

ŠIAULIŲ UNIVERSITETAS
TECHNOLOGIJOS, FIZINIŲ IR BIOMEDICINOS MOKSLŲ FAKULTETAS
INFORMATIKOS IR MATEMATIKOS KATEDRA

DEIVIDAS LESUTIS
(Finansų matematikos studijų programa, 612G17002)

**VIDINĖS MIGRACIJOS
XXI A. PRADŽIOS LIETUVOJE
STATISTINĖ ANALIZĖ**

Bakalauro darbas

Darbo vadovas

doc. dr. V. Kanišauskas

ŠIAULIAI, 2017

Patvirtinu, kad bakalauro darbas yra originalus, neturintis plagiato požymių.

.....

(Parašas)	(Vardas ir pavardė)	(Data)
-----------	---------------------	--------

TURINYS

ĮVADAS	5
1 TEORINĖ DALIS	7
1.1 PAGRINDINĖS MATEMATINĖS STATISTIKOS SĄVOKOS	7
1.2 POLINOMINIS SKIRSTINYS	7
1.3 POLINOMINIO SKIRSTINIO TAIKYMAS	8
1.4 SERIJŲ KRITERIJUS	9
1.5 LAIKO EILUTĖS	10
1.5.1 AUTOREGRESINIAI PROCESAI	10
1.5.2 AUTOREGRESINIŲ PROCESŲ PARAMETRŲ VERTINIMAS	11
1.5.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS	11
1.6 KORELIACINĖ ANALIZĖ	12
1.6.1 EMPIRINIS KORELIACIJOS KOEFICIENTAS	12
1.6.2 HIPOTEZĖ APIE KORELIACIJOS KOEFICIENTĄ	13
1.6.3 TIESINĖ REGRESIJA	14
2 PRAKTINĖ DALIS	18
2.1 DUOMENYS. PRADINĖ STATISTINĖ ANALIZĖ LIETUVOS MASTU	18
2.2 VIDINĖS MIGRACIJOS LIETUVOS APSKRITYSE FAKTO NUSTATYMAS	21
2.2.1 GYVENTOJŲ PASISKIRSTYMAS APSKRITYSE	21
2.2.2 VIDINĖS MIGRACIJOS 2001–2006 M. FAKTO NUSTATYMAS	23
2.2.3 VIDINĖS MIGRACIJOS 2015–2016 M. FAKTO NUSTATYMAS	24
2.2.4 VIDINĖS MIGRACIJOS POKYČIŲ LIETUVOS APSKRITYSE ATSPINDYS LIETUVOS ŽEMĖLAPYJE	25
2.3 PRIKLAUSOMYBĖ TARP VIDINĖS IR IŠORINĖS MIGRACIJŲ	28
2.3.1 LIETUVOS MASTU, PAGAL 2001–2016 M. DUOMENIS	28
2.3.2 LIETUVOS MASTU PAGAL APSKRITIS	30
2.3.3 PRIKLAUSOMYBĖ TARP MIGRANTŲ IR LIETUVOS GYVENTOJŲ	34
2.4 KOMPLEKSIŠKAS LAIKO EILUČIŲ TAIKYMAS	36
2.4.1 PRIELAUDŲ LAIKO EILUTĖMS TAIKYTI PATIKRINIMAS	36
2.5 LAIKO EILUČIŲ TAIKYMAS VILNIAUS APSKRİČIAI	40
2.5.1 TRENDŲ ELIMINAVIMAS IR SURADIMAS	40
2.5.2 AR(1) SURADIMAS	41
2.6 ATVYKUSIŲ Į APSKRITIS IR IŠVYKUSIŲ IŠ APSKRİČIŲ GYVENTOJŲ ANALIZĖ	43
2.6.1 ATVYKUSIŲ Į APSKRITIS GYVENTOJŲ ANALIZĖ	44
2.6.2 IŠVYKUSIŲ IŠ APSKRİČIŲ GYVENTOJŲ ANALIZĖ	47
DARBO REZULTATAI	52

LITERATŪRA	53
SANTRAUKA	54
SUMMARY	55
PRIEDAI	57

IVADAS

Temos aktualumas. Migracija – tai kėlimasis iš vienos gyvenamosios vietos į kitą. Žmonių migracija vyksta nuo seniausių laikų. Šiais laikais žmonės tapo dar judresni – išaugo jų mobilumas. Šiuolaikinėje visuomenėje migruoti šalies viduje ar už jos ribų dažniausiai reiškia norą pagerinti gyvenimo sąlygas ar tiesiog plėsti pasaulėžiūrą pakeitus gyvenamąją vietą. Lietuvai įstojus į Europos Sąjungą 2004 metais gyventojų išorinės migracijos skaičiai išaugo dėl ranka pasiekiamos galimybės išvykti svetur. Šiuo metu šalies išorinės migracijos rodikliai yra vieni didžiausių Europos Sąjungoje. Ne tik išorinė migracija gali padėti pagerinti gyvenimo sąlygas. Migruojant šalies viduje iš kaimų ar mažesnių miestų į didesnius galima rasti daugiau darbo vietų bei patrauklesnius darbo užmokesčius, geresnes mokymosi sąlygas ir t.t.

Migracija nėra neigiamas reiškinys, bet sąvoką padalijus į išorinę bei vidinę migraciją, išorinė Lietuvai nėra tokia jau teigiama. Dėl išorinės migracijos kinta gyventojų sudėtis, sensta visuomenė, prarandama daug jaunų bei gabių darbingo amžiaus žmonių.

Migracija gali būti tiek savanoriška, tiek priverstinė. Savanoriškai migruojama norint įgyti aukštąjį išsilavinimą, rasti geriau apmokamą darbą, gyventi palankesnio klimato sąlygomis ar dėl didesnio paslaugų pasirinkimo. Priverstinai migruojama dėl socialinių ar ekonominių priežasčių, stichinių nelaimių, bado, rasinės bei tautinės diskriminacijos ar karo. Dėl priverstinės migracijos atsiranda vis daugiau pabėgėlių.

Dažniausiai migraciją Lietuvoje tyrinėja socialinių mokslų atstovai, o fizinių mokslų atstovų darbų yra mažiau. Statistikos departamentas leidžia prieiti prie nemažo kiekio duomenų, susijusių su migracija. Matematinės statistikos metodais ištirsime tam tikrus dėsningumus, susijusius su Lietuvos gyventojų migracija, todėl darbas yra aktualus.

Tyrimo objektas. Statistikos departamento statistiniai ir demografiniai duomenys apie Lietuvos gyventojų migraciją.

Chronologinės ribos. 2001–2016 metai.

Tyrimo tikslas: Atlikti Lietuvos XXI a. pradžios vidinės migracijos statistinę ir lyginamąją analizę.

Tyrimo uždaviniai:

1. Taikant vidinės migracijos matematinį modelį nustatyti, ar Lietuvos apskrityse 2001–2016 m. laikotarpiu vyko vidinė migracija.

2. Lietuvos žemėlapyje pavaizduoti vidinės migracijos pokyčius Lietuvos apskrityse tiriamu laikotarpiu.
3. Nustatyti statistinę priklausomybę tarp vidinės ir išorinės migracijų.
4. Nustatyti, ar egzistuoja prielaidos taikyti laiko eilučių modelius statistiniams duomenims apie migrantus apskrityse. Jei prielaidos patenkinamos, pritaikyti tuos modelius.
5. Nustatyti, kaip atvykstantys į apskritis žmonės veikia vidinės migracijos procesą, ir parodyti jį Lietuvos žemėlapyje.
6. Nustatyti, kaip išvykstantys iš apskričių žmonės veikia vidinės migracijos procesą, ir parodyti jį Lietuvos žemėlapyje.

Tyrimo metodai. Taikomi matematinės statistikos, lyginamasis, vizualinis–dedukcinis ir laiko eilučių metodai.

Darbo struktūra. Darbas yra sudarytas iš įvado, dviejų skyrių, darbo rezultatų apibendrinimo, santraukos, literatūros sąrašo ir priedo.

Pirmajame skyriuje, pavadintame „Teorinė dalis“, pateikti pagrindiniai matematiniai metodai, taikomi darbe. Tai – pagrindinės matematinės statistikos sąvokos, polinominis skirstinys, jo taikymas migracijai, serijų kriterijus apie duomenų atsitiktinumą ir nepriklausomumą, laiko eilučių teorijos elementai, statistinė hipotezė apie koreliacijos koeficientą. Skyriaus pabaigoje pateiktos pagrindinės migracijos sąvokos ir migracijų rūšys.

Antrasis skyrius pavadintas „Praktinė dalis“. Skyrių sudaro smulkesnės dalys, kuriose pateiktas darbo tyrimas ir gauti rezultatai.

Pirmojoje dalyje pateikiami pirminiai vidinių ir išorinių migrantų duomenys. Tie duomenys pavaizduoti lentelėse ir rastos jų statistinės charakteristikos. Antrojoje skyriaus dalyje nustatinėjamas faktas apie migraciją Lietuvos apskrityse. Nustačius šį faktą, gyventojų skaičiaus pokyčiai Lietuvos apskrityse pavaizduoti Lietuvos žemėlapyje, iš kurio galima matyti, kaip juda gyventojai. Trečiojoje dalyje ieškomas empirinis koreliacijos koeficientas tarp vidinių bei išorinių migrantų, sudarytas grafikas ir rasta tiesinė lygtis. Ketvirtojoje skyriaus dalyje tikrinama hipotezė apie tiesinės priklausomybės egzistavimą. Penktojoje dalyje ieškoma tiesinės regresijos lygtis vidinių ir išorinių migrantų duomenims nuo 2001 iki 2016 metų. Šeštojoje skyriaus dalyje kompleksiskai taikoma laiko eilučių teorija: tikrinama, ar yra prielaidų laiko eilutėms taikyti, ar egzistuoja trendas tiriamuose duomenyse, atliekamas rasto trendo eliminavimas iš duomenų ir AR(1) proceso lygties radimas.

1 TEORINĖ DALIS

1.1 PAGRINDINĖS MATEMATINĖS STATISTIKOS SĄVOKOS

Pagrindinės matematinės sąvokos, formulės ir apibrėžimai paimti iš [1].

Apibrėžimas. Imtis – stebėtų rezultatų sąrašas.

Apibrėžimas. Variacinė eilutė – imties reikšmių išdėstymas nemažėjančia 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. Hipotezė vadiname kiekvieną nepatvirtintą teiginį apie nežinomą populiacijos požymio skirstinį ar jo parametrus arba kelių populiacijų nepriklausomumą. 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

Sakykime (žr. [4]), 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 $X = (X_1, \dots, X_k)$ tikimybiniis skirstinys 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 paprastąją 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:

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

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 tikėtino 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 $\chi^2_{1-\alpha}(k-1)$ 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 [2] straipsnio.

1.4 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. Norint įsitikinti, kaip yra iš tikrųjų, tikrinama hipotezė H_0 : imties elementai $X_1, \dots, X_n$ yra atsitiktiniai ir nepriklausomi. Tokiai hipotezei tikrinti taikomas serijų kriterijus su empirine mediana.

Pagal šį kriterijų pirmiausia užrašomi imties $X_1, \dots, X_n$ elementai didėjančia tvarka.

Pagal formulę

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

randama imties empirinė mediana M_e . Po to M_e reikšmė lyginama su pradinės imties elementais $X_1, \dots, X_n$. Rašomas „+“, jei $X_1 > M_e$, ir „–“, jei $X_1 < M_e$. Jei reikšmės sutampa, nerašoma nieko. Ši procedūra tęsiama, kol patikrinami visi imties elementai. Gaunama „+“–ų ir „–“–ų eilutė. Joje esančių „+“ skaičius žymimas raide k_1 , o „–“ skaičius – k_2 . Bendras visų serijų skaičius žymimas 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 tarpu, kai N didelis, jis nurodo tarp imties elementų esančią dėsningą kaitą. Abiem šiais atvejais pažeidžiamas imties elementams keliamas atsitiktinumo 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:

Serijų kriterijaus aprašas paimtas iš [1] knygos.

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

1.5.2 AUTOREGRESINIŲ PROCESŲ PARAMETRŲ VERTINIMAS

Tarkime, kad turime autoregresijos proceso $\{X_t\}$, išreiškiamo (1) lygtimi, stebinius $X_1, X_2, \dots, X_n$, ir žinoma lygties eilė p . Reikia gauti nežinomus parametrus $a_1, a_2, \dots, a_p, \sigma^2$.

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 įrašome į (1) 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.

1.5.3 LAIKO EILUČIŲ IŠLYGINIMO METODAS

Esminį vaidmenį sprendžiant išskyrimo ir įvertinimo uždavinį trendinės $f_{TR}(t)$, sezoninės $\mu(t)$ ir ciklinės $\psi(t)$ sudaromosios išskaidyme yra pradinis analizės etapas, kuriame išaiškinamas pats buvimo ar nebuvimo faktas neatsitiktinės (ir priklausančios nuo laiko t) komponentės išskaidyme. Iš esmės, tikrinama statistinė hipotezė $H_0: MX(t) = a = const..$ Konstruojamas įvertis (aproksimacija) nežinomi integralinei neatsitiktinei daliai $f(t) = \chi(A) \cdot f_{TR}(t) + \chi(B) \cdot \mu(t) + \chi(C) \cdot \psi(t)$, t.y. sprendžiamas analizuojamos laiko eilutės $X(t)$ palyginimas (eliminavimas atsitiktinės liekanos $a(t)$). 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 darome tikrindami nepriklausomumą). T.y. vietoj kiekvieno pradinės eilės $X(t_1), \dots, X(t_n)$ nario $X(t)$, $t = 1, 2, \dots, n$, rašome „+“, jei $X(t) > X_{med}^{(n)}$, ir „–“, 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 tikrą pastovų lygį a (t.y., jei teisinga hipotezė $H_0: MX(t) = a = const.$), tai plusų ir minusų kaita 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 $\tau_{0,05}(n)$, pasiskirsčiusiu $\tau_{0,05}(n) \approx [1,43 \cdot \ln(n + 1)]$ (čia $[x]$ žymi „sveiką x dalį”),
- įverčiais iš viršaus ir apačios tikimybėms $P\{\gamma(n) < \gamma_{0,95}(n); \tau(n) > \tau_{0,05}(n)\}$.

Tai leidžia suformuluoti tokią 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.

1.6 KORELIACINĖ ANALIZĖ

Pagrindinė šio skyrelio informacija pateikta remiantis [1] knyga.

1.6.1 EMPIRINIS KORELIACIJOS KOEFICIENTAS

Sakykime, 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 vadinama imties taškų (X_i, Y_i) , $i=1, \dots, n$, atidėtų stačiakampėje koordinatinių sistemoje, visuma.

Apibrėžimas. Dvimatės imties empiriniu skirstiniu vadinamas sąlyginis skirstinys 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

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

čia

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

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

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

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 .

1.6.2 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 yra labai svarbu žinoti, ar tarp dydžių X_i ir Y_i yra (tiesinė) priklausomybė ar ne, tai pirmiausia tikrinama hipotezė $H_0: \rho = 0$.

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

Išskirsime kelis atvejus.

- 1) Tegu imties tūris n yra pakankamai didelis. Tada empirinis koreliacijos koeficientas r turi asimptotiškai normalųjį skirstinį su vidurkiu $\mathbf{M}r = \rho$ ir dispersija $\mathbf{D}r = \frac{(1-\rho^2)^2}{n}$.

Todėl, kai didelis imties tūris n ir teisinga hipotezė $H_0: \rho = 0$, statistika

$$r\sqrt{n} \sim N(0,1).$$

Tuo remiantis pasirenkamas reikšmingumo lygmuo α ir hipotezei $H_0: \rho = 0$ tikrinti apibrėžiama kritinė sritis:

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

- 2) Kai 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 sudaroma kritinė sritis:

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

1.6.3 TIESINĖ REGRESIJA

Tiesinės regresijos teorija paimta iš [1] knygos.

