

ŠIAULIŲ UNIVERSITETAS
REGIONŲ PLĖTROS INSTITUTAS

DEIVIDAS LESUTIS

(Matematikos studijų programa, valstybinis kodas: 6211AX010)

NEIGIAMŲ FAKTORIŲ ĮTAKA
VIDINEI LIETUVIŲ MIGRACIJAI XXI AMŽIUJE

Magistro darbas

Vadovas
doc. dr. V. Kanišauskas

ŠIAULIAI, 2020

TURINYS

TURINYS.....	2
ĮVADAS.....	4
1 TEORINĖ DALIS	7
1.1 PAGRINDINĖS MATEMATINĖS STATISTIKOS SĄVOKOS	7
1.2 POLINOMINIS SKIRSTINYS	7
1.3 POLINOMINIO SKIRSTINIO TAIKYMAS	8
1.4 KULBAKO-LEIBLERO INFORMACIJA	9
1.5 STATISTINĖS HIPOTEZĖS IR JŲ KLAIDOS TIKIMYBĖS	9
1.6 SERIJŲ KRITERIJUS	10
1.7 LAIKO EILUTĖS.....	11
1.7.1 AUTOREGRESINIAI PROCESAI.....	11
1.7.2 AUTOREGRESINIŲ PROCESŲ PARAMETRŲ VERTINIMAS.....	12
1.7.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS.....	12
1.8 KORELIACINĖ ANALIZĖ	14
1.8.1 EMPIRINIS KORELIACIJOS KOEFICIENTAS	14
1.8.2 HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTĄ.....	15
1.8.3 TIESINĖ REGRESIJA	16
1.9 TIESINĖ REGRESIJA SU PROGRAMA „SPSS“	18
1.10 DUOMENŲ NORMALUMAS SU PROGRAMA „SPSS“	21
2. TYRIMAS. PRAKTINĖ DALIS.....	22
2.1 DUOMENYS. PRADINĖ STATISTINĖ ANALIZĖ LIETUVOS MASTU	22
2.1.1 DUOMENŲ NORMALUMAS.....	29
2.1.2 DUOMENŲ NEPRIKLAUSOMUMAS IR ATSITIKTINUMAS	30
2.1.3 PRIKLAUSOMYBĖ TARP NEIGIAMŲ FAKTORIŲ IR VIDINĖS MIGRACIJOS LIETUVOS MASTU, PAGAL 2002-2018 M. DUOMENIS	31
2.2 DUOMENŲ APSKRITYSE GRUPAVIMAS	40
2.3 FAKTORIŲ PASISKIRSTYMAS GRUPĖSE	47
2.4 FAKTORIŲ POKYČIŲ GRUPĖSE ATVAIZDAVIMAS LIETUVOS ŽEMĖLAPYJE.....	52
2.4.1 ATVYKĖ.....	52
2.4.2 IŠVYKĖ	53
2.4.3 BENDRAS JUDĖJIMAS	54
2.4.4 NUSIKALTIMAI	55
2.4.5 SAVIŽUDYBĖS	56
2.4.6 BEDARBIAI	57
2.5 FAKTORIŲ ĮTAKOS GYVENTOJŲ JUDĖJIMUI GRUPĖSE ATVAIZDAVIMAS LIETUVOS ŽEMĖLAPYJE	58
2.6 KOMPLEKSIŠKAS LAIKO EILUČIŲ TAIKYMAS.....	63

2.6.1	I GRUPĖS NUSIKALTIMŲ DUOMENYS	63
2.6.2	KITOS GRUPIŲ CHARAKTERISTIKOS KURIOMS TAIKYTINA LAIKO EILUČIŲ TEORIJA	66
3.	DARBO REZULTATAI	69
	LITERATŪRA	70
	SANTRAUKA	71
	SUMMARY	72
	PRIEDAI	74

IVADAS

Temos aktualumas. Vidinė migracija – tai gyvenamosios vietos keitimas šalies viduje. Tokia migracija stipriai pasireiškia tarp miesto ir kaimo, didelių ir mažų miestelių. Pati migracija vyksta jau nuo seniausių laikų, žmonės keliavo ieškodami geresnių gyvenimo sąlygų. Migracija gali būti savanoriška (kai ieškoma geresnių darbo sąlygų, bandant įgyti išsilavinimą, dėl geresnio klimato ar aplinkos) ir priverstinė (karai, badas, gamtos stichijos, diskriminacija, religinės ar politinės priežastys). O kokios priežastys verčia žmones migruoti į kitas vietas Lietuvoje? Išskyrėme kelis neigiamus faktorius:

1. Savižudybė – tai savo paties gyvybės atėmimas. Priežasčių dėl kurių žmogus nusprendžia atimti sau gyvybę gali būtų daug: stresas, depresija, psichikos ligos, alkoholizmas, finansinės bėdos, bedarbystė, žalingi įpročiai, emocinės problemos. Asmens savižudybė paliečia šeimą, draugus ir visus žmones, kurie jį pažinojo. Toks įvykis gali giliai sukrėsti artimuosius ir pakeisti jų gyvenimą ilgam laikui – sukelti neviltį, depresiją, norą gyventi kitaip, pakeisti savo aplinką, gyvenamą vietą ar darbovietę. Tai gali iš esmės pakeisti draugų bei artimųjų gyvenimus.

2. Nusikalstamumas – tai procesas, kurį sudaro visuma baudžiamų veiklų, padarytų tam tikroje teritorijoje per tam tikrą laikotarpį. Jeigu nusikaltimų daroma daug, galima galvoti, kad šalyje kažkas negerai – prasti valdžios sprendimai, didelis nedarbas, mažas uždarbis, mažai galimybių tobulėti. Savaimė suprantama, kad esant dideliame nusikaltimų skaičiui vietovėje, darosi nesaugu gyventi, todėl gali tekti keisti gyvenamą vietą į saugesnę.

3. Bedarbystė. Bedarbiu yra laikomas nedirbantis darbingo amžiaus asmuo, nesimokantis dieninėje mokymosi įstaigoje, užsiregistravęs gyvenamosios vietos valstybinėje darbo biržoje kaip ieškantis darbo. Didelis bedarbių skaičius yra labai svarbus socialiniu atžvilgiu. Kuo šalyje daugiau bedarbių, tuo žmonėms sunkiau gyventi, rasti darbą. Esant ilgam nedarbo laikui, prasideda finansinės problemos. Pasiekus tokią būseną žmogui gali atsirasti noras migruoti ar net emigruoti svetur, kad pagerintų savo finansinę būklę. Nuo 2007 m. prasidėjusi pasaulinė ekonominė krizė smarkiai išaugino bedarbių skaičių Lietuvoje. Pasiekus piką 2010 metais, situacija ėmė gerėti.

Šiame magistro darbe patikrinsime, kaip savižudybės, nusikalstamumas ir bedarbystė veikia vidinę migraciją Lietuvoje.

Tyrimo objektas. Statistikos departamento statistiniai ir demografiniai duomenys apie Lietuvos gyventojų migraciją, nedarbo lygį ir nusikalstamumą bei duomenys iš Higienos instituto Sveikatos informacijos centro apie Lietuvos gyventojų savižudybes.

Chronologinės ribos. 2002–2018 metai.

Tyrimo tikslas: Ištirti neigiamų faktorių įtaką Lietuvos gyventojų migracijai šalies viduje.

Tyrimo uždaviniai:

1. Nustatyti statistinę priklausomybę tarp neigiamų faktorių ir vidinių migrantų skaičių Lietuvos mastu.
2. Sugrupuoti Lietuvos apskričių nagrinėjamus duomenis į tris grupes pagal dydį. Nustatyti statistinę priklausomybę tarp neigiamų faktorių ir vidinės migracijos skaičių nustatytų grupių viduje ir tarp grupių.
3. Taikant vidinės migracijos matematinį modelį nustatyti, ar vidinių migrantų ir neigiamų faktorių duomenų grupėse pokyčiai yra reikšmingi.
4. Lietuvos žemėlapyje pavaizduoti vidinės migracijos ir neigiamų faktorių pokyčius pagal grupes, rasti ryšius tarp jų.
5. Lietuvos žemėlapyje pavaizduoti, kaip neigiami faktoriai veikia žmonių vidinę migraciją pagal grupes.
6. Nustatyti, ar prielaidos taikyti laiko eilučių modelius statistiniams duomenims apie vidinius migrantus ir neigiamus faktorius grupėse yra tenkinamos. Jei prielaidos tenkinamos, pritaikyti tuos modelius.

Tyrimo metodai. Taikomi tikimybiniai, matematinės statistikos, lyginamosios analizės, vizualinis–dedukcinis ir laiko eilučių metodai.

Darbo struktūra. Darbas yra sudarytas iš įvado, dviejų skyrių, santraukos, literatūros sąrašo ir priedų.

Pirmajame skyriuje pateikiama pagrindinė teorija ir naudojamos sąvokos magistro darbe.

Antrajame skyriuje atliekamas tyrimas, kurio pradžioje pateikiami vidinių migrantų, savivadybių, bedarbių ir nusikaltimų skaičiai Lietuvoje. Su tais duomenimis 2.1 skyrelyje atliekamas pradinis tyrimas – duomenų apdorojimas, randamos pagrindinės statistinės charakteristikos, priklausomybės tarp duomenų ir hipotezių tikrinimas. Toliau 2.2 skyrelyje dešimties Lietuvos apskričių duomenys yra grupuojami ir tyrimas tęsiamas su grupių duomenimis. Tiems duomenims tikrinamos hipotezės, ieškomos priklausomybės bei tiesinės regresijos lygtys. Taip pat nustatoma kaip vidiniai migrantai ir neigiami faktoriai kas 4 metus yra pasiskirstę grupėse, tikrinama hipotezė apie polinominio skirstinio parametrus, taikoma Kulbako-Leiblero informacija ir jų ryšys su klaidų tikimybėmis. 2.3 skyrelyje nagrinėjamas tiriamų faktorių pasiskirstymas grupėse kas 4 metus, kuris 2.4 skyrelyje yra atvaizduojamas Lietuvos žemėlapyje. Toliau 2.5 skyrelyje, pasitelkus Lietuvos žemėlapi, randami ryšiai tarp

neigiamų faktorių ir vidinių migrantų duomenų. Galiausiai, 2.6 skyrelyje, tikrinamos prielaidos laiko eilučių modeliams taikyti grupių duomenims. Duomenims, kuriems taikytina laiko eilučių teorija, randamos AR(4) proceso lygtys su trendais.

Magistrinio darbo pabaigoje pateikiama santrauka, literatūros sąrašas ir priedai.

1 TEORINĖ DALIS

1.1 PAGRINDINĖS MATEMATINĖS STATISTIKOS SĄVOKOS

Pagrindinės matematinės sąvokos, formulės ir apibrėžimai parašyti remiantis [1,6,8] knygomis.

Apibrėžimas. Imtis – stebėtų rezultatų sąrašas.

Apibrėžimas. Variacinė eilutė - imties reikšmių išdėstymas didėjančia tvarka. T.y. $x_1^* \leq x_2^* \leq \dots \leq x_n^*$, čia $x_{min} = x_1^*$ - mažiausias elementas, o $x_{max} = x_n^*$ - didžiausias imties elementas.

Apibrėžimas. Vidurkis – tai vidutinė statistinės eilutės reikšmė apskaičiuojama formule:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

Čia n - imties dydis, $x_1, x_2, \dots, x_n$ – statistinės eilutės elementai.

Apibrėžimas. Mediana – vidurinė variacinės eilutės reikšmė, apskaičiuojama formule:

$$M_e = \begin{cases} X_{(\frac{n+1}{2})}, & \text{jei } n = 2l + 1, \\ \frac{1}{2} \left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right), & \text{jei } n = 2l. \end{cases}$$

Apibrėžimas. Imties dispersija parodo sklaidą apie vidurkį ir randama pagal formulę:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Apibrėžimas. Imties standartinis nuokrypis yra dažniausiai taikomas sklaidos matas.

$$S = \sqrt{S^2}.$$

Apibrėžimas. Hipoteze vadiname bet kokią teiginį apie tikrąjį imties skirstinį. Kai kalbame apie hipotezes visada turime ne tik patį teiginį – nulinę hipotezę H_0 , bet ir alternatyvų teiginį – alternatyvią hipotezę (alternatyvą) H_1 .

1.2 POLINOMINIS SKIRSTINYS

Pagal J. Kruopį [1] sakykime, kad turime nepriklausomų eksperimentų seką tokią, kad kiekviename bandyme gali įvykti nesutaikomi įvykiai $A_1, \dots, A_k$, kur $k > 2$. Įvykio A_i tikimybė kiekviename bandyme yra lygi $p_i > 0$, $p_1 + p_2 + \dots + p_k = 1$, ir nepriklauso nuo kitų bandymų rezultatų. Atlikus m bandymų, įvykusių įvykių A_i skaičių žymime X_i , $i = 1, 2, \dots, k$. Tada atsitiktinio vektoriaus $X = (X_1, \dots, X_k)$ tikimybinius skirstinius vadinamas polinominiu ir žymimas $X \sim P(m, p_1, \dots, p_k)$. Jį nusako tikimybės:

$$P(X_1 = m_1, X_2 = m_2, \dots, X_k = m_k) = \frac{m!}{m_1! \dots m_k!} p_1^{m_1} \dots p_k^{m_k},$$

čia $m \geq m_i \geq 0$ ir $m_1 + \dots + m_k = m$.

1.3 POLINOMINIO SKIRSTINIO TAIKYMAS

Tarkime, kad turime paprastą atsitiktinę imtį $X^n = (X_1, X_2, \dots, X_n)$, kur $X_1, X_2, \dots, X_n$ yra nepriklausomi atsitiktiniai elementai, simbolizuojantys apskrities ar kitos didelės teritorijos gyventojus. Bendrą gyventojų skaičių žymime n . Tegu A_j simbolizuoja j -ąją Šiaulių (ar kitą) apskritį, kai $j = 1, \dots, k$; k – apskričių skaičius. Pažymėkime Y_j gyventojų, patekusių į A_j apskritį, skaičių:

$$Y_j = \sum_{i=1}^n 1(X_i \in A_j), \quad j = 1, \dots, k.$$

Vietoje pradinės imties $X^n = (X_1, X_2, \dots, X_n)$ gauname imtį $Y^k = (Y_1, Y_2, \dots, Y_k)$. Atsitiktinis vektorius $Y^k = (Y_1, Y_2, \dots, Y_k)$ turi polinominį skirstinį $Y^k \sim P = (p_1, p_2, \dots, p_k, n)$, t. Y.

$$P(Y_1 = n_1, Y_2 = n_2, \dots, Y_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

čia $n_1 + \dots + n_k = n$ ir $p_1 + p_2 + \dots + p_k = 1$, $p_i, i = 1, \dots, k$ yra tikimybės.

Yra žinoma, kad parametrų p_i maksimalaus tikėtumo įverčiai yra pavidalo $\frac{Y_i}{n}$.

Tikriname statistinę hipotezę $H_0: p_i = p_{i0}, i = 1, \dots, k$, turėdami imtį $Y^k = (Y_1, Y_2, \dots, Y_k)$.

Taikome Chi-kvadrato kriterijų. Sudarome Pirsono statistiką:

$$\chi^2 = \sum_{i=1}^k \frac{(Y_i - n \cdot p_{i0})^2}{n \cdot p_{i0}} = \sum_{i=1}^k \frac{n}{p_{i0}} \cdot \left(\frac{Y_i}{n} - p_{i0} \right)^2,$$

kuri asimptotiškai turi χ^2 skirstinį su $k-1$ laisvės laipsnių, kai hipotezė H_0 teisinga.

Todėl, parinkę reikšmingumo lygmenį α , hipotezei H_0 tikrinti konstruojame tokią kritinę sritį:

$$W = \{Y^n : \chi^2 > \chi_{1-\alpha}^2(k-1)\},$$

čia $\chi_{1-\alpha}^2(k-1)$ yra χ^2 skirstinio su $k-1$ laisvės laipsnių $1-\alpha$ eilės kvantilis.

Jei imties realizacija $y^k = (y_1, y_2, \dots, y_k)$ tokia, kad

$$\chi^2 = \chi^2(y^k) > \chi^2_{1-\alpha}(k-1),$$

tada hipotezę H_0 atmetame, o jei

$$\chi^2 \leq \chi^2_{1-\alpha}(k-1),$$

tada hipotezę priimame.

Pateiktas polinominio skirstinio taikymo metodas ir jo aprašymas paimti iš [9] straipsnio.

1.4 KULBAKO-LEIBLERO INFORMACIJA

Tarkime turime dvi hipotezes H_1, H_2 kurios aprašo dvi c – informacijos kiekių tikimybes (c grupių arba kategorijų),

$$H_i: P = P_i, \quad P_i = (p_{i1}, p_{i2}, \dots, p_{ic}), \quad p_{i1} + p_{i2} + \dots + p_{ic} = 1, \quad i = 1, 2.$$

Tada Kulbako-Leiblero informacija, atitinkanti tikimybinius matus P_1 ir P_2

yra

$$I(1:2) = p_{11} \log \frac{p_{11}}{p_{21}} + p_{12} \log \frac{p_{12}}{p_{22}} + \dots + p_{1c} \log \frac{p_{1c}}{p_{2c}}.$$

$I(1:2)$ nusako pirmos rūšies klaidos tikimybės $\alpha_1(\delta_n)$ eksponentinį mažėjimą, kai fiksuota antros rūšies klaidos tikimybė $\alpha_2(\delta_n) = \alpha_2 \in (0,1)$, δ_n – statistinis kriterijus, t.y.

$$\alpha_1(\delta_n) \sim e^{-nI(1:2)}, \quad n \rightarrow \infty.$$

Analogiškai

$$I(2:1) = p_{21} \log \frac{p_{21}}{p_{11}} + p_{22} \log \frac{p_{22}}{p_{12}} + \dots + p_{2c} \log \frac{p_{2c}}{p_{1c}}.$$

$I(2:1)$ nusako antros rūšies klaidos tikimybės $\alpha_2(\delta_n)$ eksponentinį mažėjimą, kai fiksuota pirmos rūšies klaidos tikimybė $\alpha_1(\delta_n) = \alpha_1 \in (0,1)$, t.y.

$$\alpha_2(\delta_n) \sim e^{-nI(2:1)}, \quad n \rightarrow \infty.$$

Sumarinė informacija $J(1,2)$ apibrėžiama formule:

$$J(1,2) = I(1:2) + I(2:1).$$

Apibrėžtos informacijos yra teigiamos: $I(1:2) \geq 0$, $I(2:1) \geq 0$, $J(1,2) \geq 0$.

Skyrelio informacija paimta iš [10] (74 psl.) knygos.

1.5 STATISTINĖS HIPOTEZĖS IR JŲ KLAIDOS TIKIMYBĖS

Teorija parašyta remiantis [8] knyga.

Tarkime turime binarinį statistinį eksperimentą $\mathcal{E} = (\mathcal{X}, \mathcal{B}, \{P_0, P_1\})$ su imtimi $X \in \mathcal{X}$. Iš anksto nėra žinoma, kuris tikrasis imties X skirstinys P_0 ar P_1 .

Apibrėžimas. Bet koks teiginys apie tikrąją imties X skirstinį P^X vadinamas hipoteze. Kadangi šiuo atveju tėra dvi galimybės, tai tikriname tik dvi hipotezes:

$$H_0: P^X = P_0, \quad H_1: P^X = P_1,$$

(užrašas $P^X = P_i$ reiškia teiginį: tikrasis X skirstinys P^X yra P_i).

Apibrėžimas. Hipotezė, sudaryta iš vienos galimybės vadinama paprastąja.

Apibrėžimas. Hipotezių H_0 ir H_1 tikrinimo statistiniu kriterijumi vadiname kiekvieną statistiką $\delta = \delta(X)$, įgyjančią dvi reikšmes 0 ir 1. Kada $\delta(X)$ įgyja reikšmę 0, tada hipotezė H_0 priimama (H_1 atmetama), o kai įgyja reikšmę 1 - H_1 priimama, o H_0 atmetama.

Šitaip apibrėžtas statistinis kriterijus $\delta(X)$ galimas X reikšmes $\mathcal{X}$ padalija į dvi nesikertančias aibes $\mathcal{X}_0 = \mathcal{X} \setminus W$, kur priimama hipotezė H_0 priimama, o H_1 atmetama, ir $\mathcal{X}_1 = W$, kur priimama H_1 , o H_0 atmetama.

Apibrėžimas. Sritis W , kurioje priimama hipotezė H_1 , o H_0 atmetama vadinama kritine sritimi.

Apibrėžimas. Skaičius α_0 (atitinkamai α_1), vadinamas kriterijaus $\delta = \delta(X)$ 1-osios (atitinkamai 2-osios) rūšies klaidos tikimybe, žymi tikimybę atmesti hipotezę H_0 (atitinkamai H_1), kada ji teisinga.

1.6 SERIJŲ KRITERIJUS

Atsitiktinė imtis $X^n = (X_1, \dots, X_n)$ pagal prielaidą yra sudaryta iš nepriklausomų atsitiktinių dydžių. Tačiau konkrečiame modelyje imties elementai $X_1, \dots, X_n$ gali būti ir net visai ne atsitiktiniai ir priklausomi. Norėdami įsitikinti, kaip yra iš tikrųjų, tikriname hipotezę H_0 : imties elementai $X_1, \dots, X_n$ yra atsitiktiniai ir nepriklausomi. Tokiai hipotezei tikrinti taikysime serijų kriterijų su empirine mediana.

Pagal šį kriterijų pirmiausia užrašome imties $X_1, \dots, X_n$ elementus didėjančia tvarka.

Pagal formulę

$$M_e = \begin{cases} X_{(\frac{n+1}{2})}, & \text{jei } n = 2l + 1, \\ \frac{1}{2} \left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right), & \text{jei } n = 2l, \end{cases}$$

randame imties empirinę medianą M_e . Po to M_e reikšmę lyginame su pradinės imties elementais $X_1, \dots, X_n$. Rašome „+“ jei $X_1 > M_e$ ir „-“, jei $X_1 < M_e$, jei reikšmės sutampa, nerašome nieko. Šią procedūrą tęsiame, kol patikriname visus imties elementus. Gauname „+“-ų ir „-“-ų eilutę. Joje esančių „+“ skaičių žymėsime raide k_1 , o „-“ skaičių - k_2 . Bendrą visų serijų skaičių žymėsime raide N . $N = N(X^n)$ bus serijų kriterijaus statistika.

Mažas imties serijų skaičius N rodo, kad imties elementai yra giminingi, lyg būtų iš vienos šeimos. Tuo metu, kai N didelis, jis nurodo tarp imties elementų esamą dėsningą kaitą. Abiem šiais atvejais pažeidžiamas imties elementams keliamas atsitiktinumumo ir nepriklausomumo reikalavimas. Tuo remiantis hipotezei H_0 tikrinti galima susidaryti vieną iš trijų alternatyvų:

$H_1^{(1)}$: imties elementai priklausomi ir neatsitiktiniai.

Hipotezei H_0 tikrinti, priklausomai nuo alternatyvos, turime kritines sritis:

$W_1 = \left\{ X^n: N \leq N_{\frac{\alpha}{2}}(A, n_1, n_2) \text{ arba } N \geq N_{\frac{\alpha}{2}}(V, n_1, n_2) \right\}$, kai alternatyva $H_1^{(1)}$, arba čia $n_1 = \max(k_1, k_2)$, $n_2 = \min(k_1, k_2)$, $N_{\alpha}(A, n_1, n_2) = N_{\alpha}(\text{Apatinė}, n_1, n_2)$ ir $N_{\alpha}(V, n_1, n_2) = N_{\alpha}(\text{Viršutinė}, n_1, n_2)$ -tam tikros kritinės reikšmės, priklausančios nuo ktiretijos reikšmingumo lygmens α ir n_1, n_2 reikšmių, atitinkančių nagrinėjamą imtį. Serijų kriterijaus aprašas paimtas iš knygos [1].

1.7 LAIKO EILUTĖS

Laiko eilučių teorijos pagrindinės sąvokos ir formulės parašytos remiantis [11] knyga.

1.7.1 AUTOREGRESINIAI PROCESAI

Apibrėžimas. Atsitiktinė seka $\{X_t\}$ yra vadinama pirmosios eilės autoregresijos procesu, arba AR(1), jeigu kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + \varepsilon_t,$$

čia dydis a yra konstanta, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Apibrėžimas. Atsitiktinė seka $\{X_t\}$ yra vadinama antrosios eilės autoregresijos procesu, arba AR(2), jeigu kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + \varepsilon_t,$$

čia a_1, a_2 , yra konstantos, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Apibrėžimas. Atsitiktinė $\{X_t\}$ seka yra vadinama p – tos eilės autoregresijos procesu arba AR(p), jei kiekvienas X_t tenkina skirtuminę lygtį

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + \dots + a_p \cdot X_{t-p} + \varepsilon_t, \quad (1.2)$$

čia $a_1, a_2, \dots, a_p$ yra pastovieji dydžiai, o $\{\varepsilon_t\}$ – atsitiktinė paklaidų seka.

Padauginę abi lygties (1.2) puses iš X_{t-r} , turime:

$$X_t \cdot X_{t-r} = a_1 \cdot (X_{t-1} \cdot X_{t-r}) + a_2 \cdot (X_{t-2} \cdot X_{t-r}) + \dots + a_p \cdot (X_{t-p} \cdot X_{t-r}) + (X_{t-r} \cdot \varepsilon_t),$$

čia $r = 0, 1, 2, \dots$

Iš šios lygybės kartu su (1.2) gauname:

[illegible]

Ši lygtis vadinama Yule – Walker lygtimi ir naudojama AR koeficientų ir koreliacinės funkcijos ryšiui išreikšti.

1.7.2 AUTOREGRESINIŲ PROCESŲ PARAMETRŲ VERTINIMAS

Tarkime, kad turime autoregresijos proceso $\{X_t\}$, išreiškiamo (1.2) lygtimi, stebinius $X_1, X_2, \dots, X_n$

ir žinoma lygties eilė p . Reikia gauti nežinomų parametrų $a_1, a_2, \dots, a_p, \sigma^2$.

Žinome, kad

$$\hat{\gamma}(p) = \frac{1}{n} \cdot \sum_{j=1}^{n-k} (X_{j+k} - \bar{X}) \cdot (X_j - \bar{X}),$$

čia

$$\bar{X} = \frac{1}{n} \cdot \sum_{j=1}^n X_j.$$

Iš lygčių sistemos apskaičiuojame įverčius $\hat{\gamma}(0), \hat{\gamma}(1), \dots, \hat{\gamma}(p)$ ir įstatome į (1.2) lygtį. Gauname lygtį su nežinomaisiais parametrais $a_1, a_2, \dots, a_p, \sigma^2$. Tuomet surandame $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ įverčius.

Teorija parašyta remiantis [11] knyga.

1.7.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS

Teorija parašyta remiantis [4] knyga.

Esminį vaidmenį sprendžiant išskyrimo ir įvertinius trendinės $f_{TR}(t)$, sezoninės $\mu(t)$ ir ciklinės $\psi(t)$ sudaromosios išskaidyme yra pradinis analizės etapas kuriame išaiškinamas pats faktas buvimo ar nebuvimo neatsitiktinės (ir priklausančios nuo laiko t) komponentės išskaidyme, iš esmės kalba eina apie statistinės hipotezės $H_0: MX(t) = a = const.$ patikrinimą. Konstruojamas įvertis (aproksimacija) dėl nežinomos integralinės neatsitiktinės dalies $f(t) = \chi(A) \cdot f_{TR}(t) + \chi(B) \cdot \mu(t) + \chi(C) \cdot \psi(t)$, t.y. sprendžiamas sulyginimas

(eliminavimas atsitiktinės liekanos $a(t)$) analizuojamos laiko eilutės $X(t)$. Surašome analizuojamos laiko eilutės $X(t_1), \dots, X(t_n)$ narius didėjančia tvarka (variacione eile). Jai apibrėžkime medianą $X_{med}^{(n)}$ pagal formulę

$$X_{med}^{(n)} = \begin{cases} X_{(\frac{n+1}{2})}, & \text{jei } n = 2l + 1, \\ \frac{1}{2} \left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right), & \text{jei } n = 2l \end{cases}$$

Po to sudarome serijas iš plusų ir minusų kaip tikrindami nepriklausomumą. Atskiru atveju grįžtant prie pradinės eilės $X(t_1), \dots, X(t_n)$ vietoj kiekvieno nario $X(t)$, $t=1, 2, \dots, n$, rašome +, jei $X(t) > X_{med}^{(n)}$ ir minusą (-), jei $X(t) < X_{med}^{(n)}$ (laiko eilutės nariai kurie yra lygūs $X_{med}^{(n)}$ neimami).

