduojamos 4 paveiksle.

Rodiklių *gyvulininkystė* ir *derlius* indeksai rodo, kuriuose šalies regionuose žemės ūkio gamyba efektyviausia, bei stygių – skirtumų tarp esamo lygio ir galimo pasiekti (maksimalios reikšmės 1). Marijampolės apskrityje aukštas ir augalininkystės, ir gyvulininkystės lygis. Alytaus, Tauragės, Vilniaus apskrityse *gyvulininkystės* indekso stygius yra kur kas mažesnis negu *derliaus*. Kauno, Šiaulių apskrityse *gyvulininkystės* ir *derliaus* indeksų stygiai panašūs. Utenos ap-



4 pav. Žemės ūkio rodiklių indeksai

5 lentelė. Paslaugos

Paslaugos	Trumpiniai
Asmeninių ir buitinių daiktų taisymas	<i>taisymas</i>
Nekilnojamojo turto operacijos	<i>n. turtas</i>
Mašinų ir įrengimų be operatoriaus ar vairuotojo ir asmeninių bei buitinių daiktų nuoma	<i>nuoma</i>
Kompiuteriai ir su jais susijusi veikla	<i>kompiuteriai</i>
Mokslu tiriamieji ir taikomieji darbai	<i>m. darbai</i>
Verslo veiklos	<i>verslas</i>
Švietimas	<i>švietimas</i>
Atliekų ir šiukšlių šalinimas	<i>šalinimas</i>
Poilsio, pramogų organizavimas ir sportas	<i>poilsis</i>

skirtyje žemės ūkio gamybos efektyvumas mažiausias pagal abu požymius.

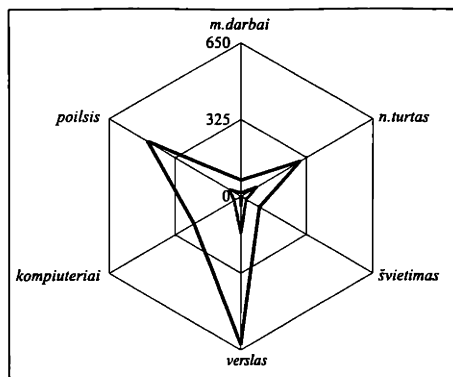
Paslaugų įmonių pajamų tyrimas

Šalyje vis labiau plečiasi teikiamų paslaugų sąrašas. Tai lemia didelę paslaugų sektoriaus dalį bendrojo vidaus produkto (BVP), užimtumo struktūroje. Išankstiniais duomenimis, 1999 m. paslaugų srityje buvo sukurta 60,1 proc. bendrosios pridėtinės vertės (1998 m. – 57,2 proc.), o užimti gyventojai paslaugų sektoriuje sudarė 53,4 proc. (1998 m. – 51,5 proc.) visų užimtųjų. Todėl paslaugų sektoriaus požymius būtina įtraukti į ekonominio regionų išsivystymo tyrimą. Paslaugų įmonės skirstomos pagal veiklas. Kadangi veiklos rūšių pavadinimai ilgi, tekste naudojami rodiklių trumpiniai. Lentelėje iš-

vardytos paslaugų įmonių grupės ($p = 9$), kurių rodikliai yra tiriami.

Analizuojami požymiai – pajamos, gautos už suteiktas paslaugas. Palyginti apskritis naudojami santykiniai dydžiai – pajamos, tenkančios vienam gyventojui, litais. Kai kurių paslaugų rodiklių skirtumai parodyti 5 paveiksle, atidėtos didžiausia reikšmė, šalies vidurkis ir mažiausia reikšmė.

