
Rangų nustatymo pagal kartoschemas metodikos optimizavimas ir taikymas

Audrius Kryžanauskas

El.paštas: audrius.kryzanauskas@geo.lt

ĮVADAS

Geografiniuose tyrimuose dažnai tenka naudoti kartogramų metodu sudarytas kartoschemas, prie kurių nepateikiami papildomi statistiniai duomenys. Nereitai patiems tenka sudarinėti norimų pavaizduoti reiškinių kartogramas. Tokių informacijos šaltinių vertinimas dažniausiai būna vizualinis, nestandartizuotas, todėl galimi subjektyvūs sprendimai.

Kartais pravartu žinoti reiškinių intensyvumo administraciniame vienetė vieta, lyginant su kitais administraciniais vienetais, t. y. rangą. Tokia informacija yra išsamesnė už patį kartogramos vaizdą, kur administracinio vieneto rangą vertinant vizualiai galima sužinoti tik tam tikrose ribose. Rangus pravartu turėti vykdant teritorinį rajonavimą, klasifikavimą, siekiant išsiaiškinti ryšį tarp tam tikrų reiškinių ir pan.

Pagrindinis tyrimo tikslas buvo sukurti ir optimizuoti rangų nustatymo metodiką nagrinėjant Lietuvos socialinius ekonominius reiškinius vaizduojančias kartogramas. Siekiama visapusiškai išsiaiškinti ir išspręsti problemas, kurios iškyla tiek nustatant teritorinių vienetų rangus, tiek ieškant ryšio tarp jų. Rangų metodika yra matematiškai apibrėžta ir pranašesnė už ekspertinių kartogramų lyginimą bei vertinimą, nes operuojama skaitinėmis reikšmėmis.

Tiriant tam tikrą žemėlapių analizės būdą, svarbu atsižvelgti ir į vaizdavimo būdo tikslumą. Nustatinėjant rangus buvo nagrinėjamos dviejų rūšių matematiškai apibrėžtos legendos skalės. Siekta nustatyti, kokių rūšių skalės yra patogesnės tyrimams atlikti naudojant rangus.

TYRIMO METODIKA

Tyrimas susideda iš dviejų pagrindinių uždavinių blokų: 1 – rangavimo taikymo kartometrijoje optimizavimo, 2 – rangų koreliacijos koeficiento (toliau – RKK) panaudojimo.

Sprendžiant pirmojo bloko uždavinius iškyla trys pagrindinės problemos: rago nustatymas pagal kartogramą, legendos skalės parinkimas, nevienodo tiriamo teritorinio vieneto dydis. Šioje dalyje sukuriamas kartometrinė metodika, pritaikyta tirti Lietuvos kartogramas.

Sprendžiant visas problemas GIS sudaromos įvairiatemės kartogramos, kurioms naudojami 1999 m. statistikos duomenys apie socialinius ir ekonominius reiškinius Lietuvoje (Lietuvos apskritys, 2000). Kartografinis pagrindas – duomenų bazė GDB-200. Ši duomenų bazė buvo kiek pertvarkyta, „Adm“ sluoksnyje generalizavus ribas 1:3 500 000 masteliui ir pašalinus miestų plotų teritorijas. Taigi lieka tik administracinių rajonų plotų teritorijos. Jos yra senojo skirstymo, nes pagal naujojo skirstymo administracinius vienetus statistikos duomenys dar nepateikiami. Kartogramos sudaromos ArcView 3.0a programa.

Pagal pirminius statistikos duomenis nustatomas tikrasis rangas (toliau – TR), kuris vėliau padeda kurti rangų nustatymo metodiką. Ieškoma optimaliausia principinė schema kartometriniam rangui (toliau – KR) nustatyti. KR nustatomas pagal sudarytas kartogramas, po to lyginamas su TR ir aiškinamas, koks yra efektyviausias rangų nustatymas. Būdo patikimumas įvertinamas apskaičiavus vidutinę kvadratinę klaidą (toliau – VKK) tarp KR, nustatyto pagal tam tikrą metodiką, ir TR:

$$VKK = \sqrt{\frac{(TR - KR)^2}{n}};$$

čia n – rajonų skaičius (44).

Galutinai suformulavus rangų nustatymo metodiką, daromos išvados apie legendos skalių patikimumą (lyginamos tik dvi matematiškai apibrėžtos skalės).

Antrajame uždavinių bloke, remiantis kartogramomis, siekiama nustatyti ryšio stiprumą tarp įvairių reiškinių. Ši dalis yra RKK taikymo pavyzdys.

(pvz., nustatomas įvairių socialinių reiškinių ryšys su gimstamumu).