Gauta plusų ir minusų eilutė charakterizuojama bendru serijų skaičiumi $\gamma(n)$ ir pačios ilgiausios serijos ilgiu $\tau(n)$. Čia serija laikoma kartu einanti plusų arba minusų grupė, pavyzdžiui seriją gali sudaryti tik vienas + arba vienas – ir tada jos ilgis lygus vienetui. Akivaizdu, kad jei nagrinėjame seką $X(t_1), \dots, X(t_n)$ sudarytą iš statistiškai nepriklausomų stebėjimų, atsitiktinai varijuojančių apie tam tikro pastovaus lygio a (t.y. jei teisinga hipotezė $H_0: MX(t) = a = const.$), tai kaita plusų ir minusų nurodytoje sukonstruotoje jų eilutėje turi būti daugiau ar mažiau atsitiktinė, t.y. ta seka privalo neturėti labai ilgos plusų arba minusų serijos einančios iš eilės, ir atitinkamai bendras serijų skaičius $\gamma(n)$ negali būti per daug mažas. Šiame kriterijuje tikslinga nagrinėti statistinę porą $(\gamma(n); \tau(n))$. Todėl mes turime užfiksuoti ir sužinoti atsitiktinių dydžių $(\gamma(n); \tau(n))$ dvimatį skirstinį galiojantį su sąlyga, kad tikrinama hipotezė $H_0: MX(t) = a = const.$ teisinga. Čia nurodysime apytikslį kriterijų ir jo sukonstravimui pasinaudosime:

- penkių procentų taško įverčiu $\tau_{0,05}(n)$ pasiskirsčiusio $\tau_{0,05}(n) \approx [1,43 \cdot \ln(n + 1)]$ (čia $[x]$ žymi “sveiką x dalį”)
- įverčiais iš viršaus ir apačios dėl tikimybių $P\{\gamma(n) < \gamma_{0,95}(n); \tau(n) > \tau_{0,05}(n)\}$.

Tai leidžia suformuluoti sekančią apytikslę taisyklę (tikrinant hipotezę $H_0: MX(t) = a = const.$):
jei nors viena iš nelygybių

$$\gamma(n) > \left\lceil \frac{1}{2} (n + 2 - 1,96\sqrt{n-1}) \right\rceil$$

$$\tau(n) < [1,43 \ln(n + 1)]$$

negalioja, tada hipotezė atmetama su tikimybe α , esančia tarp 0,05 ir 0,0975.

1.8 KORELIACINĖ ANALIZĖ

Iš knygų [6,8] bus pateikta pagrindinė šio skyrelio informacija.

1.8.1 EMPIRINIS KORELIACIJOS KOEFICIENTAS

Sakykim, kad turime dvimatę atsitiktinę imtį $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, kurios elementai (X_i, Y_i) , turintys tą patį dvimatį skirstinį $P_{\theta}^{(2)}, \theta \in \Theta$, paimti iš nepriklausomų dvimačių atsitiktinių vektorių sekos $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n), \dots$. Šią imtį pirmiausia charakterizuoja išsibarstymo diagrama ir dvimatis empirinis skirstinys.

Apibrėžimas. Dvimatės imties išsisklaidymo diagrama vadiname imties taškų $(X_i, Y_i), i=1, \dots, n$, atidėtų stačiakampėje koordinačių sistemoje, visumą.

Apibrėžimas. Dvimatės imties empiriniu skirstiniu vadiname sąlyginį skirstinį diskretaus atsitiktinio vektoriaus (X, Y) , įgyjančio reikšmes $(X_i, Y_i), i=1, \dots, n$ su tikimybe $\frac{1}{n}$, esant sąlygai, kad pasirodė $(X_1, Y_1), \dots, (X_n, Y_n)$.

To vektoriaus skirstinio skaitinės charakteristikos vadinamos dvimatės imties empirinėmis charakteristikomis.

Vadinasi, teorinio dvimatės atsitiktinės imties $(X_1, Y_1), \dots, (X_n, Y_n)$ koreliacijos koeficiento

$$\rho = \frac{M(X_i - MX_i)(Y_i - MY_i)}{\sqrt{DX_i DY_i}},$$

empirinis analogas yra empirinis koreliacijos koeficientas, apskaičiuojamas pagal formulę

$$r = \frac{m_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

kur

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

$$S_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, S_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2},$$

$$m_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Koreliacijos koeficientas ρ apibūdina tiesinę priklausomybę tarp dydžių X_i ir Y_i .

Kadangi teorinio (tikrojo) koreliacijos koeficiento ρ tarp X_i ir Y_i nežinome, tai vietoj jo naudojame jo įvertį – empirinį koreliacijos koeficientą r .

1.8.2 HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTĄ

Empirinis koreliacijos koeficientas r yra atsitiktinis dydis, todėl turėdami jo reikšmę, apskaičiuotą pagal vieną dvimatės imties realizaciją, negalime būti visiškai tikri, kad ji pakankamai artima tikrajai koreliacijos koeficiento ρ reikšmei. Šiuo atveju daug patikimiau yra patikrinti statistinę hipotezę apie galimą ρ reikšmę. Kadangi praktiškai labai svarbu yra žinoti, ar tarp dydžių X_i ir Y_i yra (tiesinė) priklausomybė ar ne, tai pirmiausia tikrinsime hipotezę $H_0: \rho = 0$.

Tarkime, kad turime dvimatę atsitiktinę imtį $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, kurios elementai (X_i, Y_i) turi dvimatį normalųjį skirstinį

$$N_2 \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} \right)$$

su parametrais $a_1, a_2, \sigma_X^2, \sigma_Y^2, \rho$, kur ρ – koreliacijos koeficientas tarp X_i ir Y_i .

Išskirsime kelis atvejus.

- 1) atvejis. Kada imties tūris $n > 10$, o $|r|$ artimas vienetui, tada taikoma R. A. Fišerio transformacija

$$Arthr = \frac{1}{2} \ln \frac{1+r}{1-r}$$

Šiuo atveju, jei teisinga hipotezė $H_0: \rho = \rho_0$, kur ρ_0 – duotas skaičius, tai

$$Arthr \sim N \left(Arth\rho_0, \frac{1}{n-3} \right).$$

Todėl

$$Z = \sqrt{n-3} (Arthr - Arth\rho_0) \sim N(0,1),$$

Kai H_0 teisinga.

Tuo remiantis hipotezei $H_0: \rho = \rho_0$ tikrinti su reikšmingumo lygmeniu α sudarome kritinę sritį:

$$W_1 = \left\{ (X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}} \right\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq \rho_0.$$

- 2) atvejis. Kada imties tūris n mažas, statistika

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

turi Stjudento skirstinį su $n-2$ laisvės laipsnių, jei hipotezė $H_0: \rho = 0$ teisinga. Tuo remiantis parenkamas reikšmingumo lygmuo α ir hipotezei $H_0: \rho = 0$ tikrinti apibrėžiama kritinė sritis:

$$W_1 = \left\{ (X_i, Y_i), i = 1 \dots, n: |T| > t_{1-\frac{\alpha}{2}}(n-2) \right\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq 0.$$

1.8.3 TIESINĖ REGRESIJA

Tiesinės regresijos teorija parašyta remiantis iš [2,8] knygomis.

Sakykim, kad turime du atsitiktinius dydžius X ir Y . Apskaičiavę jų koreliacijos koeficientą gauname, kad jis artimas ± 1 . Tai reiškia, kad tarp dydžių X ir Y egzistuoja beveik tiesinė priklausomybė, kuri užrašoma per sąlyginį vidurkį

$$M(Y|X = x) = \beta_0 + \beta_1 x. \quad (1)$$

(1) formulė nusako atsitiktinio dydžio Y tiesinės regresijos lygtį X atžvilgiu. Skaičiai β_0 ir β_1 vadinami tiesinės regresijos parametrais.

Paprastai atsitiktinių dydžių X ir Y tikimybiniai skirstiniai tiksliai nežinomi, todėl tiesinės regresijos lygtį $y = \beta_0 + \beta_1 x$ įvertiname pagal (X, Y) stebėjimo rezultatus $(x_1, y_1), \dots, (x_n, y_n)$, kurie teoriškai yra realizacija atsitiktinės dvimatės imties $(X_1, Y_1), \dots, (X_n, Y_n)$, kur $M(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i, i = 1, \dots, n$. Toliau daroma prielaida, kad

$$(Y_i|X_i = x_i) \sim N(\beta_0 + \beta_1 x_i, \sigma^2),$$

kuri reiškia, kad

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (2)$$

Čia $\varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, n$, - nepriklausomi atsitiktiniai dydžiai.

Kadangi kiekvienai fiksuotai X_i reikšmei x_i mes turime normalųjį atsitiktinio dydžio išsibastymą, tai pradinės dvimatės imties nagrinėjimas ekvivalentus nagrinėjimui dvimatės imties $(X_1, Y_1), \dots, (X_n, Y_n)$, kurioje $X_1, \dots, X_n$ yra neatsitiktiniai dydžiai, o atsitiktiniai dydžiai Y_i užrašomi (2) pavidalu. Kitaip sakant, turime iš anksto žinomų skaičių rinkinį $(X_1, \dots, X_n)$ ir atsitiktinę nepriklausomų dydžių imtį $(Y_1, \dots, Y_n)$, kur

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \dots, n, \quad (3)$$

(3) formulė nusako tiesinės regresijos modelį.

Tiesinės regresinės analizės uždavinys yra toks: naudojantis imtimi $(X_1, Y_1), \dots, (X_n, Y_n)$

- 1) gauti „geriausius“ (3) modelio nežinomų parametrų β_0, β_1 ir σ^2 taškinis ir intervalinius įvertinius,
- 2) patikrinti statistines hipotezes apie parametrus β_0 ir β_1 ,
- 3) nustatyti, ar yra modelis gerai suderintas su imties duomenimis (modelio adekvatumo patikrinimas)

Tiesinės regresijos parametrų β_0 ir β_1 įverčius $\widehat{\beta}_0$ ir $\widehat{\beta}_1$ randame mažiausių kvadratų metodu, minimizuodami dydį

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2 \rightarrow \min.$$

Panaudoję būtiną ekstremumo sąlygą $\frac{\partial S^2}{\partial \widehat{\beta}_1} = 0$ ir $\frac{\partial S^2}{\partial \widehat{\beta}_0} = 0$, gauname

$$\begin{cases} \sum_{i=1}^n (Y_i - \widehat{\beta}_1 X_i - \widehat{\beta}_0) X_i = 0, \\ \sum_{i=1}^n (Y_i - \widehat{\beta}_1 X_i - \widehat{\beta}_0) = 0. \end{cases}$$

Išsprendę šią sistemą, turime įverčius

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}, \quad \widehat{\beta}_1 = \frac{Q_{X,Y}}{Q_X},$$

kur

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

$$Q_X = \sum_{i=1}^n (X_i - \bar{X})^2,$$

$$Q_{X,Y} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

(3) prielaidos dėka galima įsitikinti, kad

$$\frac{S^2(n-2)}{\sigma^2} \sim C(n-2),$$

t.y. turi χ^2 skirstinį su $n-2$ laisvės laipsnių.

Remdamiesi šiuo faktu, pasirinkę pasiklovimo tikimybę $1-\alpha$, randame parametrų β_0 ir β_1 pasikliautinus atsitiktinius intervalus.

Tiesinės regresijos modelis vadinamas nereikšmingu, jei parametras $\beta_1 = 0$. Hipotezei $H_0: \beta_1 = 0$ patikrinti galima remiantis statistika

$$F = \frac{(\widehat{\beta}_1)^2 Q_X}{S^2},$$

kuri turi Fišerio skirstinį su 1 ir $n-2$ laisvės laipsnių, jei hipotezė H_0 teisinga. Pasirenkame reikšmingumo lygmenį α ir hipotezei $H_0: \beta_1 = 0$ tikrinti apibrėžiame kritinę sritį

$$W_1 = \left\{ (X_i, Y_i), i = 1, \dots, n: F < F_{\frac{\alpha}{2}}(1, n-2) \text{ arba } F > F_{1-\frac{\alpha}{2}}(1, n-2) \right\},$$

$$\text{jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \beta_1 \neq 0,$$

$$W_2 = \{ (X_i, Y_i), i = 1, \dots, n: F < F_{\alpha}(1, n-2) \},$$

$$\text{jei } H_0 \text{ alternatyva yra } H_1^{(2)}: \beta_1 < 0,$$

$$W_3 = \{ (X_i, Y_i), i = 1, \dots, n: F > F_{1-\alpha}(1, n-2) \},$$

$$\text{jei } H_0 \text{ alternatyva yra } H_1^{(3)}: \beta_1 > 0,$$

čia $F_p(1, n-2)$ – Fišerio skirstinio su 1 ir $n-2$ laisvės laipsnių p eilės kvantilis.

Kada hipotezė $H_0: \beta_1 = 0$ atmetama, tiesinės regresijos modelis laikomas reikšmingu.

Turėdami regresijos tiesės parametrų įverčius, galime nustatyti, ar gautoji lygtis $y = \hat{\beta}_0 + \hat{\beta}_1 x$ teisingai aprašo imties duomenis, t.y. ar ji adekvati jiems.

Tam tikslui apskaičiuojamos statistikos

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2,$$

kur S_1^2 charakterizuoja duomenų nukrypimą nuo vidurio, o S^2 – duomenų nukrypimą nuo regresijos linijos taškų. Sudarome statistiką

$$F_1 = \frac{S_1^2 - S^2}{S^2},$$

kuri turi Fišerio skirstinį su 1 ir $n-2$ laisvės laipsnių, kada teisinga (3) prielaida. Pasirenkame reikšmingumo lygmenį α ir apskaičiuojame $1-\alpha$ eilės Fišerio skirstinio su 1 ir $n-2$ laisvės laipsnių kvantilį $F_{1-\alpha}(1, n-2)$. Jei dvimatė imtis $(X_1, Y_1), \dots, (X_n, Y_n)$ tokia, kad

$$F_1 > F_{1-\alpha}(1, n-2),$$

tada regresijos lygtis laikoma adekvačia imties duomenims.

1.9 TIESINĖ REGRESIJA SU PROGRAMA „SPSS“

Pagrindinė šio skyrelio teorija parašyta remiantis [3,5] šaltiniais.

Tikrinant ar daugiamatė tiesinės regresijos lygtis gali būti naudojama prognozavimui, ar ji yra tikslinga – svarbu atkreipti dėmesį į kai kuriuos rodiklius kuriuos gauname skaičiuodami tiesinę regresiją programoje „SPSS“.

Pirmiausias rodiklis į kurį reikia atkreipti dėmesį yra koreliacijos koeficientai. Visi nepriklausomi kintamieji turi koreliuoti su priklausomu kintamuoju, o tarpusavyje stipriai koreliuoti neturėtų.

Antra, reikia atkreipti dėmesį į „Model Summary“ lentelės rezultatus. Koreguotas determinacijos koeficientas (Adjusted R Square) parodo kiek % modelis paaiškina pasirinkto kintamojo reikšmių sklaidos apie vidurkį tiesine regresija nepriklausomų kintamųjų atžvilgiu. Kuo šis rodiklis didesnis, tuo modelis tinkamesnis. Taip pat ir determinacijos koeficientas (R Square) rodo tą patį. Jeigu imtis gerokai didesnė už nepriklausomus kintamuosius, dažniausiai žiūrima į determinacijos koeficiento reikšmę. Mažiausia reikšmė tinkamam modeliui yra $R^2 \geq 0,2$. Taip pat lentelėje yra standartinės regresijos paklaidos (Std. Error of the Estimate) dydis, kuo jis mažesnis, tuo geriau regresijos modelis apibūdina duomenis.

Durbin-Watson statistika rodo ar yra autokoreliacija, jei šis rodiklis yra tarp 1,5 ir 2,5 dažniausiai tariama, kad autokoreliacijos nėra.

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1					

Sekanti lentelė yra "Anova". Ieškant daugiamačių regresijos tikrinama hipotezė apie regresijos netiesiškumą. Jei lentelėje reikšmingumo lygmuo (Sig.) yra $<0,05$, hipotezė atmetama ir tai reiškia, kad regresija yra tiesinė – bent vienas koeficientas nelygus nuliui ir regresijos modelis bent jau iš dalies tinka prognozėms.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression					
	Residual					
	Total					

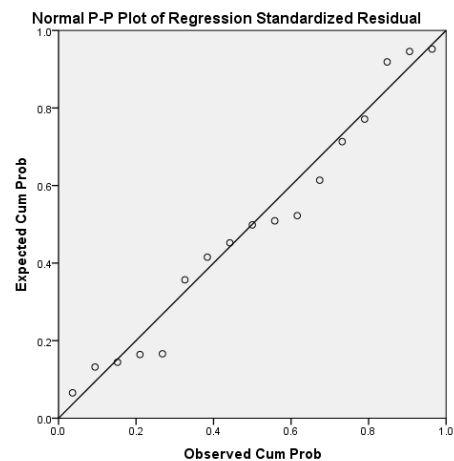
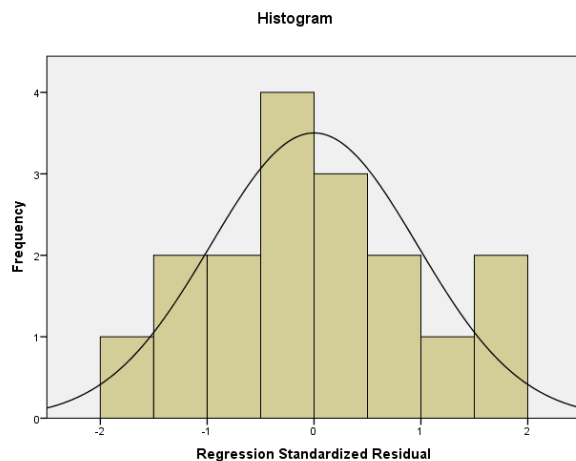
Lentelėje „Coefficients“ skiltyje B „Unstandardized Coefficients“ gauname regresijos lygties koeficientus. Jeigu koeficientas reikšmingas, lentelės skiltyje „Sig.“ reikšmė turi būti $<0,05$, tokiu atveju kintamasis statistiškai reikšmingas ir yra tinkamas prognozavimui. Kitu atveju jis nėra tinkamas. Standartizuotų koeficientų (Standardized Coefficients) reikšmės dydis parodo, kuris nepriklausomas kintamasis yra svarbiausias. Dispersijos mažėjimo daugiklis (VIF) parodo, ar nepriklausomi kintamieji turi multikolinearumo problemą. Jei „VIF“ > 4 – multikolinearumas yra. Tą patį parodo „Tolerance“ jei reikšmė mažesnė nei 0,25.

Coefficients ^a								
Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta				Tolerance	VIF
1 (Constant)								

Lentelėje „Residual Statistics“ reiktų atkreipti dėmesį į Kuko matą (Cook's Distance), jis parodo ar duomenyse yra išskirčių – jei matas viršija 1, tai nėra gerai.

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value					
...	...	...	...	...	...
Cook's Distance					
Centered Leverage Value					

Informacija apie modelio standartizuotųjų liekamųjų paklaidų normalumą yra pateikiama dviejuose grafikuose „Histogram“ ir „P-P plot“. Histogramos grafike yra nubraižyta ir normalioji kreivė, t.y. tai, į ką turėtų būti panaši histograma, jeigu liekamosios paklaidos būtų normalios. Antrajame grafike atidėti standartizuotųjų liekamųjų paklaidų ir normaliojo atsitiktinio kintamojo santykiniai procentiniai dažniai. Kuo taškai arčiau nubrėžtos tiesės (idealiu atveju visi taškai yra ant tiesės), tuo duomenys normalesni.



Standartizuotosios liekamosios paklaidos naudojamos patikrinimui, ar Y normaliai pasiskirstęs. Normalumui ištirti naudojamas Šapiro – Vilko (Shapiro – Wilk) testas. Jei lentelėje „Tests of Normality“ „Sig.“ Reikšmė $> 0,05$ teigiame, kad standartizuotosios liekamosios paklaidos yra normalios. Taip pat pateikiamas Kolmogorovo-Smirnov testas, lygiai taip pat jei „Sig.“ reikšmė $> 0,05$ teigiame, kad standartizuotosios liekamosios paklaidos yra normalios.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual						

1.10 DUOMENŲ NORMALUMAS SU PROGRAMA „SPSS“

Skyrelis parašytas remiantis [3,5] šaltiniais.

Patikrinti, ar duomenys yra pasiskirstę normaliai, naudojamas Šapiro – Vilko testas. Jei lentelėje „Tests of Normality“ „Sig.“ Reikšmė $> 0,05$ teigiame, kad duomenys pasiskirstę normaliai. Taip pat pateikiamas Kolmogorovo-Smirnov testas, lygiai taip pat jei „Sig“ reikšmė $> 0,05$ teigiame, kad duomenys pasiskirstę normaliai.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual						

2. TYRIMAS. PRAKTINĖ DALIS

2.1 DUOMENYS. PRADINĖ STATISTINĖ ANALIZĖ LIETUVOS MASTU

Neigiamų faktorių, kurie gali daryti įtaką žmonių migracijai yra įvairių. Šiame tyrime pasirinkome nedarbo, nusikaltimų bei savižudybių duomenis, kurie bus laikomi neigiamais. Esant dideliame nedarbe žmonės iš apskrities išvyksta tai reiškia, kad toje apskrityje dides išvykusių skaičius, o nedarbui mažėjant dides atvykusių skaičius arba mažes išvykusių žmonių kiekis. Kitas neigiamas faktorius yra nusikalstamumas – jam padidėjus žmonės galimai ieškos saugesnės vietos gyventi, o tai turėtų atsispindėti išvykusių iš apskrities gyventojų skaičiuje, o nusikalstamumui mažėjant, atvykusių skaičius dides. Savižudybės Lietuvoje yra didžiulė problema, 2016 metų duomenimis 100 tūkst. gyventojų teko 25,7 savižudybės. Savižudybė gali padaryti įtaką žmonių judėjimui, tarkime, po artimojo netekties šeima nori pakeisti gyvenamąją vietą. Surinktus duomenis apie šalies viduje migravusius žmones (vadinsime „vidiniai migrantai“), nusikaltimus, nedarbą, savižudybes 2002 – 2018 metais panagrinėsime plačiau (duomenys naudojami remiantis [7,12] šaltiniais).

Pirmiausia turime gyventojų atvykusių bei išvykusių iš apskričių skaičius.

1 lentelė. Išvykusių iš apskričių gyventojų skaičiai nuo 2002 iki 2018 metų

Apskritis	#	2002	2003	2004	2005	2006	2007	2008	2009	2010
Vilniaus	1	6360	10066	10113	8990	8634	9543	9434	8228	9185
Alytaus	2	2192	3734	4080	3962	3229	3209	3347	2990	2988
Kauno	3	8409	12343	11595	10520	10077	11098	11704	9928	9307
Klaipėdos	4	3845	6435	6680	6006	6322	6873	7030	6085	5652
Marijampolės	5	2202	3058	3360	3141	3299	3373	3381	3138	3030
Panevėžio	6	4120	5677	5752	4646	4705	5089	4952	4787	4691
Šiaulių	7	5091	6856	7280	6491	6647	7717	7305	6429	6360
Tauragės	8	1723	2261	2748	2567	2432	2734	2516	2193	2112
Telšių	9	2601	3851	4052	3496	3568	3983	3935	3306	3316
Utenos	10	2754	3544	3713	3150	2824	2964	3328	2859	3155
	Σ	39297	57825	59373	52969	51737	56583	56932	49943	49796

Apskritis	#	2011	2012	2013	2014	2015	2016	2017	2018
Vilniaus	1	9595	11810	11233	12476	12722	13912	14349	16106
Alytaus	2	3294	3562	3138	3425	3384	3736	3555	3916
Kauno	3	9995	11212	10995	12182	12138	13310	13617	14721
Klaipėdos	4	6088	6480	6441	7368	7356	8694	9120	10156
Marijampolės	5	3122	3791	3470	3630	3498	3723	4423	4611
Panevėžio	6	4695	5559	5274	5346	5242	5961	5895	6401
Šiaulių	7	6648	7246	6955	7247	7393	7946	7919	8789
Tauragės	8	2282	2867	2588	2592	2606	2942	2882	3198
Telšių	9	3243	3844	3473	3709	3621	4623	4181	4591
Utenos	10	3237	3682	3676	3440	3466	3725	3476	3687
	Σ	52199	60053	57243	61415	61426	68572	69417	76176

Sudėję visų apskričių išvykusių skaičius 2002 metais gauname bendrą išvykusių gyventojų skaičių į kitas apskritis Lietuvoje tais metais. Atlikę tą patį veiksmą visu tiriamu laikotarpiu gausime išvykusių skaičius Lietuvos mastu 2002 – 2018 metais.

2 lentelė. Atvykusių į apskritis gyventojų skaičiai nuo 2002 iki 2018 metų

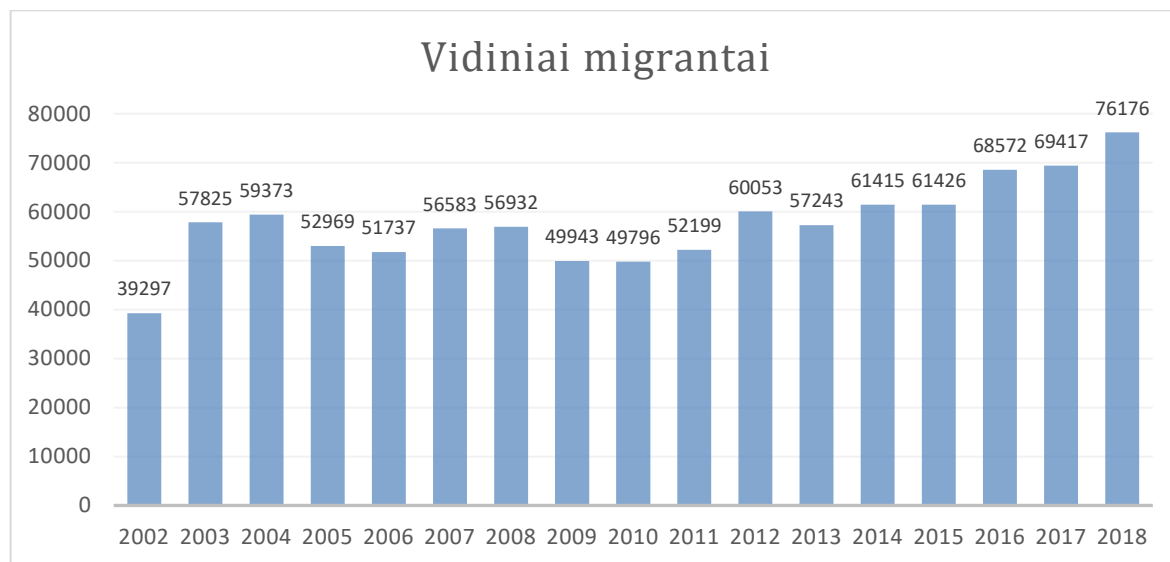
Apskritis	#	2002	2003	2004	2005	2006	2007	2008	2009	2010
Vilniaus	1	8207	14138	14051	12516	11391	12920	13387	11986	11879
Alytaus	2	1945	3202	3495	3343	2720	2696	2687	2431	2550
Kauno	3	7644	11035	11158	10496	10363	11395	11988	10136	9669
Klaipėdos	4	3924	6977	7299	6641	6798	7711	7957	6654	6400
Marijampolės	5	2113	2835	2993	2736	2915	3019	2806	2611	2622
Panevėžio	6	3905	5021	4814	3896	3957	4207	3979	3943	3906
Šiaulių	7	4587	5780	6238	5500	5682	6453	6069	5490	5504
Tauragės	8	1664	2080	2318	1971	2022	2123	1839	1708	1553
Telšių	9	2593	3592	3710	3137	3338	3525	3434	2687	2831
Utenos	10	2715	3165	3297	2733	2551	2534	2786	2297	2882
	Σ	39297	57825	59373	52969	51737	56583	56932	49943	49796
Apskritis	#	2011	2012	2013	2014	2015	2016	2017	2018	
Vilniaus	1	12827	17169	15289	15919	17172	19093	19005	20629	
Alytaus	2	2880	2730	2523	2767	2789	2951	2779	3280	
Kauno	3	10153	11504	11786	13002	12706	14058	14722	16030	
Klaipėdos	4	6691	6968	7231	8424	8133	9411	10122	11268	
Marijampolės	5	2679	2926	2763	3009	2841	3020	3553	3727	
Panevėžio	6	3788	4407	4256	4310	4080	4495	4377	4925	
Šiaulių	7	5847	5993	5863	6083	6146	6742	6604	7295	
Tauragės	8	1803	2298	1948	1991	1880	2189	2083	2204	
Telšių	9	2759	3190	2631	2945	2829	3535	3227	3616	
Utenos	10	2772	2868	2953	2965	2850	3078	2945	3202	
	Σ	52199	60053	57243	61415	61426	68572	69417	76176	

Atlikę tokius pat veiksmus kasmet kaip ir su išvykusių skaičiais, gauname atvykusių į apskritis skaičius Lietuvos mastu 2002 – 2018 metais. Sudarome naują atvykusių ir išvykusių iš apskričių gyventojų skaičiaus sumos kiekvienais metais lentelę.