Iš 5 paveikslu matyti, kad pajamos atskiruose regionuose pasiskirsčiusios labai nevienodai. Palyginti su kitomis apskritimis Vilniaus apskrityje paslaugų sektorius labai išvystytas. Pavyzdžiui, Vilniaus apskrityje pajamos gautos už poilsio, pramogų organizavimą ir sportą vienam gyventojui sudarė 462,4 litų, o Tauragės apskrityje – tik 1,1 lito. Pajamos gautos už kitas su-



5 pav. Paslaugų rodiklių skirtumai

6 lentelė. Paslaugų rodiklių koreliacijos koeficientai

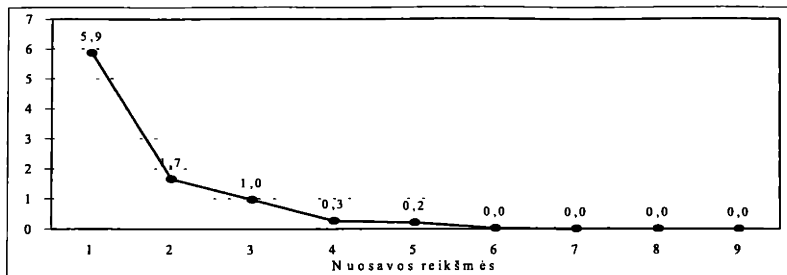
	Taisymas	N. turtas	Nuoma	Kompiuteriai	M. darbai	Verslas	Švietimas	Šalinimas	Poilsis
Taisymas	1	0,15	0,75	0,07	-0,07	0,14	0,12	-0,17	0,10
N. turtas		1	0,51	0,95	0,71	0,94	0,90	-0,50	0,94
Nuoma			1	0,37	0,09	0,41	0,41	-0,21	0,39
Kompiuteriai				1	0,69	0,96	0,84	-0,35	0,99
M. darbai					1	0,85	0,92	-0,65	0,63
Verslas						1	0,95	-0,50	0,93
Švietimas							1	-0,65	0,80
Šalinimas								1	-0,30
Poilsis									1

teiktas paslaugas taip pat rodo regionų išsivystymo skirtumus. Kadangi rodiklių yra devyni ($p = 9$), reikia iš jų atrinkti informatyviausius, labiausiai tinkamus charakterizuoti regiono išsivystymą. Tikėtina, kad kai kurie paslaugų sektoriaus požymiai yra tarpusavyje koreliuoti, jų kitimą lemia tie patys veiksniai. Suskaičiuojama pradinių duomenų koreliacinė matrica.

Pastebima stipri tiesinė požymių *n.turtas*, *kompiuteriai*, *m.darbai*, *verslas*, *švietimas*, *poilsis* priklausomybė. Ši rodiklių grupė rodo gyvenimo kokybę, gyventojų galimybes įsigyti nekasdienį turtą, vertybes. Kiti tarpusavyje koreliuoti

rodikliai – *taisymas* ir *nuoma*. Jie rodo kasdienių, buitinių daiktų naudojimą. Rodiklis *šalinimas* nuo kitų požymių mažai priklausomas. Tai, kas išmetama, jau nebėra vertybė ir nerodo gyventojų gerovės. Mažinti pradinių kintamųjų skaičių taikomas pagrindinių komponenčių metodas. Reikšmingiausias komponentės atrinkamos pagal nuosavų reikšmių dydį.

Kaip matyti iš 6 paveikslėlio, nuosavų reikšmių, didesnių už 1, yra dvi, taigi paslaugų įmonių pajamas galima nusakyti dviem kintamaisiais. Pirmos komponentės nuosava reikšmė lygi 5,9, antros – 1,7. Jų suminė dispersija paaiškina