Sakykime, kad turime du atsitiktinius dydžius X ir Y . Jei jų koreliacijos koeficientas artimas ± 1 , tai reiškia, kad tarp dydžių X ir Y egzistuoja beveik tiesinė priklausomybė, kuri užrašoma per sąlyginį vidurkį

$$M(Y|X = x) = \beta_0 + \beta_1 x. \quad (2)$$

(2) 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

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (3)$$

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

Kadangi kiekvienai fiksuotai X_i reikšmei x_i 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 (3) išraiška. Kitaip sakant, turime iš anksto žinomų skaičių rinkinį $(X_1, \dots, X_n)$ ir atsitiktinę nepriklausomų dydžių imtį $(Y_1, \dots, Y_n)$, kur

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \dots, n. \quad (4)$$

(4) 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“ (4) modelio nežinomų parametrų β_0, β_1 ir σ^2 taškinius 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 įverčiai $\widehat{\beta}_0$ ir $\widehat{\beta}_1$ randami mažiausių kvadratų metodu, minimizuojant dydį

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

Panaudojus būtiną ekstremumo sąlygą $\frac{\partial S^2}{\partial \beta_1} = 0$ ir $\frac{\partial S^2}{\partial \beta_0} = 0$, gaunama

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

Išsprendus šią sistemą, gaunami įverčiai

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

čia

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

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

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

(4) prielaidos dėka galima įsitikinti, kad

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

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

Remiantis šiuo faktu, pasirinkus pasiklovimo tikimybę $1-\alpha$, randami parametru β_0 ir β_1 pasikliautiniai atsitiktiniai intervalai.

Tiesinės regresijos modelis vadinamas nereikšmingu, jei parametras $\beta_1 = 0$. Hipotezę $H_0: \beta_1 = 0$ 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

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

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

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

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

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

jei H_0 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.

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

Turėdami regresijos tiesės parametru 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

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

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

čia S_1^2 charakterizuoja duomenų nukrypimą nuo vidurio, o S^2 – duomenų nukrypimą nuo regresijos linijos taškų. Sudaroma statistika

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

kuri turi Fišerio skirstinį su 1 ir $n-2$ laisvės laipsnių, kai teisinga (4) prielaida. Pasirenkamas reikšmingumo lygmuo α ir apskaičiuojamas $1-\alpha$ eilės Fišerio skirstinio su 1 ir $n-2$ laisvės laipsnių kvantilis $F_{1-\alpha}(1, n-2)$. 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.

2 PRAKTINĖ DALIS

2.1 DUOMENYS. PRADINĖ STATISTINĖ ANALIZĖ LIETUVOS MASTU

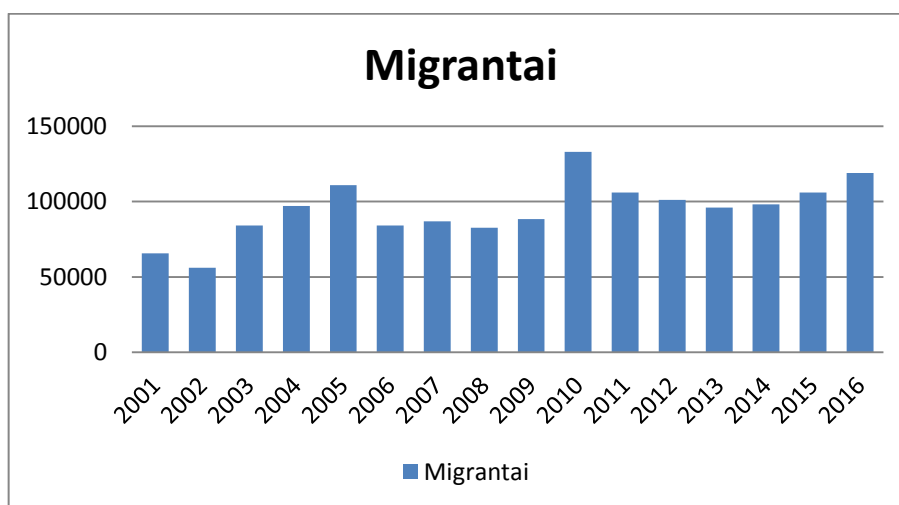
Lietuvos statistikos departamentas pateikia daug demografinių duomenų apie Lietuvos gyventojus, t.y. mirčių, gimimų, santuokų, migrantų ir t.t. skaičius (žr. [5]). Šiame darbe nagrinėjami migrantų duomenys. Skirstysime migrantus į dvi rūšis: vidinius migrantus ir išorinius (tarptautinius) migrantus. Vidiniai migrantai – tai tie, kurie juda šalies viduje, o išoriniai – kurie išvyksta iš Lietuvos ir apsigyvena užsienyje. Iš Lietuvos statistikos departamento bus paimti duomenys apie vidinius ir išorinius migrantus tiek Lietuvos, tiek ir Lietuvos apskričių mastu, o vėliau – duomenys ir apie atvykstančius ir išvykstančius iš Lietuvos apskričių žmones nagrinėjamu laikotarpiu – 2001–2016 metais.

Pirmiausia nagrinėsime migrantus Lietuvos mastu. Remiantis Lietuvos statistikos departamento duomenimis, 2001–2016 metais buvo tiek migrantų:

2001	2002	2003	2004	2005	2006	2007	2008
65732	56016	84108	97064	110854	84127	86966	82682

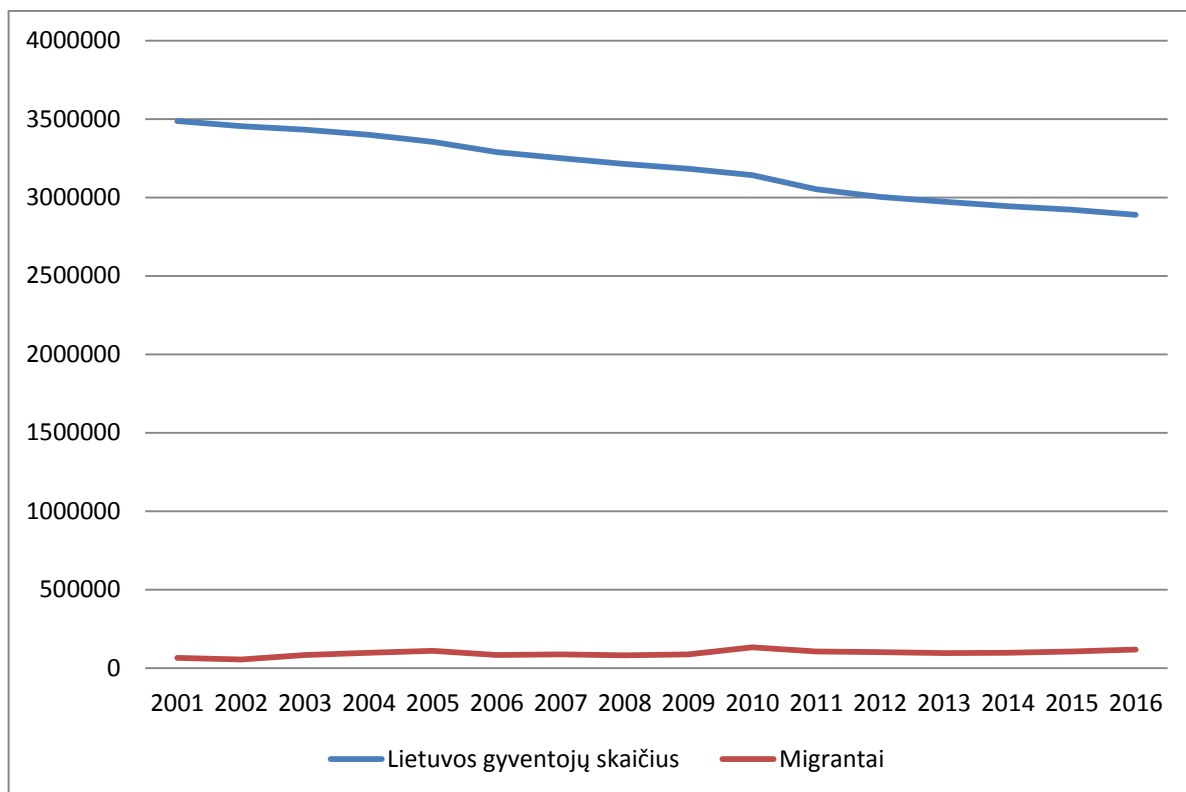
2009	2010	2011	2012	2013	2014	2015	2016
88443	132953	106062	101153	96061	98036	105959	118905

Akivaizdu, kad tokia lentelė yra mažai informatyvi. Daug patogiau naudoti diagramą (žr. 1 pav.).



1 pav. Lietuvos migrantų skaičius nuo 2001 iki 2016 metų

Migrantų ir Lietuvos gyventojų skaičių nuo 2001 iki 2016 metų pavaizduokime viename grafike, kad matytume kitimą laikui bėgant.



2 pav. Migrantų ir Lietuvos gyventojų skaičiai nuo 2001 iki 2016 metų

Iš 2 pav. matome, kad bendras Lietuvos migrantų skaičius, kurį sudaro vidinių ir išorinių migrantų suma, didėja, o bendras Lietuvos gyventojų skaičius gana sparčiai mažėja. Praktiškai per 15 metų Lietuva prarado apie 0,5 milijonų gyventojų. Jei ta tendencija išliks, tai po 90 metų lietuvių neliks.

Norėdami susidaryti aiškesnį vaizdą apie migraciją, atskirkime vidinius migrantus nuo išorinių (1 lentelė). Su turimais duomenimis atliksime pradinę statistinę analizę.

1 lentelė. Migrantų ir Lietuvos gyventojų skaičiai nuo 2001 iki 2016 metų

	2001	2002	2003	2004	2005	2006	2007	2008
Lietuvos gyventojų skaičius	3486998	3454637	3431497	3398929	3355220	3289835	3249983	3212605
Vidiniai migrantai	37891	39297	57825	59373	52969	51737	56583	56932
Išoriniai migrantai	27841	16719	26283	37691	57885	32390	30383	25750

	2009	2010	2011	2012	2013	2014	2015	2016
Lietuvos gyventojų skaičius	3183856	3141976	3052588	3003641	2971905	2943472	2921262	2888558
Vidiniai migrantai	49943	49796	52199	60053	57243	61415	61426	68572
Išoriniai migrantai	38500	83157	53863	41100	38818	36621	44533	50333

Kiekvienai šių duomenų eilutei surasime vidurkį, standartinę nuokrypį ir medianą.

Pirmą duomenų eilutę (Lietuvos gyventojų skaičius nuo 2001 metų iki 2016 metų) pažymėkime X . Pirmiausia randame variacinę eilutę:

2888558; 2921262; 2943472; 2971905; 3003641; 3052588; 3141976; 3183856; 3212605; 3249983; 3289835; 3355220; 3398929; 3431497; 3454637; 3486998.

Remiantis tuo, gauname, kad

$$\text{Mediana: } M_e = \frac{1}{2} \left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1} \right) = 3198231.$$

$$\text{Vidurkis: } \bar{X} = \frac{1}{16} \sum_{i=1}^{16} X_i = 3186685.$$

$$\text{Dispersija: } S^2 = \frac{1}{16} \sum_{i=1}^{16} (X_i - \bar{X})^2 = 39299697.$$

$$\text{Standartinis nuokrypis: } S = \sqrt{S^2} = 198242.$$

Analogiškai apskaičiuojamos ir kitų 1 lentelėje pateiktų duomenų (eilučių) statistinės charakteristikos. Gauti rezultatai pateikiami 2 lentelėje.

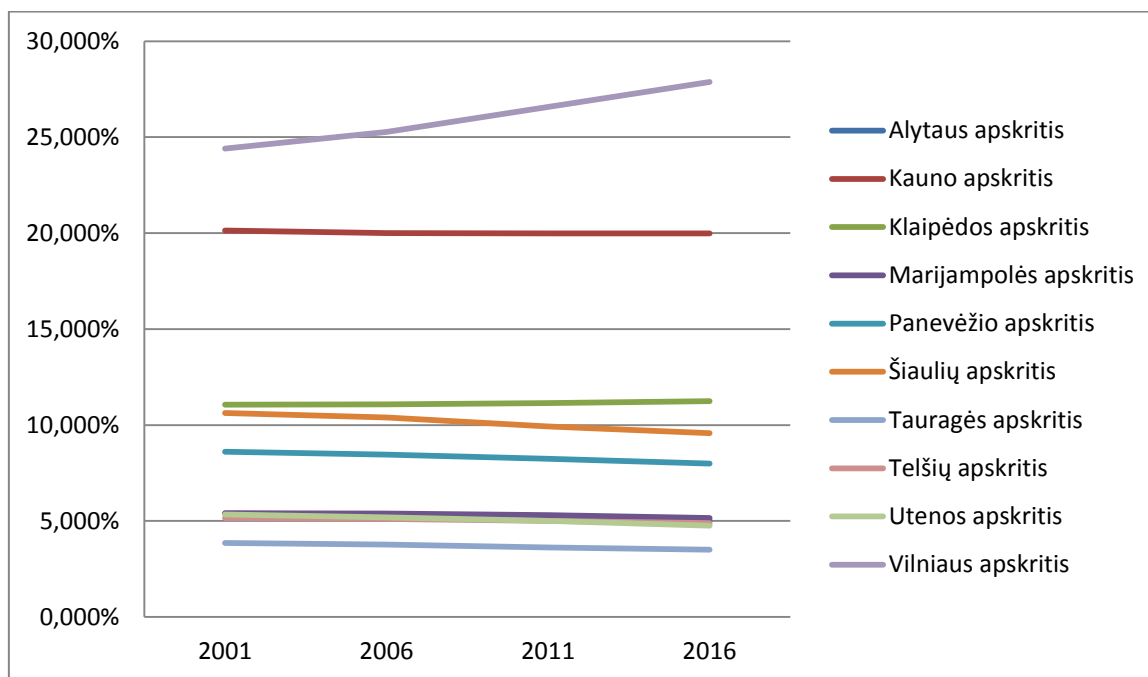
	Mediana	Vidurkis	Dispersija	Standartinis nuokrypis
Lietuvos gyventojų skaičius	3198231	3186685	39299697	198242
Vidiniai migrantai	56757	54578	58892588	7674
Išoriniai migrantai	38096	40117	234672102	15319

Išvada: Nagrinėjamu laikotarpiu Lietuvoje buvo vidutiniškai 3,186 milijono gyventojų. Kiekvienais metais šalies viduje migruodavo 54,578 tūkstančių gyventojų. Išorinių migrantų buvo vidutiniškai 40,117 tūkstančių kasmet. Didžiausias bendras migrantų (vidinių ir išorinių) skaičius užfiksuotas 2010 metais, kuris siekė 132953.

2.2 VIDINĖS MIGRACIJOS LIETUVOS APSKRITYSE FAKTO NUSTATYMAS

2.2.1 GYVENTOJŲ PASISKIRSTYMAS APSKRITYSE

Migracijos procesus veikia daug faktorių kurie atsispindi Statistikos departamento demografiniuose duomenyse kiekvienais metais. Tie duomenys apima skaičius, kiek Lietuvoje buvo gyventojų, kiek gyventojų buvo Lietuvos apskrityse, kiek gyventojų atvyko bei išvyko iš apskričių į kitas apskritis ar užsienį, kiek gimė ir mirė, kiek buvo santuokų ir t.t. Visa tai aprėpti sudėtinga, todėl geriausia nustatyti, kaip kito bendras gyventojų skaičius visose Lietuvos apskrityse laikui bėgant. Tam tikslui tinka polinominio skirstinio modelis: jį nusako k atskirų dėžių ir n vienodų rutulių, kurie gali patekti į bet kurią (k -ąją) dėžę su tikimybėmis p_k . Pagal analogiją galima laikyti, kad Lietuvos gyventojai yra rutuliai, o k dėžių – tai k apskričių, į kurias patenka (tiksliau, yra) gyventojai. Mes nežinome, su kokiomis tikimybėmis p_k patenka gyventojai į k apskritį, bet turime tos tikimybės įvertį, kurį atitinka apskrities gyventojų procentinė dalis visos Lietuvos gyventojų atžvilgiu. Praėjus metams ar didesniai laiko tarpui, jei tie gyventojų dalies apskrityse procentai pasikeičia nedaug, tai galima teigti, kad didelis žmonių judėjimas tarp Lietuvos apskričių nevyksta. Norint būti tikresniu, tuo klausimu yra gerai taikyti statistines hipotezes apie polinominio skirstinio parametrus, kaip rekomendavo V. Kanišauskas straipsnyje „Vidinės migracijos matematinis modelis“. Naudojantis tomis rekomendacijomis darbe bus bandoma išsiaiškinti, ar migracija tarp apskričių vyko, dalinant laiko skalę į 5 metų laikotarpius, nors akivaizdu, kad vidinė migracija vyko ir kasmet – tai matyti iš pradinių duomenų ir 1 patikrinto fakto 2015–2016 metais, kuris parodytas darbe.