3 lentelė. Išvykusių ir atvykusių į apskritis gyventojų skaičių sumos nuo 2002 iki 2018 metų.

Metai	Išvykę	Atvykę	Metai	Išvykę	Atvykę
2002	39297	39297	2008	56932	56932
2003	57825	57825	2009	49943	49943
2004	59373	59373	2010	49796	49796
2005	52969	52969	2011	52199	52199
2006	51737	51737	2012	60053	60053
2007	56583	56583	2013	57243	57243
2008	56932	56932	2014	61415	61415
2009	49943	49943	2015	61426	61426
2010	49796	49796	2016	68572	68572
2011	52199	52199	2017	69417	69417
2012	60053	60053	2018	76176	76176

Iš 3 lentelės duomenų galime matyti, kad suminės atvykusių ir išvykusių gyventojų reikšmės Lietuvos mastu kasmet yra vienodos. Žiūrint į šiuos rezultatus Lietuvos mastu, gyventojai neišvyksta iš šalies, o keičia gyvenamąją vietą Lietuvoje (nesvarbu išvyko ar atvyko), todėl suminius rezultatus kasmet laikome vidinių migrantų skaičiumi. Sudarome diagramą:



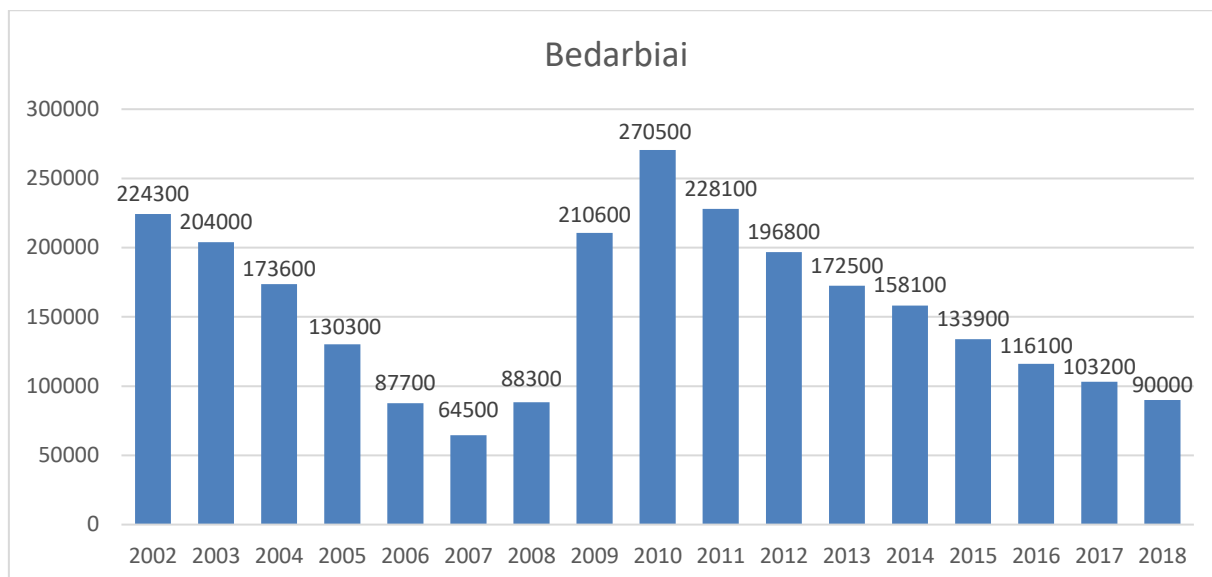
1 pav. Gyventojų migravusių šalies viduje skaičius nuo 2002 iki 2018 metų

Analogiškai sudėdami apskričių duomenis kasmet, sudarysime bedarbių, savižudybių bei nusikaltimų lenteles ir diagramas Lietuvos mastu.

4 lentelė. Bedarbių skaičiai apskrityse nuo 2002 iki 2018 metų

Apskritis	#	2002	2003	2004	2005	2006	2007	2008	2009	2010
Vilniaus	1	60700	50200	45200	35800	20900	19000	26500	61500	71100
Alytaus	2	13000	10600	11400	5900	3200	2300	3200	12300	13400
Kauno	3	46700	40100	33100	28300	19100	12800	18000	39800	50700
Klaipėdos	4	22400	22000	21700	12500	12000	6400	11700	24300	32000
Marijampolės	5	9100	6500	5200	2300	1800	1300	2300	7800	11300
Panevėžio	6	18300	15400	15800	13700	9100	7800	6900	17300	27400
Šiaulių	7	24300	30500	19600	16700	10000	6400	8700	22200	27500
Tauragės	8	5100	6000	4900	3800	2100	1400	2500	5600	5700
Telšių	9	12900	10500	7500	6200	4900	3500	4400	11900	16000
Utenos	10	11800	12200	9200	5100	4600	3600	4100	7900	15400
Σ		224300	204000	173600	130300	87700	64500	88300	210600	270500

Apskritis	#	2011	2012	2013	2014	2015	2016	2017	2018
Vilniaus	1	62300	56000	41400	36200	32500	24500	21700	21400
Alytaus	2	12400	11000	12100	12900	11100	8600	7100	5300
Kauno	3	38400	31100	29900	26600	21000	19200	16000	17100
Klaipėdos	4	22100	15300	12200	13400	10400	8400	10300	7700
Marijampolės	5	9600	9700	10500	9300	7400	6900	6000	5000
Panevėžio	6	23900	18500	17500	15300	13400	12500	9800	7600
Šiaulių	7	24300	22600	20500	18000	17100	14800	12300	10100
Tauragės	8	5800	6500	4600	5300	4300	4000	4200	3500
Telšių	9	12800	9700	9300	8800	7300	8100	6500	4700
Utenos	10	16500	16400	14500	12300	9400	9100	9300	7600
Σ		228100	196800	172500	158100	133900	116100	103200	90000

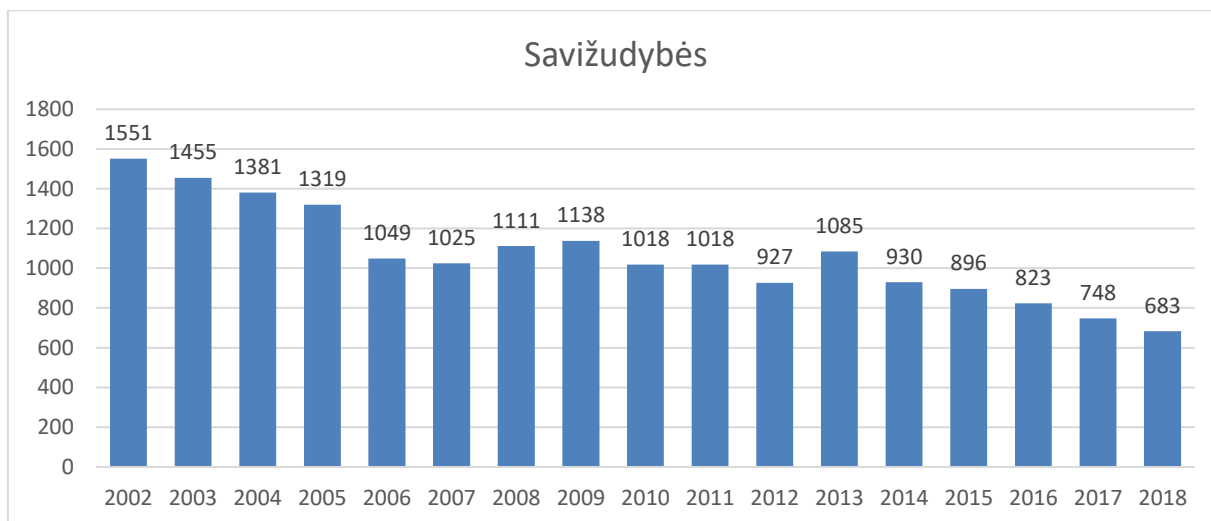


2 pav. Lietuvos bedarbių skaičius nuo 2002 iki 2018 metų

2 pav. pateiktoje diagramoje matome, kad nuo 2002 iki 2007 metų nedirbančių asmenų sumažėjo itin daug. Nuo 2007 metų prasidėjusi pasaulinė ekonominė krizė Lietuvoje taip pat pasijuto – bankrutavo ar stabdė veiklą daug įmonių, buvo atleidinėjami darbuotojai. Krizė pasiekė aukščiausią tašką 2010 metais, po to bedarbių ėmė mažėti.

5 lentelė. Savižudybių skaičiai apskrityse nuo 2002 iki 2018 metų

Apskritis	#	2002	2003	2004	2005	2006	2007	2008	2009	2010
Vilniaus	1	281	267	270	249	204	219	228	230	198
Alytaus	2	117	89	101	101	87	70	80	61	75
Kauno	3	288	254	250	242	186	204	194	202	196
Klaipėdos	4	152	152	128	128	115	99	91	99	99
Marijampolės	5	84	93	87	96	68	63	78	68	57
Panevėžio	6	169	145	124	107	97	101	105	117	114
Šiaulių	7	196	178	184	148	121	106	122	149	97
Tauragės	8	69	73	69	67	45	50	52	57	53
Telšių	9	89	106	76	81	70	48	58	69	46
Utenos	10	106	98	92	100	56	65	103	86	83
Σ		1551	1455	1381	1319	1049	1025	1111	1138	1018
Apskritis	#	2011	2012	2013	2014	2015	2016	2017	2018	
Vilniaus	1	145	164	215	160	186	178	162	168	
Alytaus	2	77	58	76	73	76	47	34	44	
Kauno	3	209	170	184	182	147	159	155	131	
Klaipėdos	4	93	86	95	83	71	82	56	65	
Marijampolės	5	80	67	86	71	69	55	52	40	
Panevėžio	6	102	106	104	99	78	86	81	72	
Šiaulių	7	137	117	134	107	99	92	79	64	
Tauragės	8	42	38	57	45	50	37	41	20	
Telšių	9	64	55	51	60	59	25	36	32	
Utenos	10	69	66	83	50	61	62	52	47	
Σ		1018	927	1085	930	896	823	748	683	



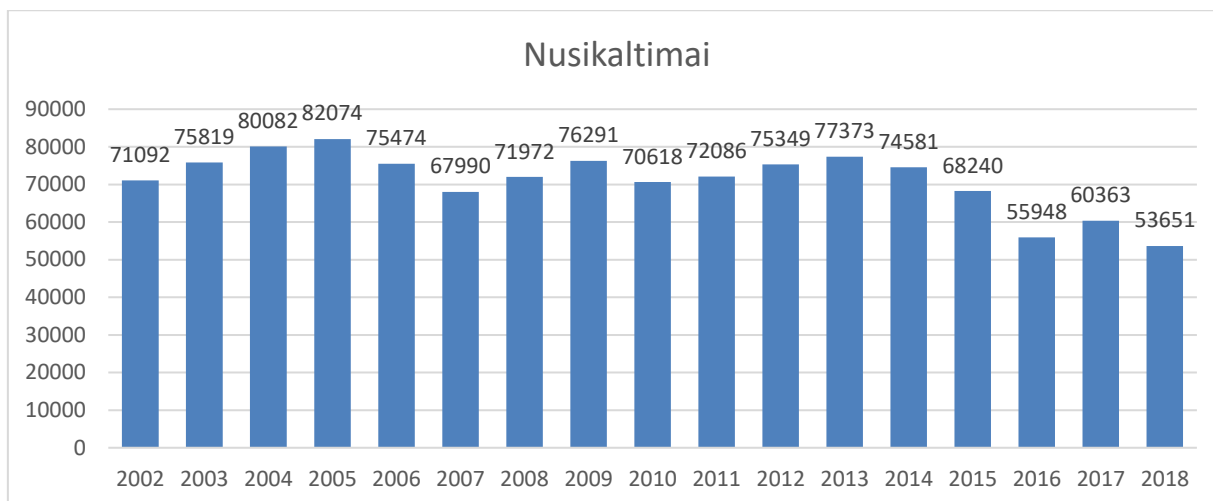
3 pav. Savižudybių skaičius Lietuvoje nuo 2002 iki 2018 metų

Savižudybių skaičius Lietuvoje tiriamu laikotarpiu mažėja, tai galime matyti 3 pav. Galimai savižudybių mažėjimui įtakos turi gerėjančios gyvenimo sąlygos šalyje. Savižudybės nuo 2013 iki 2018 mažėjo kaip ir nedarbas. Gali būti, kad sumažėjęs nedarbas turėjo įtakos ir lėmė savižudybių skaičiaus sumažėjimą.

6 lentelė. Nusikaltimų skaičiai apskrityse nuo 2002 iki 2018 metų

Apskritis	#	2002	2003	2004	2005	2006	2007	2008	2009	2010
Vilniaus	1	20737	24172	27600	28176	26749	24905	27556	28554	24885
Alytaus	2	2498	2503	2563	2731	2702	2233	2496	2661	2792
Kauno	3	14477	15233	16225	18736	16014	13752	14005	15343	14697
Klaipėdos	4	9476	9410	9600	9147	8044	7991	8059	8487	7576
Marijampolės	5	2877	2994	2635	2852	3105	2700	2898	2911	3056
Panevėžio	6	6192	6538	6968	6199	5562	4641	4790	4930	4633
Šiaulių	7	6608	6939	6974	6409	6053	5435	5274	6162	5829
Tauragės	8	2315	2106	1937	2072	2087	1770	2019	2158	2198
Telšių	9	2721	2635	2417	2500	2294	2068	2184	2275	2272
Utenos	10	3191	3289	3163	3252	2864	2495	2691	2810	2680
Σ		71092	75819	80082	82074	75474	67990	71972	76291	70618

Apskritis	#	2011	2012	2013	2014	2015	2016	2017	2018
Vilniaus	1	25014	26039	23863	22983	20928	17615	18349	17797
Alytaus	2	2792	3201	3204	3241	2930	2239	2245	2167
Kauno	3	14051	13533	15115	15260	14567	12396	14083	11625
Klaipėdos	4	8082	8322	9307	7884	7430	6237	6700	5369
Marijampolės	5	3351	3424	3793	3526	3322	2657	2773	2321
Panevėžio	6	4624	4801	5219	5820	4951	4580	4705	4089
Šiaulių	7	6352	7478	8213	7415	6703	4979	5209	4588
Tauragės	8	2129	2444	2494	2063	2147	1444	1505	1353
Telšių	9	2630	3267	3137	3420	2844	2014	2210	2095
Utenos	10	3061	2840	3028	2969	2418	1787	2584	2247
Σ		72086	75349	77373	74581	68240	55948	60363	53651



4 pav. Nusikaltimų skaičius Lietuvoje nuo 2002 iki 2018 metų

Nusikaltimų skaičiai Lietuvoje ryškių pokyčių turi nedaug, didžiausi buvo 2005 ir 2018 metais.

Su turimais duomenimis atliksime pradinę statistinę analizę. (Lietuvos gyventojų skaičius 7 lentelėje paimtas iš 4 priedo).

7 lentelė. Neigiamų faktorių, migrantų ir Lietuvos gyventojų skaičiai nuo 2002 iki 2018 metų

	2002	2003	2004	2005	2006
Lietuvos gyventojų skaičius	3454637	3431497	3398929	3355220	3289835
Vidiniai migrantai	39297	57825	59373	52969	51737
Nusikaltimai	71092	75819	80082	82074	75474
Bedarbiai	224300	204000	173600	130300	87700
Savižudybės	1551	1455	1381	1319	1049
	2007	2008	2009	2010	2011
Lietuvos gyventojų skaičius	3249983	3212605	3183856	3141976	3052588
Vidiniai migrantai	56583	56932	49943	49796	52199
Nusikaltimai	67990	71972	76291	70618	72086
Bedarbiai	64500	88300	210600	270500	228100
Savižudybės	1025	1111	1138	1018	1018
	2012	2013	2014	2015	2016
Lietuvos gyventojų skaičius	3003641	2971905	2943472	2921262	2888558
Vidiniai migrantai	60053	57243	61415	61426	68572
Nusikaltimai	75349	77373	74581	68240	55948
Bedarbiai	196800	172500	158100	133900	116100
Savižudybės	927	1085	930	896	823
	2017	2018			
Lietuvos gyventojų skaičius	2847904	2808901			
Vidiniai migrantai	69417	76176			
Nusikaltimai	60363	53651			
Bedarbiai	103200	90000			
Savižudybės	748	683			

Kiekvienai šių duomenų eilutei surasime vidurkį, standartinį nuokrypį ir medianą. Kaip tai daroma bus pademonstruota su pirma duomenų eilute (Lietuvos gyventojų skaičius nuo 2002 metų iki 2018 metų), kuri pažymima X .

Pirmiausia randame variacinę eilutę:

2808901; 2847904; 2888558; 2921262; 2943472; 2971905; 3003641; 3052588; 3141976;
3183856; 3212605; 3249983; 3289835; 3355220; 3398929; 3431497; 3454637.

Remiantis tuo, gauname, kad

Mediana: $M_e = 3141976$.

Vidurkis: $\bar{X} = \frac{1}{17} \sum_{i=1}^{17} X_i = 3126868,765$.

Dispersija: $S^2 = \frac{1}{17} \sum_{i=1}^{17} (X_i - \bar{X})^2 = 43251463397$.

Standartinis nuokrypis: $S = \sqrt{S^2} = 207969,862$.

Analogiškai apskaičiuosime kitas statistines charakteristikas. Apskaičiuoti duomenys pateikiami lentelėje:

8 lentelė. Statistinės charakteristikos

	Mediana	Vidurkis	Dispersija	Standartinis nuokrypis
Lietuvos gyventojų skaičius	3141976	3126868,765	43251463397	207969,862
Vidiniai migrantai	57243	57703,294	69673417,384	8347,060
Nusikaltimai	72086	71117,824	59301138,263	7700,723
Bedarbiai	158100	156029,412	3412643253	58417,833
Savižudybės	1025	1068,059	55143,232	234,826

Išvada: Nagrinėjamu laikotarpiu Lietuvoje kasmet buvo vidutiniškai 3,126 milijonai gyventojų. Kiekvienais metais šalies viduje migruodavo 57,703 tūkstančiai gyventojų. Nusikaltimų kasmet įvykdyta vidutiniškai 71,117 tūkstančiai. Darbo neturinčių žmonių vidutiniškai 156 tūkstančiai kasmet. Šalyje užfiksuojama vidutiniškai 1068 savižudybės kiekvienais metais.

2.1.1 DUOMENŲ NORMALUMAS

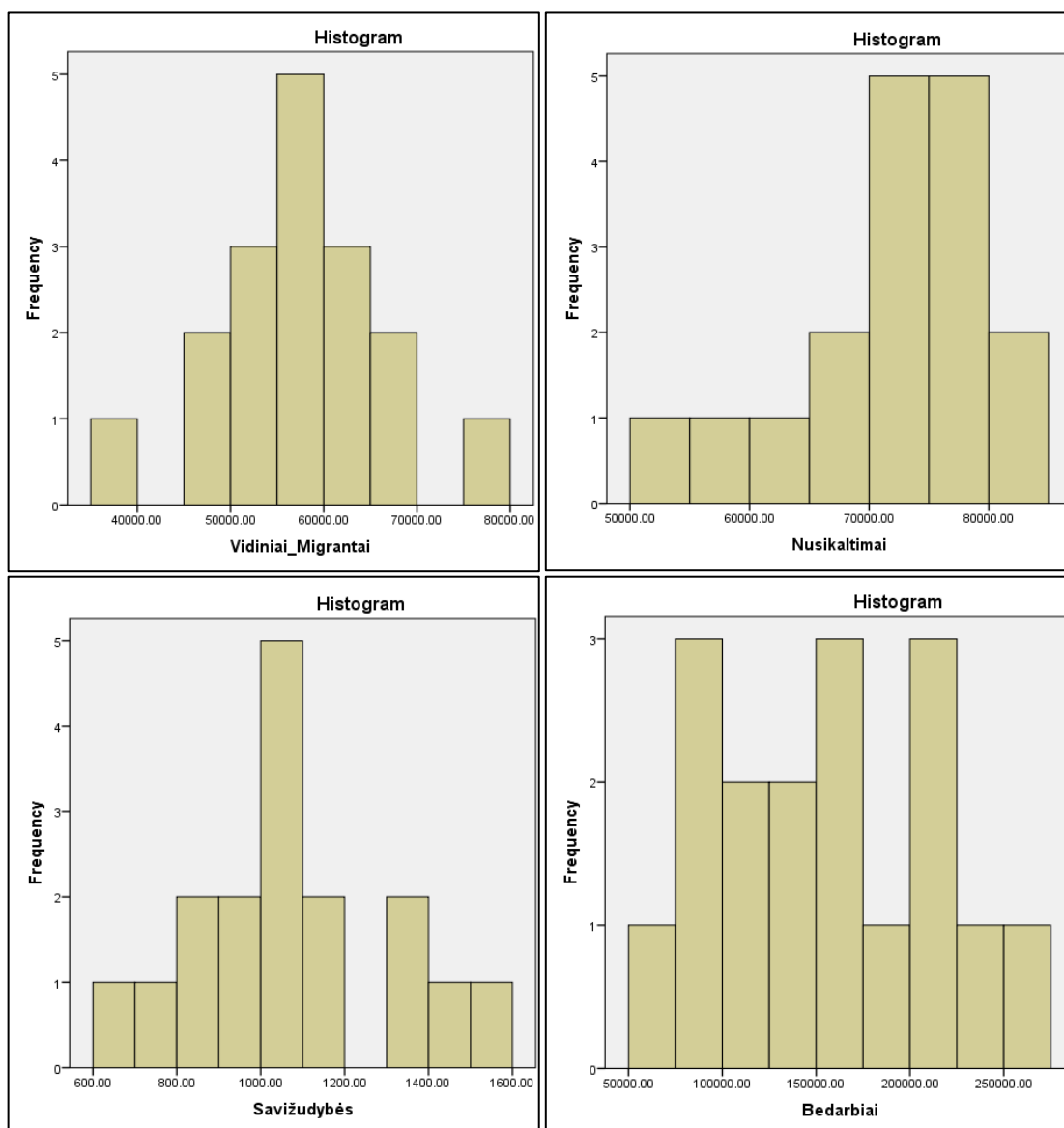
Programos SPSS pagalba patikrinsime ar vidinių migrantų ir neigiamų faktorių duomenys pasiskirstę normaliai.

9 lentelė. Duomenų normalumo tikrinimo rezultatai

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Vidiniai_Migrantai	.156	17	.200*	.966	17	.741
Nusikaltimai	.181	17	.143	.898	17	.063
Savižudybės	.151	17	.200*	.956	17	.554
Bedarbiai	.114	17	.200*	.957	17	.576

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Gauti 9 lentelės rezultatai rodo, kad vidinių migrantų, nusikaltimų, savižudybių ir bedarbių duomenys yra pasiskirstę normaliai, nes Šapiro – Vilko “Sig.” reikšmės yra didesnės už 0,05. Iš pateiktų grafikų duomenų normalumas taip pat matosi.

2.1.2 DUOMENŲ NEPRIKLAUSOMUMAS IR ATSITIKTINUMAS

Naudodami serijų kriterijų patikrinsime H_0 : duomenys nepriklausomi ir atsitiktiniai. Pradėsime nuo vidinių migrantų duomenų:

39297; 57825; 59373; 52696; 51737; 56583; 56932; 49943; 49796; 52199; 60053; 57243; 61415; 61426; 68572; 69417; 76176.

Surandame medianą: $M_e = X_{\frac{n+1}{2}} = 57243$.

Lyginame

(-)(+ +)(- - - - - -)(+ + + + +)

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$N = N(\text{serijų skaičius}) = 4$.

$k_1 = 8, k_2 = 8$.

$n_1 = 8, n_2 = 8$.

$\alpha = 0,1$.

$N_{0,05}(A, 8, 8) = 5$.

$N_{0,05}(V, 8, 8) = 13$.

Jeigu $N_{0,05}(A, 8, 8) < N < N_{0,05}(V, 8, 8)$, tada H_0 teisinga. Kadangi $5 \nless 4 < 13$ H_0 atmetame. Tai reiškia, kad duomenys priklausomi ir neatsitiktiniai.

Toliau tą patį procesą analogiškai taikome kiekvienam faktoriui, rezultatus pateikiame 10 lentelėje.

10 lentelė. Hipotezės (duomenys nepriklausomi ir atsitiktiniai) tikrinimas faktoriams

Faktorius	H_0 priimta / atmesta
Bedarbiai	-
Nusikaltimai	+
Savižudybės	-

Vidinių migrantų, bedarbių ir savižudybių duomenys Lietuvos mastu yra priklausomi ir neatsitiktiniai, tai reiškia, kad galima taikyti laiko eilučių teoriją.

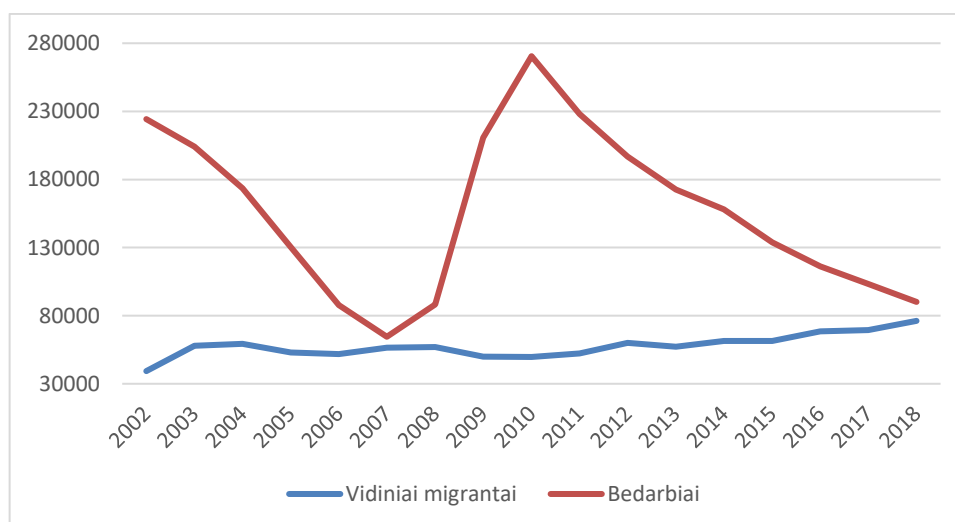
2.1.3 PRIKLAUSOMYBĖ TARP NEIGIAMŲ FAKTORIŲ IR VIDINĖS MIGRACIJOS LIETUVOS MASTU, PAGAL 2002-2018 M. DUOMENIS

Norint ištirti priklausomybę tarp neigiamų faktorių ir vidinės migracijos Lietuvos mastu, surašome duomenis į lentelę, kad viskas būtų aiškiau.