6 pav. Komponentių nuosavos reikšmės

7 lentelė. Pradinių rodiklių komponentių svoriai

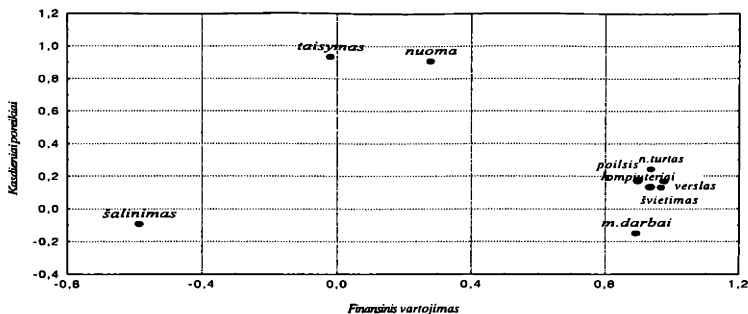
	Taisy- mas	N. turtas	Nuoma	Kompiu- teriai	M. dar- bai	Veršlas	Švieti- mas	Šalini- mas	Poilsis
Finansinės galimybės	-0,02	0,94	0,28	0,93	0,89	0,97	0,96	-0,59	0,90
Kasdieniai poreikiai	0,93	0,25	0,91	0,14	-0,15	0,16	0,13	-0,09	0,17

83,5 proc. bendros kintamųjų dispersijos. Nuo devynių ($p = 9$) rodiklių galima pereiti prie dviejų ($p' = 2$), nes likusios komponentės sudaro palyginti mažą bendros dispersijos dalį (16,5%). Kai jau žinomas informatyviausių rodiklių skaičius, nagrinėjama duomenų struktūra. Apskaičiuojami pradinių rodiklių pagrindinių komponentių svoriai.

Pirma komponentė turi didelius svorius rodiklyje *n.turtas*, *kompiuteriai*, *m.darbai*, *veršlas*, *švietimas* ir *poilsis*. Minėti rodikliai rodo gyventojų galimybes įsigyti ilgalaikį turtą, įgyti naujų žinių, maloniai leisti laisvalaikį, todėl komponentę interpretuojame kaip *finansines galimybes*. Kintamasis *finansinės galimybės* turi nedidelius svorius požymyje *taisyimas* ir *nuoma*, tačiau antros komponentės svoriai juose yra dideli. Asmeniniai ir buitiniai daiktai naudojami dažnai, todėl ši komponentė įvardija-

ma kaip *kasdieniai poreikiai*. Apibendrintų kintamųjų svoriai rodiklyje *šalinimas* maži. Sviurių pasiskirstymas vaizduojamas 7 paveiksle.

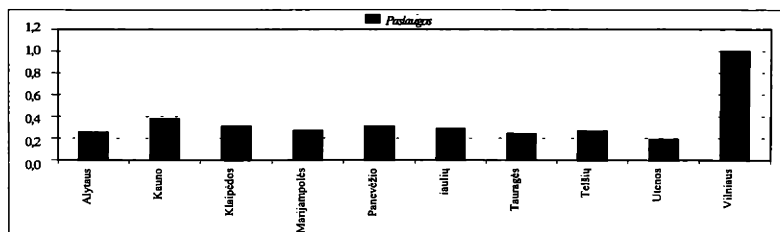
Iš 7 paveikslu matyti susidariusios pradinių rodiklių grupės bei ryšiai su komponentėmis *finansinės galimybės* ir *kasdieniai poreikiai*. Kadangi tiriant gyvenimo lygį svarbesni yra tie rodikliai, kurie rodo gyventojų materialinę gerovę, tai rodiklius *taisyimas*, *nuoma*, *šalinimas* atmesime, kartu panaikinsime šių rodiklių poveikį bendram kintamajam *finansinės galimybės*. Pakartotinai pagrindinių komponentių metodu tiriami rodikliai *n.turtas*, *kompiuteriai*, *m.darbai*, *veršlas*, *švietimas*, *poilsis* ($p = 6$). Suskaičiuojamos komponentių nuosavos reikšmės ir atrenkama informatyviausia, kurią pavadiname *paslaugos* ($p' = 1$). Suskaičiuojami komponentės svoriai pradinuose rodikliuose.



