

VILNIAUS UNIVERSITETO
MATEMATIKOS IR INFORMATIKOS FAKULTETO
MATEMATINĖS STATISTIKOS KATEDRA

Baigiamasis bakalauro darbas

Migracijos dinamikos statistinis tyrimas

Statistical Analysis of Migration Dynamics

Eglė Čepaitytė

VILNIUS 2016

Darbo vadovas: doc. dr. Rūta Levulienė

Darbo recenzentas: doc. dr. Pranas Vaitkus

Darbo gynimo data _____

Darbo įvertinimas _____

Registravimo Nr. _____

Atidavimo į katedra data _____

Turinys

1 ĮVADAS	2
2 TEORINĖ DALIS	3
2.1 Polinominė regresija	3
2.2 Pažingsninė regresija	3
2.3 Durbino ir Watsono testas	4
2.4 Andersono ir Darlingo testas	5
2.5 Regresinė panelinių duomenų analizė	6
2.6 Klasterinė analizė	7
3 PRAKTINĖ DALIS	9
3.1 Statistiniai duomenys	9
3.2 Migracijai įtaką darantys veiksniai	9
3.3 Lietuvos piliečių emigracijos kryptys	23
3.4 Veiksniai darantys įtaką migracijai Lietuvoje	28
3.5 Lietuvos savivaldybių klasterizacija pagal nedarbą ir migraciją	32
4 IŠVADOS	37
LITERATŪRA	40

1 ĮVADAS

Migracija egzistuoja nuo žmonijos atsiradimo laikų, žmonės keisdavo gyvenamąją vietą ieškodami geresnių gyvenimo sąlygų. Mūsų dienomis paprastas keliavimas, galimybė susirasti gyvenamąją vietą ir darbą internetu, greita žinių ir informacijos sklaida lemia itin didelius migracijos mastus. Žmonės migruoja dėl asmeninių, politinių ir ekonominių priežasčių. Šiame darbe nagrinėjamos pastarosios.

Kadangi ekonominių rodiklių yra labai daug, analizei pasirenkami tik penki svarbiausi: bendrasis vidaus produktas (BVP), nedarbo lygis, kainų indeksas, namų ūkio grynosios metinės pajamos ir dvidešimties procentų turtingiausių žmonių pajamų ir aštuoniasdešimties neturtingiausiųjų pajamų santykis. Šie rodikliai pasirenkami todėl, kad jie yra aktualūs gyventojams, valstybinėms ir privačioms įmonėms ir organizacijoms, dažnai minimi žiniasklaidoje bei jais remiantis lyginamas pragyvenimo lygis skirtingose valstybėse. Turtingiausių ir neturtingiausių gyventojų pajamų santykio negalima priskirti prie dažnai cituojamų, tačiau šis rodiklis įdomus tuo, kad jis gali padėti įvertinti kitus rodiklius, pavyzdžiui, vidutinės pajamas. Jei atskirtis tarp turtingųjų ir neturtingųjų yra labai didelė, tai gali lemti aukštesnes vidutinės pajamas, nors didžioji dalis valstybės gyventojų turi mažas pajamas. Nagrinėjant rodiklių visumą, galima daryti tikslesnes išvadas apie nagrinėjamą situaciją.

Turimi duomenys yra 2004-2014 metų apie trisdešimt dvi Europos valstybes.

Darbo tikslas yra ištirti ar nagrinėjami ekonominiai rodikliai turi įtakos migracijai ir jei turi, kokią įtaką jai daro.

Darbo uždaviniai:

- Patikrinti ar yra priklausomybė tarp migracijos rodiklio ir kitų nagrinėjamų rodiklių;
- Ištirti į kurias šalis emigruoja daugiausiai Lietuvos piliečių;
- Suklasterizuoti valstybes, į kurias emigruoja Lietuvos piliečiai;
- Patikrinti, kurie ekonominiai rodikliai įtakoja migraciją Lietuvoje;
- Suskirstyti Lietuvos savivaldybes atsižvelgiant į nedarbą ir migraciją.