11 lentelė. Migrantų ir neigiamų faktorių skaičiai 2002 – 2018 metais

	Y (Vidiniai migrantai)	X (Bedarbiai)	U (Nusikaltimai)	Z (Savižudybės)
2002	39297	224300	71092	1551
2003	57825	204000	75819	1455
2004	59373	173600	80082	1381
2005	52969	130300	82074	1319
2006	51737	87700	75474	1049
2007	56583	64500	67990	1025
2008	56932	88300	71972	1111
2009	49943	210600	76291	1138
2010	49796	270500	70618	1018
2011	52199	228100	72086	1018
2012	60053	196800	75349	927
2013	57243	172500	77373	1085
2014	61415	158100	74581	930
2015	61426	133900	68240	896
2016	68572	116100	55948	823
2017	69417	103200	60363	748
2018	76176	90000	53651	683

Pirmiausia matome bedarbių ir vidinių migrantų skaičius 2002-2018 metų laikotarpiu. Kad galėtume susidaryti geresnį vaizdą apie duomenis, sudarome diagramą:



5 pav. Vidinių migrantų ir bedarbių skaičius Lietuvoje nuo 2002 iki 2018 metų

Iš 5 pav. matome, kad vidinių migrantų duomenys yra ganėtinai tolygūs, o bedarbių skaičius atvirkščiai – turi ryškių nuosmūkių bei pakilimų. Patikrinsime ar yra statistinė priklausomybė tarp šių duomenų, rasime empirinį koreliacijos koeficientą pasinaudoję formule:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

$$\text{čia } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Duomenis iš 11 lentelės naudojame tolimesniems skaičiavimams.

$$\begin{aligned} \bar{Y} &= \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{17} (39297 + 57825 + 59373 + 52969 + 51737 + 56583 + 56932 \\ &\quad + 49943 + 49796 + 52199 + 60053 + 57243 + 61415 + 61426 + 68572 \\ &\quad + 69417 + 76176) = 57703,294. \end{aligned}$$

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{17} (224300 + 204000 + 173600 + 130300 + 87700 + 64500 + 88300 \\ &\quad + 210600 + 270500 + 228100 + 196800 + 172500 + 158100 + 133900 \\ &\quad + 116100 + 103200 + 90000) = 156029,412. \end{aligned}$$

$$\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = -4521532947,059.$$

$$\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = 8289492120,835.$$

Tada viską surašome į formulę ir gauname

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{-4521532947,059}{8289492120,835} = -0,545.$$

Empirinį koreliacijos koeficientą vidinių migrantų bei bedarbių duomenims nuo 2002 iki 2018 metų gavome -0,545 ir tai rodo vidutinę priklausomybę tarp duomenų.

Norėdami visiškai įsitikinti rezultato teisingumu, tikriname atitinkamą statistinę hipotezę, kad koreliacija yra teigiama.

Tikriname statistinę hipotezę $H_0: \rho = 0$, imdami reikšmingumo lygmenį $\alpha=0,05$.

Sudarome statistiką:

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

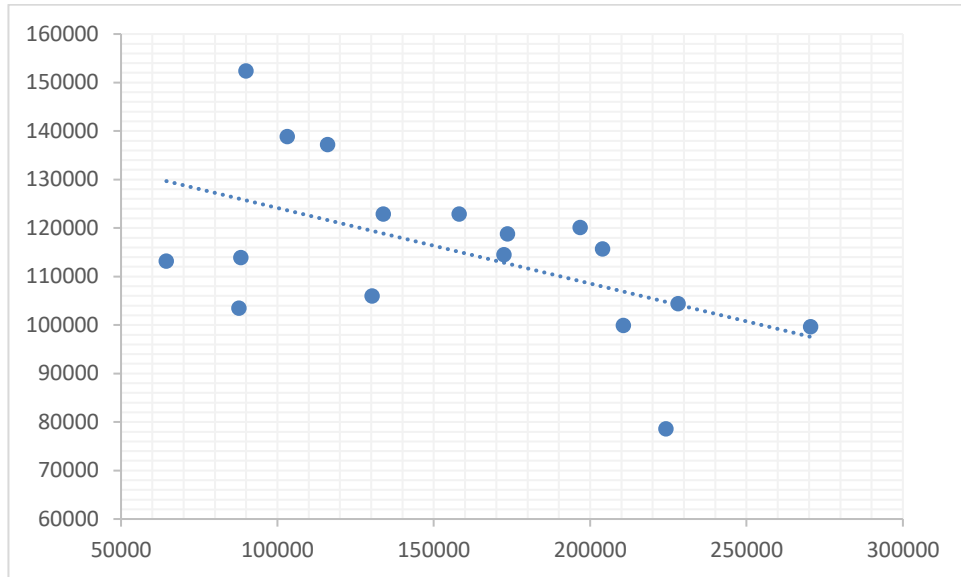
Įrašome duomenis:

$$T = \frac{-0,545\sqrt{17-2}}{\sqrt{1-(-0,545)^2}} = -2,551;$$

Kadangi $|T| = 2,551 > 2,131 = t_{0,975}(15)$, tai hipotezė H_0 atmetama.

Išvada: Priklausomybė tarp vidinių migrantų ir bedarbių, pagal 2002-2018 m. duomenis, yra vidutinė.

Toliau programos „Excel“ pagalba sudarome grafiką ir randame tiesinės regresijos lygtį.



6 pav. Vidinių migrantų ir bedarbių skaičius Lietuvoje nuo 2002 iki 2018 metų

Iš 6 pav. galime matyti, kad priklausomybė tarp migrantų ir bedarbių nėra stipri, nes didžioji dalis taškų išsibarstę toli nuo tiesės.

Tiesinę priklausomybę aprašome tokia lygtimi:

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x + \varepsilon, \varepsilon \sim N(0, \hat{\sigma}^2), \hat{\sigma}^2 = S^2.$$

$$n = 17.$$

$$\bar{X} = 156029,412.$$

$$\bar{Y} = 57703,294.$$

$$Q_X = \sum_{i=1}^n (X_i - \bar{X})^2 = 3412643253.$$

$$Q_{X,Y} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = -265972526.$$

$$\widehat{\beta}_1 = \frac{Q_{X,Y}}{Q_X} = -0,078.$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \cdot \bar{X} = 69863,820.$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 \cdot X_i)^2 = 55470106,316.$$

$$S = \sqrt{S^2} = 7447,826.$$

Apskaičiavę gauname tiesinės regresijos lygtį

$$y = 69863,820 - 0,078x + \varepsilon, \varepsilon \sim N(0; 5547016,316).$$

Analogiškai atlikę skaičiavimus gauname empirinius koreliacijos tarp visų charakteristikų. Rezultatus pateikiame 12 lentelėje.

12 lentelė. Empiriniai koreliacijos koeficientai tarp vidinių migrantų ir neigiamų faktorių duomenų

	X (Bedarbiai)	U (Nusikaltimai)	Z (Savižudybės)
Y (Vidiniai migrantai)	-0,545	-0,645	-0,728
X (Bedarbiai)		0,392	0,418
U (Nusikaltimai)			0,690

Iš 12 lentelės matome, kad didžiausias empirinis koreliacijos koeficientas yra tarp savižudybių ir vidinių migrantų $r = -0,728$.

Kadangi empirinis koreliacijos koeficientas yra pakankamai didelis, patikriname statistinę hipotezę $H_0: \rho = -0,89$, imame reikšmingumo lygmenį $\alpha = 0,05$. Panaudosime statistiką:

$$Arthr = \frac{1}{2} \ln \frac{1+r}{1-r}.$$

Irašome duomenis

$$Arthr = \frac{1}{2} \ln \frac{1 - 0,728}{1 + 0,728} = -0,924$$

$$Arth(-0,89) = \frac{1}{2} \ln \frac{1 - 0,89}{1 + 0,89} = -1,422$$

$$Z = \sqrt{n-3} (Arthr - Arth(-0,9))$$

$$Z = \sqrt{17-3} \cdot (-0,924 + 1,422) = 1,861$$

Kadangi $|Z| = 1,861 \leq 1,96 = u_{0,975}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = -0,89$ teisingumui.

Išvada: priklausomybė tarp vidinių migrantų ir savižudybių pagal 2002-2018 metų duomenis Lietuvos mastu yra didelė.

Randame tiesinės regresijos lygtį tarp vidinių migrantų ir savižudybių:

$$\bar{Z} = 1068,059.$$

$$\bar{Y} = 57703,294.$$

$$Q_Z = \sum_{i=1}^n (Z_i - \bar{Z})^2 = 55143,232.$$

$$Q_{Z,Y} = \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) = -1427527,606.$$

$$\widehat{\beta}_1 = \frac{Q_{Z,Y}}{Q_Z} = -25,888.$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \cdot \bar{Z} = 85352,806.$$

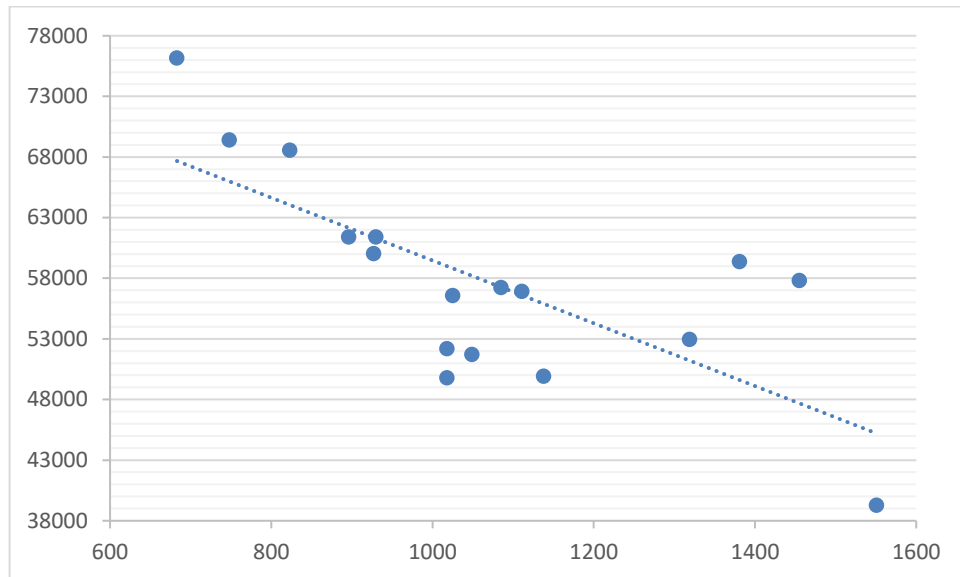
$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 \cdot Z_i)^2 = 37080525,037.$$

$$S = \sqrt{S^2} = 6089,378.$$

Apskaičiavę gauname tiesinės regresijos lygtį

$$y = 85352,806 - 25,888z + \varepsilon, \varepsilon \sim N(0; 37080525,037).$$

Kadangi empirinis koreliacijos koeficientas gautas didelis, galima teigti, kad tiesinė priklausomybė tarp šių duomenų didelė. Šitie duomenys pavaizduoti 7 paveiksle.



7 pav. Vidinių migrantų ir savižudybių skaičius Lietuvoje nuo 2002 iki 2018 metų

Didėjant savižudybių skaičiui, vidinė migracija mažėja. Iš tiesinės regresijos lygties matome, kad koeficientas prie z (savižudybių skaičius) yra $-25,888$ – kiekviena savižudybė tiek sumažina vidinių migrantų skaičių pridėjus konstantą. 7 pav. esanti regresijos kreivė rodo, kad kuo didesnis savižudybių skaičius, tuo mažiau vidinių migrantų.

Empirinis koreliacijos koeficientas vidinių migrantų ir nusikaltimų duomenims gautas $r = -0,645$, tai yra didesnis nei vidutinio stiprumo neigiamas ryšys (vienam didėjant – mažėja kitas). Įsitikinkime gautu rezultatu tikrindami hipotezę $H_0: \rho = 0$, kad populiacijos koreliacijos koeficientas lygus 0. Reikšmingumo lygmenį imame $\alpha = 0,05$.

Sudarome statistiką

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

$$T = \frac{-0,645\sqrt{17-2}}{\sqrt{1-(-0,645)^2}} = -3,269;$$

Kadangi $|T| = 3,269 > 2,131 = t_{0,975}(15)$, tai hipotezė H_0 atmetama.

Išvada: Priklausomybė tarp vidinių migrantų ir nusikaltimų, pagal 2002-2018 m. duomenis, yra vidutinė.

Randame tiesinės regresijos lygtį:

$$\bar{U} = 71117,824.$$

$$\bar{Y} = 57703,294.$$

$$Q_U = \sum_{i=1}^n (U_i - \bar{U})^2 = 59301138,263.$$

$$Q_{U,Y} = \sum_{i=1}^n (U_i - \bar{U})(Y_i - \bar{Y}) = -41460755,654.$$

$$\widehat{\beta}_1 = \frac{Q_{U,Y}}{Q_U} = -0,699.$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \cdot \bar{U} = 107425,758.$$

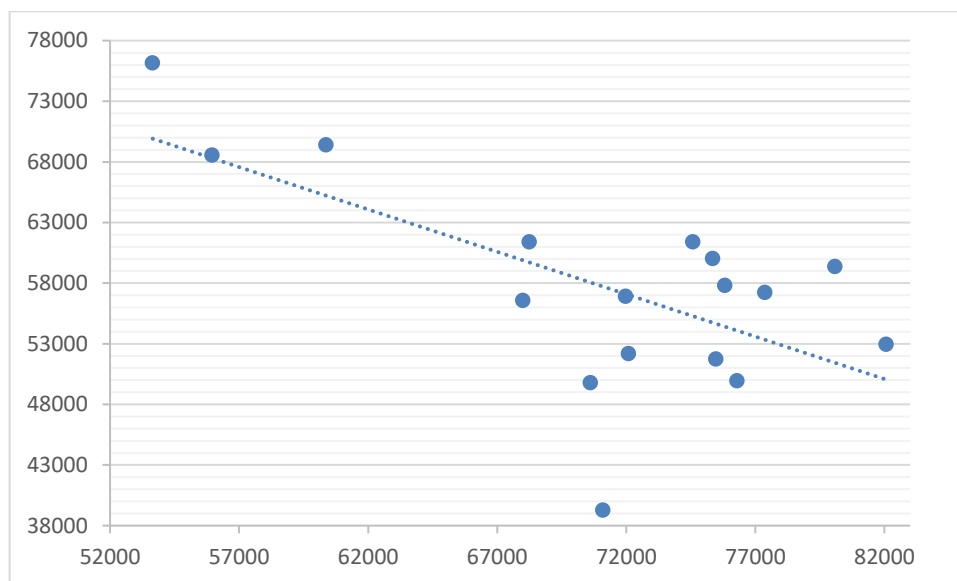
$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 \cdot U_i)^2 = 46110658,322.$$

$$S = \sqrt{S^2} = 6790,483.$$

Apskaičiavę gauname tiesinės regresijos lygtį

$$y = 107425,758 - 0,699u + \varepsilon, \varepsilon \sim N(0; 46110658,322).$$

Sudarome nusikaltimų ir vidinių migrantų grafiką:



8 pav. Vidinių migrantų ir nusikaltimų skaičius Lietuvoje nuo 2002 iki 2018 metų

Taip pat galime parašyti daugiamatę tiesinės regresijos lygtį programos SPSS pagalba.

Iš 12 lentelės žinome, kad vidinių migrantų ir neigiamų faktorių koreliacijos koeficientai yra didesni nei vidutiniai. Tačiau tarp savižudybių ir nusikaltimų koreliacijos koeficientas gautas $r=0,69$, o tai gali kelti problemų norint sudaryti tinkamą daugiamatės tiesinės regresijos lygtį, nes rekomenduojama, kad nepriklausomi faktoriai tarpusavyje turėtų ne didesnę koreliacijos koeficientą nei $|r|=0,7$. Panagrinėsime regresijos skaičiavimo rezultatus.

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.791 ^a	.625	.539	5841.71428	1.206

a. Predictors: (Constant), Bedarbiai, Nusikaltimai, Savižudybės

b. Dependent Variable: Vidiniai_Migrantai

Matome, kad „R Square“ yra 0,625, tai reiškia, kad pagal turimus duomenis 62,5% vidinių migrantų pokyčių susiję su nusikaltimų, savižudybių bei bedarbių skaičių pokyčiais. Regresijos paklaida yra labai didelė 5841,714, tai jau verčia abejoti modelio tinkamumu. Durbino-Watsono statistika 1,206 – galimai duomenų liekamosios paklaidos koreliuoja (yra autokoreliacija).

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	740814961.380	3	246938320.460	7.236	.004 ^b
	Residual	443633134.149	13	34125625.704		
	Total	1184448095.529	16			

a. Dependent Variable: Vidiniai_Migrantai

b. Predictors: (Constant), Savižudybės, Bedarbiai, Nusikaltimai

“ANOVA” lentelės rezultatai rodo, kad modelis iš dallies tinkamas prognozavimui “Sig.” reikšmė mažesnė už 0,05.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	98249.386	13651.930	7.197	.000		
	Bedarbiai (x)	-.038	.027	-.264	.186	.805	1.243
	Nusikaltimai (u)	-.239	.257	-.220	.371	.511	1.955
	Savižudybės (z)	-16.565	8.545	-.466	.075	.499	2.006

a. Dependent Variable: Vidiniai_Migrantai

Pagal koeficientų lentelę gauta lygtis atrodo taip:

$$y = 98249,386 - 0,38x - 0,239u - 16,565z.$$

Matome, kad pagal gautą lygtį didžiąsą įtaką vidiniams migrantams turi savižudybių skaičius, koeficientas (-16,565). Bedarbių ir nusikaltimų koeficientai gerokai mažesni. Deja ši daugiamatė tiesinės regresijos lygtis nėra taikytina, nes priimtos visos nulinės hipotezės apie

koeficientų lygybę nuliui (Sig. stulpelio reikšmės koeficientams yra $> 0,05$). Standartizuotų koeficientų skiltyje taip pat galima matyti, kad svarbiausias rodiklis modelyje yra savižudybių skaičius – gautas koeficientas $-0,466$ daugiau nei dvigubai didesnis už nusikaltimų kuris yra $-0,22$ ir daugiau nei pusantro karto didesnis už bedarbių koeficientą $-0,264$. Darome išvadą, kad prognozavimui tinka jau anksčiau apskaičiuota tiesinės regresijos lygtis tarp vidinių migrantų ir savižudžių:

$$y = 85352,806 - 25,888z + \varepsilon, \varepsilon \sim N(0; 37080525,037).$$

Atliekame tyrimą su “SPSS”

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.728 ^a	.530	.499	6089.37805	1.001

a. Predictors: (Constant), Savižudybės

b. Dependent Variable: Vidiniai_Migrantai

“R Square” rodo, kad 53% vidinių migrantų pokyčių susiję su savižudybių pokyčiais. Nors paklaida yra didelė, tolimesni rezultatai parodys, kad modelis yra tinkamas.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	628240219.978	1	628240219.978	16.943	.001 ^b
	Residual	556207875.551	15	37080525.037		
	Total	1184448095.52	16			
		9				

a. Dependent Variable: Vidiniai_Migrantai

b. Predictors: (Constant), Savižudybės

“ANOVA” lentelės rezultatai rodo, kad modelis reikšmingas – “Sig.” $< 0,05$ ir yra gerokai reikšmingesnis už anksčiau gautą rezultatą kai buvo ieškoma regresijos lygtis su visais neigiamais faktoriais.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	85352.806	6877.784		12.410	.000		
Savižudybės	-25.888	6.289	-.728	-4.116	.001	1.000	1.000

a. Dependent Variable: Vidiniai_Migrantai

Galima matyti, kad savižudybių skaičius yra statistiškai reikšmingas šiame modelyje - Sig. reikšmė <0,05, VIF < 4.

Pažiūrėkime į Kuko mato reikšmes, kurios esant tinkamam modeliui turėtų būti ≤ 1 .

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	45201.0898	67671.5547	57703.2941	6266.18016	17
Std. Predicted Value	-1.995	1.591	.000	1.000	17
Standard Error of Predicted Value	1480.729	3377.392	2006.329	598.435	17
Adjusted Predicted Value	44850.9609	65314.7852	57474.1174	6203.04651	17
Residual	-9203.19824	10138.69629	.00000	5896.01494	17
Std. Residual	-1.511	1.665	.000	.968	17
Stud. Residual	-1.560	1.883	.018	1.054	17
Deleted Residual	-9806.25000	12974.03809	229.17671	7012.61622	17
Stud. Deleted Residual	-1.647	2.082	.036	1.109	17
Mahal. Distance	.005	3.981	.941	1.180	17
Cook's Distance	.000	.496	.102	.155	17
Centered Leverage Value	.000	.249	.059	.074	17

a. Dependent Variable: Vidiniai_Migrantai

Kuko mato reikšmės atitinka reikalavimus. Lieka patikrinti standartizuotų liekanų normalumą.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	.136	17	.200*	.945	17	.385

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Šapiro-Vilko testo rezultatai rodo, kad standartizuotos liekanos yra pasiskirsčiusios normaliai. Galime teigti, kad tiesinės regresijos modelis nusakantis vidinių migrantų skaičių pokyčius kintant savižudžių skaičiams modelis yra tinkamas.

Patikriname tiesinės regresijos modelį 2005 metų duomenims. Naudojame skaičius iš 11 lentelės:

$$y = 85352,806 - 25,888z + \varepsilon,$$

$$y = 85352,806 - 25,888 \cdot 1319 + \varepsilon,$$

Gauname

$$y = 51206,534 + \varepsilon$$

Išvada: 2005 metais vidinių migrantų skaičius Lietuvoje buvo 51737. Įrašius 2005 metų savižudžių skaičių Lietuvoje į gautą lygtį gavome $y = 51206,534 + \varepsilon$. Matosi, kad su paklaida lygtis yra teisinga.

2.2 DUOMENŲ APSKRITYSE GRUPAVIMAS

Perėjus į apskričių lygį jau galime matyti atvykusius, išvykusius ir bendrą judėjimo skaičių kiekvienoje apskrityje (žr. 1 priedą). Kad galėtume lengviau tirti duomenis, sugrupuosime 10 apskričių į 3 grupes.

Apskritis grupuojame pagal atvykusių į ją skaičių. Atimdami viso laikotarpio vidurkį kiekvienai apskričiai iš atvykusiųjų skaičiaus gauname tokią lentelę:

12 lentelė. Atvykusių į apskritį skirtumas nuo tos apskrities vidurkio

	Apskrities nr.									
	1	2	3	4	5	6	7	8	9	10
2002	-6356,412	-864,882	-3993,941	-3641,235	-779,235	-345,941	-1405,706	-316,824	-558,706	-143,412
2003	-425,412	392,118	-602,941	-588,235	-57,235	770,059	-212,706	99,176	440,294	306,588
2004	-512,412	685,118	-479,941	-266,235	100,765	563,059	245,294	337,176	558,294	438,588
2005	-2047,412	533,118	-1141,941	-924,235	-156,235	-354,941	-492,706	-9,824	-14,706	-125,412
2006	-3172,412	-89,882	-1274,941	-767,235	22,765	-293,941	-310,706	41,176	186,294	-307,412
2007	-1643,412	-113,882	-242,941	145,765	126,765	-43,941	460,294	142,176	373,294	-324,412
2008	-1176,412	-122,882	350,059	391,765	-86,235	-271,941	76,294	-141,824	282,294	-72,412
2009	-2577,412	-378,882	-1501,941	-911,235	-281,235	-307,941	-502,706	-272,824	-464,706	-561,412
2010	-2684,412	-259,882	-1968,941	-1165,235	-270,235	-344,941	-488,706	-427,824	-320,706	23,588
2011	-1736,412	70,118	-1484,941	-874,235	-213,235	-462,941	-145,706	-177,824	-392,706	-86,412
2012	2605,588	-79,882	-133,941	-597,235	33,765	156,059	0,294	317,176	38,294	9,588
2013	725,588	-286,882	148,059	-334,235	-129,235	5,059	-129,706	-32,824	-520,706	94,588
2014	1355,588	-42,882	1364,059	858,765	116,765	59,059	90,294	10,176	-206,706	106,588
2015	2608,588	-20,882	1068,059	567,765	-51,235	-170,941	153,294	-100,824	-322,706	-8,412
2016	4529,588	141,118	2420,059	1845,765	127,765	244,059	749,294	208,176	383,294	219,588
2017	4441,588	-30,882	3084,059	2556,765	660,765	126,059	611,294	102,176	75,294	86,588
2018	6065,588	470,118	4392,059	3702,765	834,765	674,059	1302,294	223,176	464,294	343,588
Spalvų reikšmės	((-6356) - (-4286))		((-4286) - (-1923))		((-1923) - 101]		(101-2171]	(2171 - 4242]	(4242-6312]	

Sugrupuojame apskritis pagal tai, kiek skirtingų spalvų jos turi. Į I grupę patenka apskritys turinčios (1-2] spalvas, į II grupę (2-4] spalvas, o į III (4-6] spalvas.

Pagal tokį grupavimą turime:

I grupė – 2 (Alytus), 5 (Marijampolė), 6 (Panevėžys), 7 (Šiauliai), 8 (Tauragė), 9 (Telšiai), 10 (Utena),

II grupė – 4 (Klaipėda),

III grupė – 1 (Vilnius), 3 (Kaunas).

Po to sudėjome kiekvienos grupės rodiklius ir gauname naujas duomenų lenteles (žr. 5 priedą).

Kiekvieną iš gautų grupių charakteristikų sunormuojame pagal grupės vidutinį gyventojų skaičių tenkantį 100 tūkst. gyventojų. Turime tokius normavimo dydžius:

Grupė	I	II	III
Normavimo dydis	226,320	59,239	246,008

Po normavimo turime naujus grupių duomenis, kuriuos naudosime tolimesniems skaičiavimams.

13 lentelė. I grupės sunormuoti duomenys

Sunormuota I Grupė						
	Atvykę (X1)	Išvykę (Y1)	Bendras judėjimas (XY1)	Nusikaltimai (N1)	Savižudybės (S1)	Bedarbiai (B1)
2002	86,258	91,388	177,646	116,658	3,667	417,550
2003	113,445	128,053	241,498	119,318	3,455	405,178
2004	118,703	136,908	255,611	117,784	3,239	325,203
2005	103,022	121,302	224,324	114,948	3,093	237,274
2006	102,443	117,992	220,435	108,992	2,404	157,741
2007	108,506	128,442	236,947	94,300	2,223	116,207
2008	104,277	127,094	231,371	98,763	2,642	141,834
2009	93,527	113,565	207,091	105,633	2,682	375,574
2010	96,536	113,344	209,880	103,658	2,320	515,641
2011	99,540	117,183	216,724	110,193	2,523	465,270
2012	107,865	134,990	242,855	121,310	2,240	417,108
2013	101,348	126,255	227,602	128,526	2,611	393,248
2014	106,354	129,856	236,210	125,724	2,231	361,877
2015	103,460	129,065	232,524	111,855	2,174	309,296
2016	114,926	144,291	259,217	87,045	1,785	282,785
2017	112,973	142,855	255,828	93,810	1,657	243,902
2018	124,819	155,501	280,320	83,333	1,410	193,531

14 lentelė. II grupės sunormuoti duomenys

Sunormuota II Grupė						
	Atvykę (X2)	Išvykę (Y2)	Bendras judėjimas (XY2)	Nusikaltimai (N2)	Savižudybės (S2)	Bedarbiai (B2)
2002	66,240	64,907	131,147	159,962	2,566	378,130
2003	117,777	108,628	226,405	158,848	2,566	371,377
2004	123,213	112,764	235,977	162,056	2,161	366,313
2005	112,105	101,386	213,491	154,409	2,161	211,010
2006	114,756	106,720	221,476	135,789	1,941	202,569
2007	130,168	116,022	246,189	134,894	1,671	108,037
2008	134,320	118,672	252,992	136,042	1,536	197,505
2009	112,325	102,720	215,044	143,267	1,671	410,203
2010	108,037	95,410	203,447	127,889	1,671	540,185
2011	112,949	102,770	215,720	136,431	1,570	373,065
2012	117,625	109,388	227,013	140,482	1,452	258,276
2013	122,065	108,729	230,794	157,109	1,604	205,946
2014	142,204	124,378	266,581	133,088	1,401	226,203
2015	137,291	124,175	261,467	125,424	1,199	175,560
2016	158,865	146,762	305,627	105,285	1,384	141,799
2017	170,867	153,953	324,820	113,101	0,945	173,872
2018	190,213	171,441	361,654	90,633	1,097	129,982

15 lentelė. III grupės sunormuoti duomenys

Sunormuota III Grupė						
	Atvykę (X3)	Išvykę (Y3)	Bendras judėjimas (XY3)	Nusikaltimai (N3)	Savižudybės (S3)	Bedarbiai (B3)
2002	64,433	60,035	124,467	143,141	2,313	436,570
2003	102,326	91,090	193,416	160,177	2,118	367,061
2004	102,472	88,241	190,713	178,144	2,114	318,282
2005	93,541	79,306	172,848	190,693	1,996	260,560
2006	88,428	76,058	164,486	173,827	1,585	162,596
2007	98,838	83,904	182,742	157,137	1,719	129,264
2008	103,147	85,924	189,071	168,941	1,715	180,888
2009	89,924	73,802	163,726	178,437	1,756	411,774
2010	87,590	75,168	162,759	160,897	1,602	495,105
2011	93,411	79,631	173,043	158,795	1,439	409,335
2012	116,553	93,582	210,135	160,856	1,358	354,053
2013	110,057	90,355	200,412	158,442	1,622	289,827
2014	117,561	100,232	217,793	155,454	1,390	255,276
2015	121,451	101,053	222,505	144,284	1,354	217,472
2016	134,756	110,655	245,410	121,992	1,370	177,636
2017	137,097	113,679	250,776	131,833	1,289	153,247
2018	149,015	125,309	274,324	119,598	1,215	156,499

Toliau skaičiavimus atliksime šioms sunormuotiems grupių duomenims.