RKK galima taikyti ne tik nustatant vienalaikių reiškinių ryšio stiprumą. Nustačius ryšio stiprumą tarp to paties reiškinio skirtingu metu, galima daryti išvadas apie reiškinio pastovumą. Iš esmės antrojoje tyrimo dalyje, nustačius ryšio stiprumą tarp reiškinių, siekiama išsiaiškinti galimas išvadas.

Reikia pastebėti, kad nors kartograma yra reiškinio vidutinio intensyvumo vaizdavimo būdas kokio nors teritorinio vieneto ribose (Salyščev, 1971), RKK skaičiuoti verta tik tarp tų kartogramų, kur vaizduojamas santykinis reiškinio intensyvumas, nes kitaip ryšio stiprumo skaičiavimas netenka prasmės (pvz., ir taip aišku, kad gimusiųjų skaičius didesnis bus ten, kur daugiau gyventojų).

RANGŲ KORELIACIJOS ESMĖ IR TAIKYMAS GEOGRAFIJOJE

Rangai – tai skaičiai nuo vieno iki n , kuriais pakeičiamos tikrosios reiškinio reikšmės jų didėjimo arba mažėjimo tvarka. Šiame darbe reiškinio reikšmės bus pateikiamos rangais tik jų didėjimo tvarka.

Kaip pagrindinis rangų taikymo pavyzdys čia aprašomas RKK ir nagrinėjamos su juo susijusios problemos. Šis koeficientas skaičiuojamas tarp dviejų reiškinių, kurių rangai administracinio vieneto ribose yra žinomi.

Yra du RKK – Spearmano ir Kendallio. Spearmano RKK yra išvestinis iš Pearsono koreliacijos koeficiento, kuris praktikoje naudojamas dažniausiai. Skaičiuojant koeficientą dvi atitinkamų rangų grupės yra lyginamos tarpusavyje, jų skirtumas keliamas kvadratu ir sumuojami visų matavimų skirtumų kvadratai. Po to rezultatas standartizuojamas taip, kad jo ribos būtų -1 ir $+1$ (Kruopis, 1993). Spearmano RKK apskaičiuojamas pagal formulę:

$$r_s = 1 - 6 \sum_{i=1}^n \frac{(r(X_i) - r(Y_i))^2}{n(n^2 - 1)} = 1 - 6 \sum_{i=1}^n \frac{d_i^2}{n(n^2 - 1)};$$

čia $r(X_i)$ ir $r(Y_i)$ – tiriamųjų reiškinių X ir Y i-ojo bandymo rangas, $d_i = r(X_i) - r(Y_i)$ – i-ojo bandymo rangų skirtumas, n – bandymų (mūsų atveju teritorinių vienetų) skaičius.

Alternatyvus Spearmano koeficientui yra Kendallio RKK, tačiau jis paremtas visai kitokiais skaičiavimo principais. Šis koeficientas (r_k) yra pagrįstas kiekvienos stebėjimų poros lyginimu. Tarkime, kad stebėjimų duomenys pakeisti rangais ir stebėjimo duomenų X_i rangai surašyti didėjančia tvarka. Apskaičiuojame inversijų skaičių s kėlinyje Y_1, \dots, Y_n . Inversija vadiname kiekvieną atvejį, kai prieš mažesnę skaičių parašytas didesnis. Pavyzdžiui, 5, 3, 1, 4, 2 kėlinyje yra 7-ios inversijos. Inversijų skaičius kinta

nuo 0 iki $n(n-1)/2$; šie ribiniai inversijų skaičiai atitinkamai yra 1, 2, ..., n ir $n, n-1, \dots, 2, 1$ kėliniuose.

Kendallio RKK apibrėžiamas lygybe:

$$r_k = 1 - \frac{4s}{n(n-1)};$$

čia s – inversijų skaičius, n – teritorinių vienetų skaičius.

Kendallio ir Spearmano būdu apskaičiuotos reikšmės skiriasi labai nedaug. Dėl paprastesnių skaičiavimų darbe skaičiuojamas tik Spearmano RKK.

Tiek geografijoje, tiek matematikoje RKK nėra plačiai taikomas. Taip yra todėl, kad neverta tikrųjų reiškinio reikšmių keisti jų rangais. Geriau tarp pačių tikrųjų reikšmių skaičiuoti Pearsono koreliacijos koeficientą, kuris patikimiau įvertina ryšį. Labai nedažni atvejai, kada nežinomos tikrosios reiškinio reikšmės, tačiau žinomas jų rangas.

Rangų koreliacijos koeficiento taikymą geografijoje aprašo G. M. Barberis (Barber, 1988), M. K. Bočarovas (Bočarov, 1971) ir kt. Šiuose darbuose pateikiamas tik koeficientų skaičiavimo principas. Jokie tyrimai, naudojant šiuos koeficientus kaip efektyviausius, neatliekami.