Darbas sudarytas iš teorinės ir praktinės dalių. Teorinėje dalyje aprašomi praktinėje dalyje taikomi metodai: polinominė regresija, pažingsninė regresija, paneliniai duomenys, jungusis ir fiksuotas modeliai, skirti jiems tirti, Durbino ir Vatsono, Andersono ir Darlingo testai, klasterinė analizė. Praktinėje dalyje, taikant teorijoje aprašytus modelius, analizuojami turimi duomenys.

Skaičiavimams atlikti naudojamas statistinis paketas R. Grafikai braižomi su programomis R ir MS Excel.

2 TEORINĖ DALIS

2.1 Polinominė regresija

Polinominės regresijos modelį $Y_i = \alpha + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \epsilon_i$ galima suvesti į daugialypės regresijos modelį. Tai padaryti galima atlikus kintamųjų pakeitimą: $x_1 = x, x_2 = x^2, \dots, x_k = x^k$. Tuomet gaunamas daugialypės regresijos modelis:

$$Y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon_i \quad (2.1)$$

čia $\epsilon_i \sim N(0, \sigma^2)$ yra atsitiktinė paklaida, x_k - nepriklausomi kintamieji, β_k - nežinomi parametrai, $k=1, \dots, p$.

Parametrų įverčiai randami mažiausių kvadratų metodu. Pažymėkime:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad x_k = \frac{1}{n} \sum_{i=1}^n x_{k,i}, \quad \bar{Y}_i = \hat{\alpha} + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p.$$

Parametrų įverčiai randami iš lygties $\hat{\beta} = C^{-1}d$, kur $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_p \end{pmatrix}$, $C = \|c_{kl}\|$, $k, l = 1, \dots, p$,

$$d = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_p \end{pmatrix}.$$

Matricos C elementai randami taip: $c_{kl} = \sum_{i=1}^n (x_{k,i} - \bar{x}_k)(x_{l,i} - \bar{x}_l)$, $l, p = 1, \dots, p$.

Vektoriaus d elementų radimas: $d_k = \sum_{i=1}^n (x_{k,i} - \bar{x}_k)Y_i$

Įvertis α randamas išsprendžiant lygybę: $\hat{\alpha} = \bar{Y}_i - \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$

Regresijos modelio prielaidos:

1. ϵ_i normaliai pasiskirstę atsitiktiniai dydžiai, t.y. $\epsilon_i \sim N(0, \sigma^2)$;
2. Visų ϵ_i vidurkiai lygūs nuliui, t.y. $E\epsilon_i = 0$
3. Visų ϵ_i dispersijos vienodos, t.y. $D\epsilon_i = \sigma^2$
4. Visi ϵ_i nepriklausomi.

(Žr. [1])

2.2 Pažingsninė regresija

Dažnai taikant regresinę analizę neturima pakankamai informacijos apie tai, kurie iš kintamųjų $X_1, X_2, \dots, X_m$ labiau tinka atsitiktiniam dydžiui Y prognozuoti. Be to, galutinai parinktas modelis yra tuo vertingesnis, kuo mažesniu parametru skaičiumi jis nusakomas. Stengiamasi, kad kintamųjų skaičiaus $X_1, X_2, \dots, X_m$ sumažinimas iš esmės nepablogintų prognozės tikslumo.

Matematinėje statistikoje taikoma daug įvairių metodų, leidžiančių sumažinti vektoriaus $(X_1, X_2, \dots, X_m)^T$ matavimo dydį iš esmės nepabloginant prognozės tikslumo. Pateiksiu vieną iš dažniausiai naudojamų metodų.

Tarkime, vektorius $(X_1, X_2, \dots, X_m)^T$ pasiskirstęs pagal $m+1$ -matį normalųjį dėsnį. Stebint šį atsitiktinį vektorių gauta didumo n imtis

$$(Y_i, X_{1i}, X_{2i}, \dots, X_{mi})^T, \quad i = 1, 2, \dots, n. \quad (2.2)$$

Kadangi tiesinio prognozavimo tikslumo matas yra suvestinis koreliacijos koeficientas, tai būtų normalu imti tokį kintamųjų $X_1, X_2, \dots, X_m$ poaibį, kad dauginio koreliacijos koeficiento įvertis būtų kuo didesnis. Tačiau tokio metodo negalima realizuoti praktiškai, jeigu kintamųjų skaičius k yra pakankamai didelis. Vienas iš paprasčiausių ir dažniausiai praktiškai taikomų metodų yra pažingsninė regresija.