3 pav. Procentinis gyventojų pasiskirstymas apskrityse

3 pav. puikiai matyti, kad Vilniaus apskrityje gyventojų skaičius visos Lietuvos atžvilgiu didėja sparčiausiai, jei tiksliau, tai nuo 2001 iki 2016 metų skaičius padidėjo 3,48 %. Kitaip tariant, Vilniaus apskritis auga kitų apskričių sąskaita. Įdomu tai, kad Kauno apskrityje gyventojų skaičius Lietuvos atžvilgiu tiriamu laikotarpiu išlieka stabilus, t.y. apie 20 %. Galima manyti, kad Kauno savivaldybė puikiai tvarkosi, priima teisingus sprendimus gyventojų atžvilgiu, todėl matome šį stabilumą. Klaipėdos apskrityje skaičius, kad ir nežymiai, bet per tiriamąjį laikotarpį išaugo ir tikėtina, kad augs toliau. Visose kitose apskrityse gyventojų skaičius mažėja. Sparčiausias mažėjimas matomas Šiaulių apskrityje, per tiriamą laikotarpį gyventojų skaičius sumažėjo 1,05 %. Galime teigti, kad Šiaulių apskritis „maitina“ Vilniaus apskritį. Tikėtina, kad didelę dalį šio skaičiaus sudaro studentai, kurie renkasi studijas Vilniuje. 3 lentelėje pateikti konkretūs skaičiai kiekvienai apskričiai.

3 lentelė. Procentinis gyventojų apskrityse pasiskirstymas.

Apskrities Nr.	Apskrities pav.	2001 m.	Δ	2006 m.	Δ	2011 m.	Δ	2016 m.
1	Alytaus apskritis	5,39%	-0,05%	5,34%	-0,15%	5,19%	-0,17%	5,02%
2	Kauno apskritis	20,13%	-0,13%	20,00%	-0,01%	19,99%	0,00%	19,99%
3	Klaipėdos apskritis	11,07%	0,01%	11,08%	0,06%	11,14%	0,10%	11,24%
4	Marijampolės apskritis	5,41%	-0,02%	5,39%	-0,08%	5,31%	-0,15%	5,16%
5	Panevėžio apskritis	8,61%	-0,15%	8,46%	-0,22%	8,24%	-0,24%	8,00%
6	Šiaulių apskritis	10,62%	-0,23%	10,39%	-0,46%	9,93%	-0,36%	9,57%
7	Tauragės apskritis	3,85%	-0,08%	3,77%	-0,15%	3,62%	-0,12%	3,50%
8	Telšių apskritis	5,16%	-0,06%	5,10%	-0,10%	5,00%	-0,11%	4,89%
9	Utenos apskritis	5,34%	-0,15%	5,19%	-0,19%	5,00%	-0,25%	4,75%
10	Vilniaus apskritis	24,40%	0,87%	25,27%	1,31%	26,58%	1,30%	27,88%
Gyventojai visoje Lietuvoje		3486998		3289835		3052588		2888558

2.2.2 VIDINĖS MIGRACIJOS 2001–2006 M. FAKTO NUSTATYMAS

Matematiškai patikrinsime, ar gyventojų kiekio Alytaus, Kauno, Klaipėdos, Marijampolės, Panevėžio, Šiaulių, Tauragės, Telšių, Utenos ir Vilniaus apskrityse procentinės dalies pokyčiai, laikui pasislinkus iš momento $t_1 = 2001$ į $t_2 = 2006$, yra statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,01$. Tam tikslui tikriname statistinę hipotezę $H_0^{(1)}$: $p_1 = 0,0539$, $p_2 = 0,2013$, $p_3 = 0,1107$, $p_4 = 0,0541$, $p_5 = 0,0861$, $p_6 = 0,1062$, $p_7 = 0,0385$, $p_8 = 0,0516$, $p_9 = 0,0534$, $p_{10} = 0,2440$. Skaičiai 0,0539; 0,2013; 0,1107; 0,0541; 0,0861; 0,1062; 0,0385; 0,0516; 0,0534; 0,2440 gaunami Lietuvos apskričių procentines dalis 2001 m. padalinus iš 100. 2006 m. p_i reikšmės virto $\hat{p}_i$, kai $i = 1, \dots, 10$. Mūsų atveju 2006 m. apskričių procentinės dalys, padalintos iš 100 bus: $\hat{p}_1 = 0,0534$, $\hat{p}_2 = 0,2000$, $\hat{p}_3 = 0,1108$, $\hat{p}_4 = 0,0539$, $\hat{p}_5 = 0,0846$, $\hat{p}_6 = 0,1039$, $\hat{p}_7 = 0,0377$, $\hat{p}_8 = 0,0510$, $\hat{p}_9 = 0,0519$, $\hat{p}_{10} = 0,2527$. Naujoje situacijoje $n = n(2006) = 3289835$ (visų Lietuvos gyventojų skaičius 2006 m.). Čia duomenys paimti iš 3 lentelės. Viską surašome į Pirsono statistiką ir gauname

$$\chi^2 = 3289835 \left(\frac{(0,0534 - 0,0539)^2}{0,0539} + \frac{(0,2000 - 0,2013)^2}{0,2013} + \frac{(0,1108 - 0,1107)^2}{0,1107} + \frac{(0,0539 - 0,0541)^2}{0,0541} + \frac{(0,0846 - 0,0861)^2}{0,0861} + \frac{(0,1039 - 0,1062)^2}{0,1062} + \frac{(0,0377 - 0,0385)^2}{0,0385} + \frac{(0,0510 - 0,0516)^2}{0,0516} + \frac{(0,0519 - 0,0534)^2}{0,0534} + \frac{(0,2527 - 0,2440)^2}{0,2440} \right) = 1532,2318.$$

Į formulę $\chi^2_{1-\alpha}(k-1)$ įrašę $k = 10$ ir $\alpha = 0,01$ gauname $\chi^2_{0,99}(9) = 27,9$ (reikšmė paimta iš kvantilių lentelės [1]). Mūsų atveju $\chi^2 = 1532,2318 > \chi^2_{0,99}(9) = 27,9$, todėl hipotezę $H_0^{(1)}$ apie tai, kad tarp Alytaus, Kauno, Klaipėdos, Marijampolės, Panevėžio, Šiaulių, Tauragės, Telšių, Utenos ir Vilniaus apskričių migracija nevyksta, atmetame. Taigi, **migracija tarp apskričių vyko** ir situacija per 5 metus kardinaliai pasikeitė, gyventojų skaičiaus pokyčiai buvo reikšmingi.

2.2.3 VIDINĖS MIGRACIJOS 2015–2016 M. FAKTO NUSTATYMAS

Papildomai patikrinsime, ar migracija tarp apskričių vyko 2015–2016 metais. Tam tikslui tikriname statistinę hipotezę apie polinomino skirstinio parametrus $H_0^{(2)}$: $p_1 = 0,0506$, $p_2 = 0,1995$, $p_3 = 0,1120$, $p_4 = 0,0519$, $p_5 = 0,0805$, $p_6 = 0,0964$, $p_7 = 0,0352$, $p_8 = 0,0491$, $p_9 = 0,479$, $p_{10} = 0,2764$, $\hat{p}_1 = 0,0502$, $\hat{p}_2 = 0,1999$, $\hat{p}_3 = 0,1124$, $\hat{p}_4 = 0,0516$, $\hat{p}_5 = 0,0800$, $\hat{p}_6 = 0,0957$, $\hat{p}_7 = 0,0350$, $\hat{p}_8 = 0,0489$, $\hat{p}_9 = 0,0475$, $\hat{p}_{10} = 0,2788$. $n = n(2016) = 2888558$. Duomenis surašome į Pirsono statistiką ir gauname

$$\begin{aligned} \chi^2 = 2888558 & \left(\frac{(0,0502 - 0,0506)^2}{0,0506} + \frac{(0,1999 - 0,1995)^2}{0,1995} + \frac{(0,1124 - 0,1120)^2}{0,1120} \right. \\ & + \frac{(0,0516 - 0,0519)^2}{0,0519} + \frac{(0,0800 - 0,0805)^2}{0,0805} + \frac{(0,0957 - 0,0964)^2}{0,0964} \\ & + \frac{(0,0350 - 0,0352)^2}{0,0352} + \frac{(0,0489 - 0,0491)^2}{0,0491} + \frac{(0,0475 - 0,0479)^2}{0,0479} \\ & \left. + \frac{(0,2788 - 0,2764)^2}{0,2764} \right) = 130,5857. \end{aligned}$$

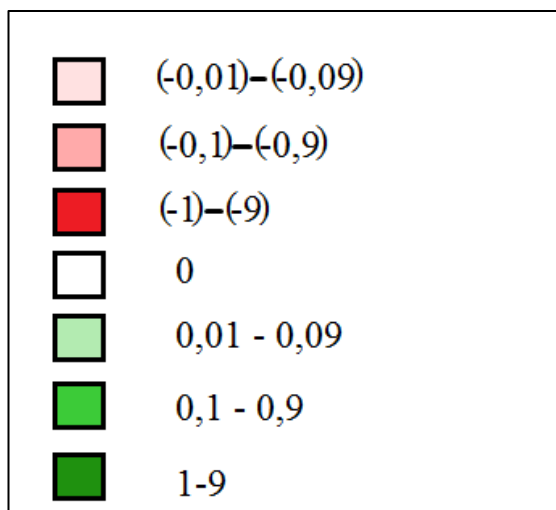
Hipotezė $H_0^{(2)}$ taip pat atmetama, kadangi $\chi^2 = 130,5857 > \chi^2_{0,99}(9) = 27,9$.

Galime daryti išvadą, kad **vidinė migracija 2015–2016 m. vyko**. Kadangi statistika $\chi^2 = 130,5857$ gana didelė, o pokyčiai tarp Lietuvos apskričių gyventojų kiekvienais metais nuo 2001 iki 2016 m. yra panašūs, tai ir analogiškos statistinės hipotezės bus atmestos, kas patvirtins vidinės migracijos kardinalius pokyčius, vykstančius tarp Lietuvos apskričių kasmet.

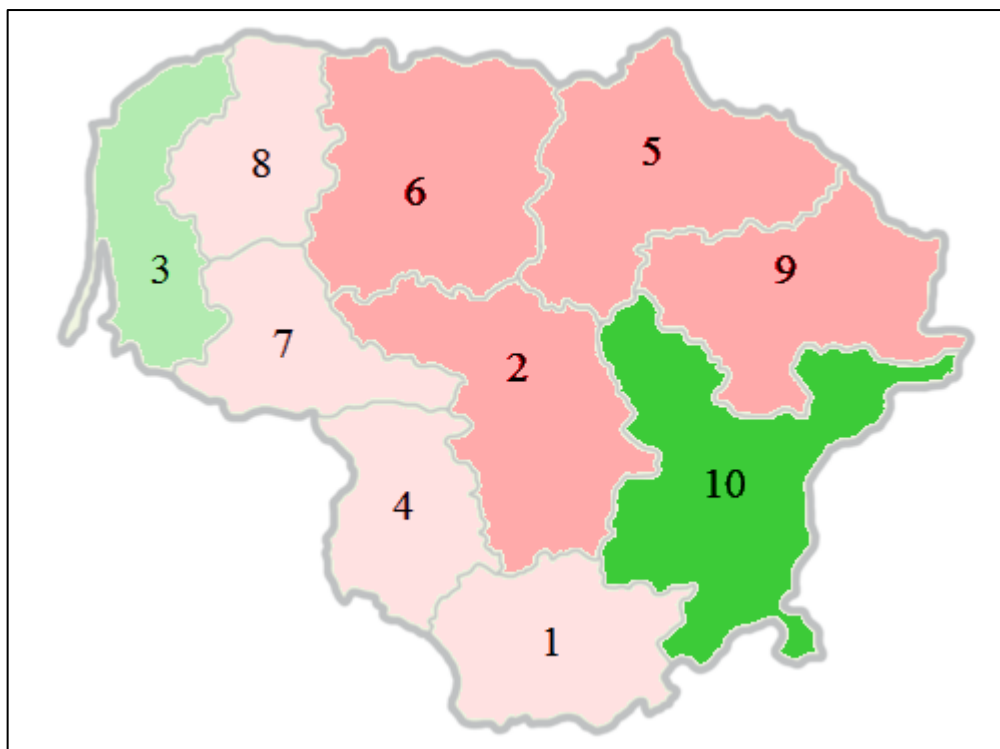
2.2.4 VIDINĖS MIGRACIJOS POKYČIŲ LIETUVOS APSKRITYSE ATSPINDYS LIETUVOS ŽEMĖLAPYJE

Kaip vyko vidinės migracijos procesai Lietuvos apskrityse, geriausia pavaizduoti Lietuvos žemėlapyje. Tai ir bus padaryta.

Žemėlapiuose raudona spalva reiškia gyventojų mažėjimą, o žalia atvirkščiai – didėjimą. Kuo ryškesnė (tamsesnė) spalva, tuo pokytis didesnis. 4 pav. pateikta spalvų skalė su reikšmėmis.