Su reikšmingumo lygmeniu $\alpha = 0,1$ patikrinsime hipotezę H_0 : duomenys nepriklausomi ir atsitiktiniai. Rezultatus pateikiame 16 lentelėje.

16 lentelė. H_0 tikrinimas grupių duomenims

	H_0 : duomenys nepriklausomi ir atsitiktiniai. Priimta / atmesta		
Grupė	I	II	III
Atvykę	+	+	-
Išvykę	+	+	-
Bendras judėjimas	+	+	-
Nusikaltimai	-	+	-
Savižudybės	+	-	+
Bedarbiai	-	+	-

Programos „SPSS“ pagalba patikriname duomenų normalumą grupių duomenims, su reikšmingumo lygmeniu $\alpha = 0,05$. Rezultatus pateikiame 17 lentelėje.

17 lentelė. Duomenų normalumas grupėse

	Teigiama / neigiama		
Grupė	I	II	III
Atvykę	+	+	+
Išvykę	+	+	+
Bendras judėjimas	+	+	+
Nusikaltimai	+	+	+
Savižudybės	+	+	+
Bedarbiai	+	+	+

Matome, kad visi grupių duomenys yra pasiskirstę normaliai. Įdomu patikrinti priklausomybę tarp grupių duomenų. Sekančiose lentelėse, kuriose skaičiuojamas empirinis koreliacijos koeficientas įvedame spalvas, kad lengviau matytusi priklausomybių didumas.

0 – 0,49	0,50 – 0,79	0,80 – 1
----------	-------------	----------

Paskaičiavę empirinį koreliacijos koeficientą grupių duomenims gavome tokius rezultatus:

18 lentelė. Empiriniai koreliacijos koeficientai grupėse

I grupė					
	Išvykę (Y1)	Bendras judėjimas (XY1)	Nusikaltimai (N1)	Savižudybės (S1)	Bedarbiai (B1)
Atvykę (X1)	0,936	0,976	-0,383	-0,459	-0,415
Išvykę (Y1)		0,990	-0,434	-0,678	-0,405
Bendras judėjimas (XY1)			-0,420	-0,601	-0,416
Nusikaltimai (N1)				0,627	0,531
Savižudybės (S1)					0,325
II grupė					
	Išvykę (Y2)	Bendras judėjimas (XY2)	Nusikaltimai (N2)	Savižudybės (S2)	Bedarbiai (B2)
Atvykę (X2)	0,996	0,999	-0,807	-0,766	-0,635
Išvykę (Y2)		0,999	-0,814	-0,745	-0,640
Bendras judėjimas (XY2)			-0,811	-0,757	-0,638
Nusikaltimai (N2)				0,798	0,458
Savižudybės (S2)					0,487
III grupė					
	Išvykę (Y3)	Bendras judėjimas (XY3)	Nusikaltimai (N3)	Savižudybės (S3)	Bedarbiai (B3)
Atvykę (X3)	0,991	0,998	-0,643	-0,739	-0,605
Išvykę (Y3)		0,997	-0,674	-0,689	-0,604
Bendras judėjimas (XY3)			-0,658	-0,718	-0,606
Nusikaltimai (N3)				0,522	0,319
Savižudybės (S3)					0,487

Gauti rezultatai rodo, kad didžiausios priklausomybės yra tarp judėjimo visose grupėse $|r| > 0,9$. I grupėje tarp savižudybių, išvykusių, bendro judėjimo ir nusikaltimų gauti stipresni nei vidutiniai koreliacijos koeficientai $|r| > 0,6$, o tarp bedarbių ir nusikaltimų gautas koreliacijos koeficientas arčiau vidutinio $r = 0,531$. II grupėje priklausomybė tarp visų judėjimo charakteristikų ir nusikaltimų stipri, koreliacijos koeficientai $|r| > 0,8$. Tarp savižudybių ir visų judėjimo charakteristikų bei nusikaltimų $|r| > 0,7$, o tarp bedarbių ir judėjimo charakteristikų $|r| > 0,6$. III grupėje judėjimo charakteristikų, bedarbių ir nusikaltimų bei savižudybių koreliacijos koeficientai šiek tiek didesni nei. SPSS pagalba randame tiesinės regresijos lygtis charakteristikoms, kai $|r| > 0,5$.

19 lentelė. Tiesinės regresijos lygtys grupėse

Grupė	Tiesinės regresijos lygtis
I	$X1 = 28,359 + (0,61) * Y1$
I	$X1 = 14,735 + (0,391) * XY1$
I	$Y1 = -14,734 + (0,609) * XY1$
I	$Y1 = 167,047 + (-16,095) * S1$
I	$XY1 = 158,583 + (62,879) * S1$
I	$N1 = 74,469 + (13,596) * S1$
I	$N1 = 89,579 + (0,06) * B1$
II	$X2 = -4,247 + (1,139) * Y2$
II	$X2 = -2,264 + (0,534) * XY2$
II	$X2 = 281,155 + (-1,127) * N2$
II	$X2 = 204,653 + (-45,743) * S2$
II	$X2 = 166,119 + (-0,146) * B2$
II	$Y2 = 2,264 + (0,466) * XY2$
II	$Y2 = 251,034 + (-0,993) * N2$
II	$Y2 = 181,267 + (-38,912) * S2$
II	$Y2 = 149,650 + (-0,129) * B2$
II	$XY2 = 532,189 + (-2,120) * N2$
II	$XY2 = 385,920 + (-84,655) * S2$
II	$XY2 = 315,769 + (-0,275) * B2$
II	$N2 = 78,739 + (34,136) * S2$
III	$X3 = -8,204 + (1,276) * Y3$
III	$X3 = -4,078 + (0,563) * XY3$
III	$X3 = 215,509 + (-0,696) * N3$
III	$X3 = 185,315 + (-47,925) * S3$
III	$X3 = 137,752 + (-0,111) * B3$
III	$Y3 = 4,078 + (0,437) * XY3$
III	$Y3 = 178,595 + (-0,566) * N3$
III	$Y3 = 146,989 + (-34,727) * S3$
III	$Y3 = 114,090 + (-0,086) * B3$
III	$XY3 = 394,102 + (-1,262) * N3$
III	$XY3 = 332,305 + (-82,653) * S3$
III	$XY3 = 251,842 + (-0,197) * B3$
III	$N3 = 105,190 + (31,280) * S3$

Taip pat apskaičiuojame empirinius koreliacijos koeficientus tarp grupių duomenų. Rezultatai pateikti 20 lentelėje.

20 lentelė. Empiriniai koreliacijos koeficientai tarp grupių

I ir II grupė						
	Atvykę	Išvykę	Bendras judėjimas	Nusikaltimai	Savižudybės	Bedarbiai
Atvykę	0,803	0,818	0,811	-0,462	-0,335	-0,490
Išvykę	0,920	0,927	0,924	-0,618	-0,614	-0,597
Bendras judėjimas	0,887	0,898	0,893	-0,565	-0,512	-0,563
Nusikaltimai	-0,605	-0,621	-0,613	0,796	0,473	0,354
Savižudybės	-0,820	-0,814	-0,818	0,918	0,925	0,542
Bedarbiai	-0,478	-0,478	-0,479	0,334	0,224	0,783
I ir III grupė						
	Atvykę	Išvykę	Bendras judėjimas	Nusikaltimai	Savižudybės	Bedarbiai
Atvykę	0,792	0,823	0,807	-0,368	-0,342	-0,577
Išvykę	0,936	0,932	0,937	-0,454	-0,594	-0,636
Bendras judėjimas	0,893	0,903	0,899	-0,426	-0,503	-0,622
Nusikaltimai	-0,432	-0,432	-0,433	0,533	0,395	0,489
Savižudybės	-0,783	-0,745	-0,768	0,607	0,939	0,561
Bedarbiai	-0,300	-0,301	-0,301	0,055	0,134	0,900
II ir III grupė						
	Atvykę	Išvykę	Bendras judėjimas	Nusikaltimai	Savižudybės	Bedarbiai
Atvykę	0,945	0,951	0,949	-0,737	-0,761	-0,621
Išvykę	0,946	0,954	0,951	-0,731	-0,704	-0,605
Bendras judėjimas	0,948	0,955	0,952	-0,736	-0,738	-0,615
Nusikaltimai	-0,583	-0,611	-0,597	0,733	0,517	0,379
Savižudybės	-0,716	-0,706	-0,712	0,814	0,935	0,455
Bedarbiai	-0,705	-0,706	-0,706	0,511	0,474	0,935

20 lentelėje galime matyti, kad daugiausiai žalios spalvos langelių, kai $|r| > 0,8$ yra tarp I ir III grupių charakteristikų, o tarp I ir III jų mažiausiai. Geltona spalva pažymėtų langelių daugiausia tarp II ir III grupių.

2.3 FAKTORIŲ PASISKIRSTYMAS GRUPĖSE

Sekančiam žingsiniui atlikti sudarysime lentelę, kurioje bus galima pamatyti neigiamų faktorių ir judėjimo charakteristikų procentinį pasiskirstymą ir jų pokyčius grupėse kas 4 metus.

21 lentelė. Faktorių procentinės dalys grupėse

Atvykę									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	39,76%	-6,24%	33,52%	-0,48%	33,04%	-3,99%	29,05%	-2,15%	26,90%
II	30,54%	7,01%	37,55%	-0,57%	36,98%	1,86%	38,84%	2,15%	40,99%
III	29,70%	-0,77%	28,93%	1,05%	29,98%	2,13%	32,11%	0,00%	32,11%
Σ	216,931		305,627		292,163		366,118		464,047
Išvykę									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	42,24%	-3,02%	39,23%	0,69%	39,92%	-3,29%	36,63%	-2,25%	34,38%
II	30,00%	5,48%	35,48%	-1,88%	33,60%	1,48%	35,09%	2,82%	37,91%
III	27,75%	-2,46%	25,29%	1,19%	26,47%	1,80%	28,28%	-0,57%	27,71%
Σ	216,329		300,771		283,922		354,466		452,251
Bendras judėjimas									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	41,00%	-4,65%	36,35%	0,08%	36,43%	-3,65%	32,78%	-2,19%	30,59%
II	30,27%	6,25%	36,52%	-1,21%	35,32%	1,68%	37,00%	2,47%	39,47%
III	28,73%	-1,60%	27,13%	1,13%	28,25%	1,97%	30,22%	-0,29%	29,94%
Σ	433,261		606,398		576,085		720,584		916,297
Nusikaltimai									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	27,79%	-1,75%	26,04%	0,38%	26,41%	3,94%	30,35%	-1,96%	28,39%
II	38,11%	-5,67%	32,44%	0,15%	32,59%	-0,46%	32,13%	-1,25%	30,87%
III	34,10%	7,42%	41,53%	-0,53%	41,00%	-3,47%	37,53%	3,21%	40,74%
Σ	419,761		418,608		392,444		414,267		293,564
Savižudybės									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	42,91%	-2,38%	40,53%	0,95%	41,48%	2,95%	44,43%	-6,56%	37,87%
II	30,02%	2,71%	32,74%	-2,85%	29,88%	-1,99%	27,90%	1,58%	29,48%
III	27,06%	-0,33%	26,73%	1,91%	28,64%	-0,96%	27,68%	4,97%	32,65%
Σ	8,546		5,930		5,592		5,023		3,722
Bedarbiai									
Grupė	2002	Δ	2006	Δ	2010	Δ	2014	Δ	2018
I	33,89%	-3,72%	30,17%	3,08%	33,25%	9,66%	42,91%	-2,59%	40,32%
II	30,69%	8,05%	38,74%	-3,91%	34,83%	-8,01%	26,82%	0,26%	27,08%
III	35,43%	-4,33%	31,09%	0,83%	31,92%	-1,65%	30,27%	2,33%	32,60%
Σ	1232,250		522,907		1550,931		843,355		480,012

Matematiškai patikriname, ar atvykusiųjų į I grupę procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2002$ į $t_2 = 2006$ yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,01$. Reikšmingumui nustatyti tikriname statistinę hipotezę $H_0^{(1)}: p_1 = 0,3976, p_2 = 0,3054, p_3 = 0,2970$. Skaičiai p_1, p_2, p_3 gaunami atvykusiųjų grupėse procentines dalis 2002 metais padalinus iš 100. 2006 m. p_i reikšmės pasikeitė į $\hat{p}_i$, čia $i = 1, 2, 3$. Turimu atveju 2006 m. duomenys bus $\hat{p}_1 = 0,3352, \hat{p}_2 = 0,3755, \hat{p}_3 = 0,2893$. Naujoje situacijoje $n = n(2006) = 305,627$ (visų atvykusių į grupes gyventojų skaičius tenkantis 100 tūkst. gyventojų). Viską surašome į Pirsono statistiką ir gauname

$$\chi^2 = 305,627 \left(\frac{(0,3352-0,3976)^2}{0,3976} + \frac{(0,3755-0,3054)^2}{0,3054} + \frac{(0,2893-0,2970)^2}{0,2970} \right) = 7,972.$$

Į formulę $\chi^2_{1-\alpha}(k-1)$ įrašome $k = 3$, ir $\alpha = 0,01$ gauname $\chi^2_{0,99}(2) = 13,8$ (reikšmė paimta iš kvantilių lentelės[8]). Mūsų atveju $\chi^2 = 7,972 < \chi^2_{0,99}(9) = 27,9$ todėl hipotezę $H_0^{(1)}$ apie tai, kad į I, II ir III grupes atvykusių gyventojų skaičius nepakito nuo 2002 iki 2006 metų, priimame. Tai reiškia, kad šitai charakteristikai per 4 metus pokyčiai nebuvo reikšmingi. Kitaip sakant, proporcijos grupėse po 4 metų stipriai nepakito. Kitoms charakteristikoms analogiškai atliekami skaičiavimai ir pateikiami 22 lentelėje.

22 lentelė. Hipotezės apie polinominį skirstinį taikymo rezultatai

$\alpha = 0,01$	H_0 priimta / atmesta			
Atvykę	+	+	+	+
Išvykę	+	+	+	+
Bendras judėjimas	+	+	+	+
Nusikaltimai	+	+	+	+
Savižudybės	+	+	+	+
Bedarbiai	-	+	-	+
$\alpha = 0,1$	H_0 priimta / atmesta			
Atvykę	-	+	+	+
Išvykę	+	+	+	+
Bendras judėjimas	-	+	+	+
Nusikaltimai	-	+	+	+
Savižudybės	+	+	+	+
Bedarbiai	-	-	-	+

Prieš tai atliktą tyrimą taip pat buvo galima atlikti panaudojus Kulbako-Leiblero informaciją tarp hipotezės tiriamų skirstinių. Jei ta informacija lygi nuliui, tai pagrindinėje hipotezėje ir alternatyvoje esantys skirstiniai (atitinkami tankiai arba diskretūs skirstiniai, priklausomai nuo to situacijos) sutampa. Jei Kulbako-Leiblero informacija artima nuliui, tai

skirstiniai esantys hipotezėse yra labai artimi. Pritaikome Kulbako-Leiblero informacijos kiekių teoriją 2002-2006 metų atvykusių į grupes procentinėms dalims-atitinkamiems Polinominiams skirstiniams, esančioms hipotezėse $H_0^{(1)}$ ir $H_1^{(1)}$, gauname Kulbako-Leiblero informacijas artimas nuliui:

$$I(1:2) = 0,398 \cdot \log \frac{0,398}{0,305} + 0,335 \cdot \log \frac{0,335}{0,375} + 0,297 \cdot \log \frac{0,297}{0,289} = 0,005,$$

$$I(2:1) = 0,305 \cdot \log \frac{0,305}{0,398} + 0,375 \cdot \log \frac{0,375}{0,335} + 0,289 \cdot \log \frac{0,289}{0,297} = 0,006,$$

$$J(1:2) = I(1:2) + I(2:1) = 0,011.$$

Analogiškai atliekami skaičiavimai visoms nagrinėjamoms charakteristikoms, surandant Kulbako-Leiblero informacijas tarp atitinkamų kitų Polinominių skirstinių.

23 lentelė. Kulbako-Leiblero informacija

Atvykę				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,005	0,0001	0,002	0,0006
I(2:1)	0,006	0,0001	0,002	0,0006
J(1:2)	0,011	0,0002	0,003	0,0012
Išvykę				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,003	0,0004	0,001	0,001
I(2:1)	0,003	0,0004	0,001	0,001
J(1:2)	0,006	0,0008	0,002	0,002
Bendras judėjimas				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,004	0,0002	0,0013	0,0007
I(2:1)	0,004	0,0002	0,0013	0,0007
J(1:2)	0,008	0,0004	0,0026	0,0014
Nusikaltimai				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,0054	0,00003	0,002	0,001
I(2:1)	0,0055	0,00003	0,002	0,001
J(1:2)	0,0109	0,00006	0,004	0,002
Savižudybės				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,001	0,001	0,001	0,004
I(2:1)	0,001	0,001	0,001	0,004
J(1:2)	0,002	0,002	0,002	0,008
Bedarbiai				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
I(1:2)	0,006	0,002	0,01	0,001
I(2:1)	0,006	0,002	0,01	0,001
J(1:2)	0,012	0,004	0,02	0,002

Išvada. Kuo mažesnės Kulbako-Leiblero informacijos tarp Polinominių skirstinių tikrinamose hipotezėse, tuo daugiau šansų, kad atitinkama skirstinių panašumo hipotezė bus priimta.

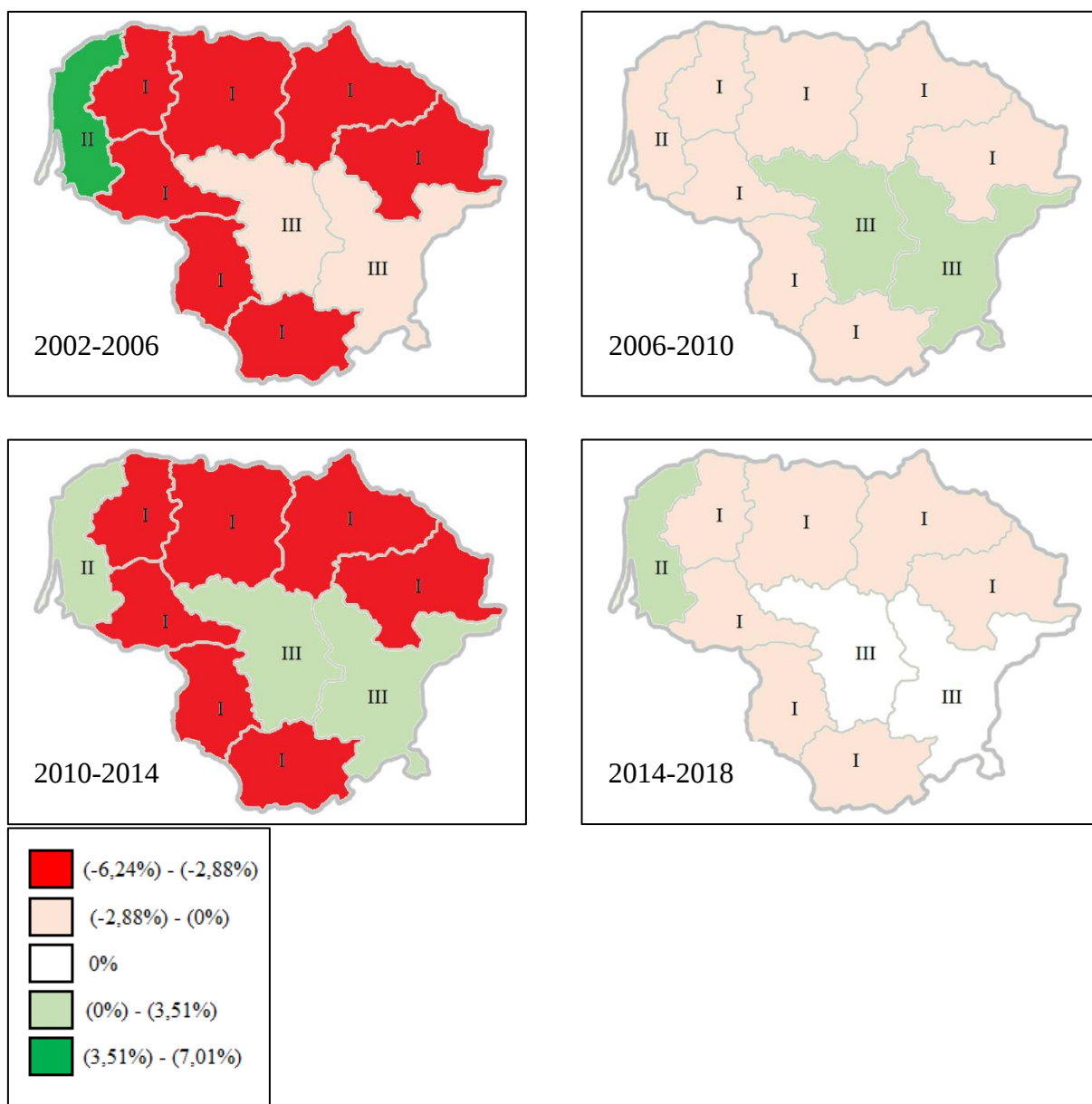
Iš informacinės Kulbako teorijos [10], žinome, kad esant dideliui n $I(1:2)$ nusako pirmos rūšies klaidos tikimybės $\alpha_1(\delta_n)$ eksponentinį mažėjimą, kai fiksuota antros rūšies klaidos tikimybė $\alpha_2(\delta_n) = \alpha_2 \in (0,1)$, δ_n – statistinis kriterijus, t.y. $\alpha_1(\delta_n) \sim e^{-nI(1:2)}$, $I(2:1)$ nusako antros rūšies klaidos tikimybės $\alpha_2(\delta_n)$ eksponentinį mažėjimą, kai fiksuota pirmos rūšies klaidos tikimybė $\alpha_1(\delta_n) = \alpha_1 \in (0,1)$, t.y. $\alpha_2(\delta_n) \sim e^{-nI(2:1)}$, o $I(1:2)$ yra suminė informacija. Tai žinant galima apytiksliai apskaičiuoti pirmos ir antros rūšies klaidos tikimybes, tikrinant atitinkamas hipotezes. Kulbako-Leiblero informacijos taikymas leidžia išvengti sudetingų skaičiavimų, kurie gaunami tiesiogiai skaičiuojant pirmos ir antros rūšies klaidų tikimybes. Kartu reikia atkreipti dėmesį, kad 24 lentelėje pirmos ir antros rūšies klaidų tikimybės tarpusavį nesusijusios (jos iš skirtingų modelių). Pvz., 2002-2006 m. pirmos rūšies tikimybė 0,217 reiškia, kad atmesti nulinę hipotezę, kai ji teisinga yra maža tikimybė – tai rodo ir ankstesnis tyrimas, kai ta hipotezė buvo priimta. Kadangi tyrimo duomenys grupėse buvo sunormuoti, padarant juos atitinkančius 100 tūkst. gyventojų, tai naujoje situacijoje n gavosi daug mažesnis, pvz., savižudybių atveju n (sveikoji dalis) svyruoja tarp 3 ir 8, o tai palyginus su labai dideliu n (artimu begalybei) – labai maži skaičiai. Kartu mažos Kulbako-Leiblero informacijos (pagal pradinę prielaidą jos turi būti didesnės už nulį) tarp atitinkamų skirstinių (23 lentelė) labai iškreipia vaizdą, kas neleidžia rasti net pakankamas apytiksliai pirmos ir antros rūšies klaidų tikimybes. Tai galima stebėti 24 lentelėje, kur buvo pritaikytos $\alpha_1(\delta_n) \sim e^{-nI(1:2)}$ ir $\alpha_2(\delta_n) \sim e^{-nI(2:1)}$ formulės. Kartu 24 lentelė rodo, kad norint pritaikyti minėtas apytiksliai pirmos ir antros rūšies klaidos tikimybes, formulėse ne tik n turi būti labai didelis, bet ir atitinkamos Kulbako-Leiblero informacijos turi būti ženkliai didesnės už nulį.

24 lentelė. Pirmos ir antros rūšies klaidų apytikslės tikimybės, gautos pritaikius Kulbako-Leiblero informacijas

Atvykę				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,217	0,971	0,481	0,629
Antros rūšies klaidos tikimybė	0,272	0,970	0,557	0,693
Išvykę				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,406	0,893	0,702	0,636
Antros rūšies klaidos tikimybė	0,523	0,887	0,753	0,702
Bendras judėjimas				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,088	0,891	0,486	0,400
Antros rūšies klaidos tikimybė	0,177	0,886	0,562	0,486
Nusikaltimai				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,123	0,988	0,437	0,746
Antros rūšies klaidos tikimybė	0,123	0,988	0,456	0,661
Savižudybės				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,994	0,994	0,995	0,985
Antros rūšies klaidos tikimybė	0,991	0,994	0,994	0,980
Bedarbiai				
	2002 - 2006	2006 - 2010	2010 - 2014	2014 - 2018
Pirmos rūšies klaidos tikimybė	0,043	0,045	0,0002	0,619
Antros rūšies klaidos tikimybė	0,001	0,351	0,0000002	0,430

2.4 FAKTORIŲ POKYČIŲ GRUPĖSE ATVAIZDAVIMAS LIETUVOS ŽEMĖLAPYJE

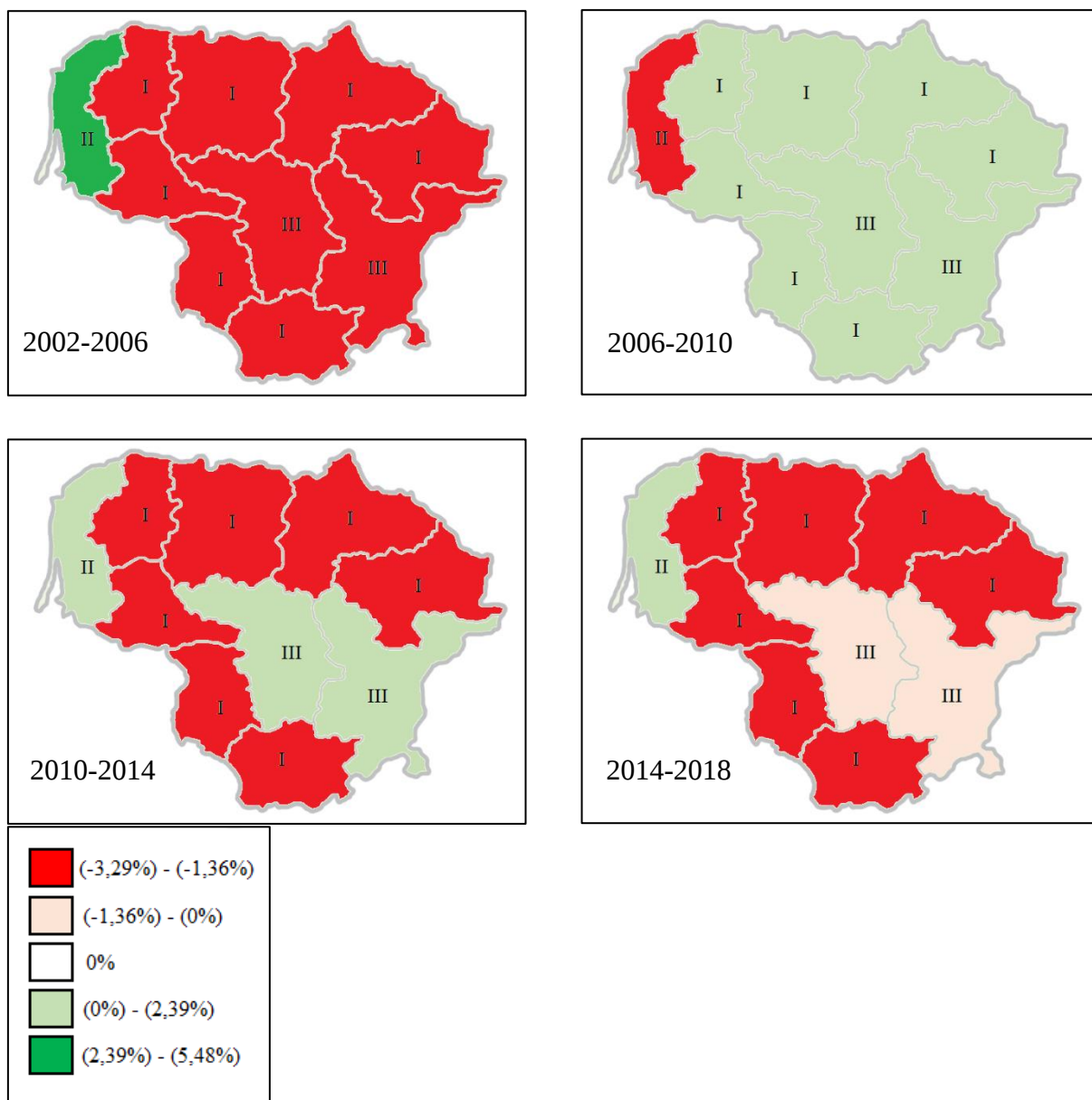
2.4.1 ATVYKĘ



Matome, kad 2002 – 2006 metų laikotarpį atvykusių labiausiai sumažėjo I grupėje, III grupėje taip pat sumažėjo, bet ne tiek daug. Daugiausiai gyventojų atvyko į II grupę. Visa tai galėjo būti susiję su Lietuvos įstojimu į Europos Sąjungą 2004 metais. Viskas yra atvirkščiai 2006 – 2010 metų laikotarpiu, čia atvykusių skaičius augo III grupėje, likusiose mažėjo. Tai galima paaiškinti 2010 m. pilnai pradėjus funkcionuoti Lietuvoje studijų finansavimo krepšelių sistemai, kuri iškaisvino jaunimo prisirašimą prie vietos ir jie pajudėjo į didžiuosius Lietuvos miestus – Kauną ir Vilnių. Dar didesnis mažėjimas atvykusių į I grupę matomas

2010 – 2014 metais. Tai vėl paaiškinama toliau egzistuojančia krepšelių finansavimo studijų sistema Lietuvoje – jaunimą studijomis labiausiai traukė sostinė – Vilnius, ir gretimas Kauno regionas, nes čia po studijų galima buvo rasti gerai apmokamą darbą. Paskutiniu tiriamu laikotarpiu atvykusių į III grupę skaičius nepakito, augo II, o I mažėjo. I ir II grupėse ankstesniais laikotarpiais buvo per daug sumažėję žmonių ir jų trūkumas atsispindėjo atsiradusiose naujose darbo vietose šiuose regionuose.