Nulinis žingsnis. Parenkame du reikšmingumo lygmenis P^* ir P , $P^* > P$, randame koreliacijos koeficientų tarp atsitiktinių dydžių Y ir X_i įverčius $\rho(Y, X_i) = \rho_{YX_i} = \rho_{0i}$, $i = 1, \dots, k$.

Pirmas žingsnis. Išrenkame maksimalų įvertį

$$\max_{1 \leq i \leq k} \rho_{0i} = \rho_{0i_1}.$$

Kintamasis X_{i_1} įtraukiamas į kintamųjų Y prognozuoti sąrašą, jeigu teisinga nelygybė:

$$F_{0i_1} = \rho_{0i_1}^2(n-2)/(1-\rho_{0i_1}^2) > F_P(1, n-2). \quad (2.3)$$

Priešingu atveju daroma išvada, kad Y prognozavimas pagal $X_1, X_2, \dots, X_k$ negalimas.

Antras žingsnis. Randame dalinių koreliacijų koeficientų įverčius

$$\rho_{YX-iX_{i_1}} = \rho_{0i_2(i_1)}. \quad (2.4)$$

iš jų išrenkame maksimalų:

$$\max_{i \neq i_1} \rho_{0i(i_1)} = \rho_{0i_2(i_1)}. \quad (2.5)$$

Kintamasis X_{i_2} įtraukiamas į sąrašą, jeigu teisinga nelygybė

$$\rho_{0i_2(o_1)}^2(n-3)/(1-\rho_{0i_2(i_1)}^2) > F_P(1, n-3). \quad (2.6)$$

Priešingu atveju sąraše paliekamas tik kintamasis X_{i_1} . Jeigu (1.20) nelygybė teisinga, tai tikrinama, ar reikia išbraukti iš sąrašo X_{i_1} , kai įrašytas kintamasis X_{i_2} . Kintamasis X_{i_1} išbraukiamas, jeigu teisinga nelygybė

$$\rho_{0i_2(o_1)}^2(n-3)/(1-\rho_{0i_2(i_1)}^2) < F_P(1, n-3) \quad (2.7)$$

Taigi po antrojo žingsnio atsitiktiniam dydžiui Y prognozuoti gali likti tik kintamasis X_{i_1} ((1) nelygybė neteisinga); tik kintamasis X_{i_2} (1), (2) nelygybės tenkinamos); vektorius (X_{i_1}, X_{i_2}) ((1) nelygybė teisinga, o (2) neteisinga). Pirmuoju atveju procesas užbaigiamas; antruoju ir trečiuoju atvejais pereinama prie trečio žingsnio.

Trečiasis, ketvirtasis, ... žingsniai. Rekurentiškai kartojamas antrasis žingsnis. Tiksliau, remiantis dalinių koreliacijos įverčių reikšmėmis, pagal taisykles, analogiškas (1), daroma išvada, ar reikia papildyti kintamųjų sąrašą dar vienu kintamuoju. Įrašius naują kintamąjį, tikrinama, ar nereikia kurio nors kintamojo iš sąrašo išbraukti. Procesas užbaigiamas, kai nelygybės, atitinkančios (1), (2), netenkinamos. (Žr. [2])

2.3 Durbino ir Watsono testas

Durbino ir Watsono testas naudojamas tam, kad būtų nustatyta ar liekamosios paklaidos nėra autokoreliuotos.

Regresijos modelio paklaidas sieja ryšys: $\epsilon = \rho\epsilon_{i-1} + z_i$ čia $z_i \sim N(0, \sigma^2)$ ir $z_1, z_2, \dots$ yra nepriklausomi atsitiktiniai dydžiai. Paklaidos e_i nekoreliuoja, jeigu $\rho = 0$.