7 pav. Komponentų svoriai pradinuose rodikliuose

8 lentelė. Komponentės paslaugos svoriai pradinuose rodikliuose

	N. turtas	Kompiuteriai	M. darbai	Verklas	Švietimas	Poilsis
Paslaugos	0,96	0,96	0,84	0,99	0,95	0,94



8 pav. Paslaugų rodiklių indeksas

Iš 8 lentelės matyti, kad yra stiprus tiriamų rodiklių ir komponentės ryšys (reikšmės didesnės už 0,8). Patikrinti, ar komponentė pakankamai tiksliai aprašo pradinis kintamuosius, skaičiuojami skirtumai tarp pradinės (9 lentelė) ir atgamtos koreliacinių matricių.

Pradinių rodiklių koreliacijos koeficientai atkurti norimu tikslumu (reikšmės mažesnės už absoliutų dydį 0,2). Taigi komponentė *paslaugos* yra tinkama nusakyti paslaugų įmonių paja-

mas ir bus taikoma toliau tiriant regionų išsivystymo lygį.

Įvertinti ir palyginti apskričių skirtumus skaičiuojamas komponentės *paslaugos* indeksas. Indeksas pasiskirstymas pagal apskritis vaizduojamas 8 paveiksle.

Matyti, kad Vilniaus apskrityje paslaugų sektorius yra kur kas labiau išvystytas negu kitose apskrityse. Paslaugų srities rodikliai yra vieni iš svarbiausių veiksmų, lemiančių regionų ekono-

1 lentelė. Skirtumai tarp pradinių ir atgamtųjų koreliacijos koeficientų

	N. turtas	Kompiuteriai	M. darbai	Verslas	Švietimas	Poilsis
V. turtas	0,07	0,02	-0,11	-0,01	-0,02	0,04
Kompiuteriai		0,07	-0,12	0,00	-0,08	0,09
M. darbai			0,29	0,01	0,12	-0,16
Verslas				0,01	0,00	0,00
Švietimas					0,09	-0,10
Poilsis						0,12

0 lentelė. Ekonominių rodiklių koreliacijos koeficientai

	Užmo-kestis	Pel-nas	Materialinės investicijos	Užsienio investicijos	Der-lius	Gyvuli-ninkystė	Pramo-nė	Sta-tyba	Pre-kyba	Pas-laugos
Užmokestis	1	0,89	0,80	0,75	-0,63	-0,58	0,28	0,91	0,59	0,63
Pelnas		1	0,88	0,87	-0,50	-0,37	0,21	0,88	0,69	0,71
Materialinės investicijos			1	0,68	-0,45	-0,41	0,39	0,82	0,57	0,45
Užsienio investicijos				1	-0,40	0,04	-0,05	0,67	0,80	0,91
Derlius					1	0,48	-0,20	-0,51	-0,08	-0,22
Gyvininkystė						1	-0,33	-0,53	0,18	0,18
Pramonė							1	0,53	0,19	-0,14
Statyba								1	0,62	0,56
Prekyba									1	0,90
Paslaugos										1

inio ir socialinio išsivystymo netolygumus. Komponentę *paslaugos* įtraukiame į ekonominių rodiklių rinkinį.

Apibendrintų ekonominių faktorių sudarymas

Kai jau rasti žemės ūkio efektyvumą ir paslaugų įmonių pajamas nusakantys bendri kintamieji, atitinkamai – *derlius* ir *paslaugos*, galima įvertinti apskričių ekonominę gerovę. Kiti rodikliai įtraukti į ekonominių rodiklių rinkinį: *užmokestis*, *pelnas*, *materialinės investicijos*, *užsienio investicijos*, *gyvulininkystė*, *pramonė*, *statyba*, *prekyba*. Iš viso dešimt požymių ($p = 10$). Nustatyti jų tarpusavio ryšius apskaičiuojami koreliacijos koeficientai.