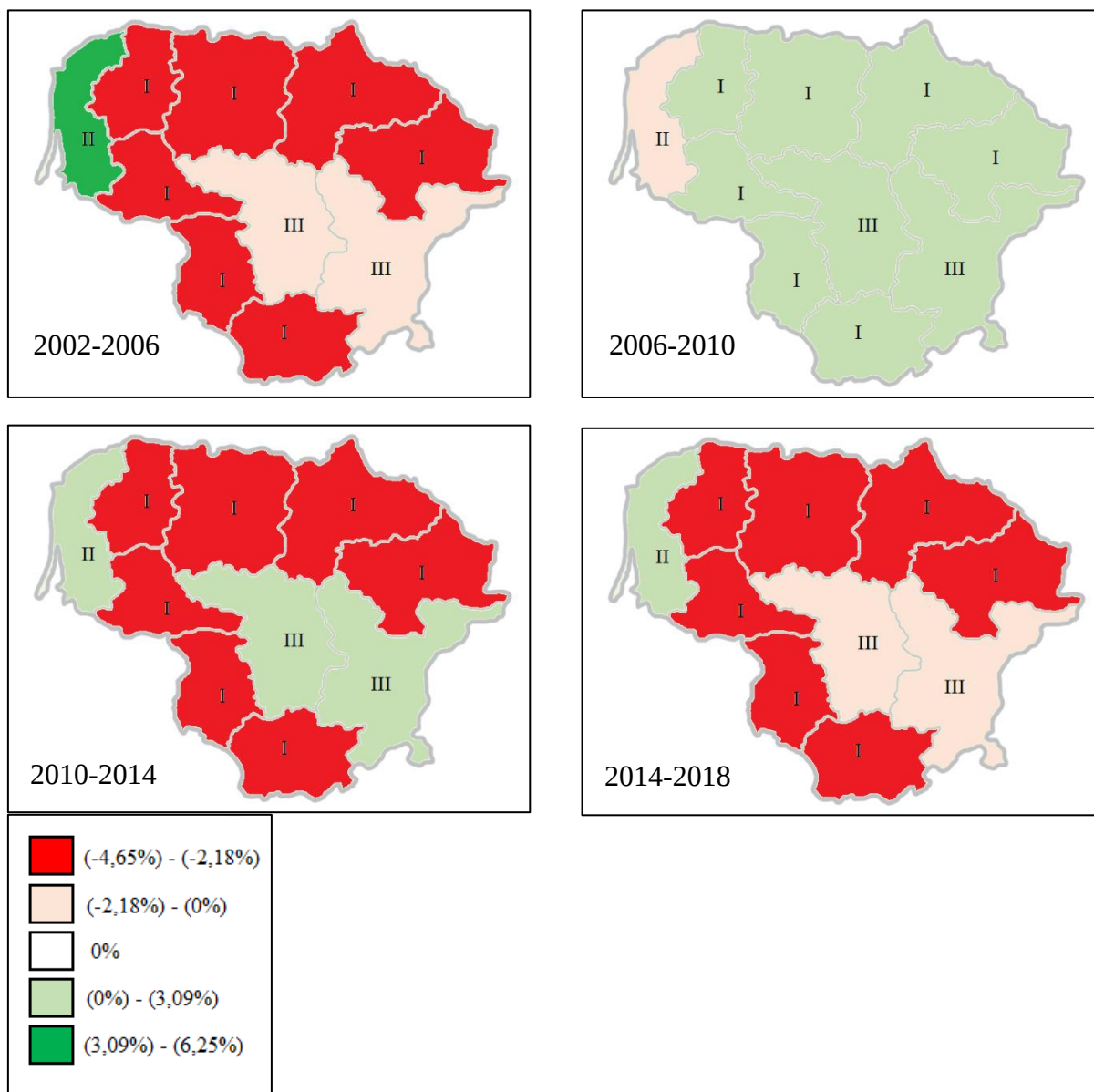
2.4.2 IŠVYKĘ



Išvykusių skaičius 2002 – 2006 metais sumažėjo I ir III grupėse, o II grupėje pokytis daugiau nei 2 %. Sekančiu laikotarpiu išvykusių iš II grupės sumažėjo, likusiose grupėse išvykusių padaugėjo. 2010 – 2014 II ir III grupėse išvykusių skaičius padidėjo, bet mažėjo I

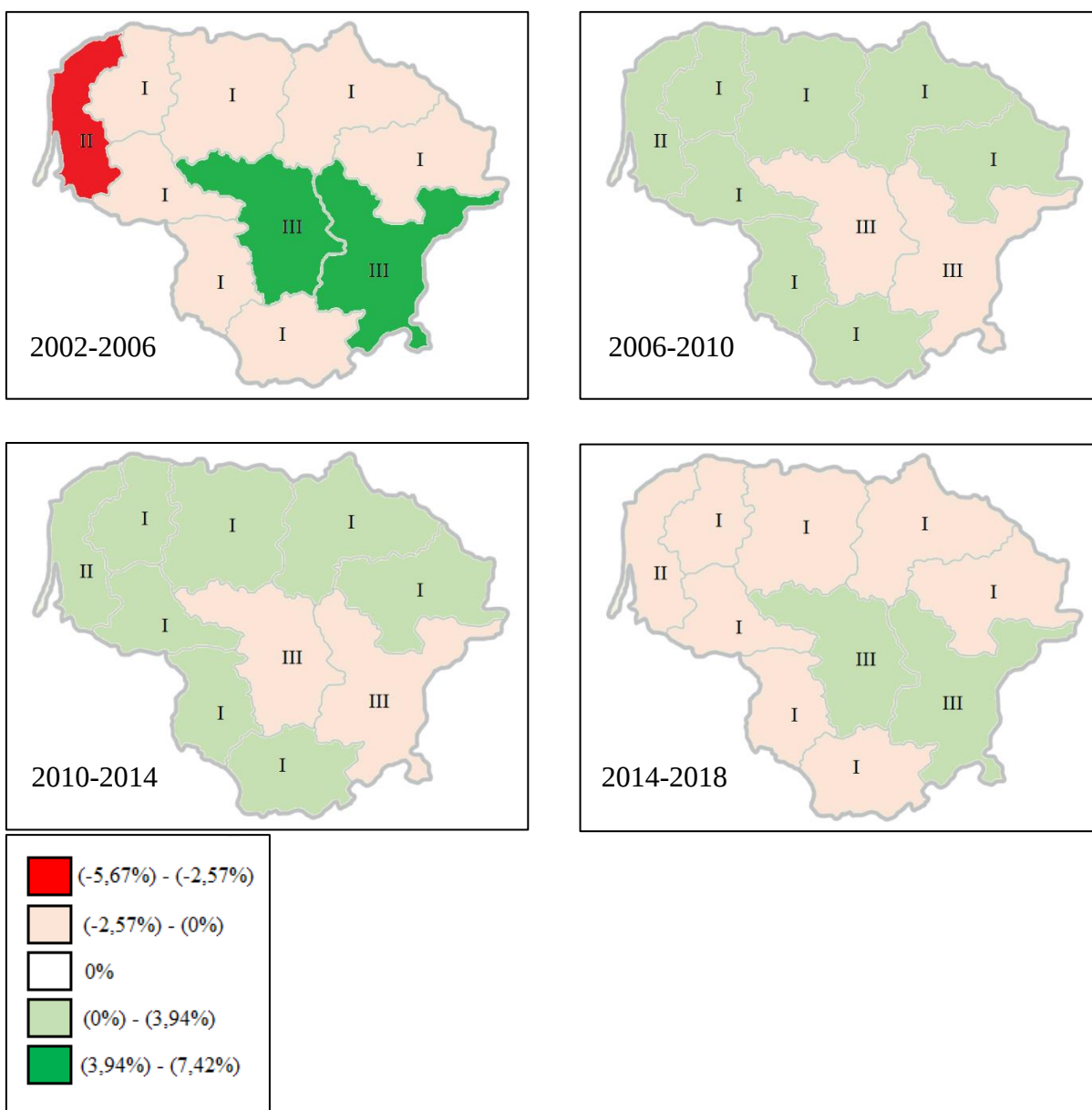
grupėje. Paskutiniu tiriamu laikotarpiu II grupėje išvykusių skaičius padidėjo, III grupėje sumažėjo iki 1,36 %, o I grupėje mažėjo dar daugiau – iki 3,29 %.

2.4.3 BENDRAS JUDĖJIMAS



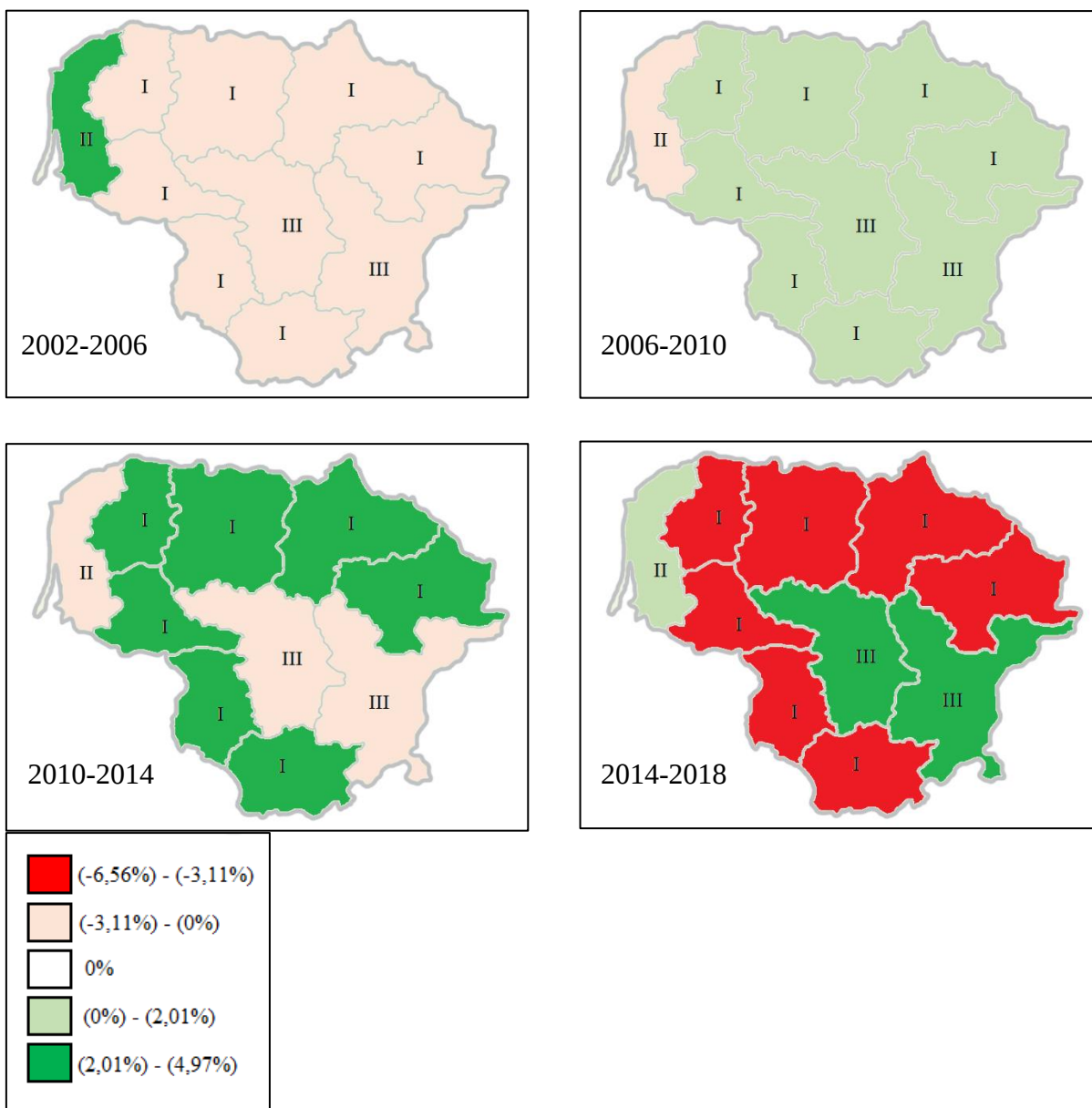
Bendras judėjimas tai atvykusių ir išvykusių skaičius, jei jis mažėja galime teigti, kad žmonės ne taip intensyviai keičia gyvenamąją vietą – „susitupi“, o jei didėja viskas yra atvirkščiai. Matome, kad 2002 – 2006 metais bendras judėjimas mažėjo I ir III grupėse, o II didėjo. Pereinant į sekantį laikotarpį žemėlapis „pažaliavo“ I ir III grupėse judėjimas didėjo, o II mažėjo. Sekančiame laikotarpyje I grupėje mažėjimas buvo didžiausias, o II ir III grupėse nedidelis augimas. 2014 – 2018 metai labai panašūs į 2002 – 2006 metų laikotarpį, tik čia II grupėje nebe toks intensyvus judėjimo padidėjimas.

2.4.4 NUSIKALTIMAI



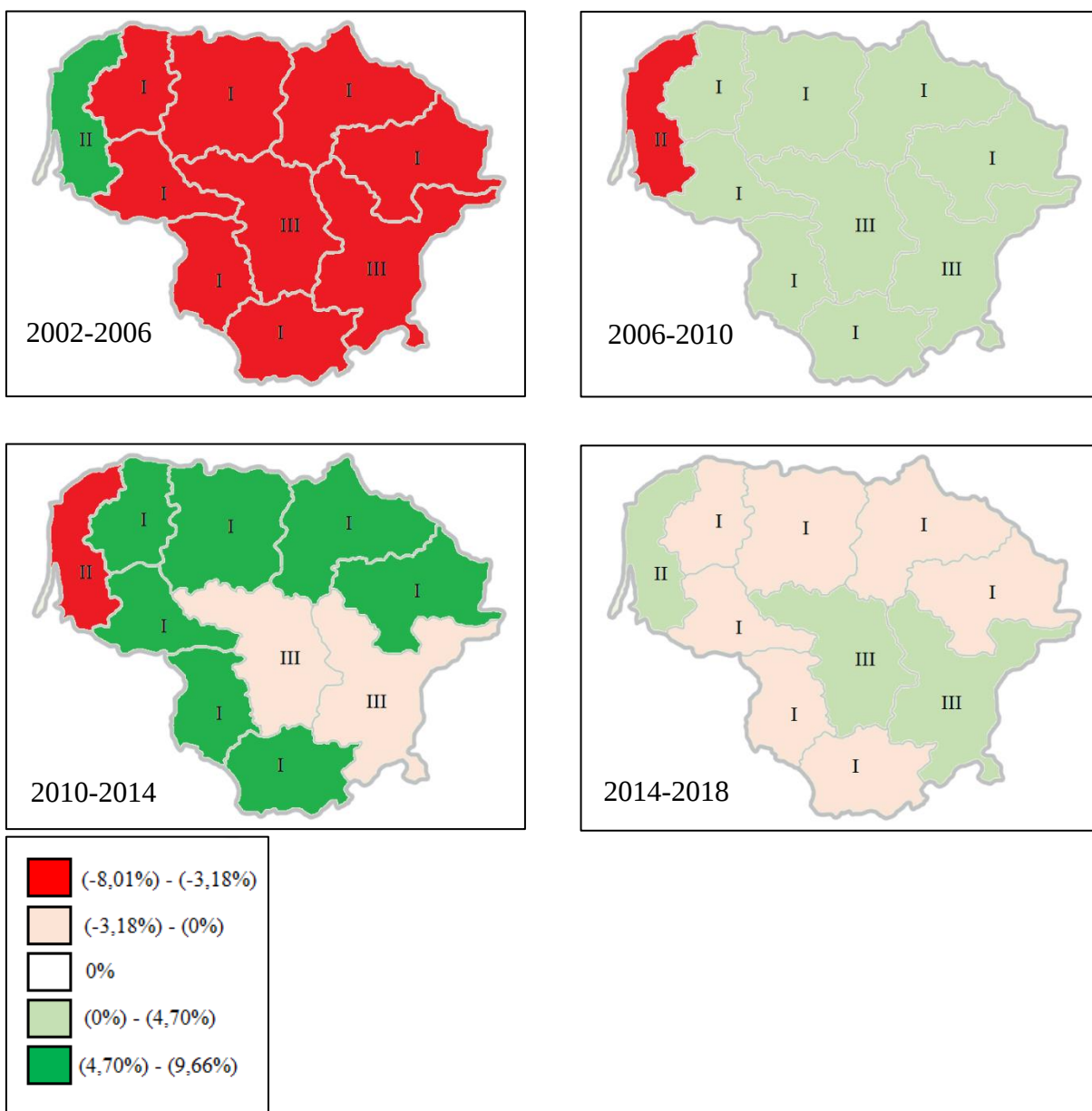
Nusikaltimų skaičius labiausiai augo III grupėje 2002 – 2006 metais, I grupėje šiek tiek mažėjo, o II mažėjo daugiau nei 5%. 2006 – 2010 ir 2010 - 2014 metų laikotarpiu III grupėje nusikaltimų skaičius mažėjo, visose likusiose grupėse didėjo iki beveik 4%. Paskutiniu tiriamu laikotarpiu III grupėje nusikaltimų skaičius didėjo, likusiose grupėse buvo nedidelis mažėjimas.

2.4.5 SAVIŽUDYBĖS



Savižudybių skaičius I ir III grupėse sumažėjo iki 3 % 2002 – 2006 metais, II grupėje padidėjo ženkliai. 2006 – 2010 metais tik II grupėje užfiksuotas savižudybių skaičiaus sumažėjimas – visur kitur didėjo. 2010 – 2014 II ir III grupėse savižudybių skaičius mažėjo, bet ne itin daug, o I grupėje išaugo stipriai. Paskutiniu tiriamu laikotarpiu savižudybių skaičius labiausiai sumažėjo I grupėje, nedaug padidėjo II, o III išaugo labai daug.

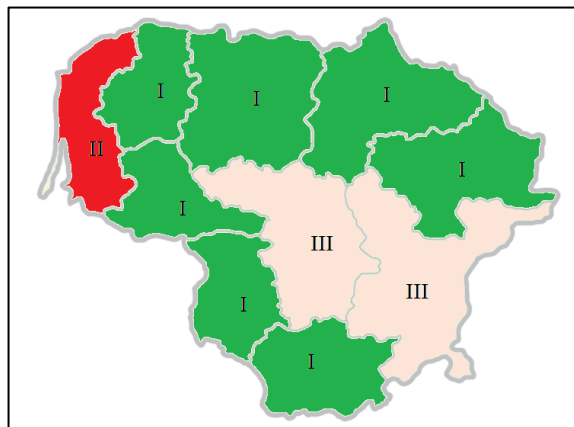
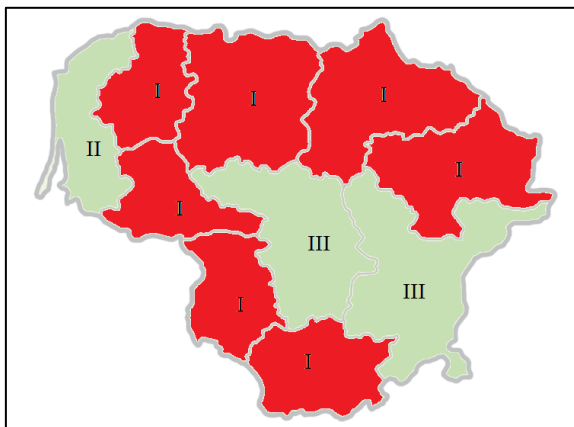
2.4.6 BEDARBIAI



Bedarbių skaičiaus augimas dižiausias II grupėje 2002 – 2006 metais, o I ir II sumažėjo iki 8%. 2006 – 2010 metų laikotarpiu bedarbių skaičius augo I ir III grupėse, II mažėjo. I grupėje stipriai išaugo iki beveik 10% 2010 – 2014 metais, likusiose grupėse mažėjo. Paskutiniu tiriamu laikotarpiu I grupėje mažėjo, II ir III padidėjo iki 4%.

2.5 FAKTORIŲ ĮTAKOS GYVENTOJŲ JUDĖJIMUI GRUPĖSE ATVAIZDAVIMAS LIETUVOS ŽEMĖLAPYJE

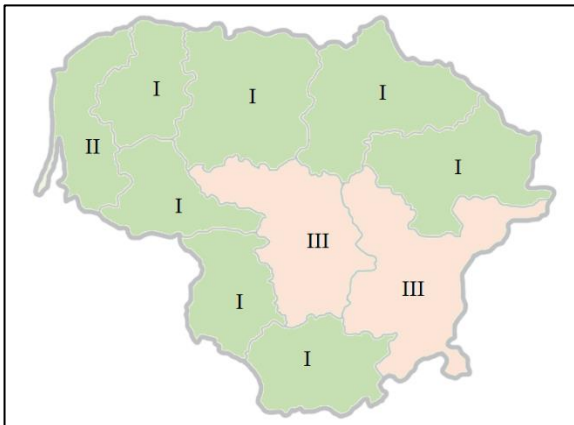
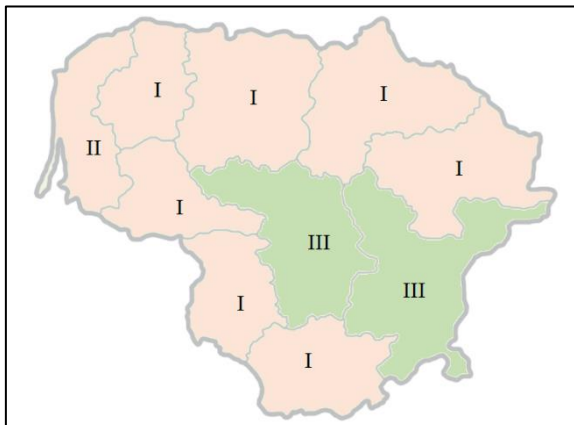
Įdomu pamatyti, kaip neigiami faktoriai veikia žmonių judėjimą grupėse.



9 pav. Atvykusių ir bedarbių pokyčiai 2010 -2014 metais

Matome, kad padidėjus bedarbių skaičiui I grupėje sumažėjo atvykusių į ją skaičius, o II ir III grupėse situacija geresnė - bedarbių skaičiui sumažėjus, atvykusių padidėjo.

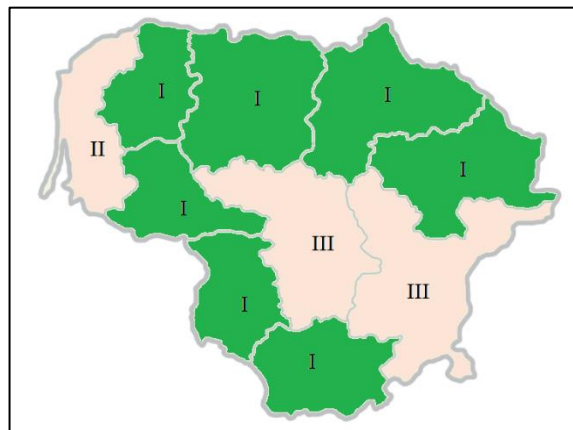
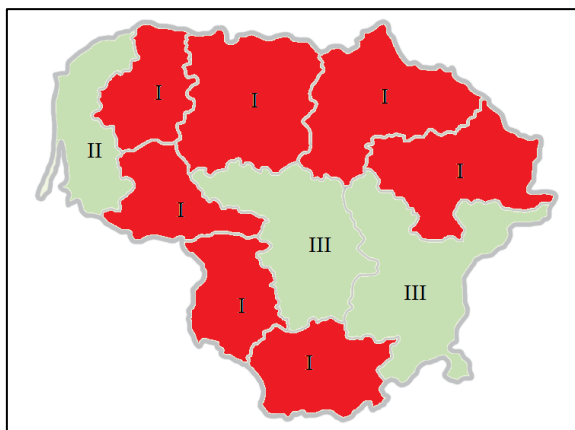
Išvada. Kai atvykusių žmonių į grupes mažėja, bedarbių skaičius tose grupėse didėja.



10 pav. Atvykusių ir nusikaltimų pokyčiai 2006 -2010 metais

10 pav. matome, kad padidėjus nusikaltimų skaičiui I ir II grupėse atvykusių skaičius į tas grupes galimai siekiant užtikrinti savo saugumą – sumažėjo. III grupėje viskas atvirkščiai, ten nusikaltimų sumažėjo, tai atspindi atvykusių skaičiaus padidėjimą.

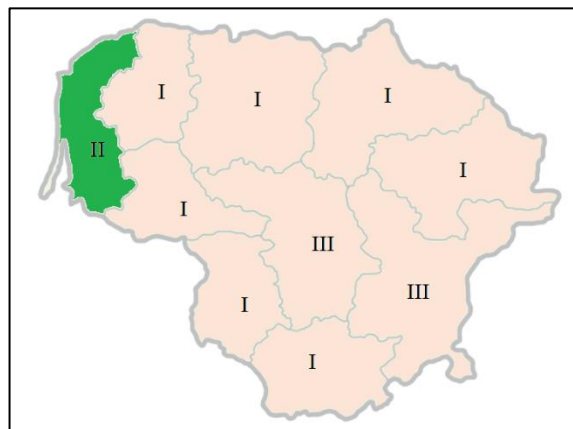
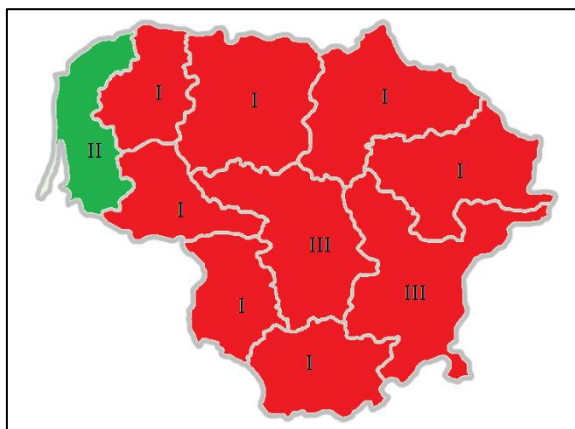
Išvada. Kai nusikaltimų skaičius grupėse didėja, atvykstančių žmonių į tas grupes mažėja.