Kartografijoje (kartometrijoje) tikslinga taikyti RKK, kai reiškinys vaizduojamas kartogramų metodu. Iš kartogramos praktiškai neįmanoma gauti tikrųjų reiškinio reikšmių, nes jos patenka į tam tikrą intervalą. A. M. Berliant pasišūlė taikyti RKK, kai reikia nustatyti ryšio stiprumą tarp reiškinių, vaizduojamų kartogramomis. Kaip pavyzdį jis ištyrė ryšio stiprumą tarp Nachičevanio srities gyventojų skaičiaus ir lovų skaičiaus ligoninėse tuose pačiuose administraciniuose vienetuose. Koreliacijos koeficientas 0,8 apibūdino gerą ryšį tarp šių reiškinių ir tai reikė, kad gyventojai yra gana tolygiai aprūpinti ligoninėmis (Berliant, 1986). Tačiau šis mokslininkas ištyrė tik 5 teritorinius vienetus, kuriuose reiškiniai buvo vaizduojami 4 intervalų skalėje. Tai reiškia, kad tik du teritoriniai vienetai pateko į vieną intervalą. Taigi nėra rangų problemos. A. M. Berliant pasišūlė tokią išėitį: jeigu du rajonai patenka į vieną intervalą ir skiria kelias vietas (tarkime 3 ir 4), jiems suteikiamas vidutinis tų vietų rangas (šiuo atveju 3,5). Tuo tarpu jeigu 44 Lietuvos rajonus kartografuotume 7 intervalų skale, gali pasitaikyti, jog 6 rajonams reikės suteikti vienodą rangą, o tai jau sukelia didelius netikslumus. Dėl to viena svarbiausių sprendžiamų problemų yra rango nustatymo problema.

RANGŲ NUSTATYMO PROBLEMA

Rangų nustatymas yra svarbiausia tyrimo problema, sprendžiama šiame darbe. RKK skaičiavimo principai, kai žinomi rangai, yra aiškiai suformuluoti, ir

RKK skaičiavimas remiasi paprasčiausiu formulės naudojimu. Tačiau gauti korektiškus reiškinį rangus pagal žemėlapi yra gana sudėtinga, o neišsprendus šios problemos, neįmanomas ir pats RKK skaičiavimas.

Nagrinėjamos kartogramos, kurių legendoje duomenys skaidomi į 7 intervalus. Taip dažniausiai kartografuojami Lietuvos socialiniai ekonominiai reiškiniai. Nustatant administracinio vieneto rangą patogiu, kad legendoje būtų kuo daugiau intervalų. Šiuo atveju, kai turime 7 intervalus, duomenys tolygiausiai gali pasiskirstyti taip: 5 atvejais į tą patį intervalą pateks po 6 rajonus ir 2 atvejais – po 7 rajonus. Rangų nustatymo problema iškyla siekiant paskirstyti juos tarp į vieną intervalą patenkančių rajonų.

Yra du gana paprasti rangų nustatymo būdai. Pirmasis jų pasiūlytas A. M. Berlianto (Berliant, 1986), kai į tą patį intervalą patenkančioms teritoriniams vienetams suteikiamas vidutinis jų rangas. Pavyzdžiui: tarkime, į aukščiausią – pirmąjį – intervalą patenka 4 rajonai, o į antrąjį – 3. Skaičiuojama tokia tvarka: pirmojo intervalo rajonų rangas = $(1+2+3+4)/4 = 2,5$, antrojo intervalo rajonų rangas = $(5+6+7)/3 = 6$. Skirstant į rangus šiuo būdu gaunamos didelės vidutinės kvadratinės klaidos, lyginant su tikruoju rangų, tačiau lyginti du taip sutvarkytus žemėlapius yra gana korektiška. Šio metodo patikimumas vertinamas jau ne pagal vidutinę kvadratinę klaidą, o lyginant koreliacijos koeficientus, apskaičiuotus naudojant tikrąjį ir kartometrinių rangų.

Antrasis būdas, siūlomas darbo autoriaus, yra „geografiškesnis“ už pirmąjį, bet reikalauja daugiau skaičiavimų. Šiuo būdu nustatant rangus į reiškinį žvelgiama kaip į vientisą teritorinę sistemą. Atkreipiamas dėmesys į aplinkinių rajonų poveikį, kuris priklauso nuo kaimyninių rajonų „svorio“ vidurkio (toliau – KSV) ir papildomų pataisymų (jie bus aptarti vėliau).