Rodiklių prasmė ir 10 lentelėje pateikiami koreliacijos koeficientai rodo, kad ekonominius

rodiklius galima grupuoti atsižvelgiant į tarpusavio ryšius. Vieną grupę sudarytų rodikliai *užmokestis*, *pelnas*, *materialinės investicijos*, *užsienio investicijos*, *prekyba* ir *paslaugos*, antrą – *gyvulininkystė*, *derlius*, *pramonė*, *statyba*. Maksimalaus tikėtimumo metodu patikrinsime, ar pakanka dviejų apibendrintų faktorių aprašyti pradinius kintamuosius. Suskaičiuojami apibendrintų faktorių svoriai pradinuose rodikliuose.

Iš 11 lentelės matyti, kad pirmas apibendrintas faktorius nusako apskrities investicinį patraukumą, verslo sėkmę ir finansinius reikalus, jį interpretuojame kaip *finansinius išteklius*. Antaras rodo gamybinį veiklos rūšių pasiekimus, todėl jį interpretuojame kaip *gamybos efektyvumą*. Patikrinti ar gerai apibendrinti faktoriai ap-

11 lentelė. Apibendrintų faktorių svoriai pradinuose rodikliuose

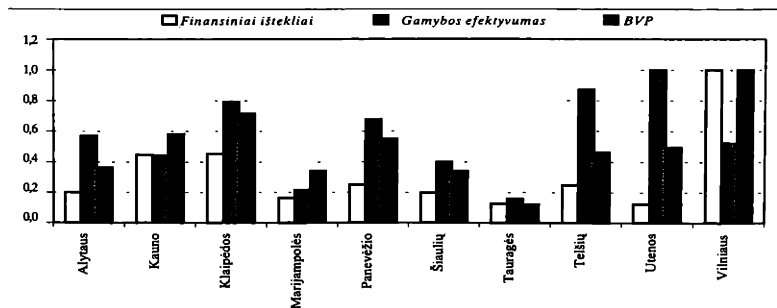
	Finansiniai ištekliai	Gamybos efektyvumas
Užmokestis	0,65	0,71
Pelnas	0,78	0,53
Materialinės investicijos	0,59	0,59
Užsienio investicijos	0,94	0,14
Derlius	-0,26	-0,55
Gyvulininkystė	0,16	-0,90
Pramonė	-0,02	0,47
Statyba	0,61	0,70
Prekyba	0,87	-0,01
Paslaugos	0,93	-0,05

rašo pradinius kintamuosius skaičiuojamos liekanos – skirtumai tarp pradinės (10 lentelė) ir atgamtos koreliacinių matricių.

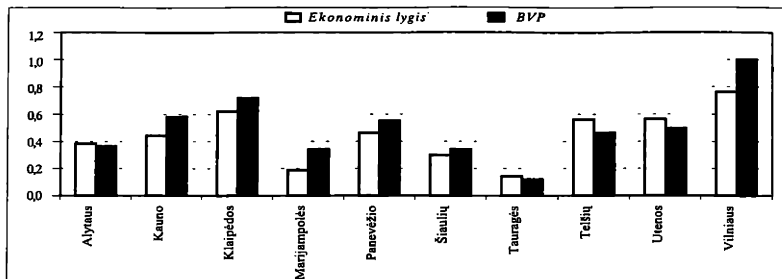
Iš 12 lentelės matyti, kad apibendrinti faktoriai pakankamai tiksliai aprašo pradinius rodiklius, jų pagrindu atkurti pradinių rodiklių tarpusavio ryšiai mažai skiriasi nuo pradinių. Truputėlį prasčiau atkurta priklausomybė tarp rodiklių pramonė ir prekyba (skirtumas didesnis už 0,2). Patikrinome hipotezę, kad liekanos yra nereikšmingos ir artimos 0. Sudaryti apibendrinti faktoriai ($p' = 2$) tinka nusakyti pra-