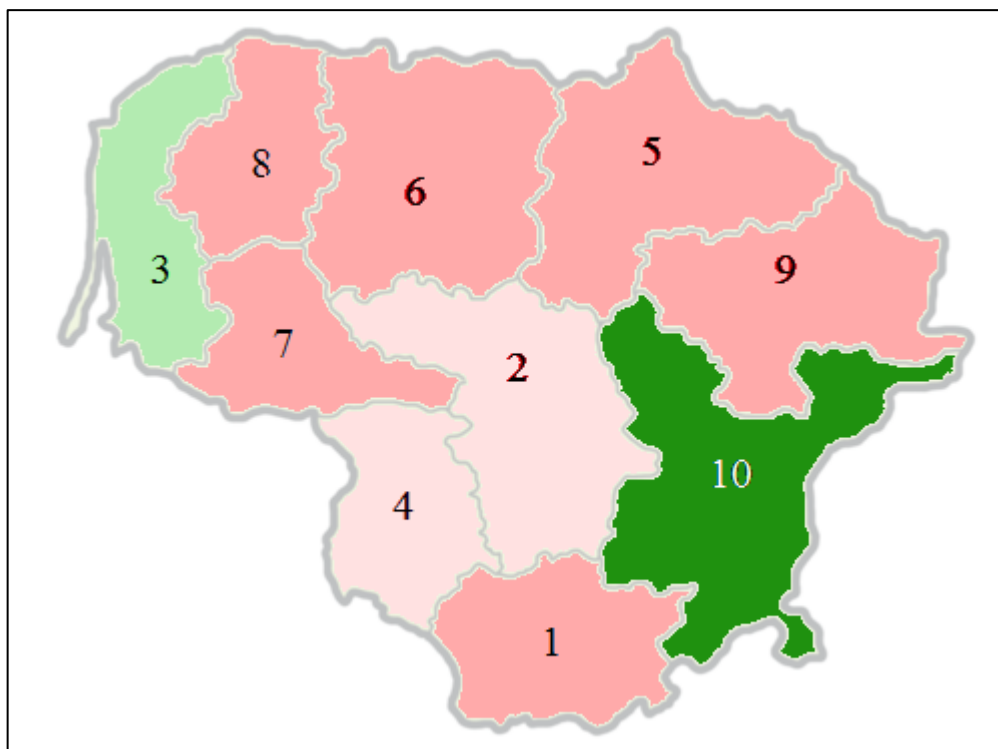


4 pav. Spalvų reikšmės Lietuvos žemėlapyje



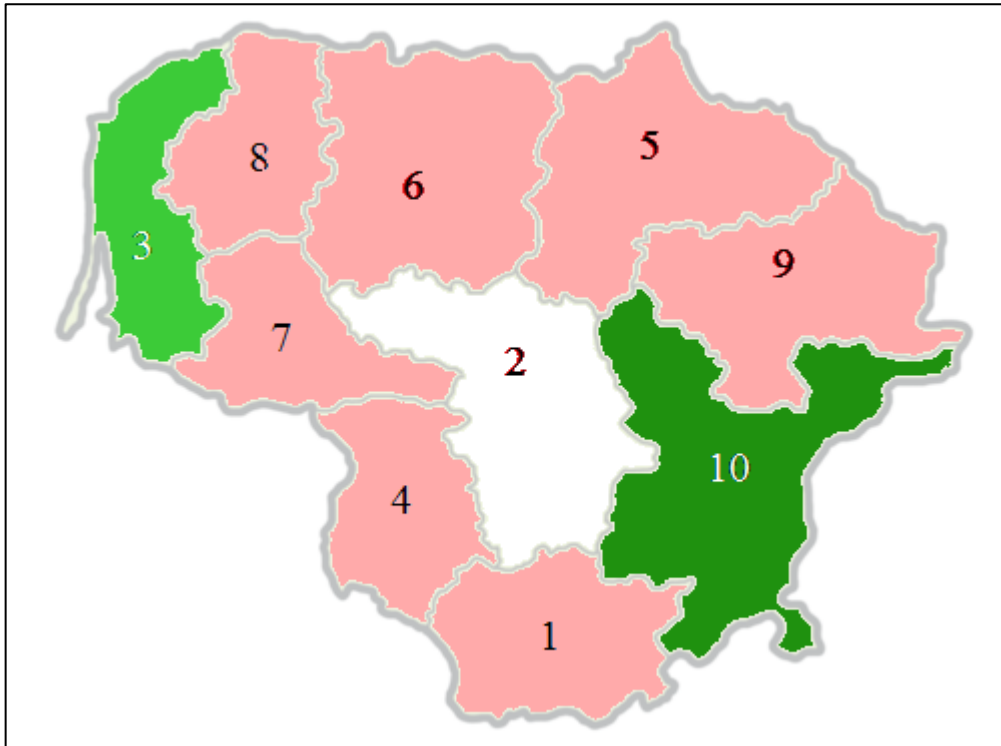
5 pav. Migracijos apskrityse pokytis 2001–2006 metais

Iš 5 pav. galime matyti, kad Lietuvos gyventojai labiausiai vyko į Vilniaus apskritį. Klaipėda žmones domino šiek tiek mažiau. Įdomu tai, kad Kauno apskrities gyventojai ganėtinai intensyviai iškeliavo iš savo apskrities. Neatsilieka ir Šiaulių, Panevėžio bei Utenos apskritys. O likusiose Lietuvos apskrityse gyventojų migracija šiek tiek mažesnė nei anksčiau minėtuose apskrityse.



6 pav. Migracijos apskrityse pokytis 2006–2011 metais

6 pav. akivaizdžiai matyti, kad Kauno apskrities gyventojai migruoja gerokai mažiau, ten situacija gerėja. Klaipėdos apskritis, nors ir ženkliai mažiau nei Vilniaus, vis dar traukia gyventojus, o likusioje Lietuvoje stipriai migruojama iš apskričių.



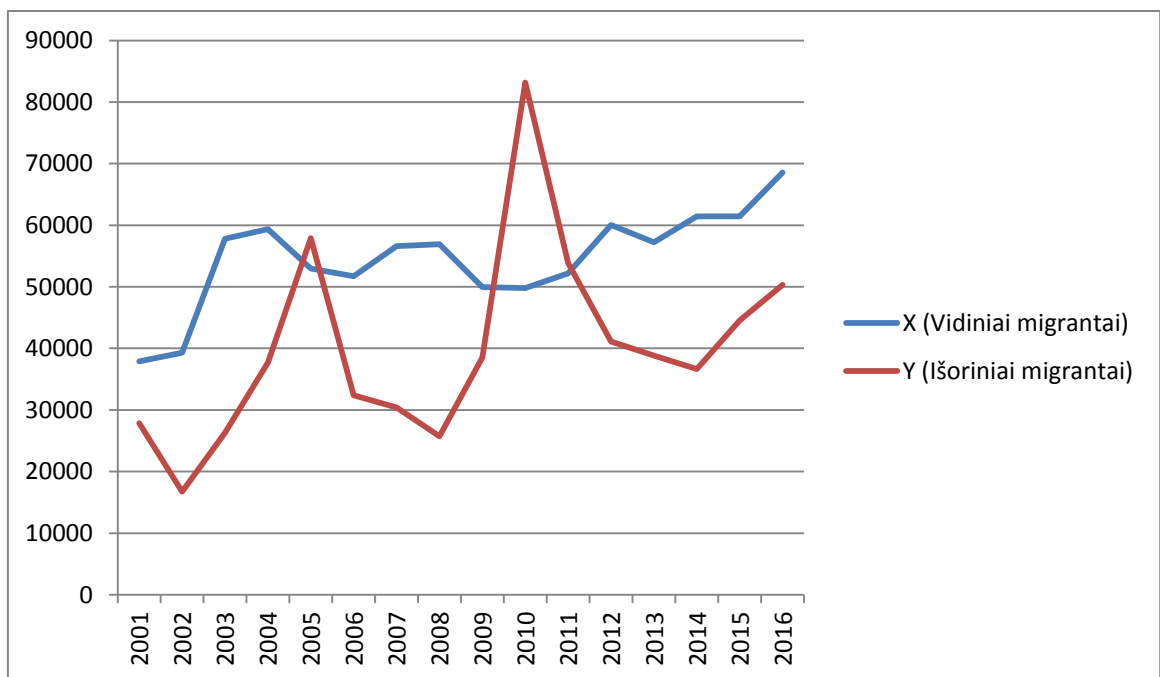
7 pav. Migracijos apskrityse pokytis 2011–2016 metais

2011–2016 metais Kauno apskrities gyventojų migracija stabilizavosi, o tai rodo, kad galimybė pristabdyti migraciją yra. Kauno apskrities valdantieji ėmėsi teigiamų veiksmų, kurie paskatino žmones nekeisti gyvenamosios vietos. Vilniaus bei Klaipėdos apskritys toliau traukia žmones, kai tuo tarpu iš likusių apskričių gyventojai migruoja.

2.3 PRIKLAUSOMYBĖ TARP VIDINĖS IR IŠORINĖS MIGRACIJŲ

2.3.1 LIETUVOS MASTU, PAGAL 2001–2016 M. DUOMENIS

Statistikos departamentas pateikia duomenis atskirai apie vidinius ir išorinius migrantus. Turime vidinių bei išorinių migrantų skaičius nuo 2001 iki 2016 metų. Kad galėtume susidaryti geresnį vaizdą apie tuos duomenis, brėžiame grafiką (žr. 8 pav.).



8 pav. Vidinių ir išorinių migrantų dinamika 2001–2016 metais Lietuvoje

Iš 8 pav. matome, kad vidinių migrantų duomenys maždaug tolygūs, kai tuo tarpu išorinių migrantų duomenys 2005 ir 2010 metais stipriai pakyla. Patikrinsime, ar yra statistinė priklausomybė tarp vidinės bei išorinės migracijos Lietuvos mastu per visą laikotarpį (2001–2016 m.). Rasime empirinį koreliacijos koeficientą pagal formulę:

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

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

4 lentelė. Vidinių ir išorinių migrantų skaičiai 2001–2016 metais

	X (Vidiniai migrantai)	Y (Išoriniai migrantai)
2001	37891	27841
2002	39297	16719
2003	57825	26283
2004	59373	37691
2005	52969	57885
2006	51737	32390
2007	56583	30383
2008	56932	25750
2009	49943	38500
2010	49796	83157
2011	52199	53863
2012	60053	41100
2013	57243	38818
2014	61415	36621
2015	61426	44533
2016	68572	50333

Duomenis iš 4 lentelės naudojame tolimesniems skaičiavimams.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{16} (37891 + 39297 + 57825 + 59373 + 52969 + 51737 + 56583 + 56932 + 49943 + 49796 + 52199 + 60053 + 57243 + 61415 + 61426 + 68572) = 54578,375.$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{16} (27841 + 16719 + 26283 + 37691 + 57885 + 32390 + 30383 + 25750 + 38500 + 83157 + 53863 + 41100 + 38818 + 36621 + 44533 + 50333) = 40116,688.$$

$$\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = 366067081,875.$$

$$\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = 1880966389,948.$$

Tada viską surašome į formulę ir gauname

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{366067081,875}{1880966389,948} = 0,195.$$

Empirinį koreliacijos koeficientą vidinių bei išorinių migrantų duomenims nuo 2001 iki 2016 metų gavome mažesnį nei 0,2, ir tai rodo, kad priklausomybė tarp tų duomenų nedidelė. Norėdami visiškai įsitikinti rezultato teisingumu, tikriname statistinę hipotezę $H_0: \rho = 0$. Sudarome statistiką:

$$r\sqrt{n} \sim N(0,1);$$

$$r\sqrt{n} = 0,195 \cdot \sqrt{16} = 0,78.$$

Tuo remdamiesi hipotezei $H_0: \rho = 0$ tikrinti su reikšmingumo lygmeniu $\alpha=0,01$ sudarome kritinę sritį:

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

Kadangi $|r|\sqrt{n} = 0,78 \leq 2,576 = u_{0,995}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = 0$ teisingumui.

2.3.2 LIETUVOS MASTU PAGAL APSKRITIS

Lietuvos gyventojai migruoja iš savo apskrities kiekvienais metais. Todėl yra įdomu ištirti, ar vidinių ir išorinių migrantų skaičiai Lietuvos apskrityse yra tarpusavyje susiję, t.y. priklausomi. Tam tikslui pagal Lietuvos statistikos departamento duomenis išrašome vidinių ir išorinių migrantų skaičius tyrimui (5–8 lent.).

2001 m.

5 lentelė. Vidinių ir išorinių migrantų skaičiai 2001 metais

2001		
Apskrities Nr.	X	Y
1	2291	1018
2	8160	5674
3	3727	4263
4	1714	1109
5	4320	2301
6	4894	3247
7	1509	1077
8	2822	1485
9	2693	1614
10	5761	6053

Atliekami skaičiavimai:

$$\bar{X} = 3789,1, \bar{Y} = 2784,1.$$

$$\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = 31793607,9,$$

$$\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = 35896113,12,$$

$$r = \frac{31793607,9}{35896113,12} = 0,886.$$

Norėdami visiškai įsitikinti rezultato teisingumu, tikriname statistinę hipotezę $H_0: \rho = 0,9$ taikydami R. A. Fišerio transformaciją:

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

$$Arthr = \frac{1}{2} \ln \frac{1+0,886}{1-0,886} = 1,40$$

$$Arth0,9 = \frac{1}{2} \ln \frac{1+0,9}{1-0,9} = 1,47$$

$$Z = \sqrt{n-3}(Arthr - Arth0,9)$$

$$Z = \sqrt{10-3} \cdot (1,40 - 1,47) = -0,19$$

Tuo remdamiesi hipotezei $H_0: \rho = 0,9$ tikrinti su reikšmingumo lygmeniu $\alpha=0,01$ sudarome kritinę sritį:

$$W_1 = \{(X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}}\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq 0,9.$$

Kadangi $|Z| = 0,19 \leq 2,576 = u_{0,995}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = 0,9$ teisingumui.

Išvada: priklausomybė pagal apskritis tarp išorinių ir vidinių migrantų pagal 2001 m. duomenis yra didelė.

2006 m.

6 lentelė. Vidinių ir išorinių migrantų skaičiai 2006 metais

2006		
Apskritis Nr.	X	Y
1	3229	1149
2	10077	6577
3	6322	4689
4	3299	1669
5	4705	2645
6	6647	4356
7	2432	1414
8	3568	2613
9	2824	1773
10	8634	5205

$$\bar{X} = 5173,7; \bar{Y} = 3209; r = \frac{43220417}{44353129,51} = 0,975.$$

2006 metų vidinių ir išorinių migrantų duomenims tikriname hipotezę $H_0: \rho = 0,9$.

$$Arthr = \frac{1}{2} \ln \frac{1 + 0,975}{1 - 0,975} = 2,18$$

$$Arth0,9 = \frac{1}{2} \ln \frac{1 + 0,9}{1 - 0,9} = 1,47$$

$$Z = \sqrt{n-3}(Arthr - Arth0,9)$$

$$Z = \sqrt{10-3} \cdot (2,18 - 1,47) = 1,88$$

Tuo remdamiesi hipotezei $H_0: \rho = 0,9$ tikrinti su reikšmingumo lygmeniu $\alpha = 0,01$ sudarome kritinę sritį:

$$W_1 = \left\{ (X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}} \right\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq 0,9.$$

Kadangi $|Z| = 1,88 \leq 2,576 = u_{0,995}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = 0,9$ teisingumui.

Išvada: priklausomybė pagal apskritis tarp išorinių ir vidinių migrantų pagal 2006 m. duomenis yra didelė.

2011 m.

7 lentelė. Vidinių ir išorinių migrantų skaičiai 2011 metais

2011		
Apskritis Nr.	X	Y
1	3294	2868
2	9995	11820
3	6088	6992
4	3122	2495
5	4695	3553
6	6648	7079
7	2282	1848
8	3243	2679
9	3237	2309
10	9595	12220

$$\bar{X} = 5219,9; \bar{Y} = 5386,3; r = \frac{98158097,30}{99029533,16} = 0,991.$$

2011 metų migrantų duomenims patikrinkime statistinę hipotezę $H_0: \rho = 0,99$.

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

$$Arthr = \frac{1}{2} \ln \frac{1 + 0,991}{1 - 0,991} = 2,70$$

$$Arth_{0,99} = \frac{1}{2} \ln \frac{1 + 0,99}{1 - 0,99} = 2,65$$

$$Z = \sqrt{n - 3} (Arthr - Arth_{0,99})$$

$$Z = \sqrt{10 - 3} \cdot (2,70 - 2,65) = 0,13$$

Tuo remdamesi hipotezei $H_0: \rho = 0,99$ tikrinti su reikšmingumo lygmeniu $\alpha = 0,01$ sudarome kritinę sritį:

$$W_1 = \left\{ (X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}} \right\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq 0,99.$$

Kadangi $|Z| = 0,13 \leq 2,576 = u_{0,995}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = 0,99$ teisingumui.

Išvada: priklausomybė pagal apskritis tarp išorinių ir vidinių migrantų pagal 2011 m. duomenis yra didelė.

2016 m.

8 lentelė. Vidinių ir išorinių migrantų skaičiai 2016 metais

2016		
Apskritis Nr.	X	Y
1	3736	2602
2	13310	9891
3	8694	6512
4	3723	2933
5	5961	3736
6	7946	5742
7	2942	1744
8	4623	2710
9	3725	2312
10	13912	12151

$$\bar{X} = 6857,2; \bar{Y} = 5033,3; r = \frac{239549910}{240313261} = 0,989.$$

2011 metų migrantų duomenims patikrinkime statistinę hipotezę $H_0: \rho = 0,99$.