„Svoris“ – tai intervalo, į kurį patenka rajonas, numeris pradedant nuo žemiausio intervalo. KSV apskaičiuojamas kaimyninių rajonų „svorių“ sumą padalijus iš kaimyninių rajonų skaičiaus. Taigi, nustatant rajono rangą, pirmiausia atsižvelgiama į paties rajono reiškinio intervalą – kuo jis aukštesnis, tuo rajono rangas aukštesnis. Vėliau patekę į tą patį intervalą rajonai tarpusavyje suskirstomi į rangus pagal kaimyninių rajonų „svorių“ vidurkį. Kaip matome, pagal šią metodiką kiekvienas rajonas gauna individualų rangą. Svarbiausias veiksnys metodikoje yra KSV, todėl, siekiant optimizuoti rango nustatymą, pirmiausia reikia teisingai įvertinti teritorinę kaimyninių rajonų įtaką.

Įvertinant kiekvieno kaimyninio rajono „svorį“, būtų pravartu įvertinti ir to svorio reikšmę rajonui, kaip svarbų rodiklį pasitelkus abiejų rajonų ribą. To-

kiu atveju, rajonui galutinai skaičiuojant KSV, reikėtų kaip koeficientą prie kiekvieno „svorio“ prirašyti rajonų ribos ir tiriamojo rajono perimetro santykį. Įvedus šią sąlygą skaičiavimai tampa labai sudėtingi, todėl darbe šios sąlygos atsisakyta.

Galimi įvairūs kaimyninių rajonų įtakos (KI) pataisymai atsižvelgiant į papildomas sąlygas. Jeigu tam tikra dalis kaimyninių rajonų „svorių“ skiriasi per tam tikrą vienetą nuo paties rajono „svorio“, galima daryti išvadą, kad bendri tiriamojo reiškinio ypatumai yra labiau susitelkę pačiame rajone, ir kaimyniniai rajonai tokiam rajonui turi mažesnę įtaką nei kitiems, kur šios sąlygos netenkinamos. Nustačius įtakos pataisą, ją galima pridėti arba atimti iš KSV taikant tam tikrą koeficientą.

Taigi kaimyninių rajonų įtaka KI apskaičiuojama pagal formulę:

$$KI = KSV \pm P;$$

čia KSV – kaimyninių rajonų „svorio“ vidurkis, P – įtakos pataisa.

$$KSV = \frac{\sum_{i=1}^n KS_i}{n};$$

čia KS_i – i-ojo kaimyninio rajono „svoris“, n – kaimyninių rajonų skaičius.

P skaičiuojamas tuo atveju, kai kaimyninių rajonų svoris b(%) yra mažesnis arba didesnis už tiriamojo rajono svorį d vienetų. Pavyzdžiui, Vilniaus rajono „svoris“ yra 6, nes jis pagal intensyvumą patenka į šeštąjį intervalą (1 pav., b). Kaimyninio Trakų rajono „svoris“ yra 1. Koeficientas d Vilniaus ir Trakų rajonų atveju yra $d = 6 - 1 = 5$.

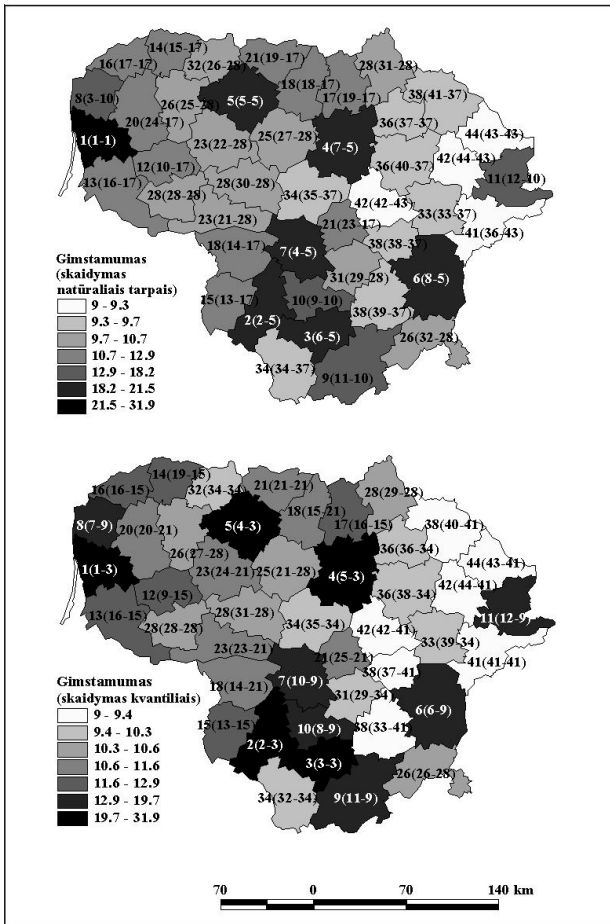
$$b = \frac{n_d}{n} \cdot 100\%;$$

čia n_d – kaimyninių rajonų skaičius, kurių svoris už tiriamojo rajono svorį mažesnis d vienetų. Tokiu atveju P apskaičiuojamas pagal formulę:

$$P = k \frac{\sum_{a=1}^b |RS - KS_a|}{n_d};$$

čia RS – tiriamojo rajono „svoris“, k – pataisos koeficiento įtaka, KS_a – a-ojo rajono, tenkinančio minėtą sąlygą, „svoris“.