2 lentelė. Skirtumai tarp pradinių ir atgamtų koreliacijos koeficientų

	Užmokestis	Pelnas	Materialinės investicijos	Užsienio investicijos	Derlius	Gyvulininkystė	Pramonė	Statyba	Prekyba	Paslaugos
Užmokestis	0,10	-0,03	-0,03	0,01	-0,05	-0,03	-0,04	0,00	0,00	0,03
Pelnas		0,12	0,07	0,03	0,01	0,00	-0,02	0,00	-0,02	-0,02
Materialinės investicijos			0,30	0,02	0,05	0,05	0,11	0,02	0,03	-0,09
Užsienio investicijos				0,10	-0,06	0,02	-0,09	-0,03	-0,05	0,00
Derlius					0,63	0,00	0,07	0,05	0,14	0,00
Gyvulininkystė						0,17	0,11	0,01	0,03	-0,02
Pramonė							0,78	0,20	0,21	-0,08
Statyba								0,16	0,06	0,00
Prekyba									0,24	0,04
Paslaugos										0,13



9 pav. Ekonominio lygio faktorių ir BVP indeksai



10 pav. Regionų ekonominio lygio ir BVP indeksų pasiskirstymas

dinius rodiklius ($p = 10$). Patikrinti, ar nauji kintamieji gerai charakterizuoja ekonominį gyvenimą apskrityse, palyginsime faktorius *finansiniai ištekliai*, *gamybės efektyvumas* su BVP vienam gyventojui. Nagrinėjami rodikliai perskačiuojami į indeksus. Taip galime įvertinti, kuriose apskrityse BVP didesnę įtaką turi finansinis faktorius, o kuriose – gamybinis. Gauti rezultatai parodyti 9 paveiksle.

Matyti, kad Vilniaus apskrityje didelę įtaką ekonominiam gyvenimui daro *finansiniai ištekliai*. Tuo tarpu Utenos, Telšių apskrityse svarbesnė gamybės komponentė. Čia palyginti su kitomis apskritimis efektyvios yra pramonės ir statybos veiklos.

LITERATŪRA

1. Lietuvos apskritys. Ekonominė ir socialinė raida. Vilnius: Statistikos departamentas, 1999.
2. Lietuvos apskritys. Ekonominė ir socialinė raida. Vilnius: Statistikos departamentas, 1998.
3. Lietuvos socialinė raida. Vilnius: Statistikos departamentas, 2000.
4. Pranešimas apie žmogaus socialinę raidą Lietuvoje 1999. Vilnius, JTVP, 1999.
5. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. Москва: ЮНИТИ, 1998.

Nustatyti bendrą ekonominį lygį skaičiuojamas *finansinių išteklių* ir *gamybės efektyvumo* indeksų aritmetinis vidurkis. *Regionų ekonominio lygio indeksas* ir *BVP* vaizduojami 10 paveiksle.

Iš 10 paveikslo matyti, kad *ekonominio lygio* ir *BVP* indeksų pasiskirstymas pagal apskritis yra panašūs ir todėl sudarytas rodiklis gana tiksliai nusako ekonominių veiklų pokyčius, įvertina regionų gerovę. Darytina išvada, kad darbe taikyti komponentinės ir faktorinės analizės metodai yra tinkami naudoti palyginti regionų išsivystymą. Sudaryti bendri kintamieji išsaugo pradinį rodiklių teikiamą informaciją, apibūdina realų ekonominį gyvenimą.

6. Лоуди Д., Максвел А. Факторный анализ как статистический метод. Лондон, 1963.

7. Elsing C. Regional Policy and Development in Lithuania. Vilnius, JTVP, 1999.

8. Human Development Report 1999. New York, UNDP, 1999.

9. Koesveld E. Human Development Across Regions in Lithuania. Vilnius, JTVP, 1999.

THE USE OF FACTOR ANALYSIS METHODS IN COMPARISON OF REGIONS

Bronislava Kaminskienė, Algimantas Misiūnas

Summary

The target of this work was to investigate the possibility to compare methods of component and factor analysis by regions. A common variable indicator, formed by applying methods of component and factor analysis, allows evaluating social and economic development level of counties. The value of economic development was compared with Gross domestic product per capita. Obtained similar findings show the fitness of mathematics-statistic models in use of regional comparative analysis.

Įteikta 2001 m. kovo mėn.