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

$$Arthr = \frac{1}{2} \ln \frac{1 + 0,989}{1 - 0,989} = 2,60$$

$$Arth0,99 = \frac{1}{2} \ln \frac{1 + 0,99}{1 - 0,99} = 2,65$$

$$Z = \sqrt{n - 3}(Arthr - Arth0,9)$$

$$Z = \sqrt{10 - 3} \cdot (2,60 - 2,65) = -0,13$$

Tuo remdamesi hipotezei $H_0: \rho = 0,99$ tikrinti su reikšmingumo lygmeniu $\alpha = 0,01$ sudarome kritinę sritį:

$$W_1 = \left\{ (X_i, Y_i), i = 1 \dots, n: |Z| > u_{1-\frac{\alpha}{2}} \right\}, \text{ jei } H_0 \text{ alternatyva yra } H_1^{(1)}: \rho \neq 0,99.$$

Kadangi $|Z| = 0,13 \leq 2,576 = u_{0,995}$, tai hipotezė H_0 priimama. Galime teigti, kad statistiniai duomenys neprieštarauja hipotezės $H_0: \rho = 0,99$ teisingumui.

Išvada: priklausomybė pagal apskritis tarp išorinių ir vidinių migrantų pagal 2016 m. duomenis yra didelė.

Bendroji išvada: Lietuvos mastu, kai duomenys išskaidyti pagal apskritis, priklausomybė tarp vidinės ir išorinės migracijų yra didelė, bet kai tie duomenys sudedami į vieną skaičių (Lietuvos mastu be apskričių), ta priklausomybė išnyksta.

2.3.3 PRIKLAUSOMYBĖ TARP MIGRANTŲ IR LIETUVOS GYVENTOJŲ

Patikrinsime, ar yra statistinė priklausomybė tarp migrantų (vidinių ir išorinių) ir visų Lietuvos gyventojų visu laikotarpiu (2001–2016 m.).

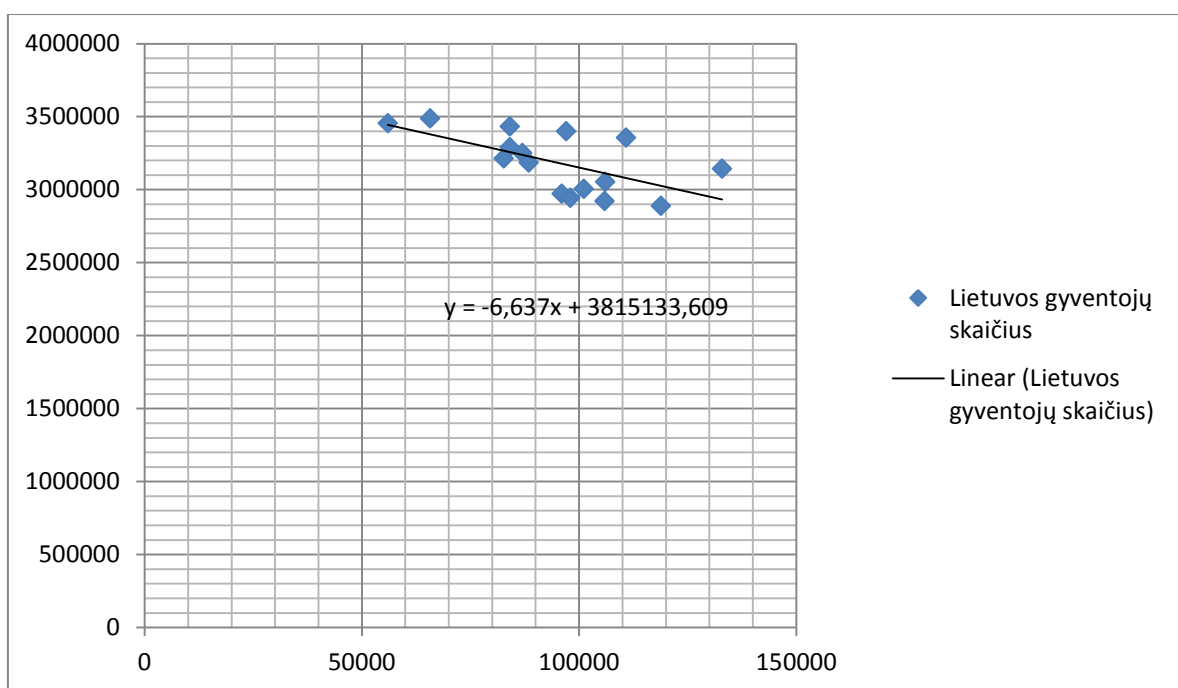
9 lentelė. Migrantų (vidinių ir išorinių) ir Lietuvos gyventojų skaičius 2001–2016 metais

Metai	X – Migrantai	Y – Nuolatinių gyventojų skaičius metų pradžioje
2001	65732	3486998
2002	56016	3454637
2003	84108	3431497
2004	97064	3398929
2005	110854	3355220
2006	84127	3289835
2007	86966	3249983
2008	82682	3212605

2009	88443	3183856
2010	132953	3141976
2011	106062	3052588
2012	101153	3003641
2013	96061	2971905
2014	98036	2943472
2015	105959	2921262
2016	118905	2888558

$$\bar{X} = 94695,063; \bar{Y} = 3186685,13; r = \frac{-36030950992,125}{58428035088,910} = -0,617.$$

Gautas empirinis koreliacijos koeficientas rodo, kad priklausomybė tarp migrantų ir Lietuvos gyventojų skaičiaus yra vidutinė. 9 lentelės duomenims MS Excel programa nubrėžiame grafiką (žr. 9 pav.).



9 pav. Lietuvos gyventojų ir migrantų dinamika

Norėdami įsitikinti, ar paskutiniame tyrime yra maža ar vidutinė priklausomybė tarp duomenų, papildomai tikriname statistinę hipotezę $H_0: \rho = 0$.

$$r\sqrt{n} = -0,617 \cdot \sqrt{16} = -2,468;$$

Tuo remdamiesi hipotezei $H_0: \rho = 0$ tikrinti su reikšmingumo lygmeniu $\alpha = 0,01$ sudarome kritinę sritį:

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

Kadangi $|r|\sqrt{n} = 2,468 < 2,576 = u_{0,995}$, tai hipotezė H_0 priimama.

Išvada: Priklausomybė tarp visų migrantų ir Lietuvos gyventojų, pagal 2001–2016 m. duomenis, yra maža.

2.4 KOMPLEKSIŠKAS LAIKO EILUČIŲ TAIKYMAS

2.4.1 PRIELAUDŲ LAIKO EILUTĖMS TAIKYTI PATIKRINIMAS

Kadangi duomenys apie migrantus statistikos departamente pateikti kasmet, galime daryti prielaidą, kad tiems duomenims naudotina laiko eilučių teorija, nes toje teorijoje svarbu, kad duomenys būtų pateikti vienodu intervalu. Tikriname hipotezę H_0 : duomenys nepriklausomi ir atsitiktiniai. Taikome serijų kriterijų su mediana (α yra duotas).

11 lentelė. Migrantai (vidiniai ir išoriniai), tūkstančiais

	Alytaus	Kauno	Klaipėdos	Marijampolės	Panevėžio	Šiaulių	Tauragės	Telšių	Utenos	Vilniaus
2001	3,309	13,834	7,990	2,823	6,621	8,141	2,586	4,307	4,307	11,814
2002	2,761	11,734	6,346	2,923	5,586	7,029	2,346	3,485	3,996	9,810
2003	4,877	17,539	10,185	4,253	8,017	10,181	3,295	5,397	5,139	15,225
2004	5,961	20,093	11,909	5,110	8,934	11,882	4,252	6,284	5,728	16,911
2005	6,460	23,077	14,788	5,979	9,406	13,623	5,074	7,353	6,016	19,078
2006	4,678	16,654	11,011	4,968	7,350	11,003	3,846	6,181	4,597	13,839
2007	4,626	17,171	11,390	4,862	7,706	11,925	3,999	6,002	4,797	14,488
2008	4,631	16,756	10,683	4,607	7,069	10,785	3,506	5,568	4,829	14,248
2009	4,638	18,232	11,810	4,936	7,694	11,879	3,591	5,461	4,785	15,417
2010	8,141	28,314	16,674	6,920	10,442	17,602	5,191	7,798	6,488	25,383
2011	6,162	21,815	13,080	5,617	8,248	13,727	4,130	5,922	5,546	21,815
2012	5,655	20,157	12,154	5,843	7,989	12,331	4,259	5,969	5,555	21,241
2013	5,051	19,296	11,864	5,280	7,573	11,748	3,815	5,374	5,545	20,515
2014	5,130	19,806	12,368	5,474	7,571	11,744	3,853	5,522	5,083	21,485
2015	5,746	20,313	13,120	5,655	8,111	12,707	4,033	5,652	5,403	25,219
2016	6,338	23,201	15,206	6,656	9,697	13,688	4,686	7,333	6,037	26,063

Šiaulių apskritis

Pradėsime nuo Šiaulių apskrities. Duomenis išrašome didėjimo tvarka:

7,029; 8,141; 10,181; 10,785; 11,003; 11,744; 11,748; 11,879; 11,882; 11,925; 12,331; 12,707; 13,623; 13,688; 13,727; 17,602.

Surandame medianą: $M_e = \frac{1}{2}(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}) = 11,881$.

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$N = N(\text{serijų skaičius}) = 8$.

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\alpha = 0,1.$$

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

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

Jeigu $N_{0,05}(A, 8, 8) < N < N_{0,05}(V, 8, 8)$, tada H_0 teisinga. Kadangi $5 < 8 < 13$, H_0 priimame. Tai reiškia, kad duomenys nepriklausomi ir atsitiktiniai, todėl laiko eilučių teorija netinka. Bet galima taikyti klasikinės matematinės statistikos teoriją (empirinis vidurkis, empirinė dispersija, empirinė moda, empirinė mediana, histograma ir t.t.).

Toliau tą patį procesą analogiškai taikome kiekvienai apskrčiai.

Alytaus apskritis

$$M_e = 5,091.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Kauno apskritis

$$M_e = 19,551.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Klaipėdos apskritis

$$M_e = 11,887.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Marijampolės apskritis

$$M_e = 5,195.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 4.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 4 < 13\} \Rightarrow H_0 \text{ atmetama.}$$

Panevėžio apskritis

$$M_e = 7,848.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Tauragės apskritis

$$M_e = 3,926.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 8.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 8 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Telšių apskritis

$$M_e = 5,787.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Utenos apskritis

$$M_e = 5,271.$$

Lyginame

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 6.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\bar{W} = \{5 < 6 < 13\} \Rightarrow H_0 \text{ priimama.}$$

Vilniaus apskritis

Bandome naudoti serijų kriterijų Vilniaus apskrities duomenims, juos išrašome didėjimo tvarka: 9,810; 11,814; 13,839; 14,248; 14,488; 15,225; 15,417; 16,911; 19,078; 20,515; 21,241; 21,485; 21,815; 25,219; 25,383; 26,063.

$$M_e = \frac{1}{2} \left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1} \right) = 17,995.$$

Lyginame:

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

H_0 : duomenys nepriklausomi ir atsitiktiniai.

$$N = N(\text{serijų skaičius}) = 4.$$

$$k_1 = 8, k_2 = 8.$$

$$n_1 = 8, n_2 = 8.$$

$$\alpha = 0,1.$$

$$N_{0,05}(4, 8, 8) = 5.$$

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

Jeigu $N_{0,05}(A, 8, 8) < N < N_{0,05}(V, 8, 8)$, tada H_0 teisinga. Kadangi įrašę skaičius gauname $5 < 4 < 13$, H_0 atmetame.

Išvada: Pritaikę serijų kriterijų kiekvienos Lietuvos apskrities duomenims gavome, kad laiko eilučių teoriją taikyti galima tik Vilniaus ir Marijampolės apskrityms.

Kadangi Vilnius yra Lietuvos sostinė, todėl mus labiau domina Vilniaus apskritis ir tolimesni skaičiavimai bus atlikti naudojant tos apskrities duomenis.

2.5 LAIKO EILUČIŲ TAIKYMAS VILNIAUS APSKRIČIAI

2.5.1 TRENDŲ ELIMINAVIMAS IR SURADIMAS

Dabar patikrinsime, ar laiko eilutė X_t turi trendą.

$$H_0: MX_t = a = \text{const.}$$

$\gamma(n)$ – bendras serijų skaičius = 4. $\tau(n)$ – ilgiausios serijos ilgis = 7.

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

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

Jei viena iš aukščiau pateiktų nelygybių negalioja, H_0 atmetame.

$$4 > \left\lceil \frac{1}{2}(16 + 2 - 1,96\sqrt{16 - 1}) \right\rceil \Leftrightarrow 4 > 5.$$

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

Tarkime, kad trendas yra tiesinis:

$$f_t = \widehat{\beta}_0 + \widehat{\beta}_1 t.$$

MS Excel programa suskaičiuojame įverčius ir gauname lygtį $y = 0,8886t + 10,732$. Atlikę skaičiavimus programoje MS Excel, eliminavome trendą ir gavome naujus duomenis.

2.5.2 AR(1) SURADIMAS

12 lentelė. Vilniaus apskrities duomenys eliminavus trendą

Metai	t	Y_t
2001	1	0,193
2002	2	-2,699
2003	3	1,827
2004	4	2,625
2005	5	3,903
2006	6	-2,225
2007	7	-2,464
2008	8	-3,593
2009	9	-3,312
2010	10	5,765
2011	11	1,308
2012	12	-0,154
2013	13	-1,769
2014	14	-1,687
2015	15	1,158
2016	16	1,113

Surasime AR(1) Vilniaus apskrities duomenims. Tarkime, turime autoregresijos procesą, išreikštą lygtimi $Y_t = a_1 \cdot Y_{t-1} + \varepsilon_t$, ir stebinius: $Y_1 = 0,193, Y_2 = -2,699, Y_3 = 1,827, Y_4 = 2,625, Y_5 = 3,903, Y_6 = -2,225, Y_7 = -2,464, Y_8 = -3,593, Y_9 = -3,312, Y_{10} = 5,765, Y_{11} = 1,308, Y_{12} = -0,154, Y_{13} = -1,769, Y_{14} = -1,687, Y_{15} = 1,158, Y_{16} = 1,113$, bei žinome lygties eilę 1 (AR(1)). Mūsų tikslas – įvertinti koeficientus $\hat{a}_1$ ir $\hat{\sigma}^2$.

Stebėjimų skaičius: $n = 16$.

Apskaičiuojame šios imties vidurkį: $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = -0,001$.