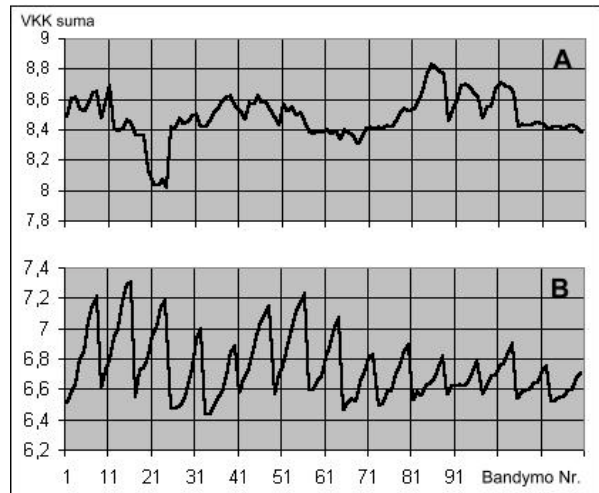
Nustačius, kad 50% kaimyninių rajonų „svoriai“ yra mažesni už paties rajono svorį 3 vienetais (t. y. 3 intervalais), galima prie KSV papildomai pridėti tų rajonų ir pagrindinių rajono svorių skirtumų vidurkį taikant koeficientą 0,5. Visi trys paminėti b, d, k rodikliai nėra galutiniai ir darbe nustatomi bandymais. Bandymai vykdomi tam tikra tvarka kompiuteriu generuojant b, d ir k rodiklius. Rodiklis b kei-



1 pav. Gimstamumas, kartografuotas dviem skirtingais būdais: *A* – skaidant natūraliais tarpais, *B* – skaidant kvantiliais. Skaičius prieš skliaustelius yra tikrasis reiškinio rangas, pirmas skaičius skliausteliuose – rangas, nustatytas pagal autoriaus metodiką, antrasis – rangas, nustatytas pagal A. M. Berlianto metodiką (Berliant, 1986)
 Fig. 1. The birthrate mapped by two different methods: *A* – by dividing data by natural breaks, *B* – by dividing data by quantiles. The number before bracket is true subject rank. The first number in brackets is the rank calculated by author's method, the second is the rank calculated by A. M. Berliant's method

čiamas nuo 0,4 iki 0,8 kas 0,1, rodiklis *d* – 2–4 intervale kas 1, rodiklis *k* – 0–0,8 intervale kas 0,1. Su kiekvienu rodiklių rinkiniu apskaičiuojami žemėlapijo kartometrinių rangai. Šie rangai lyginami su tikraisiais ir nustatoma vidutinė kvadratinė klaida. Buvo atlikti bandymai su 7-iomis kvantilių ir 7-iomis natūralių tarpų būdu braižytomis kartogramomis. Visais atvejais sumuojamos su tuo pačiu rodiklių rinkiniu gautos VKK. Jeigu šios klaidos būtų išsidėsčiusios atsitiktinai, jų grafikas bandymų atžvilgiu būtų artimas tiesei. Gauti grafikai pavaizduoti 2 paveikslėlyje.

Iš A grafiko (2 pav.) matyti, kad VKK suma kinta periodiškai. Nustatyta, kad kvantilių atveju tai lemia optimaliausia koeficiento *k* reikšmė 0,1. Skaidant na-



2 pav. Vidutinių kvadratinių klaidų sumos kaita pagal atskirus žemėlapius nustatant *d*, *b*, *k* koeficientus pagal kartogramas, kartografuotas: *A* – kvantiliais, *B* – natūralių tarpų metodu

Fig. 2. Variation of sum of mean square errors in different maps (*A* – quantile mapping, *B* – natural break mapping) by determining *d*, *b*, *k* coefficients

tūraliais tarpais (2 pav., *b*) VKK suma, esant toms pačioms *d*, *b*, *k* reikšmėms, kinta visai kitaip nei kvantilių atveju. Čia nėra ryškaus periodiškumo, tačiau akiškai išsiskiria mažiausia VKK suma. Optimalios *b*, *d*, *k* reikšmės abiem atvejais yra tos, kai VKK suma yra mažiausia. Kartografuojant kvantiliais optimalios reikšmės yra *b* = 80%, *d* = 2, *k* = 0,1. Kartografuojant natūralių tarpų metodu optimalios rodiklių reikšmės yra *b* = 60%, *d* = 2, *k* = 0,7. Kaip matome, stipriai skiriasi rodiklio *k* optimalios reikšmės abiem atvejais. Taikant RKK kartogramų nagrinėjimui, pirmiausia būtina įsitikinti, ar į kiekvieną intervalą patenka daug maž tiek pat administracinių vienetų. Jeigu taip, tada reikia taikyti pirmuosius išvardytus rodiklius, kitu atveju – antruosius *b*, *d*, *k*.