Statistinė hipotezė:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

$$\text{Kriterijaus statistika: } d = \frac{\sum_{i=2}^n (\hat{\epsilon}_i - \hat{\epsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\epsilon}_i^2}$$

Sprendimo priėmimo taisyklė. Tegul reikšmingumo lygmuo lygus 0.05. Iš V. Čekanavičiaus ir G. Murausko „Statistika ir jos taikymai“ 2 d. 13 lent. randamos atitinkamos d_L ir d_U reikšmės. Hipotezė H_0 atmetama (liekanų autokoreliacija statistiškai nereikšminga), jeigu $d_U < d < 4 - d_U$. Jokių statistinių išvadų negalima padaryti, kai $d_L \leq d \leq d_U$ arba $4 - d_U \leq d \leq 4 - d_L$.
(Žr. [1])

2.4 Andersono ir Darlingo testas

Tarkime, kad $X_1, \dots, X_n$ yra tolydžiojo atsitiktinio dydžio X imtis, $X_{(1)}, \dots, X_{(n)}$ - variancinė eilutė, $F_n(x)$ - empirinė pasiskirstymo funkcija (EDF):

$$F_n(x) = \begin{cases} 0, & \text{kai } x \leq X_{(1)}, \\ (i-1)/n, & \text{kai } X_{(i-1)} < x \leq X_{(i)}, i = 1, \dots, n, \\ 1, & \text{kai } X_{(n)} < x. \end{cases} \quad (2.8)$$

Tarkime, kad X skirstinys priklauso tolydžiųjų skirstinių šeimai P . Reikia patikrinti hipotezę H_0 , kad stebimo atsitiktinio dydžio X pasiskirstymo funkcija yra $F(x)$, t.y.

$$H_0 : MF_n(x) \equiv F(x). \quad (2.9)$$

Jeigu hipotezė H_0 teisinga, tai atsitiktinio dydžio $Y = F(X)$ skirstinys yra tolygusis intervale $(0, 1)$. Todėl statistikos grindžiamas skirtumu $F_n(X) - F(X)$, skirstinys, kai hipotezė H_0 teisinga, nepriklauso nuo $F(x)$, nes atliekant monotonišką abscisių ašies transformaciją tas skirtumas nesikeičia.

Nukrypimas nuo hipotezės matuojamas kvadratine metrika, t.y. tikrinama hipotezė H_0 su alternatyva H_1 , kuri užrašoma taip:

$$H_1 : \int_{-\infty}^{\infty} (MF_n(x) - F(x))^2 \psi(F(x)) dF(x) > 0 \quad (2.10)$$

čia $\psi(t)$ - neneigiama funkcija, apibrėžta intervale $[0, 1]$, be to, $\psi(t)$, $t\psi(t)$ ir $t^2\psi(t)$ yra integruojamos intervale $[0, 1]$. Kriterijaus statistika:

$$\omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \psi(F(x)) dF(x). \quad (2.11)$$

Dažniausiai naudojamos tokios svorio funkcijos: $\psi \equiv 1$, $\psi = (t(1-t))^{-1}$. Andersono ir Darlingo statistika A^2 (gaunama ω_n^2 paėmus $\psi = (t(1-t))^{-1}$) apibrėžiama:

$$A^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 (F(x)(1-F(x)))^{-1} dF(x) \quad (2.12)$$

A^2 apskaičiuojam taip:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \log U_{(i)} + (2n+1-2i) \log(1-U_{(i)})] \quad (2.13)$$

čia $U_{(i)} = F(X_{(i)})$.

2.5 Regresinė panelinių duomenų analizė

Paneliniai duomenys (blokuotieji duomenys arba sekiniai) - stebimas tų pačių N objektų kitimas laike T . Yra dviejų tipų blokuotieji duomenys: subalansuoti ir nesubalansuoti. Subalansuoti yra tuomet, kai visi objektai stebimi visais laiko momentais, nesubalansuoti - ne visuose laiko momentuose objektai yra stebėti. Laikoma, kad atskiro objekto reikšmės laike gali būti koreliuotos, tačiau atskiri objektai yra nepriklausomi.