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

$$\begin{cases} \hat{\gamma}(0) = 6,951, \\ \hat{\gamma}(1) = 1,123. \end{cases}$$

Gautus įverčius įstatome į Yule–Walker lygtis:

$$\begin{cases} 6,951 = a_1 \cdot 1,123 + \sigma^2, \\ 1,123 = a_1 \cdot 6,951. \end{cases}$$

Iš paskutiniosios lygties apskaičiuojame $\hat{a}_1$ įvertį, o iš pirmosios lygties išsireiškiame $\hat{\sigma}^2$ įvertį:

$$\hat{a}_1 = 0,162, \hat{\sigma}^2 = 6,769.$$

Taigi, gauname AR(1) proceso lygtį:

$$Y_t = 0,162 \cdot Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0; 6,769).$$

Gautą AR(1) proceso lygtį apjungiame su anksčiau gauta trendo lygtimi ir turime:

$$X_t = 0,8886 \cdot t + 10,732 + 0,162 \cdot (X_{t-1} - 0,8886(t-1) - 10,732) + \varepsilon_t,$$

$$\varepsilon_t \sim N(0; 6,769).$$

$$X_t = 0,8886 \cdot t + 10,732 + 0,162 \cdot X_{t-1} - 0,144(t-1) - 1,739 + \varepsilon_t,$$

$$\varepsilon_t \sim N(0; 6,769).$$

$$X_t = 0,7446 \cdot t + 9,137 + 0,162 \cdot X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0; 6,769).$$

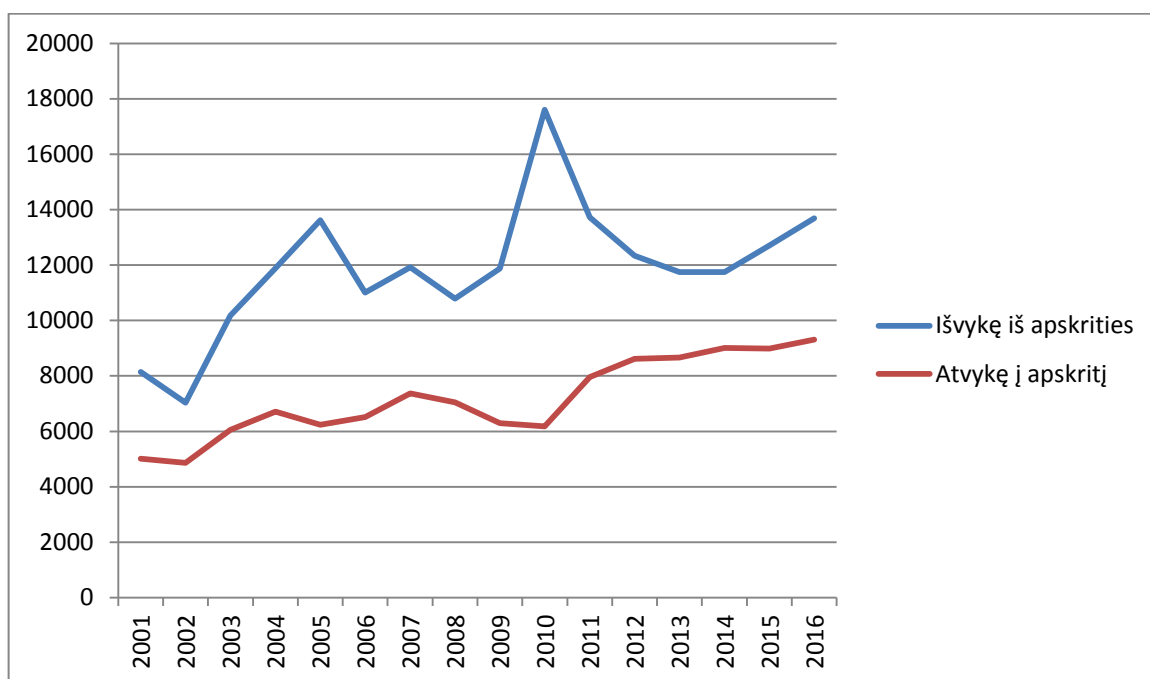
Patikrinkime, ar ši lygtis teisinga 2010 metų duomenims:

$$X_{10} = 0,7446 \cdot 10 + 9,137 + 0,162 \cdot 15,417 + 6,769 = 25,850.$$

Išvada: 11 lentelėje X_{10} (2010 metų Vilniaus apskrities migrantai)= 25,383, o surašius turimus duomenis į apskaičiuotą AR(1) proceso lygtį $X_{10} = 25,850$. Akivaizdu, kad naudodami gautą AR(1) proceso lygtį galime su nedidele paklaida apskaičiuoti duomenis.

2.6 ATVYKUSIŲ Į APSKRITIS IR IŠVYKUSIŲ IŠ APSKRIČIŲ GYVENTOJŲ ANALIZĖ

Statistikos departamente galima rasti duomenis apie atvykusius į apskritis ir išvykusius iš apskričių gyventojus. Išvykę iš apskričių gyventojai gali būti ir į užsienį, ir į kitą Lietuvos apskritį. Atvykę į apskritis žmonės gali būti iš kitų apskričių arba iš užsienio (imigravę) arba grįžę tautiečiai. Panagrinėsime šiuos duomenis.



11 pav. Išvykusių ir atvykusių į apskritis gyventojų dinamika 2001–2016 metais

Atvykusių ir išvykusių iš apskričių gyventojų duomenų pokyčiai laikui bėgant matomi 11 pav. Aiškiai matyti, kad iš apskričių gyventojų išvyksta daugiau nei į jas atvyksta. Ypač didelis išvykusių skaičius pastebimas 2010 metais.

2.6.1 ATVYKUSIŲ Į APSKRITIS GYVENTOJŲ ANALIZĖ

13 lentelė. Atvykusių į apskritis gyventojų skaičiai ir procentinės dalys

Atvykusieji								
Apskritis Nr.	2001	%	2006	%	2011	%	2016	%
1	2345	5,51%	3084	5,18%	3726	5,49%	3890	4,38%
2	8534	20,04%	11825	19,88%	13194	19,44%	17908	20,18%
3	4474	10,51%	8086	13,59%	8595	12,66%	11926	13,44%
4	1715	4,03%	3141	5,28%	3366	4,96%	3687	4,16%
5	4342	10,20%	4318	7,26%	4810	7,09%	5430	6,12%
6	5012	11,77%	6520	10,96%	7953	11,72%	9304	10,49%
7	1679	3,94%	2227	3,74%	2400	3,54%	2677	3,02%
8	2936	6,89%	3773	6,34%	3617	5,33%	4246	4,79%
9	2961	6,95%	2924	4,92%	3438	5,06%	3781	4,26%
10	8587	20,16%	13584	22,84%	16785	24,73%	25885	29,17%
Viso:	42585	100%	59482	100%	67884	100%	88734	100%

13 lentelėje turime atvykusių į apskritis gyventojų skaičius ir jų procentines dalis visų atvykusiųjų atžvilgiu. Naudodamiesi V. Kanišausko rekomendacijomis iš „Vidinės migracijos matematinio modelio“ bandysime patikrinti, ar pokyčiai tiriamu laikotarpiu tarp atvykusių į apskritis gyventojų buvo reikšmingi. Tikrinsime statistinę hipotezę H_0 : atvykusiųjų gyventojų į apskritis skaičiai nepakito per penkerius metus, kuri analogiška statistinės hipotezės apie polinominio skirstinio parametrus tikrinimą. Čia skaičiavimuose naudojamas reikšmingumo lygmuo $\alpha = 0,01$.

2001–2006 m.

$H_0: p_1 = 0,055, p_2 = 0,200, p_3 = 0,105, p_4 = 0,040, p_5 = 0,102, p_6 = 0,118, p_7 = 0,039, p_8 = 0,069, p_9 = 0,070, p_{10} = 0,202; \hat{p}_1 = 0,052, \hat{p}_2 = 0,199, \hat{p}_3 = 0,136, \hat{p}_4 = 0,053, \hat{p}_5 = 0,073, \hat{p}_6 = 0,110, \hat{p}_7 = 0,037, \hat{p}_8 = 0,063, \hat{p}_9 = 0,049, \hat{p}_{10} = 0,228.$

$n = n(2006) = 59482.$

$\chi^2 = 1918,00.$

$\chi^2 = 1918,00 > \chi_{0,99}^2(9) = 27,9$, todėl hipotezę atmetame ir galime teigti, kad atvykusiųjų nuo 2001 iki 2006 metų skaičius pakito kardinaliai.

2006–2011 m.

$H_0^{(2)}: p_1 = 0,052, p_2 = 0,199, p_3 = 0,136, p_4 = 0,053, p_5 = 0,073, p_6 = 0,110, p_7 = 0,037, p_8 = 0,063, p_9 = 0,049, p_{10} = 0,228; \hat{p}_1 = 0,055, \hat{p}_2 = 0,194, \hat{p}_3 = 0,127, \hat{p}_4 = 0,050, \hat{p}_5 = 0,071, \hat{p}_6 = 0,117, \hat{p}_7 = 0,035, \hat{p}_8 = 0,053, \hat{p}_9 = 0,051, \hat{p}_{10} = 0,247.$

$n = n(2011) = 67884.$

$\chi^2 = 340,89.$

$\chi^2 = 340,89 > \chi_{0,99}^2(9) = 27,9.$

Hipotezę atmetame, t.y. nuo 2006 iki 2011 metų atvykusiųjų skaičius pakito kardinaliai.

2011–2016 m.

$H_0^{(3)}: p_1 = 0,055, p_2 = 0,194, p_3 = 0,127, p_4 = 0,050, p_5 = 0,071, p_6 = 0,117, p_7 = 0,035, p_8 = 0,053, p_9 = 0,051, p_{10} = 0,247; \hat{p}_1 = 0,044, \hat{p}_2 = 0,202, \hat{p}_3 = 0,134, \hat{p}_4 = 0,042, \hat{p}_5 = 0,061, \hat{p}_6 = 0,105, \hat{p}_7 = 0,030, \hat{p}_8 = 0,048, \hat{p}_9 = 0,043, \hat{p}_{10} = 0,292.$

$n = n(2016) = 88734.$

$\chi^2 = 1,970.$

$\chi^2 = 1,970 < \chi_{0,99}^2(9) = 27,9.$

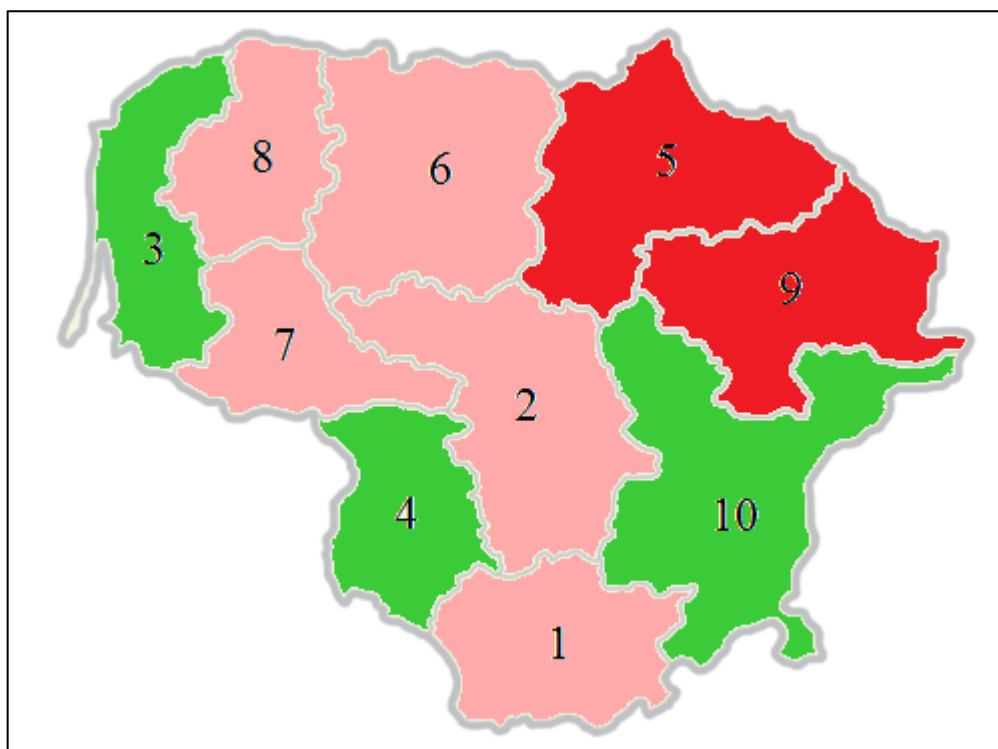
Hipotezę priimame, t.y. 2011–2016 metais atvykusiųjų į apskritis skaičius nepakito.

Atvykusių į apskritis gyventojų skaičiaus pokyčiai Lietuvos žemėlapyje

14 lentelė. Atvykusiųjų į apskritis pokytis 2001–2016 metais

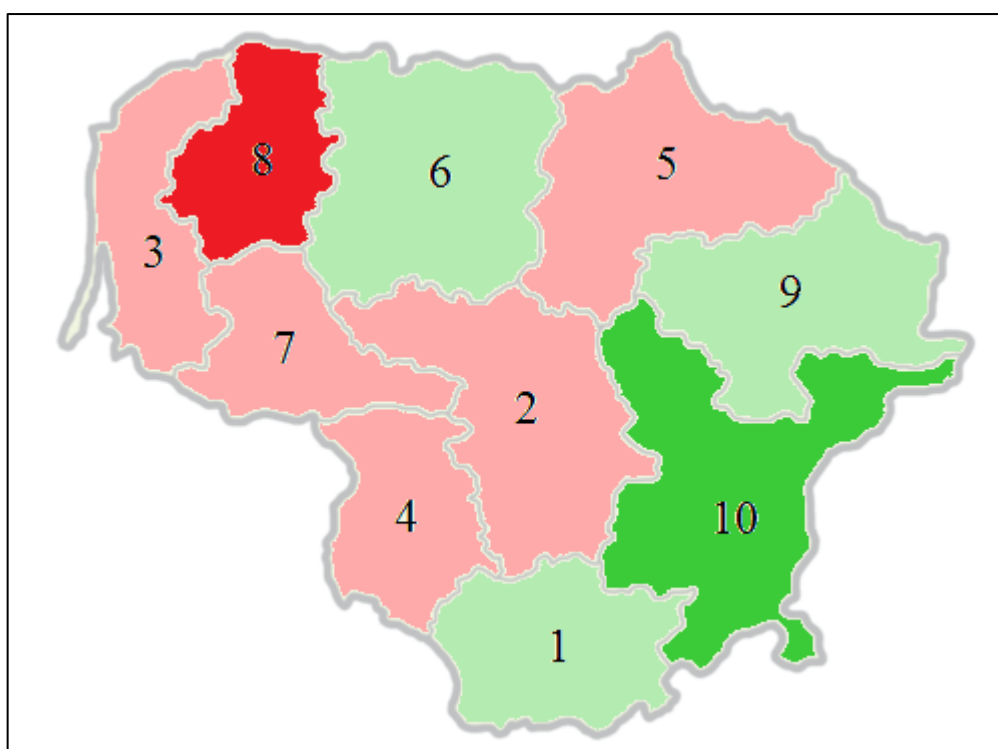
Apskritis Nr.	2001–2006 Δ	2006–2011 Δ	2011–2016 Δ
1	–0,322%	0,304%	–1,105%
2	–0,160%	–0,444%	0,746%
3	3,088%	–0,933%	0,779%
4	1,253%	–0,322%	–0,803%
5	–2,937%	–0,174%	–0,966%
6	–0,808%	0,754%	–1,230%
7	–0,199%	–0,209%	–0,519%
8	–0,551%	–1,015%	–0,543%
9	–2,037%	0,149%	–0,803%
10	2,673%	1,889%	4,445%

Šiuos atvykusių į apskritis gyventojų pokyčius būtų patogu matyti Lietuvos žemėlapyje. Naudodamiesi anksčiau pateikta spalvų skale pavaizduosime tuos pokyčius žemėlapyje. T.y. žalia spalva rodo, kad atvykstančių skaičius didėja, kuo ryškiau (tamsiau), tuo daugiau atvykstančių, o raudona reiškia, kad atvykstančių mažėja (kuo tamsiau, tuo mažėjimas spartesnis).