LEGENDOS SKALĖS PARINKIMAS

Vaizduojant reiškinį kartogramomis jo bendram supratimui (t. y. vizualiam vertinimui) ir matavimo efektyvumui svarbu, kaip duomenys skaidomi intervalais.

Apžvelgsime keletą matematiškai apibrėžtų legendos skalės parinkimo būdų. Vienas jų – kai skalė skaidoma vienodais intervalais: imama minimali ir maksimali duomenų eilutės reikšmė ir tarpas tarp jų padalijamas į pageidaujama skaičių lygių intervalų. Toks būdas pasiteisintų tik tuo atveju, kai duomenys pasiskirstę tolygiai. Tačiau jeigu dalis duomenų yra sukonzentruoti viename taške, o dalis išsibarstę, nukentčia ir žemėlapijo vaizdas, ir matavimų pagal jį efektyvumas. Šiuo būdu kartografuotos kartogramos darbe nenagrinėjamos.

Antras duomenų skaidymo būdas – kvantiliais: tai toks būdas, kai į kiekvieną intervalą patenka vienodas skaičius teritorinių vienetų (duomenų). Šis būdas efektyvus, kai norima nustatyti reiškinio rajone rangą, kadangi nėra viename intervale sukoncentruotų reikšmių. Intervaluose yra vienodas skaičius reikšmių, todėl klaidos nustatant rangus bus pasiskirsčiusios tolygiai (t. y. kuo mažiau viename intervale reikšmių, tuo mažesnis nukrypimas maksimalaus apsirikimo atveju: jeigu intervale yra 3 reikšmės, tai maksimalus tikrojo ir kartometrino rango nesutapimas gali būti 3, o jeigu yra 10 reikšmių, nesutapimas gali būti 10). Vadinasi, yra mažesnė didelių vidutinių kvadratinių klaidų tikimybė. Skirstant duomenis kvantiliais kiek nukenčia kartogramos vaizdumas. Neiškiriama keletas būdingų reikšmių viršutiniame ir apatiniame kvantilyje, todėl pagal tokią kartogramą sunku įvertinti ekstremumus.

Geografinėse informacinėse sistemose naudojamas duomenų skaidymo natūraliais intervalais būdas. Jis statistiškai paremtas Jenkso optimizacija (Using ArcView, 1996), kuri šiame darbe nebus plačiau aptarta. Pagrindinis šio metodo principas – nuokrypių nuo vidurkio sumažinimas kiekviename intervale. Taip išskiriamos būdingiausios duomenų grupės. Vaizdumo prasme šis metodas yra efektyvesnis už kvantilių metodą, nes pagrįstas logiškesniais grupavimais ir padeda išskirti būdingiausias duomenų grupes. Jeigu naudojant kvantilių metodą viršutiniame intervale neiškiriamos kelios nuo duomenų nutolusios ekstremalios reikšmės, tai natūralių tarpų atveju jos bus pavaizduotos atskirame intervale. Tačiau skirstant

Jenkso metodu į kai kuriuos intervalus pateks daugiau reikšmių nei skirstant kvantiliais, todėl galimos didesnės vidutinės kvadratinės klaidos.

Skirstant intervalais, be abejo, įmanomi dar geresni sprendimai, kurie išryškintų kai kurias reiškiniams būdingas ypatybes, tačiau šie metodai nėra matematiškai apibrėžiami ir įvardijami. Darbe lyginami tik kvantilių ir natūralių tarpų metodai, išryškinamos teigiamos ir neigiamos šių metodų pusės.

Galutinai suformulavus optimaliausią rangų nustatymo metodiką, buvo apskaičiuotos vidutinės kvadratinės klaidos tarp tikrojo ir kartometrino rango tiek skirstant natūraliais tarpais, tiek kvantiliais. Gauti rezultatai pateikiami 1 lentelėje.

Kaip matome, pagal autoriaus metodiką apskaičiuotiems rangams būdingos mažesnės vidutinės kvadratinės klaidos, lyginant su tikroju rangų. Galima daryti išvadą, kad šie rangai labiau atspindi tikrąsias reiškinio reikšmes, negu gaunami Berlianto pasiūlytu metodu. Be to, jie tinkamesni administracinių vienetų tyrimams, nes kiekvienam vienetui suteikiamas unikalus rangas.