Regresoriai gali būti trijų tipų:

- Kintantys regresoriai x_{it} - kinta ir laike, ir atskiruose objektuose;
- Laike nekintantys regresoriai $x_{it} = x_i$ su visais t ;
- Objektų atžvilgiu nekintantys regresoriai $x_{it} = x_t$ su visais i .

Panelinių duomenų modeliai

- Jungusis modelis (ang. pooled model)

Tai yra bendro laisvojo nario regresinis modelis:

$$Y_{it} = \alpha + X_{it}\beta + u_{it}, i = 1, \dots, N, t = 1, \dots, T$$

Visi stebimi objektai yra homogeniški, todėl konstanta α yra vienoda jiems visiems.

- Fiksuoto efekto modelis (ang. fixed effect model)

Stebimi objektai nėra homogeniški, todėl skiriasi jų konstantos α_i .

$$Y_{it} = \alpha_i + X_{it}\beta + u_{it}, i = 1, \dots, N, t = 1, \dots, T \quad (2.14)$$

Fiksuoto efekto modelyje individualūs efektai α_i gali koreliuoti su regresoriais X_i . α_i laikomi laisvaisiais nariais.

Kiekvienas objektas turi skirtingą laisvąjį narį α_i , tačiau krypties koeficientas β yra tas pats:

$$y_{it} = \alpha_i + X_{it}\beta + \epsilon_{it} \quad (2.15)$$

α_i įvertinio radimas: $\hat{\alpha}_i = \bar{y}_i - \bar{X}_i\beta$.

Individualūs efektai prideda priklausomam kintamajam papildomos variacijos, kuri negali būti paaiškinta turimomis kovariantėmis.

Fiksuotų efektų modelis panaikina šiuos individualius efektus:

$$y_{it} - \bar{y}_i = (\alpha_i - \bar{\alpha}_i) + (X_{it} - \bar{X}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i) \quad (2.16)$$

(Žr. [3] ir [6])

2.6 Klasterinė analizė

Klasteris yra panašių objektų grupė. Taikant klasterinę analizę, yra nustatomi objektų panašumai, pagal kuriuos jie suskirstomi į klasterius. Klasterinės analizės tikslas - suskirstyti objektus taip, kad skirtumai klasterių viduje būtų kuo mažesni, o tarp klasterių - kuo didesni. Dažniausiai naudojami panašumo matai:

1. Metriniai atstumo matai;
2. Koreliacijos koeficientai;
3. Asociatyvumo koeficientai.

Metriniai atstumo matai naudojami tada, kai objektus charakterizuojantys požymiai matuojami pagal intervalų arba santykių skalę. Daugelis metrinių atstumų matų yra metrikos. Metrika – tai skaitinė neneigiama dviejų objektų X ir Y funkcija $d(X, Y)$, tenkinanti sąlygas:

1. Simetriškumo: $d(X, Y) = d(Y, X)$;
2. Trikampio nelygybės: $d(X, Y) \leq d(X, Z) + d(Y, Z)$;
3. Netapačių objektų atskiriamumo, jei $X \neq Y$, tai $d(X, Y) \neq 0$;
4. Tapačių objektų neatskiriamumo, jei $d(X, Y) = 0$, tai X ir Y identiški.

Išvardytų sąlygų natūralumas yra suprantamas. Tai iliustruojama remiantis Euklido atstumo kvadratu - vienu iš metrinių atstumų matų.

Euklido atstumas objektams X ir Y apibrėžiamas taip:

$$D(X, Y) = \|X - Y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2} \quad (2.17)$$

Jei duomenys yra kiekybiniai, objektų panašumui įvertinti galima taikyti tiesinės koreliacijos koeficientą. Jį galima apskaičiuoti taip:

$$\rho_{XY} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} \quad (2.18)$$

Kur x_i yra X objekto i -ojo požymio reikšmė, y_j – Y objekto j -ojo požymio reikšmė, o m – matuojamų požymių skaičius. Asociatyvumo koeficientas naudojamas objektų panašumui pagal kokybinius požymius įvertinti:

$$d(X, Y) = \#(x_i \neq y_i)/m; \quad (2.19)$$

čia $\#(x_i \neq y_i)$ - nesutampančių požymių reikšmių skaičius.