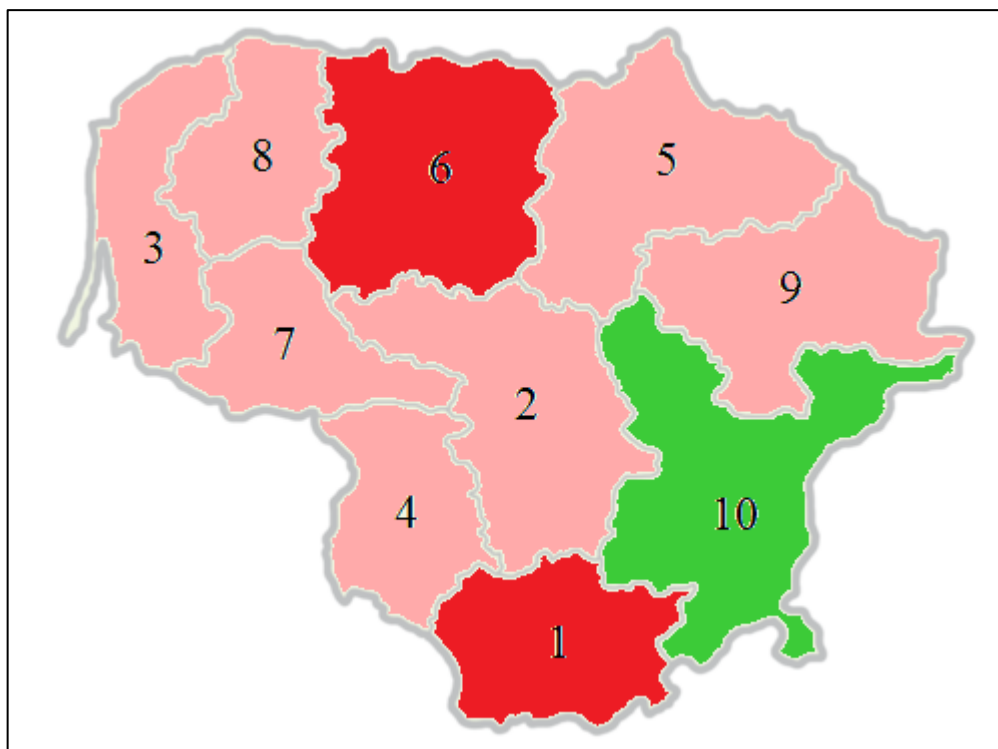
12 pav. Atvykusiųjų į apskritis pokytis 2001–2006 metais

Iš 12 pav. matome, kad nuo 2001 iki 2016 metų daugiausia gyventojų atvyko į Vilniaus, Klaipėdos ir Marijampolės apskritis. Mažiausiai – į Panevėžio ir Utenos apskritis.



13 pav. Atvykusiųjų į apskritis pokytis 2006–2011 metais

Tarp 2006 ir 2011 metų situacija pasikeitė, tai galime matyti 13 pav. Šiek tiek daugiau gyventojų atvyksta į Šiaulių, Alytaus ir Utenos apskritis. Žinoma, sostinė sutraukia daugiausia gyventojų. Likusiose apskrityse procentinis pokytis neigiamas, ypač Telšių apskrityje.



14 pav. Atvykusiųjų į apskritis pokytis 2011–2016 metais

14 pav. rodo, kad nuo 2011 iki 2016 metų atvykstančių gyventojų daugiausia Vilniaus apskrityje, o iš visur kitur atvykstančiųjų skaičius mažėja.

2.6.2 IŠVYKUSIŲ IŠ APSKRIČIŲ GYVENTOJŲ ANALIZĖ

15 lentelė. Išvykusių iš apskričių gyventojų skaičius ir procentinė dalis

Išvykusieji								
Apskritis Nr.	2001	%	2006	%	2011	%	2016	%
1	3309	5,03%	4678	5,56%	6162	5,81%	6338	5,33%
2	13834	21,05%	16654	19,80%	21815	20,57%	23201	19,51%
3	7990	12,16%	11011	13,09%	13080	12,33%	15206	12,79%
4	2823	4,29%	4968	5,91%	5617	5,30%	6656	5,60%
5	6621	10,07%	7350	8,74%	8248	7,78%	9697	8,16%
6	8141	12,39%	11003	13,08%	13727	12,94%	13688	11,51%
7	2586	3,93%	3846	4,57%	4130	3,89%	4686	3,94%
8	4307	6,55%	6181	7,35%	5922	5,58%	7333	6,17%
9	4307	6,55%	4597	5,46%	5546	5,23%	6037	5,08%
10	11814	17,97%	13839	16,45%	21815	20,57%	26063	21,92%
Viso:	65732	100%	84127	100%	106062	100%	118905	100%

Analogiškus skaičiavimus atliksime su išvykusių iš apskričių gyventojų duomenimis.

Tikrinsime hipotezę H_0 : išvykusiųjų skaičius apskrityse nepakito, kuri vėl tikrinama nadudojant statistinę hipotezę apie polinominio skirstinio parametrus.

2001–2006 m.

$H_0: p_1 = 0,050, p_2 = 0,210, p_3 = 0,122, p_4 = 0,043, p_5 = 0,101, p_6 = 0,124, p_7 = 0,039,$
 $p_8 = 0,066, p_9 = 0,066, p_{10} = 0,180, \hat{p}_1 = 0,056, \hat{p}_2 = 0,198, \hat{p}_3 = 0,131, \hat{p}_4 = 0,059,$
 $\hat{p}_5 = 0,087, \hat{p}_6 = 0,131, \hat{p}_7 = 0,046, \hat{p}_8 = 0,073, \hat{p}_9 = 0,055, \hat{p}_{10} = 0,165.$

$n = n(2006) = 84127.$

$\chi^2 = 1287,520.$

$\chi^2 = 1287,520 > \chi_{0,99}^2(9) = 27,9$, todėl hipotezę atmetame ir galime teigti, kad nuo 2001 iki 2006 metų išvykusiųjų skaičius pakito kardinaliai.

2006–2011 m.

$H_0^{(2)}: p_1 = 0,056, p_2 = 0,198, p_3 = 0,131, p_4 = 0,059, p_5 = 0,087, p_6 = 0,131, p_7 =$
 $0,046, p_8 = 0,073, p_9 = 0,055, p_{10} = 0,165, \hat{p}_1 = 0,058, \hat{p}_2 = 0,206, \hat{p}_3 = 0,123, \hat{p}_4 =$
 $0,053, \hat{p}_5 = 0,078, \hat{p}_6 = 0,129, \hat{p}_7 = 0,039, \hat{p}_8 = 0,056, \hat{p}_9 = 0,052, \hat{p}_{10} = 0,206.$

$n = n(2011) = 106062.$

$\chi^2 = 1929,956.$

$\chi^2 = 1929,956 > \chi_{0,99}^2(9) = 27,9.$

Hipotezę atmetame, t.y. nuo 2006 iki 2011 metų išvykusiųjų skaičius pakito kardinaliai.

2011–2016 m.

$H_0^{(3)}: p_1 = 0,058, p_2 = 0,206, p_3 = 0,123, p_4 = 0,053, p_5 = 0,078, p_6 = 0,129, p_7 =$
 $0,039, p_8 = 0,056, p_9 = 0,052, p_{10} = 0,206, \hat{p}_1 = 0,053, \hat{p}_2 = 0,195, \hat{p}_3 = 0,128, \hat{p}_4 =$
 $0,056, \hat{p}_5 = 0,082, \hat{p}_6 = 0,115, \hat{p}_7 = 0,039, \hat{p}_8 = 0,062, \hat{p}_9 = 0,051, \hat{p}_{10} = 0,219.$

$n = n(2016) = 118905.$

$\chi^2 = 545,955.$

$\chi^2 = 545,955 > \chi_{0,99}^2(9) = 27,9.$

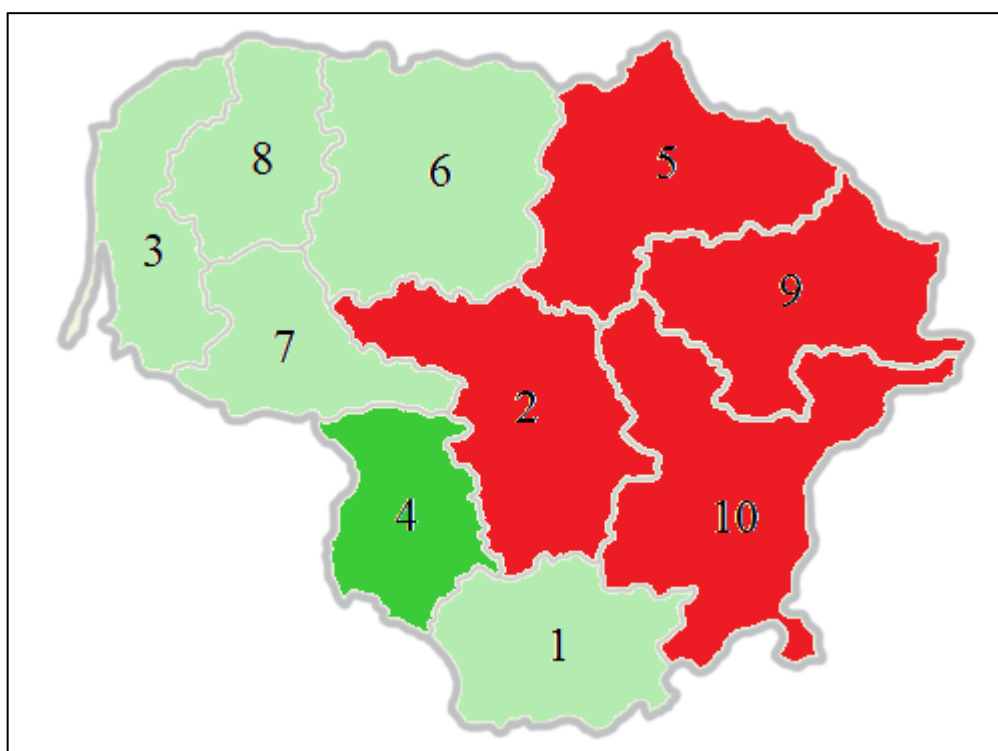
Hipotezę atmetame, t.y. išvykusiųjų skaičius 2011–2016 metais kardinaliai pakito.

Išvykusių iš apskričių gyventojų skaičiaus pokyčiai Lietuvos žemėlapyje

16 lentelė. Išvykusiųjų iš apskričių pokytis 2001–2016 metais

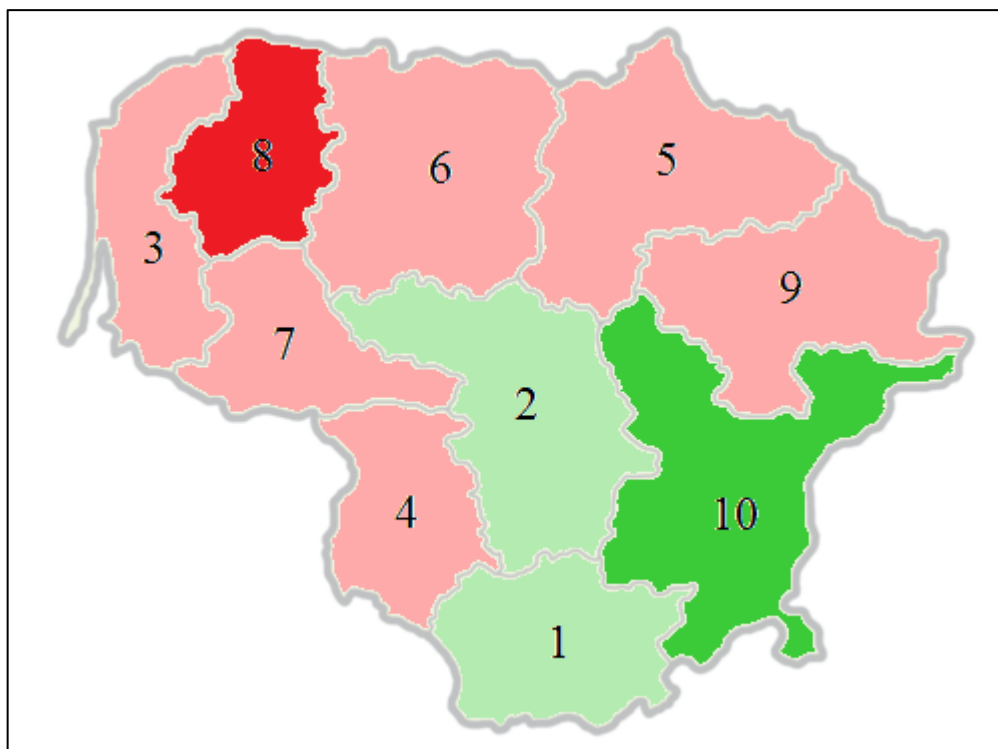
Apskritis Nr.	2001–2006 Δ	2006–2011 Δ	2011–2016 Δ
1	0,53%	0,25%	–0,48%
2	–1,25%	0,77%	–1,06%
3	0,93%	–0,76%	0,46%
4	1,61%	–0,61%	0,30%
5	–1,34%	–0,96%	0,38%
6	0,69%	–0,14%	–1,43%
7	0,64%	–0,68%	0,05%
8	0,79%	–1,76%	0,58%
9	–1,09%	–0,24%	–0,15%
10	–1,52%	4,12%	1,35%

Išvykusių iš apskričių gyventojų pokyčius, pateiktus 16 lentelėje, pavaizduosime Lietuvos žemėlapyje. Šiuo atveju, kuo spalva raudonesnė (tamsesnė, ryškesnė), tuo daugiau gyventojų išvyksta, o žalia spalva rodo, kad išvyksta mažiau.



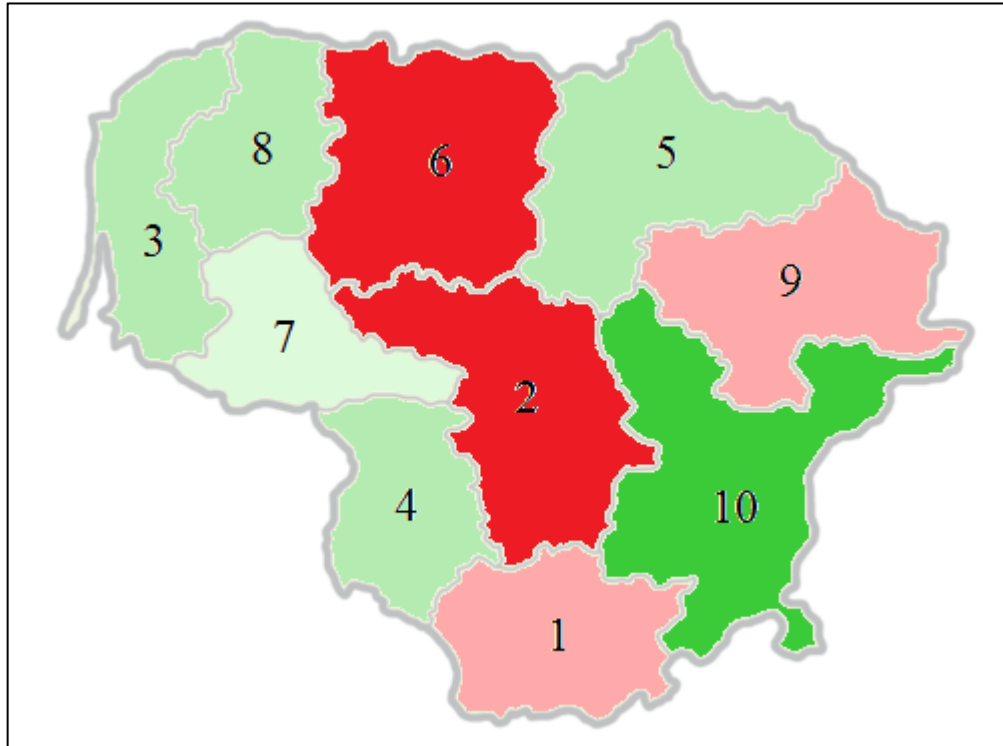
15 pav. Išvykusiųjų iš apskričių procentinis pokytis 2001–2006 metais

Galime matyti, kad nuo 2001 iki 2006 metų daugiausia gyventojų išvyko iš Panevėžio, Utenos ir Kauno apskričių. Atrodytų keista, kad iš Vilniaus apskrities išvyksta gyventojai, bet reikia nepamiršti, kad atvykstančių ten daugiausiai.



16 pav. Išvykusiųjų iš apskričių procentinis pokytis 2006–2011 metais

16 pav. akivaizdžiai matyti, kad iš Vilniaus apskrities gyventojų išvyksta mažiau nei 2001–2006 metais. Kauno ir Alytaus apskrityse taip pat pagerėjo situacija. Likusiose apskrityse išvykstančių daugėja, ypač Telšių apskrityje.