Labai aiškiai skiriasi pagal natūraliais tarpais ir kvantiliais sudarytas kartogramas atliktų matavimų tikslumas. Daug tiksliau rangai apskaičiuojami, kai duomenys suskirstyti kvantiliais.

RYŠIŲ STIPRUMO TYRIMAI

Naudojant RKK galima nustatyti ryšio stiprumą tarp skirtingų to paties laiko reiškinų ir tarp to paties reiškinio skirtingu metu. Nustatyti ryšiai tarp tirtų reiškinų pateikiami 2 lentelėje.

1 lentelė. Vidutinės kvadratinės klaidos (VKK) tarp rangų, gautos pagal skirtingus žemėlapius ir naudojant skirtingas rangų nustatymo metodikas

Table 1. Mean square errors (MSE) between ranks calculated from different maps by different ranking methods

Kartografuojamas dalykas Mapped subject	VKK pagal Berlianto rangus MSE by Berliant's ranks		VKK pagal autoriaus rangus MSE by author's ranks	
	Natūralūs tarpai Natural breaks	Kvantiliai Quantiles	Natūralūs tarpai Natural breaks	Kvantiliai Quantiles
Gimstamumas Birthrate	4,30	3,16	3,06	2,26
Pramonės darbuotojai 1999 m. Employers of industry '99	4,37	3,16	3,23	2,72
Ūkių pasiskirstymas Distribution of farms	4,13	4,12	3,39	2,90
Nedarbas 1999 m. Unemployment '99	3,87	3,16	2,47	2,37
Vidut. atlyginimas The month mean pay	4,89	5,0	3,48	1,97
Pramonės darbuotojai 1995 m. Employers of industry '99	3,69	3,16	2,98	2,50
Nedarbas 1993 m. Unemployment '93	3,66	3,16	2,60	2,20

2 lentelė. Ryšių stiprumas tarp reiškinių

Table 2. Correlations between subjects (rank correlation coefficient, RCC)

Tiriamieji dalykai Research subjects	RKK tarp tikrųjų rangų	RKK pagal autoriaus rangus RCC by Berliant's ranks		RKK pagal Berlianto rangus RCC by author's ranks	
	RCC between true ranks	Natūralūs tarpai Natural breaks	Kvantiliai Quantiles	Natūralūs tarpai Natural breaks	Kvantiliai Quantiles
Pramonės darb. 1999 m. gimstamumas Employers of industry '99 – Birthrate	0,141	0,137	0,107	0,137	0,090
Ūkių pasiskirstymas – gimstamumas Distribution of farms – Birthrate	-0,423	-0,360	-0,443	-0,020	-0,407
Nedarbas 1999 m. – gimstamumas Unemployment '99 – Birthrate	-0,044	-0,058	-0,038	-0,021	-0,027
Vidut. atlyginimas – gimstamumas Mean month pay – Birthrate	0,507	0,477	0,440	0,492	0,459
Pramonės darb. 1995 m. – pramonės darb. 1999 m. Employers of industry '95 – Employers of industry '99	0,935	0,887	0,894	0,899	0,918
Nedarbas 1993 m. – nedarbas 1999 m. Unemployment '93 – Unemployment '99	0,202	-0,041	0,162	0,261	0,192

Vertinant atskirų reiškinių tarpusavio stiprumą, su gimstamumu buvo lyginama: pramonės darbuotojų skaičius iš darbingų gyventojų skaičiaus; privačios žemės naudotojų skaičius iš darbingo amžiaus gyventojų skaičiaus; nedarbo lygis; vidutinio mėnesinio darbo užmokesčio dydis.

Buvo nustatytas RKK tarp to paties reiškinio skirtingu laiku. Jeigu koreliacijos koeficientas didelis, galime daryti išvadą, kad per tiriamąjį laikotarpį reiškinys buvo teritoriškai pastovus (tie patys rajonai turėjo daugumą tą patį rangą).

Antroje lentelėje matome, kad visi tirti reiškiniai turi mažai įtakos gimstamumui. Gimstamumas nepriklauso nuo gyventojų profesijos. Gimstamumo ir vidutinio darbo užmokesčio koreliacija jau yra pakankama, todėl galima teigti, kad didesni atlyginimą gaunantys gyventojai susilaukia daugiau vaikų.

Ištirtas dviejų laikotarpių šalies nedarbo lygis ir pramonės darbuotojų dalis iš darbingo amžiaus gyventojų. Mažas koreliacijos koeficientas tarp nedarbo lygio skirtingu metu rodo, kad šis reiškinys per tirtą laiką teritoriškai pasikeitė – pakito rajonai, kuriuose nedarbo lygis buvo didžiausias ir mažiausias. Iš to galima spręsti apie esminius pokyčius šalies ekonomikoje. Ryšys tarp pramonės darbininkų dalies iš darbingo amžiaus gyventojų skirtingu laiku yra geras (RKK = 0,9). Tai rodo, kad per tirtą laiką pramonės centrai liko tie patys.