11 pav. Atvykusių ir savižudybių pokyčiai 2010 -2014 metais

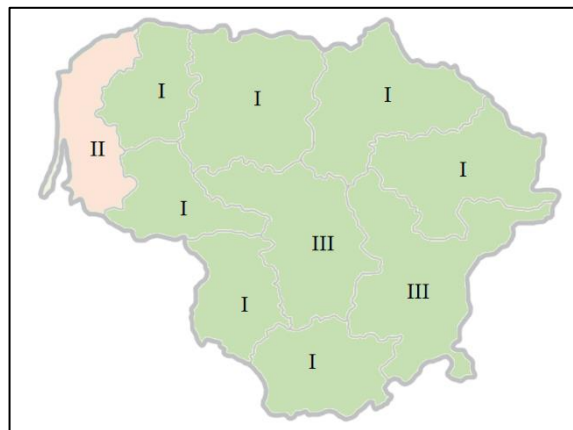
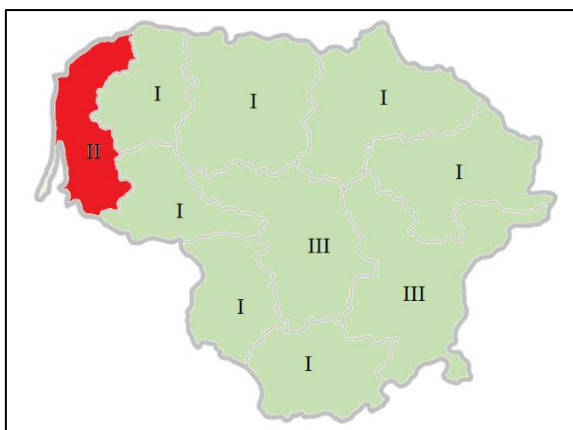
Savižudybių skaičiui padidėjus I grupėje 2010 – 2014 metais, sumažėjo atvykusių skaičius, o II ir III grupėse tais metais savižudybių skaičius sumažėjo ir padidėjo atvykusių skaičius.

Išvada. Savižudžių skaičiaus padidėjimas grupėse atspindi prastas gyvenimo sąlygas, todėl natūralu, kad ten sumažėja atvykusių žmonių skaičius.



12 pav. Išvykusių ir savižudybės pokyčiai 2002 -2006 metais

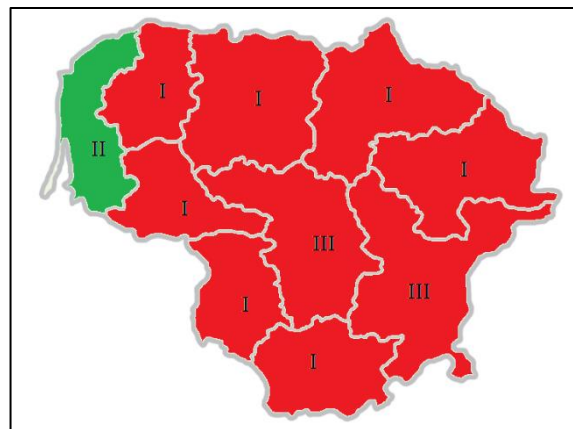
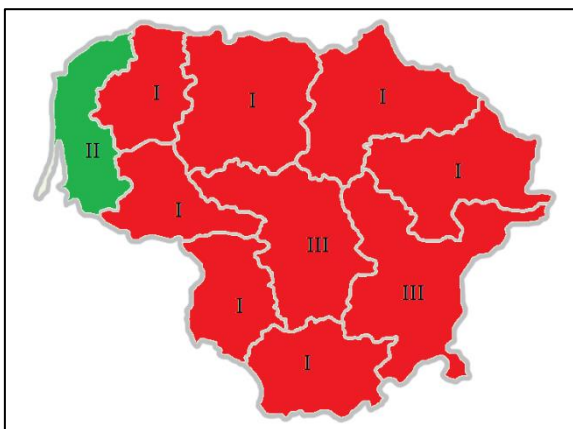
2002 – 2006 metų laikotarpiu sumažėjus savižudybių skaičiui I ir III grupėje sumažėjo ir išvykusių iš tų grupių gyventojų skaičius. II grupėje viskas atvirkščiai – ten savižudybių skaičius išaugo, tuo pat metu išvykusių skaičius padidėjo.



13 pav. Išvykusių ir savižudybių pokyčiai 2006 -2010 metais

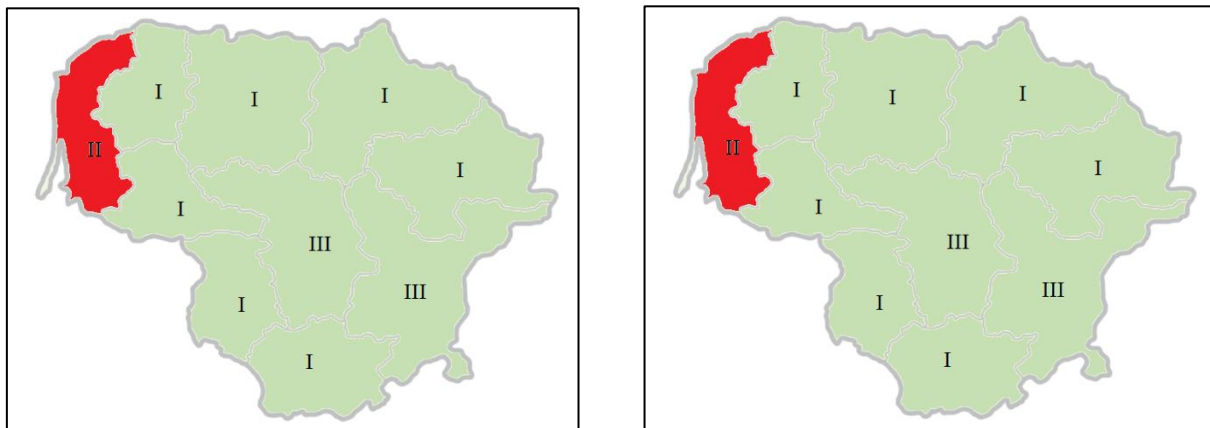
2006 – 2010 metų laikotarpiu viskas pasikeitė – I ir III grupėse išaugo savižudybių skaičius tai lėmė padidėjusį išvykusių skaičių iš jų. II grupėje išvykusių skaičius sumažėjo, taip pat ir savižudybių skaičius.

Išvada. Išvykusių žmonių ir savižudžių skaičiai tiesiogiai susiję: vieniems mažėjant, mažėja kiti, ir didėjant vieniems didėja kiti.



13 pav. Išvykusių ir bedarbių pokyčiai 2002 -2006 metais

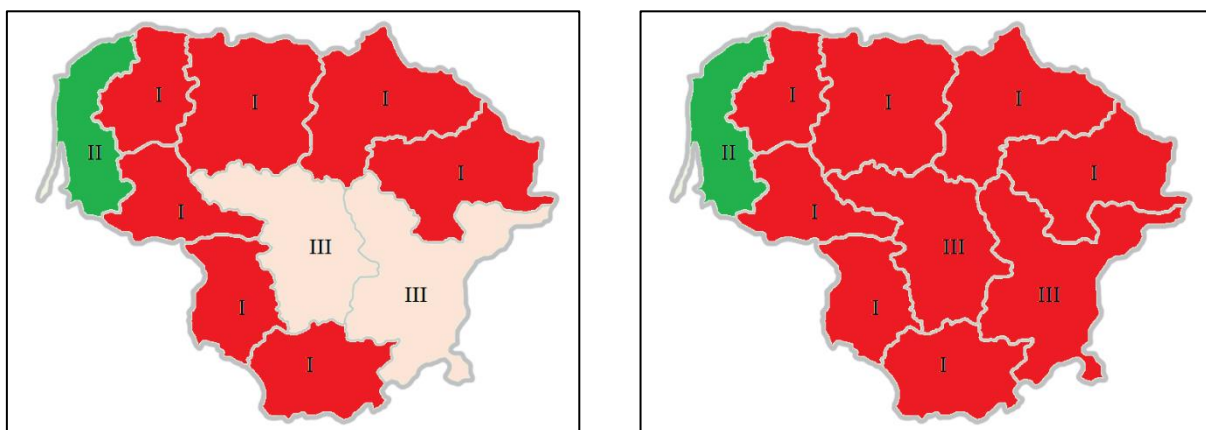
13 pav. matome, kad I ir III grupėse 2002 – 2006 metais sumažėjo bedarbystė, tai atspindi išvykusių iš tų grupių sumažėjimą. II grupėje bedarbystė padidėjo, o tuo pat metu išaugo ir išvykusių skaičius iš II grupės.



14 pav. Išvykusių ir bedarbių pokyčiai 2006 -2010 metais

2006 – 2010 metais II grupėje sumažėjus bedarbystei mažėjo išvykusių gyventojų skaičius. Likusioje Lietuvos dalyje t.y. I ir III grupėse bedarbystė padidėjo, taip pat padidėjo išvykusių skaičius. Galimai išvyko dirbti į II grupę.

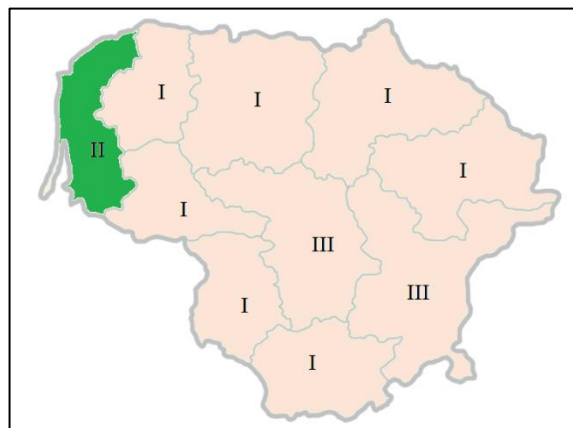
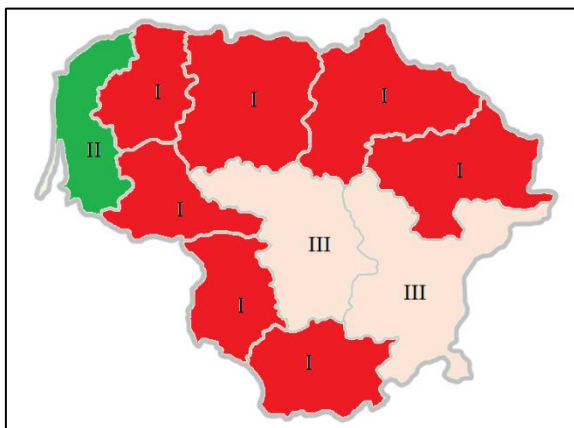
Išvada. Išvykusių žmonių skaičius tiesiogiai susijęs su bedarbių skaičiumi, t.y. didėjant vieniems didėja kiti ir atvirkščiai – mažėjant vieniems mažėja kiti.



15 pav. Bendro judėjimo ir bedarbių pokyčiai 2002 -2006 metais

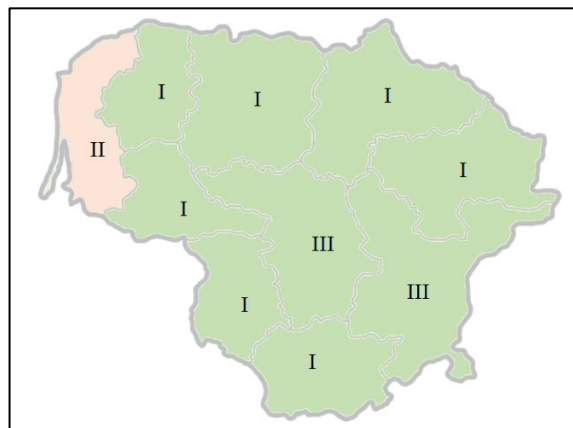
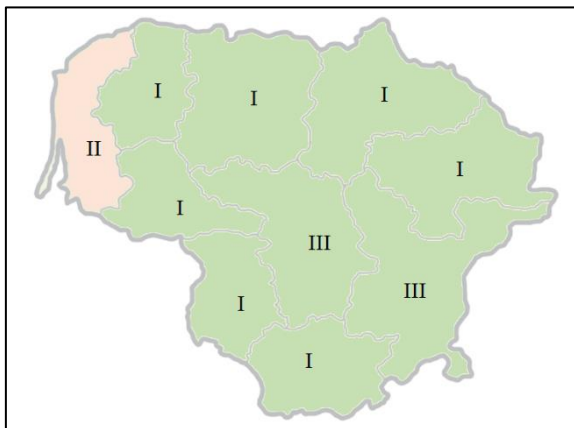
Bedarbystei sumažėjus I ir III grupėse 2002 – 2006 metais sumažėjo ir bendro judėjimo skaičiai. Žmonės nekeitė gyvenamų vietų taip intensyviai ir liko gyventi savo grupėse. II grupėje situacija kitokia, ten bedarbystė išaugo kartu su bendru judėjimu.

Išvada. Bendras žmonių judėjimas tiesiogiai susijęs su bedarbių pokyčiais.



16 pav. Bendro judėjimo ir savižudybių pokyčiai 2002 -2006 metais

Matome, kad nusižudymų skaičiui mažėjant I ir III grupėje mažėjo bendras judėjimas, II grupėje atvirkščiai – bendras judėjimas didėjo kartu su nusižudymais.



17 pav. Bendro judėjimo ir savižudybių pokyčiai 2006 -2010 metais

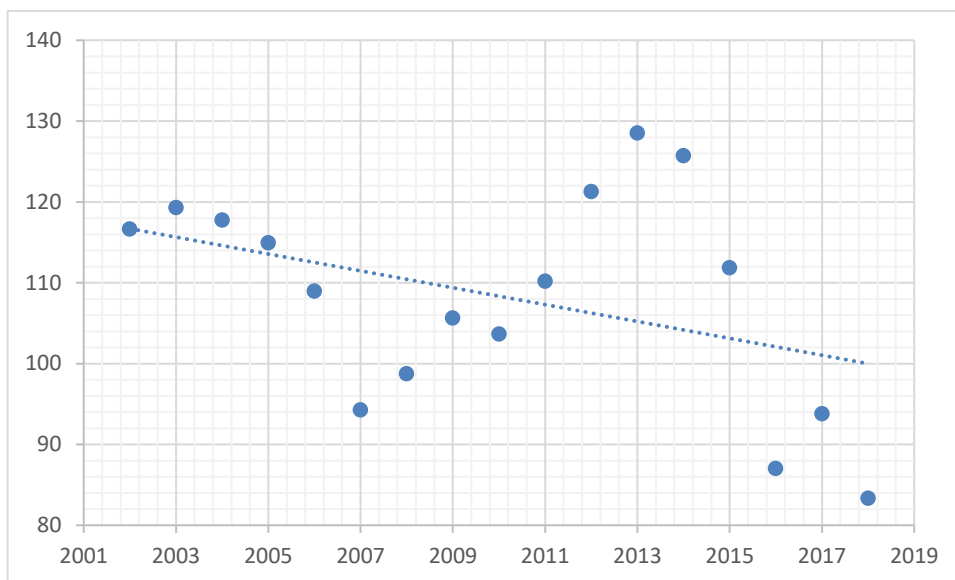
17 pav. galime matyti, kad augant savižudybių skaičiui 2006 – 2010 metais kartu augo ir bendras judėjimas tose grupėse. II grupėje savižudybių skaičius mažėjo, tai lėmė ir bendro judėjimo sumažėjimą.

Bendra išvada: gauti pokyčiai žemėlapiuose parodė, kad žmonių judėjimas yra susijęs su išskirtais neigiamais faktoriais: kai atvykusių žmonių į grupes mažėja, bedarbių tose vietose didėja, kai nusikaltimų skaičius didėja, atvykstančių žmonių į tas vietas mažėja. Savižudžių skaičiaus padidėjimas grupėse atspindi ten esančias prastas gyvenimo sąlygas, todėl natūralu, kad sumažėja atvykusių žmonių skaičius. Išvykusių žmonių ir savižudžių (ir bedarbių) skaičiai tiesiogiai susiję: vieniems mažėjant, mažėja ir kiti, o vidinis/bendras žmonių judėjimas tiesiogiai susijęs su bedarbių (ir savižudžių) pokyčiais, t.y. jei savižudybių skaičius mažėjo, tai ir bendras/vidinis žmonių judėjimas sumažėjo.

2.6 KOMPLEKSIŠKAS LAIKO EILUČIŲ TAIKYMAS

2.6.1 I GRUPĖS NUSIKALTIMŲ DUOMENYS

Pirmiausia pavaizduokime kaip atrodo I grupės nusikaltimų grafikas.



18 pav. I grupės nusikaltimų duomenys nuo 2002 iki 2018 metų

Matome, kad duomenys stipriai išsibarstę apie tiesę, todėl patikrinsime ar šių duomenų laiko eilutė turi X_t trendą. Lyginame duomenis su mediana ir sudarome “+” ir “-“ eilutę.

$$M_e = 110,193.$$

Lyginame

$$(+ + + +)(- - - -)(+ + + +)(- - -)$$

Palyginę duomenis su mediana tikriname hipotezę $H_0: MX_t = a = \text{const.}$

$\gamma(n)$ – bendras serijų skaičius = 4. $\tau(n)$ – ilgiausios serijos ilgis = 5.

$$\gamma(n) > \left\lceil \frac{1}{2}(n + 2 - 1,96\sqrt{n - 1}) \right\rceil$$

$$\tau(n) < [1,43 \ln(n + 1)]$$

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

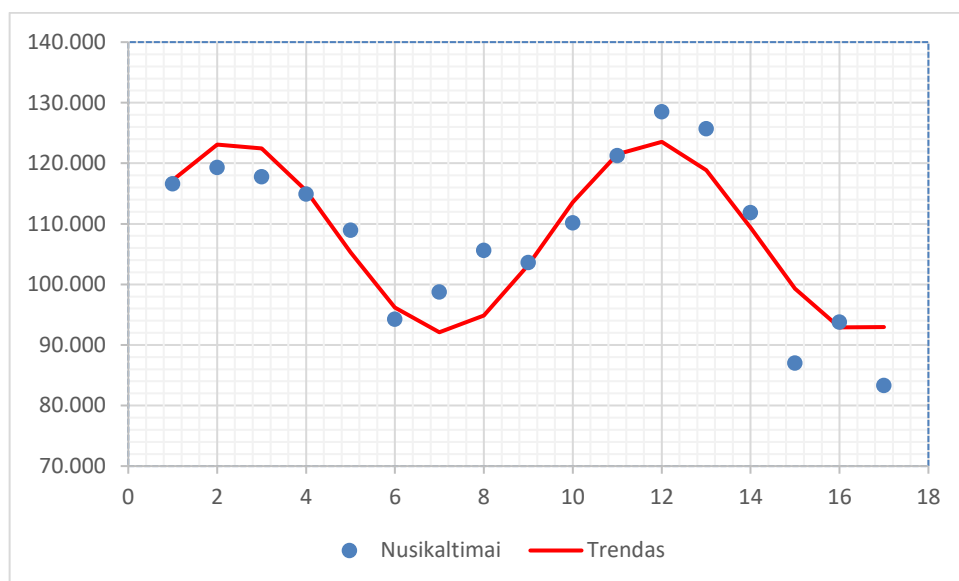
$$4 > \left\lceil \frac{1}{2}(17 + 2 - 1,96\sqrt{17 - 1}) \right\rceil \Leftrightarrow 4 > 11,16.$$

Kadangi ši lygybė negalioja, antrosios galime nebetikrinti. H_0 atmetame, o tai reiškia, kad X_t turi trendą $X_t = f_t + \varepsilon_t$.

I grupės nusikaltimų duomenų 9 pav. išsidėstymas atrodo lyg kosinusoidė, todėl darome prielaidą, kad trendas nėra tiesinis:

$$f_t = A + B \cdot \cos(\alpha \cdot t + \beta)$$

Programos MS Excel pagalba suskaičiuojame koeficientus ir gauname lygtį $f_t = 107,873 + 15,810 \cdot \cos(11,897 \cdot t + 1,609)$. Pažiūrėkime kaip atrodo gautas trendas nusikaltimų skaičiaus atžvilgiu.



19 pav. I grupės nusikaltimų duomenys ir trendas 2002 - 2018 metais

Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis.

25 lentelė. I grupės nusikaltimų duomenys eliminavus trendą

Metai	t	Y_t
2002	1	-0,546
2003	2	-3,791
2004	3	-4,661
2005	4	-0,553
2006	5	3,723
2007	6	-1,858
2008	7	6,664
2009	8	10,795
2010	9	0,460
2011	10	-3,379
2012	11	-0,180
2013	12	4,988
2014	13	6,892
2015	14	2,451
2016	15	-12,270
2017	16	0,895
2018	17	-9,630

Surasime AR(4) I grupės nusikaltimų duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi: $Y_t = a_1 \cdot Y_{t-1} + a_2 \cdot Y_{t-2} + a_3 \cdot Y_{t-3} + a_4 \cdot Y_{t-4} + \varepsilon_t$ ir stebinius: $Y_1 = -0,546, Y_2 = -3,791, Y_3 = -4,661, Y_4 = -0,553, Y_5 = 3,723, Y_6 = -1,858, Y_7 = 6,664, Y_8 = 10,795, Y_9 = 0,460, Y_{10} = -3,379, Y_{11} = -0,180, Y_{12} = 4,988, Y_{13} = 6,892, Y_{14} = 2,451, Y_{15} = -12,270, Y_{16} = 0,895, Y_{17} = -9,630$.

Mūsų tikslas, įvertinti koeficientus $\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4$ ir $\hat{\sigma}^2$.

Stebėjimų skaičius: $n = 17$

Apskaičiuojame šios imties vidurkį pagal formulę $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = -0,000002$.

Remdamiesi formule $\hat{\gamma}(k) = \frac{1}{n-k} \cdot \sum_{j=1}^{n-k} (Y_{j+k} - \bar{Y}) \cdot (Y_j - \bar{Y})$ apskaičiuojame $\hat{\gamma}(0), \hat{\gamma}(1), \hat{\gamma}(2), \hat{\gamma}(3), \hat{\gamma}(4)$ įverčius:

$$\begin{cases} \hat{\gamma}(0) = 32,299 \\ \hat{\gamma}(1) = 4,851 \\ \hat{\gamma}(2) = -0,700 \\ \hat{\gamma}(3) = -6,716 \\ \hat{\gamma}(4) = -2,784 \end{cases}$$

Turime tokias Yule – Walker lygtis, kurių pagalba rasime AR(4) proceso koeficientus:

$$\begin{cases} \hat{\gamma}(0) = a_1 \hat{\gamma}(1) + a_2 \hat{\gamma}(2) + a_3 \hat{\gamma}(3) + a_4 \hat{\gamma}(4) + \sigma^2 \\ \hat{\gamma}(1) = a_1 \hat{\gamma}(0) + a_2 \hat{\gamma}(1) + a_3 \hat{\gamma}(2) + a_4 \hat{\gamma}(3) \\ \hat{\gamma}(2) = a_1 \hat{\gamma}(1) + a_2 \hat{\gamma}(0) + a_3 \hat{\gamma}(1) + a_4 \hat{\gamma}(2) \\ \hat{\gamma}(3) = a_1 \hat{\gamma}(2) + a_2 \hat{\gamma}(1) + a_3 \hat{\gamma}(0) + a_4 \hat{\gamma}(1) \\ \hat{\gamma}(4) = a_1 \hat{\gamma}(3) + a_2 \hat{\gamma}(2) + a_3 \hat{\gamma}(1) + a_4 \hat{\gamma}(0) \end{cases}$$

Gautus įverčius įstatome į Yule – Walker lygtis:

$$\begin{cases} 32,299 = a_1 4,851 + a_2 (-0,700) + a_3 (-6,716) + a_4 (-2,784) + \sigma^2 \\ 4,851 = a_1 32,299 + a_2 4,851 + a_3 (-0,700) + a_4 (-6,716) \\ -0,700 = a_1 4,851 + a_2 32,299 + a_3 4,851 + a_4 (-0,700) \\ -6,716 = a_1 (-0,700) + a_2 4,851 + a_3 32,299 + a_4 4,851 \\ -2,784 = a_1 (-6,716) + a_2 (-0,700) + a_3 4,851 + a_4 32,299 \end{cases}$$

Programos “Mathcad” pagalba apskaičiuojame $\hat{a}$ įverčius ir $\hat{\sigma}^2$ įvertį, ir gauname AR(4) proceso lygtį:

$$Y_t = 0,142 \cdot Y_{t-1} - 0,014 \cdot Y_{t-2} - 0,199 \cdot Y_{t-3} - 0,027 \cdot Y_{t-4} + \varepsilon_t, \quad \varepsilon_t \sim N(0; 30,189)$$

Gautą AR(4) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$\begin{aligned} X_t = & 118,445 + 15,810 \cdot \cos(11,897 \cdot t + 1,609) + 0,142 \cdot X_{t-1} - 2,245 \cdot \\ & \cos(11,897 \cdot (t-1) + 1,609) - 0,014 \cdot X_{t-2} + 0,221 \cdot \cos(11,897 \cdot (t-2) + 1,609) - \\ & 0,199 \cdot X_{t-3} + 3,146 \cdot \cos(11,897 \cdot (t-3) + 1,609) - 0,027 \cdot X_{t-4} + 0,427 \cdot \\ & \cos(11,897 \cdot (t-4) + 1,609) + \varepsilon_t, \quad \varepsilon_t \sim N(0; 30,189). \end{aligned}$$

2.6.2 KITOS GRUPIŲ CHARAKTERISTIKOS KURIOMS TAIKYTINA LAIKO EILUČIŲ TEORIJA

2.2 skyriuje sužinojome, kad laiko eilučių teoriją taip pat galima taikyti I grupės bedarbių, II grupės savižudybių, III grupės atvykusiųjų, išvykusiųjų, bendro judėjimo, nusikaltimų ir bedarbių duomenims (16 lentelės rezultatai). Patikriname ar tie duomenys turi trendą.

26 lentelė. Hipotezės apie trendo egzistavimą tikrinimo rezultatai

	$H_0: MX_t = a = \text{const}$ priimta / atmesta
I grupės bedarbiai	-
II grupės savižudybės	-
III grupės atvykusieji	-
III grupės išvykusieji	-
III grupės bendras judėjimas	-
III grupės nusikaltimai	-
III grupės bedarbiai	-

Matome, kad visi duomenys turi trendą. Žemiau pateikiame apskaičiuotą kiekvienos charakteristikos trendo lygtį.