17 pav. Išvykusiųjų iš apskričių procentinis pokytis 2011–2016 metais

17 pav. šiek tiek „žalesnis“ nei 16 pav. Tai rodo, kad iš apskričių išvyksta mažiau gyventojų, galbūt dėl to, kad nebėra kam išvykti. Šiaulių ir Kauno apskrityse išvykusiųjų daugiausiai, kiek mažiau Alytaus ir Utenos apskrityse.

DARBO REZULTATAI

1. Taikant vidinės migracijos matematinį modelį darbe buvo nustatyta, kad vidinės migracijos pokyčiai Lietuvos apskrityse nuo 2001 iki 2016 metų buvo statistiškai reikšmingi su reikšmingumo lygmeniu $\alpha = 0,01$. Šie pokyčiai kas 5 metus pavaizduoti Lietuvos žemėlapyje.
2. Atlikta lyginamoji vidinės ir išorinės migracijos analizė parodė, kad vidinės ir išorinės migracijų bendriniai duomenys Lietuvos mastu nuo 2001 iki 2016 metų nekoreliuoja (priimta hipotezė $H_0: \rho = 0$, $\alpha = 0,01$). Tačiau vidinės ir išorinės migracijos apskričių duomenys labai koreliuoja: 2001 m. priimta hipotezė $H_0: \rho = 0,9$, 2006 m. priimta hipotezė $H_0: \rho = 0,9$, 2011 m. priimta hipotezė $H_0: \rho = 0,99$, 2016 m. priimta hipotezė $H_0: \rho = 0,99$ (visur $\alpha = 0,01$).
3. Darbe buvo patikrintos laiko eilučių taikomumo prielaidos visoms Lietuvos apskritims ir nustatyta, kad laiko eilutes galima taikyti tik Marijampolės ir Vilniaus apskričių migrantų (vidiniai + išoriniai) duomenims. Tas taikomumas buvo pademonstruotas Vilniaus apskrčiai, t.y. buvo siekiama nustatyti, ar nagrinėjami duomenys turi trendą, jis rastas ir po to nustatyti AR(1) modelio parametrai:
$$X_t = 0,8886 \cdot t + 10,732 + 0,162 \cdot (X_{t-1} - 0,8886(t-1) - 10,732) + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 6,769),$$
$$X_t = 0,8886 \cdot t + 10,732 + 0,162 \cdot X_{t-1} - 0,144(t-1) - 1,739 + \varepsilon_t,$$
$$\varepsilon_t \sim N(0; 6,769).$$
$$X_t = 0,7446 \cdot t + 9,137 + 0,162 \cdot X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0; 6,769).$$
4. Taikant statistines hipotezes apie polinominio skirstinio parametrus, darbe nustatyta, kad kas penkerius metus Lietuvoje išvykusių žmonių skaičiai apskrityse statistiškai reikšmingai skyrėsi ($\alpha = 0,01$), o tarp atvykusių žmonių statistiškai reikšmingi pokyčiai buvo užfiksuoti tik nuo 2001 iki 2011 metų ($\alpha = 0,01$), o nuo 2011 iki 2016 metų tokie pokyčiai statistiškai nereikšmingi ($\alpha = 0,01$). Ir atvykstančių, ir išvykstančių žmonių pokyčiai kas 5 metus pavaizduoti Lietuvos žemėlapyje.

LITERATŪRA

1. V. Kanišauskas. *Tikimybių teorijos ir matematinės statistikos pagrindai*. Šiauliai: Šiaulių universiteto leidykla, 2000.
2. V. Kanišauskas. Vidinės migracijos matematinis modelis. *Lietuvos matematikos rinkinys*, 2016. **57**: 7–12. <<https://www.mii.lt/LMR/B/2016/57B02.pdf>> [žiūrėta 2017-05-23].
3. N. Kligienė. *Įvadas į atsitiktinių sekų statistinę analizę*. Vilnius: Technika, 1998.
4. J. Kruopis. *Matematinė statistika*. Vilnius: Mokslo ir enciklopedijų leidykla, 1993.
5. Statistikos departamento duomenys. <<https://osp.stat.gov.lt/>> [žiūrėta 2017-05-23].

SANTRAUKA

Bakalauro darbo objektas yra statistikos departamento statistiniai ir demografiniai duomenys apie Lietuvos gyventojų migraciją 2001–2016 metais. Darbo tikslas – atlikti Lietuvos XXI a. pradžios vidinės migracijos statistinę ir lyginamąją analizę. Tikslui pasiekti keliami keli uždaviniai. Pirmiausia – nustatyti, ar yra pagrindo manyti, kad Lietuvos apskrityse 2001–2016 m. vyko vidinė migracija. Tai nustačius, vidinės migracijos pokyčius Lietuvos apskrityse reikia pavaizduoti Lietuvos žemėlapyje. Kitas uždavinys – nustatyti statistinę priklausomybę tarp vidinės ir išorinės migracijos duomenų. Toliau darbe taikomos laiko eilutės. Tam tikslui darbe tikrinama, ar yra prielaidos taikyti laiko eilučių modelius statistiniams duomenims apie migrantus apskrityse. Ir galiausiai, darbe siekiama nustatyti, kaip atvykstantys bei išvykstantys iš apskričių žmonės veikia vidinės migracijos procesą.

Pritaikius statistinę hipotezę apie polinominio skirstinio parametrus, darbe nustatyta, kad vidinė migracija 2001–2016 metais vyko. Kaip tai vyko, pavaizduota Lietuvos žemėlapyje.

Toliau darbe atlikta lyginamoji vidinės ir išorinės migracijos analizė. Ji parodė, kad yra tiesinė priklausomybė apskričių lygiu tarp tikrintų duomenų. Taip pat patikrinta statistinė hipotezė apie tos priklausomybės stiprumą.

Tikrinant prielaidas laiko eilutėms taikyti, darbe nustatyta, kad tik Vilniaus ir Marijampolės apskričių duomenims taikytini laiko eilučių modeliai. Nustatyta, kad Vilniaus apskrities duomenims egzistuoja trendas. Eliminavus trendą nustatyti AR(1) modelio parametrai.

Vidinės migracijos procesą įtakoja atvykusių ir išvykusių iš apskričių žmonių kiekiai, kurie analizuojami praktinėje darbo dalyje. Taikant statistines hipotezes nustatyta, kad kas penkmetį Lietuvoje išvykusių žmonių skaičiai kardinaliai skyrėsi, o tarp atvykusių žmonių kardinalūs pokyčiai buvo tik nuo 2001 iki 2011 metų, o nuo 2011 iki 2016 metų tokių pokyčių neužfiksuota. Darbe visa tai pavaizduota Lietuvos žemėlapyje.

STATISTICAL RESEARCH OF INTERNAL MIGRATION IN LITHUANIA AT THE BEGINNING OF 21st CENTURY

SUMMARY

The object of this bachelor final work is the statistical and demographical data of Statistics Lithuania about migration of Lithuanian residents over the period 2001-2016. The aim of this study is to carry out statistical and comparative analysis of internal migration in Lithuania at the beginning of the 21st century. To achieve this aim the following objectives have been set. First of all, to identify if there is a reason to think that internal migration took place in Lithuania regions between 2001 and 2016. Having identified this, changes in internal migration in Lithuania regions have to be marked in the map of Lithuania. Another objective is to identify statistical dependence between the data about internal and external migration. Next step is to apply time series. To achieve this objective, we will check if there is a reason to apply time series models to statistical data about the migrants in the regions. And finally, the work aims at finding out how people arriving at and departing from regions influence the process of internal migration.

Having applied the statistical hypothesis about multinomial distribution parameters, the final work found out that in 2001-2016 internal migration took place. The way it was is shown in the map of Lithuania.

Next part of the work carries out comparative analysis of internal and external migration. It indicated that there is a big dependence between internal and external migration, but it is noticeable only at regional level. At national level it is not seen, as it shows only total results without separate region's input.

Having checked the assumptions to apply time series, it was identified that time series models can be applied only to the data of Vilnius and Marijampole regions. It was found out that a trend exists for the data of Vilnius region. Having eliminated the trend, AR(1) model parameters were identified.

Process of internal migration is influence by the numbers of people arriving at and leaving the regions, that are analysed in the practical part of the work. Having applied statistical hypotheses, it was found out every five years the numbers of people, who departed, were significantly different, while substantial changes among the people, who arrived, were only from 2001 till 2011 and between 2011 and 2016 such changes were not recorded. All that information is presented in the map of Lithuania included in this work.

PRIEDAI

1 PRIEDAS

Gyventojų skaičiai apskrityse

Apskritys/Metai	2001	2002	2003	2004	2005	2006	2007	2008
Alytaus	187990	186484	184992	182751	179626	175813	173028	170249
Kauno	702090	695033	689787	682074	671923	658100	650408	643358
Klaipėdos	386129	382076	379878	376854	372541	364486	360434	357201
Marijampolės	188775	187281	185970	184058	181376	177439	174810	172200
Panevėžio	300305	296871	294182	290097	284918	278262	273701	268815
Šiaulių	370444	366330	362783	357441	350938	341761	335221	328699
Tauragės	134277	132982	131808	130179	127690	124020	121549	119179
Telšių	179952	178315	177139	175090	172426	167888	164963	162431
Utenos	186368	183973	181638	178637	175226	170829	167678	164167
Vilniaus	850668	845292	843320	841748	838556	831237	828191	826306

Apskritys/Metai	2009	2010	2011	2012	2013	2014	2015	2016
Alytaus	167603	164833	158403	154980	152139	149851	147809	145109
Kauno	638385	629896	610225	599638	592816	587238	583047	577358
Klaipėdos	355084	350306	340047	334744	331553	329013	327300	324618
Marijampolės	169973	167143	162205	159273	156455	154084	151793	149077
Panevėžio	264319	259374	251401	246298	242340	238748	235394	231001
Šiaulių	323353	316278	303110	295824	290471	285763	281632	276329
Tauragės	117100	114763	110511	108125	106335	104623	103083	101200
Telšių	160532	157718	152594	149843	147728	145482	143511	141293
Utenos	161062	157472	152577	149001	145760	142564	140170	137193
Vilniaus	826445	824193	811515	805915	806308	806106	807523	805380

2 PRIEDAS

		Išvykusieji emigrantai asmenys	ir Išvykusieji asmenys (vidiniai migrantai)	Atvykusieji imigrantai asmenys
Lietuvos Respublika	2001	65732	37891	42585
	2002	56016	39297	44407
	2003	84108	57825	62553
	2004	97064	59373	64926
	2005	110854	52969	59758
	2006	84127	51737	59482
	2007	86966	56583	65192
	2008	82682	56932	66229
	2009	88443	49943	56430
	2010	132953	49796	55009
	2011	106062	52199	67884
	2012	101153	60053	79896
	2013	96061	57243	79254
	2014	98036	61415	85709
	2015	105959	61426	83556
	2016	118905	68572	88734
Alytaus apskritis	2001	3309	2291	2345
	2002	2761	2192	2078
	2003	4877	3734	3333
	2004	5961	4080	3718
	2005	6460	3962	3815
	2006	4678	3229	3084
	2007	4626	3209	3111
	2008	4631	3347	3113
	2009	4638	2990	2771
	2010	8141	2988	2792
	2011	6162	3294	3726
	2012	5655	3562	3814
	2013	5051	3138	3762
	2014	5130	3425	4082
	2015	5746	3384	3990
	2016	6338	3736	3890
Kauno apskritis	2001	13834	8160	8534
	2002	11734	8409	8564
	2003	17539	12343	11898
	2004	20093	11595	12237

	2005	23077	10520	11819
	2006	16654	10077	11825
	2007	17171	11098	13032
	2008	16756	11704	14011
	2009	18232	9928	11454
	2010	28314	9307	10702
	2011	21815	9995	13194
	2012	20157	11212	15373
	2013	19296	10995	15971
	2014	19806	12182	17505
	2015	20313	12138	16975
	2016	23201	13310	17908
Klaipėdos apskritis	2001	7990	3727	4474
	2002	6346	3845	4944
	2003	10185	6435	7828
	2004	11909	6680	8320
	2005	14788	6006	7882
	2006	11011	6322	8086
	2007	11390	6873	9207
	2008	10683	7030	9319
	2009	11810	6085	7536
	2010	16674	5652	7134
	2011	13080	6088	8595
	2012	12154	6480	9551
	2013	11864	6441	10075
	2014	12368	7368	11207
	2015	13120	7356	10948
	2016	15206	8694	11926

Marijampolės apskritis	2001	2823	1714	1715
	2002	2923	2202	2176
	2003	4253	3058	2862
	2004	5110	3360	3132
	2005	5979	3141	2934
	2006	4968	3299	3141
	2007	4862	3373	3239
	2008	4607	3381	3097
	2009	4936	3138	2793
	2010	6920	3030	2794
	2011	5617	3122	3366
	2012	5843	3791	3743

	2013	5280	3470	3720
	2014	5474	3630	3990
	2015	5655	3498	3762
	2016	6656	3723	3687
Panevėžio apskritis	2001	6621	4320	4342
	2002	5586	4120	4067
	2003	8017	5677	5169
	2004	8934	5752	5079
	2005	9406	4646	4236
	2006	7350	4705	4318
	2007	7706	5089	4636
	2008	7069	4952	4390
	2009	7694	4787	4287
	2010	10442	4691	4193
	2011	8248	4695	4810
	2012	7989	5559	5585
	2013	7573	5274	5449
	2014	7571	5346	5591
	2015	8111	5242	5242
	2016	9697	5961	5430
Šiaulių apskritis	2001	8141	4894	5012
	2002	7029	5091	4864
	2003	10181	6856	6050
	2004	11882	7280	6711
	2005	13623	6491	6235
	2006	11003	6647	6520
	2007	11925	7717	7370
	2008	10785	7305	7050
	2009	11879	6429	6294
	2010	17602	6360	6174
	2011	13727	6648	7953
	2012	12331	7246	8618
	2013	11748	6955	8662
	2014	11744	7247	9012
	2015	12707	7393	8982
	2016	13688	7946	9304
Tauragės apskritis	2001	2586	1509	1679
	2002	2346	1723	1704
	2003	3295	2261	2123
	2004	4252	2748	2461
	2005	5074	2567	2160

	2006	3846	2432	2227
	2007	3999	2734	2310
	2008	3506	2516	2044
	2009	3591	2193	1872
	2010	5191	2112	1688
	2011	4130	2282	2400
	2012	4259	2867	2981
	2013	3815	2588	2713
	2014	3853	2592	2779
	2015	4033	2606	2663
	2016	4686	2942	2677
Telšių apskritis	2001	4307	2822	2936
	2002	3485	2601	2729
	2003	5397	3851	3744
	2004	6284	4052	3930
	2005	7353	3496	3412
	2006	6181	3568	3773
	2007	6002	3983	4069
	2008	5568	3935	3982
	2009	5461	3306	3105
	2010	7798	3316	3135
	2011	5922	3243	3617
	2012	5969	3844	4192
	2013	5374	3473	3650
	2014	5522	3709	4096
	2015	5652	3621	3827
	2016	7333	4623	4246
Utenos apskritis	2001	4307	2693	2961
	2002	3996	2754	3020
	2003	5139	3544	3483
	2004	5728	3713	3635
	2005	6016	3150	3095
	2006	4597	2824	2924
	2007	4797	2964	2968
	2008	4829	3328	3201
	2009	4785	2859	2617
	2010	6488	3155	3091
	2011	5546	3237	3438
	2012	5555	3682	3722
	2013	5545	3676	3855
	2014	5083	3440	4031

	2015	5403	3466	3812
	2016	6037	3725	3781
Vilniaus apskritis	2001	11814	5761	8587
	2002	9810	6360	10261
	2003	15225	10066	16063
	2004	16911	10113	15703
	2005	19078	8990	14170
	2006	13839	8634	13584
	2007	14488	9543	15250
	2008	14248	9434	16022
	2009	15417	8228	13701
	2010	25383	9185	13306
	2011	21815	9595	16785
	2012	21241	11810	22317
	2013	20515	11233	21397
	2014	21485	12476	23416
	2015	25219	12722	23355
	2016	26063	13912	25885