Klasterių sudarymo metodų yra daug. Jie skiriami pagal tai, kaip parenkami panašumo matai, atstumo tarp klasterių nustatymo kriterijai bei kokia skirstymo į klasterius strategija. Pagrindinės klasterinės analizės metodų klasės yra hierarchiniai ir nehierarchiniai. Hierarchiniai dar skirstomi į jungimo ir skaidymo. Vienas dažniausiai naudojamų nehierarchinių metodų yra k -vidurkių metodas.

Yra įvairių klasterių panašumo matų, šiame darbe naudojami trys:

1. Vienetinės jungties (artimiausio kaimyno)
 $d(U, V) = \min_{X_i \in U, Y_j \in V} d(X_i, Y_j)$, čia X_i - i -asis U objektas, Y_j - j -asis V objektas.
Šiuo metodu jungiami klasteriai, turintys artimiausius kaimynus.

2. Vidutinės jungties metodas

$$d(U, V) = \sum_{X_i \in U} \sum_{Y_j \in V} d(X_i, Y_j) / (n_U n_V), \text{ čia } n_U, n_V - \text{klasterių objektų skaičius.}$$

3. Vordo metodas

$$d(U, V) = \|\bar{U} - \bar{V}\|^2 / (1/n_U + 1/n_V)$$

Taikant Vordo metodą, imama klasterių pora ir apskaičiuojama požymių vektorių nuokrypių nuo jungtinio klasterio centro kvadratų suma. Tas pats daroma ir su kita pora ir t.t. Jungiami tie klasteriai, kuriuos sujungus nuokrypių suma mažiausia.

Klasterių skaičiaus parinkimui naudojami šie kriterijai:

1. Kubinio klasterizavimo kriterijus (CCC). Kai CCC reikšmė yra didesnė už 3, tada yra nusakomas geras klasterizavimas. Kuo didesnė CCC reikšmė, tuo geresnis klasterizavimas.
2. Pseudo F statistika (PFS). Naudojant šį kriterijų žiūrima į pseudo F reikšmę. Kuo ši reikšmė didesnė, tuo geriau nusakomas klasterizavimas.
3. Pseudo t^2 statistika. Naudojant šį kriterijų, žiūrima į pseudo t^2 reikšmę. Kuo šio kintamojo reikšmė mažesnė ir jos kritimas didesnis, nuo prieš tai buvusių klasterių skaičiaus, tuo geriau nusakomas klasterizavimas.

(Žr. [1])

3 PRAKTINĖ DALIS

3.1 Statistiniai duomenys

Duomenys surinkti iš Eurostat, Lietuvos statistikos departamento duomenų bazių ir Europos migracijos tinklo internetinio puslapio. Visi duomenys yra metiniai. Visuose duomenyse ES yra laikoma 28 dabartinės Europos Sąjungos valstybės.

Eurostat (žr. [4]) duomenys:

- Migracijos rodiklis tūkstančiui gyventojų, 2004-2014 m. Rodiklis apskaičiuojamas neto migraciją (imigracija - emigracija), padalinus iš gyventojų skaičiaus ir padauginus iš 1000.
- Nedarbo lygis (proc.), 2004-2014 m. Rodiklis apskaičiuojamas bedarbių (pagal Lietuvos darbo biržos duomenis) skaičių padalinus iš visų darbingo amžiaus (15-64 m.) žmonių skaičiaus ir padauginus iš 100.
- Bendrasis vidaus produktas (toliau - BVP) eurų vienam gyventojui, 2004-2014 m.
- Namų ūkių grynosios (atskaičius mokesčius) metinės pajamos (Eur), 2004-2014 m.
- Kainų lygio indeksai, 2004-2014 m. Vienodo prekių krepšelio kainos kiekvienoje šalyje padalinamos iš ES (28 valst.) kainų vidurkio.
- S20/S80 rodiklis, 2004-2014 m. 20% didžiausias pajamas gaunančių asmenų pajamos, padalintos 80% mažiausias pajamas gaunančių asmenų pajamų.