27 lentelė. Trendo lygtys grupių duomenims

	Trendo lygtis
I grupės bedarbiai	$f_t = 329,709 + 143,896 \cdot \cos(6,866 \cdot t + 6,283)$
II grupės savižudybės	$f_t = 2,459 - 0,086 \cdot t$
III grupės atvykusieji	$f_t = 74,878 + 3,514 \cdot t$
III grupės išvykusieji	$f_t = 66,485 + 2,600 \cdot t$
III grupės bendras judėjimas	$f_t = 141,363 + 6,114 \cdot t$
III grupės nusikaltimai	$f_t = 180,033 - 2,601 \cdot t$
III grupės bedarbiai	$f_t = 295,246 + 141,392 \cdot \cos(6,921 \cdot t + 6,048)$

Eliminavę trendą gauname AR(4) proceso lygtis:

28 lentelė. AR(4) proceso lygtys

	Trendo lygtis
I grupės bedarbiai	$Y_t = 0,669 \cdot Y_{t-1} - 1,08 \cdot Y_{t-2} + 0,259 \cdot Y_{t-3} - 0,442 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 741,145)$
II grupės savižudybės	$Y_t = 0,261 \cdot Y_{t-1} + 0,233 \cdot Y_{t-2} + 0,078 \cdot Y_{t-3} - 0,442 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 0,018)$
III grupės atvykusieji	$Y_t = 0,397 \cdot Y_{t-1} - 0,082 \cdot Y_{t-2} - 0,129 \cdot Y_{t-3} + 0,247 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 101,797)$
III grupės išvykusieji	$Y_t = 0,442 \cdot Y_{t-1} - 0,016 \cdot Y_{t-2} - 0,122 \cdot Y_{t-3} + 0,175 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 73,552)$
III grupės bendras judėjimas	$Y_t = 0,421 \cdot Y_{t-1} - 0,047 \cdot Y_{t-2} - 0,132 \cdot Y_{t-3} + 0,216 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 345,928)$
III grupės nusikaltimai	$Y_t = 0,522 \cdot Y_{t-1} - 0,22 \cdot Y_{t-2} - 0,158 \cdot Y_{t-3} + 0,233 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 142,032)$
III grupės bedarbiai	$Y_t = 0,51 \cdot Y_{t-1} - 0,569 \cdot Y_{t-2} - 0,196 \cdot Y_{t-3} + 0,127 \cdot Y_{t-4} + \varepsilon_t,$ $\varepsilon_t \sim N(0; 1088,792)$

Galiausiai apjungiame AR(4) proceso lygtis su trendo lygtimis ir gauname tokius rezultatus:

I grupės bedarbiai

$$X_t = 525,556 + 143,896 \cdot \cos(6,866 \cdot t + 6,283) + 0,669 \cdot X_{t-1} - 96,266 \cdot \cos(6,866 \cdot (t-1) + 6,283) - 1,08 \cdot X_{t-2} + 155,408 \cdot \cos(6,866 \cdot (t-2) + 6,283) + 0,259 \cdot X_{t-3} - 37,269 \cdot \cos(6,866 \cdot (t-3) + 6,283) - 0,442 \cdot X_{t-4} + 63,602 \cdot \cos(6,866 \cdot (t-4) + 6,283) + \varepsilon_t, \quad \varepsilon_t \sim N(0; 741,145)$$

II grupės savižudybės

$$X_t = 2,139 - 0,086 \cdot t + 0,261 \cdot X_{t-1} + 0,022 \cdot (t-1) + 0,233 \cdot X_{t-2} + 0,020 \cdot (t-2) + 0,078 \cdot X_{t-3} + 0,007 \cdot (t-3) - 0,442 \cdot X_{t-4} - 0,038 \cdot (t-4) + \varepsilon_t, \quad \varepsilon_t \sim N(0; 0,018)$$

III grupės atvykusieji

$$X_t = 42,456 + 3,514 \cdot t + 0,397 \cdot X_{t-1} - 1,395 \cdot (t-1) - 0,082 \cdot X_{t-2} + 6,140 \cdot (t-2) - 0,129 \cdot X_{t-3} + 9,659 \cdot (t-3) + 0,247 \cdot X_{t-4} - 0,868 \cdot (t-4) + \varepsilon_t, \quad \varepsilon_t \sim N(0; 101,797)$$

III grupės išvykusieji

$$X_t = 34,639 + 2,600 \cdot t + 0,442 \cdot X_{t-1} - 1,149 \cdot (t-1) - 0,016 \cdot X_{t-2} + 0,042 \cdot (t-2) \\ - 0,122 \cdot X_{t-3} + 0,317 \cdot (t-3) + 0,175 \cdot X_{t-4} - 0,455 \cdot (t-4) + \varepsilon_t, \\ \varepsilon_t \sim N(0; 73,552)$$

III grupės bendras judėjimas

$$X_t = 76,619 + 6,114 \cdot t + 0,421 \cdot X_{t-1} - 2,574 \cdot (t-1) - 0,047 \cdot X_{t-2} + 0,287 \cdot (t-2) \\ - 0,132 \cdot X_{t-3} + 0,807 \cdot (t-3) + 0,216 \cdot X_{t-4} - 1,321 \cdot (t-4) + \varepsilon_t, \\ \varepsilon_t \sim N(0; 345,928)$$

III grupės nusikaltimai

$$X_t = 112,161 - 2,601 \cdot t + 0,522 \cdot X_{t-1} + 1,358 \cdot (t-1) - 0,22 \cdot X_{t-2} - 0,572 \cdot (t-2) \\ - 0,158 \cdot X_{t-3} - 0,411 \cdot (t-3) + 0,233 \cdot X_{t-4} + 0,606 \cdot (t-4) + \varepsilon_t, \\ \varepsilon_t \sim N(0; 142,032)$$

III grupės bedarbiai

$$X_t = 333,037 + 141,392 \cdot \cos(6,921 \cdot t + 6,048) + 0,51 \cdot X_{t-1} - 72,110 \cdot \cos(6,921 \cdot \\ (t-1) + 6,048) - 0,569 \cdot X_{t-2} + 80,452 \cdot \cos(6,921 \cdot (t-2) + 6,048) - 0,196 \cdot X_{t-3} + \\ 27,713 \cdot \cos(6,921 \cdot (t-3) + 6,048) + 0,127 \cdot X_{t-4} - 17,957 \cdot \cos(6,921 \cdot (t-4) + \\ 6,048) + \varepsilon_t, \quad \varepsilon_t \sim N(0; 1088,792)$$

3. DARBO REZULTATAI

1. Atlikta lyginamoji vidinių migrantų ir neigiamų faktorių analizė Lietuvos mastu parodė, kad vidinių migrantų ir bedarbių bei nusikaltimų duomenys nuo 2002 iki 2018 metų koreliuoja vidutiniškai, o vidinių migrantų ir savižudybių duomenys visu tiriamu laikotarpiu koreliuoja labai stipriai ($\rho = -0,89$ teisinga su tikimybe 0,95).
2. Tyrimo metu nustatyta, kad ryšys tarp vidinių migrantų(y) ir savižudžių(x) Lietuvos mastu yra užrašomas lygtimi: $y = 85352,806 - 25,888x$.
3. Taikant statistinę hipotezę apie polinominį skirstinį darbe nustatyta, kad bedarbystės pokyčiai grupėse 2002–2006 ir 2010–2014 metais su tikimybe 0,99 buvo statistiškai reikšmingi, o 2002–2006 metais su tikimybe 0,9 atvykusiųjų, bendro/vidinio judėjimo, nusikalstamumo ir bedarbystės, 2006–2010 bei 2010–2014 metų bedarbystės pokyčiai buvo reikšmingi. Visi pokyčiai buvo iliustruoti Lietuvos žemėlapyje.
4. Gauti duomenų pokyčiai Lietuvos žemėlapiuose parodė, kad žmonių judėjimas yra susijęs su išskirtais neigiamais faktoriais: kai atvykusių žmonių mažėja, bedarbių tose vietose daugėja; kai nusikaltimų daugėja, atvykstančių žmonių į tas vietas mažėja; savižudžių skaičiaus padidėjimas atspindi prastas gyvenimo sąlygas tose vietose, todėl natūralu, kad ten sumažėja atvykusių žmonių skaičius; išvykusių žmonių ir savižudžių (ir bedarbių) skaičiai tiesiogiai susiję – vieniems mažėjant, mažėja ir kiti.
5. Nustačius, kad laiko eilučių teoriją galima taikyti I grupės nusikaltimų duomenims, rastas ir eliminuotas trendas, ir gauta AR(4) proceso lygtis: $X_t = 118,445 + 15,810 \cdot \cos(11,897 \cdot t + 1,609) + 0,142 \cdot X_{t-1} - 2,245 \cdot \cos(11,897 \cdot (t-1) + 1,609) - 0,014 \cdot X_{t-2} + 0,221 \cdot \cos(11,897 \cdot (t-2) + 1,609) - 0,199 \cdot X_{t-3} + 3,146 \cdot \cos(11,897 \cdot (t-3) + 1,609) - 0,027 \cdot X_{t-4} + 0,427 \cdot \cos(11,897 \cdot (t-4) + 1,609) + \varepsilon_t$, $\varepsilon_t \sim N(0; 30,189)$. AR(4) proceso lygtys taip pat apskaičiuotos I grupės bedarbių, II grupės savižudybių, III grupės atvykusiųjų, išvykusiųjų, bendro/vidinio judėjimo, nusikaltimų ir bedarbių duomenims.

LITERATŪRA

1. V. Bagdonavičius, J. Kruopis, *Matematinė statistika, I dalis*. Vilnius: Vilniaus universiteto leidykla, 2015.
2. V. Bagdonavičius, J. Kruopis, *Matematinė statistika, II dalis*. Vilnius: Vilniaus universiteto leidykla, 2015.
3. E. Berkman, *A conceptual guide to statistics using SPSS*. JAV, 2011.
4. V. Boguslauskas, *Ekonometrika*. Technologija, Kaunas 2007.
5. V. Čekanavičius, *Taikomoji regresinė analizė socialiniuose tyrimuose*.
<http://www.lidata.eu/index.php?file=files/mokymai/trast/trast.html&course_file=trast_turinys.html> [žiūrėta 2020-05-13]
6. V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai*. TEV leidykla, Vilnius, 2000.
7. Higienos instituto duomenys <<http://www.hi.lt/atviri-duomenys.html>> [žiūrėta 2020-04-02]
8. V. Kanišauskas, *Tikimybių teorijos ir matematinės statistikos pagrindai*. Šiauliai: Šiaulių universiteto leidykla, 2000.
9. V. Kanišauskas, *Vidinės migracijos matematinis modelis*,
<<https://www.mii.lt/LMR/B/2016/57B02.pdf>> [žiūrėta 2020-04-22]
10. S. Kullback, *Information theory and statistics*. JAV, 1978.
11. R. Leipus, *Finansinės laiko eilutės*. Vilnius 2010.
<http://web.vu.lt/mif/a.reklaite/files/2012/09/fin_laik_eil_2.pdf> [žiūrėta 2020-05-13]
12. Statistikos departamento duomenys <<https://osp.stat.gov.lt/>> [žiūrėta 2020-04-24]

SANTRAUKA

Magistro darbo objektas yra Statistikos departamento statistiniai ir demografiniai duomenys apie Lietuvos gyventojus 2002–2018 metais. Darbo tikslas – nustatyti, kaip neigiami faktoriai (nusikaltimų, savižudybių, bedarbių skaičiai) veikia gyventojų judėjimą šalies viduje. Tikslui pasiekti buvo keliami keli uždaviniai. Pirmiausia – surinkti reikiamus duomenis, rasti neigiamų faktorių priklausomybę su vidiniais migrantais Lietuvos mastu. Kitas uždavinys buvo sugrupuoti Lietuvos apskričių duomenis, nustatyti statistinę priklausomybę tarp neigiamų faktorių ir judėjimo grupių viduje bei tarp grupių. Po to, pasinaudojus hipoteze apie polinominio skirstinio parametrus, nustatyti, ar judėjimo ir neigiamų faktorių grupėse pokyčiai yra statistiškai reikšmingi kas 4 metus. Visa tai pavaizduojama Lietuvos žemėlapyje. Ir galiausiai tikrinama, ar duomenys tenkina laiko eilučių modelių taikymo prielaidas, ir jei jas tenkina – randami tie modeliai.

Atlikta lyginamoji vidinių migrantų ir neigiamų faktorių analizė Lietuvos mastu parodė, kad vidinių migrantų ir bedarbių bei nusikaltimų duomenys nuo 2002 iki 2018 metų koreliuoja vidutiniškai, o vidinių migrantų (y) ir savižudybių (x) duomenys visu tiriamu laikotarpiu koreliuoja labai stipriai ir užrašomi lygtimi: $y = 85352,806 - 25,888x$.

Tyrimo metu nustatyta, kad bedarbystės pokyčiai grupėse 2002–2006 ir 2010–2014 metais su tikimybe 0,99 buvo statistiškai reikšmingi, o 2002–2006 metais su tikimybe 0,9 atvykusiųjų, bendro/vidinio judėjimo, nusikalstamumo ir bedarbystės, o 2006–2010 bei 2010–2014 metų bedarbystės pokyčiai buvo reikšmingi.

Gauti duomenų pokyčiai Lietuvos žemėlapiuose parodė, kad žmonių judėjimas yra susijęs su išskirtais neigiamais faktoriais: kai atvykusių žmonių mažėja, bedarbių tose vietose daugėja; kai nusikaltimų daugėja, atvykstančių žmonių į tas vietas mažėja; savižudžių skaičiaus padidėjimas atspindi prastas gyvenimo sąlygas tose vietose, todėl natūralu, kad ten sumažėja atvykusių žmonių skaičius; išvykusių žmonių ir savižudžių (ir bedarbių) skaičiai tiesiogiai susiję – vieniems mažėjant, mažėja ir kiti.

Nustatyta, kad egzistuoja laiko eilučių AR(4) modeliai – jie rasti I grupės nusikaltimų ir bedarbių, II grupės savižudžių, III grupės atvykusiųjų, išvykusiųjų, bendro/vidinio judėjimo, nusikaltimų ir bedarbių duomenims.

INFLUENCE OF NEGATIVE FACTORS TO INTERNAL MIGRATION OF LITHUANIANS IN XXI CENTURY

SUMMARY

The object of this masters final work is the statistical and demographical data of Statistics Lithuania about Lithuanian residents over the period 2002-2018. The aim of this study is to determine how negative factors (crime, suicide and unemployment) affects Lithuanian residents internal movement within the country. To achieve this aim the following objectives have been set. First of all, to collect necessary data and identify statistical dependence between negative factors and internal migration nationwide. Next step is to group the data of Lithuania's regions, identify statistical dependence between negative factors and movement in the groups and between them. Another objective is to apply the statistical hypothesis about multinomial distribution and determine whether changes of movement and negative factors in groups are statistically significant every 4 years. Having identified this, changes in groups have to be marked in the map of Lithuania. And finally, check if there is a reason to apply time series models to statistical data about movement and negative factors in groups.

Comparative analysis of internal migration and negative factors indicated, that there is average dependence between internal migrants, unemployment and crime, but there is a big dependence between internal migrants (y) and suicide (x). The following linear regression was found: $y = 85352,806 - 25,888x$.

Having applied the statistical hypothesis about multinomial distribution parameters the final work found out, that in 2002-2006; 2010-2014 unemployment changes were statistically significant with probability of 0,99. In 2002-2006 changes of arrivals, general movement, crime and unemployment were statistically significant with probability of 0,9 as well as changes of unemployment in 2006-2010 and 2010-2014.

Having marked changes of data in the map of Lithuania it was found out, that negative factors does influence the movement of Lithuanian residents within the country. When there is decrease of arrivals in certain regions – unemployment increases in those same regions. When crime increases – arrivals decreases. Increased number of suicide shows bad living conditions in certain regions, this leads to decreased number of arrivals. The number of people that leaves regions where they live is directly related – when negative factors decreases, amount of people that left decreases too.

Having checked the assumptions to apply time series, it was identified that time series models can be applied to the data of I group's crime and unemployment, II group's suicide, III group's arrivals, departed, general movement, crime and unemployment. AR(4) models were found for that data.

PRIEDAI

1 PRIEDAS

Gyventojų migracija ir faktoriai apskrityse

Vilniaus apskritis (1)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	8207	6360	14567	20737	281	60700
2003	14138	10066	24204	24172	267	50200
2004	14051	10113	24164	27600	270	45200
2005	12516	8990	21506	28176	249	35800
2006	11391	8634	20025	26749	204	20900
2007	12920	9543	22463	24905	219	19000
2008	13387	9434	22821	27556	228	26500
2009	11986	8228	20214	28554	230	61500
2010	11879	9185	21064	24885	198	71100
2011	12827	9595	22422	25014	145	62300
2012	17169	11810	28979	26039	164	56000
2013	15289	11233	26522	23863	215	41400
2014	15919	12476	28395	22983	160	36200
2015	17172	12722	29894	20928	186	32500
2016	19093	13912	33005	17615	178	24500
2017	19005	14349	33354	18349	162	21700
2018	20629	16106	36735	17797	168	21400
Vidurkis	14563,412	10750,353	25313,765	23877,765	207,294	40405,882

Alytaus apskritis (2)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	1945	2192	4137	2498	117	13000
2003	3202	3734	6936	2503	89	10600
2004	3495	4080	7575	2563	101	11400
2005	3343	3962	7305	2731	101	5900
2006	2720	3229	5949	2702	87	3200
2007	2696	3209	5905	2233	70	2300
2008	2687	3347	6034	2496	80	3200
2009	2431	2990	5421	2661	61	12300
2010	2550	2988	5538	2792	75	13400
2011	2880	3294	6174	2792	77	12400
2012	2730	3562	6292	3201	58	11000
2013	2523	3138	5661	3204	76	12100
2014	2767	3425	6192	3241	73	12900
2015	2789	3384	6173	2930	76	11100
2016	2951	3736	6687	2239	47	8600
2017	2779	3555	6334	2245	34	7100
2018	3280	3916	7196	2167	44	5300
Vidurkis	2809,882	3396,529	6206,412	2658,706	74,471	9164,706

Kauno apskritis (3)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	7644	8409	16053	14477	288	46700
2003	11035	12343	23378	15233	254	40100
2004	11158	11595	22753	16225	250	33100
2005	10496	10520	21016	18736	242	28300
2006	10363	10077	20440	16014	186	19100
2007	11395	11098	22493	13752	204	12800
2008	11988	11704	23692	14005	194	18000
2009	10136	9928	20064	15343	202	39800
2010	9669	9307	18976	14697	196	50700
2011	10153	9995	20148	14051	209	38400
2012	11504	11212	22716	13533	170	31100
2013	11786	10995	22781	15115	184	29900
2014	13002	12182	25184	15260	182	26600
2015	12706	12138	24844	14567	147	21000
2016	14058	13310	27368	12396	159	19200
2017	14722	13617	28339	14083	155	16000
2018	16030	14721	30751	11625	131	17100
Vidurkis	11637,941	11361,824	22999,765	14653,647	197,235	28700,000

Klaipėdos apskritis (4)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	3924	3845	7769	9476	152	22400
2003	6977	6435	13412	9410	152	22000
2004	7299	6680	13979	9600	128	21700
2005	6641	6006	12647	9147	128	12500
2006	6798	6322	13120	8044	115	12000
2007	7711	6873	14584	7991	99	6400
2008	7957	7030	14987	8059	91	11700
2009	6654	6085	12739	8487	99	24300
2010	6400	5652	12052	7576	99	32000
2011	6691	6088	12779	8082	93	22100
2012	6968	6480	13448	8322	86	15300
2013	7231	6441	13672	9307	95	12200
2014	8424	7368	15792	7884	83	13400
2015	8133	7356	15489	7430	71	10400
2016	9411	8694	18105	6237	82	8400
2017	10122	9120	19242	6700	56	10300
2018	11268	10156	21424	5369	65	7700
Vidurkis	7565,235	6860,647	14425,882	8065,941	99,647	15576,471

Marijampolės apskritis (5)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	2113	2202	4315	2877	84	9100
2003	2835	3058	5893	2994	93	6500
2004	2993	3360	6353	2635	87	5200
2005	2736	3141	5877	2852	96	2300
2006	2915	3299	6214	3105	68	1800
2007	3019	3373	6392	2700	63	1300
2008	2806	3381	6187	2898	78	2300
2009	2611	3138	5749	2911	68	7800
2010	2622	3030	5652	3056	57	11300
2011	2679	3122	5801	3351	80	9600
2012	2926	3791	6717	3424	67	9700
2013	2763	3470	6233	3793	86	10500
2014	3009	3630	6639	3526	71	9300
2015	2841	3498	6339	3322	69	7400
2016	3020	3723	6743	2657	55	6900
2017	3553	4423	7976	2773	52	6000
2018	3727	4611	8338	2321	40	5000
Vidurkis	2892,235	3426,471	6318,706	3011,471	71,412	6588,235

Panevėžio apskritis (6)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	3905	4120	8025	6192	169	18300
2003	5021	5677	10698	6538	145	15400
2004	4814	5752	10566	6968	124	15800
2005	3896	4646	8542	6199	107	13700
2006	3957	4705	8662	5562	97	9100
2007	4207	5089	9296	4641	101	7800
2008	3979	4952	8931	4790	105	6900
2009	3943	4787	8730	4930	117	17300
2010	3906	4691	8597	4633	114	27400
2011	3788	4695	8483	4624	102	23900
2012	4407	5559	9966	4801	106	18500
2013	4256	5274	9530	5219	104	17500
2014	4310	5346	9656	5820	99	15300
2015	4080	5242	9322	4951	78	13400
2016	4495	5961	10456	4580	86	12500
2017	4377	5895	10272	4705	81	9800
2018	4925	6401	11326	4089	72	7600
Vidurkis	4250,941	5223,059	9474,000	5249,529	106,294	14717,647

Šiaulių apskritis (7)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	4587	5091	9678	6608	196	24300
2003	5780	6856	12636	6939	178	30500
2004	6238	7280	13518	6974	184	19600
2005	5500	6491	11991	6409	148	16700
2006	5682	6647	12329	6053	121	10000
2007	6453	7717	14170	5435	106	6400
2008	6069	7305	13374	5274	122	8700
2009	5490	6429	11919	6162	149	22200
2010	5504	6360	11864	5829	97	27500
2011	5847	6648	12495	6352	137	24300
2012	5993	7246	13239	7478	117	22600
2013	5863	6955	12818	8213	134	20500
2014	6083	7247	13330	7415	107	18000
2015	6146	7393	13539	6703	99	17100
2016	6742	7946	14688	4979	92	14800
2017	6604	7919	14523	5209	79	12300
2018	7295	8789	16084	4588	64	10100
Vidurkis	5992,706	7077,588	13070,294	6271,765	125,294	17976,471

Tauragės apskritis (8)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	1664	1723	3387	2315	69	5100
2003	2080	2261	4341	2106	73	6000
2004	2318	2748	5066	1937	69	4900
2005	1971	2567	4538	2072	67	3800
2006	2022	2432	4454	2087	45	2100
2007	2123	2734	4857	1770	50	1400
2008	1839	2516	4355	2019	52	2500
2009	1708	2193	3901	2158	57	5600
2010	1553	2112	3665	2198	53	5700
2011	1803	2282	4085	2129	42	5800
2012	2298	2867	5165	2444	38	6500
2013	1948	2588	4536	2494	57	4600
2014	1991	2592	4583	2063	45	5300
2015	1880	2606	4486	2147	50	4300
2016	2189	2942	5131	1444	37	4000
2017	2083	2882	4965	1505	41	4200
2018	2204	3198	5402	1353	20	3500
Vidurkis	1980,824	2543,706	4524,529	2014,176	50,882	4429,412

Telšių apskritis (9)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	2593	2601	5194	2721	89	12900
2003	3592	3851	7443	2635	106	10500
2004	3710	4052	7762	2417	76	7500
2005	3137	3496	6633	2500	81	6200
2006	3338	3568	6906	2294	70	4900
2007	3525	3983	7508	2068	48	3500
2008	3434	3935	7369	2184	58	4400
2009	2687	3306	5993	2275	69	11900
2010	2831	3316	6147	2272	46	16000
2011	2759	3243	6002	2630	64	12800
2012	3190	3844	7034	3267	55	9700
2013	2631	3473	6104	3137	51	9300
2014	2945	3709	6654	3420	60	8800
2015	2829	3621	6450	2844	59	7300
2016	3535	4623	8158	2014	25	8100
2017	3227	4181	7408	2210	36	6500
2018	3616	4591	8207	2095	32	4700
Vidurkis	3151,706	3729,000	6880,706	2528,412	60,294	8529,412

Utenos apskritis (10)						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	2715	2754	5469	3191	106	11800
2003	3165	3544	6709	3289	98	12200
2004	3297	3713	7010	3163	92	9200
2005	2733	3150	5883	3252	100	5100
2006	2551	2824	5375	2864	56	4600
2007	2534	2964	5498	2495	65	3600
2008	2786	3328	6114	2691	103	4100
2009	2297	2859	5156	2810	86	7900
2010	2882	3155	6037	2680	83	15400
2011	2772	3237	6009	3061	69	16500
2012	2868	3682	6550	2840	66	16400
2013	2953	3676	6629	3028	83	14500
2014	2965	3440	6405	2969	50	12300
2015	2850	3466	6316	2418	61	9400
2016	3078	3725	6803	1787	62	9100
2017	2945	3476	6421	2584	52	9300
2018	3202	3687	6889	2247	47	7600
Vidurkis	2858,412	3334,118	6192,529	2786,412	75,235	9941,176

2 PRIEDAS

Gyventojų skaičiai apskrityse

Apskritis Nr.					
	1	2	3	4	5
2002	845292	186484	695033	382076	187281
2003	843320	184992	689787	379878	185970
2004	841748	182751	682074	376854	184058
2005	838556	179626	671923	372541	181376
2006	831237	175813	658100	364486	177439
2007	828191	173028	650408	360434	174810
2008	826306	170249	643358	357201	172200
2009	826445	167603	638385	355084	169973
2010	824193	164833	629896	350306	167143
2011	811515	158403	610225	340047	162205
2012	805915	154980	599638	334744	159273
2013	806308	152139	592816	331553	156455
2014	806106	149851	587238	329013	154084
2015	807523	147809	583047	327300	151793
2016	805380	145109	577358	324618	149077
2017	805173	141616	569875	320507	145360
2018	805367	138095	563112	317252	141287
Apskritis Nr.					
	6	7	8	9	10
2002	296871	366330	132982	178315	183973
2003	294182	362783	131808	177139	181638
2004	290097	357441	130179	175090	178637
2005	284918	350938	127690	172426	175226
2006	278262	341761	124020	167888	170829
2007	273701	335221	121549	164963	167678
2008	268815	328699	119179	162431	164167
2009	264319	323353	117100	160532	161062
2010	259374	316278	114763	157718	157472
2011	251401	303110	110511	152594	152577
2012	246298	295824	108125	149843	149001
2013	242340	290471	106335	147728	145760
2014	238748	285763	104623	145482	142564
2015	235394	281632	103083	143511	140170
2016	231001	276329	101200	141293	137193
2017	225033	270482	98608	137769	133481
2018	218726	265467	95817	134139	129639

3 PRIEDAS

Gyventojų skaičiai grupėse

	Grupė		
	I	II	III
2002	1532236	382076	1540325
2003	1518512	379878	1533107
2004	1498253	376854	1523822
2005	1472200	372541	1510479
2006	1436012	364486	1489337
2007	1410950	360434	1478599
2008	1385740	357201	1469664
2009	1363942	355084	1464830
2010	1337581	350306	1454089
2011	1290801	340047	1421740
2012	1263344	334744	1405553
2013	1241228	331553	1399124
2014	1221115	329013	1393344
2015	1203392	327300	1390570
2016	1181202	324618	1382738
2017	1152349	320507	1375048
2018	1123170	317252	1368479

4 PRIEDAS

Gyventojų skaičius Lietuvoje

2002	3454637
2003	3431497
2004	3398929
2005	3355220
2006	3289835
2007	3249983
2008	3212605
2009	3183856
2010	3141976
2011	3052588
2012	3003641
2013	2971905
2014	2943472
2015	2921262
2016	2888558
2017	2847904
2018	2808901

Duomenys grupėse

I Grupė						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	19522	20683	40205	26402	830	94500
2003	25675	28981	54656	27004	782	91700
2004	26865	30985	57850	26657	733	73600
2005	23316	27453	50769	26015	700	53700
2006	23185	26704	49889	24667	544	35700
2007	24557	29069	53626	21342	503	26300
2008	23600	28764	52364	22352	598	32100
2009	21167	25702	46869	23907	607	85000
2010	21848	25652	47500	23460	525	116700
2011	22528	26521	49049	24939	571	105300
2012	24412	30551	54963	27455	507	94400
2013	22937	28574	51511	29088	591	89000
2014	24070	29389	53459	28454	505	81900
2015	23415	29210	52625	25315	492	70000
2016	26010	32656	58666	19700	404	64000
2017	25568	32331	57899	21231	375	55200
2018	28249	35193	63442	18860	319	43800

II Grupė						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	3924	3845	7769	9476	152	22400
2003	6977	6435	13412	9410	152	22000
2004	7299	6680	13979	9600	128	21700
2005	6641	6006	12647	9147	128	12500
2006	6798	6322	13120	8044	115	12000
2007	7711	6873	14584	7991	99	6400
2008	7957	7030	14987	8059	91	11700
2009	6654	6085	12739	8487	99	24300
2010	6400	5652	12052	7576	99	32000
2011	6691	6088	12779	8082	93	22100
2012	6968	6480	13448	8322	86	15300
2013	7231	6441	13672	9307	95	12200
2014	8424	7368	15792	7884	83	13400
2015	8133	7356	15489	7430	71	10400
2016	9411	8694	18105	6237	82	8400
2017	10122	9120	19242	6700	56	10300
2018	11268	10156	21424	5369	65	7700

III Grupė						
	Atvykę (X)	Išvykę (Y)	Bendrai (X+Y)	Nusikaltimai (N)	Savižudybės (S)	Bedarbiai (B)
2002	15851	14769	30620	35214	569	107400
2003	25173	22409	47582	39405	521	90300
2004	25209	21708	46917	43825	520	78300
2005	23012	19510	42522	46912	491	64100
2006	21754	18711	40465	42763	390	40000
2007	24315	20641	44956	38657	423	31800
2008	25375	21138	46513	41561	422	44500
2009	22122	18156	40278	43897	432	101300
2010	21548	18492	40040	39582	394	121800
2011	22980	19590	42570	39065	354	100700
2012	28673	23022	51695	39572	334	87100
2013	27075	22228	49303	38978	399	71300
2014	28921	24658	53579	38243	342	62800
2015	29878	24860	54738	35495	333	53500
2016	33151	27222	60373	30011	337	43700
2017	33727	27966	61693	32432	317	37700
2018	36659	30827	67486	29422	299	38500