IŠVADOS

1. Rango nustatymas pagal kartoschemas padidina vaizduojamų reiškinių tyrimo galimybes, nes galima

įvertinti reiškinių tarpusavio padėtį, nors keli administraciniai vienetai ir patenka į tą patį intervalą.

2. Autoriaus sukurta rangų nustatymo metodika, kai reikalingas pats rajono rangas, yra pranašesnė už A. M. Berlianto pateiktą administracinių vienetų rangų metodiką (Berliant 1986), nes šiuo būdu mažiau nukrypstama nuo tikrųjų duomenų, be to, tokias kartoschemas patogiau tyrinėti, nes kiekvienas administracinis vienetas įgyja individualų rangą.

3. Tiriant kartogramų būdu sudarytas kartoschemas, reikia atkreipti dėmesį į teritorijos visumą: administraciniai vienetai turi būti tiriami atsižvelgiant į kaimyninių teritorijų įtaką.

4. Kartometrinių tyrimų tikslumui yra svarbus vaizdavimo būdas. Nustatyta, kad kartometriškai patikimiausia informacija gaunama, kai duomenys legendos skalėje suskirstomi į vienodus pagal duomenų kiekį grupes (kvantilius). Vizualiniam kartoschemas vertinimui šis būdas ne visuomet yra optimaliausias.

5. Koreliaciniams ryšiams pagal rangus nustatyti abi metodikos tikslumo prasme pasirodė lygiavertės. Skaičiuojant rangų koreliacijos koeficientą patogiau kartoschemą suskirstyti rangais pagal A. M. Berlianto aprašytą metodiką, nes čia skaičiavimai yra daug paprastesni.

6. Nustačius koreliacinius ryšius tarp skirtingų to paties laikotarpio reiškinių, galima daryti išvadą apie reiškinių tarpusavio priklausomybę. Nustačius ryšius tarp to paties reiškinio skirtingu metu, galima spręsti, ar stipriai teritorijos atžvilgiu pakito reiškinys per tą laiką.

Literatūra

- Lietuvos apskritys 2000.* (2000). Statistikos departamentas prie Lietuvos Respublikos Vyriausybės.
- Lietuvos gyventojai pagal amžių 2000 metų pradžioje.* (2000). Statistikos departamentas prie Lietuvos Respublikos Vyriausybės.
- Kruopis J. (1993). *Matematinė statistika*. Vilnius: Mintis.
- Berliant A. M. (1986). *Obraz prostranstva: karta i informacija*. Moskva: Myslj.
- Bočarov M. K. (1971). *Metody matematičeskoj statistiki v geografii*. Moskva: Myslj.
- Salyščev K. A. (1971). *Kartografija*. Moskva: Vysšaja škola.
- Barber G. M. (1988). *Elementary statistics for geographers*. New York.
- Kendall M. G. Stuart A. (1973). *The advanced theory of statistics*. London.
- Using ArcView.* (1996). Redlands: Environmental Systems Research Institute.
- Geoduomenų bazė GDB200.* (1993–2000). Vilnius: Gis-Centras.

Audrius Kryžanauskas

OPTIMIZATION AND APPLICATION OF RANK DEFINING METHODS BY CARTOSCHEMES

S u m m a r y

In cartographical (geographical) research, maps mapped by the method of cartoshemes are used. The researcher often has to make a cartoscheme by himself. Exploring

of such cartographical sources is visual, non-standardized and we can obtain inaccurate results. In many cases exploring can be only visual and a mathematical comparison of two maps is impossible. The ranks can be used by classification of territories, planning, computing spatial correlations, etc. The general problem in ranking is that data are divided into intervals. It is difficult to rank data inside an interval.

In this article author offers the ranking method, where all mapped regions have an individual rank. The basis of this methodic is that regions inside one interval (painted with the same color on the map) are ranked depending on the interval value of neighboring regions. The influence of neighbors on the region's rank is described by three constants. These constants were calculated and optimized by comparing true and cartometrical ranks. True ranks were calculated from original statistical data used by mapping. Cartometrical ranks were defined from the map, using different combinations of the above-mentioned three constants. We took the combination of constants where the mean square error between the true and the cartometrical ranks was the least.

By comparing two mathematically described legend types – data division by natural breaks and by quantiles the conclusion was made that maps compiled by the quantile method are better suitable for rank analysis.

Calculating the rank correlation coefficient, mistakes using the author's and Berliant's ranks are similar. The author's method allows to obtain the individual rank for each region, but Berliant's method estimates just the mean rank in an interval.