Lietuvos statistikos departamento (žr. [8]) duomenys:

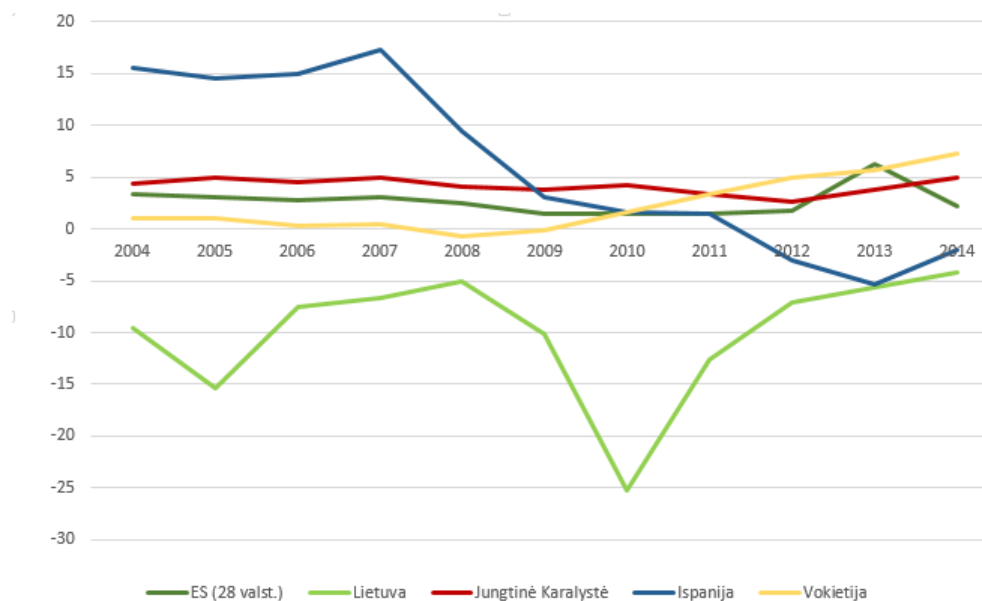
- Nedarbo lygis procentais Lietuvos savivaldybėse, 2014 m.
- Tarptautinė emigracija tūkstančiui gyventojų, 2014 m. Rodiklis apskaičiuojamas registruotų bedarbių skaičių padalinus iš darbingo amžiaus gyventojų skaičiaus.

Europos migracijos tinklo internetinio puslapio (žr. [5]) duomenys:

- Lietuvių emigrantų skaičius pagal būsimą gyvenamąją valstybę, 2010 - 2014 m.

3.2 Migracijai įtaką darantys veiksniai

Dabartiniais laikais migracija yra itin dažnas reiškinys. Migracija yra dviejų tipų: teritorinė (vidinė) ir tarptautinė (išorinė). Teritorinė migracija vyksta valstybės viduje, o tarptautinė migracija - tarp valstybių. Nagrinėjant tarptautinę migraciją, pirmiausia galima išskirti du valstybių tipus: valstybes, į kurias yra daugiau imigruojama, ir valstybes, iš kurių daugiau emigruojama. Objektiviam migracijos nagrinėjimui pasirenkamas migracijos rodiklis tūkstančiui gyventojų. Jei rodiklio reikšmė yra neigiama, reiškia emigracija viršija imigraciją ir atvirkščiai - teigiamas rodiklis parodo, kad vyrauja imigracija. Kadangi duomenyse yra 32 valstybės ir laiko eilutė sudaryta iš vienuolikos metų, grafike (3.2.1 pav.) vaizduojamos tik kelios valstybės - Lietuva, Jungtinė Karalystė, Ispanija, Vokietija ir ES (28 valst.).

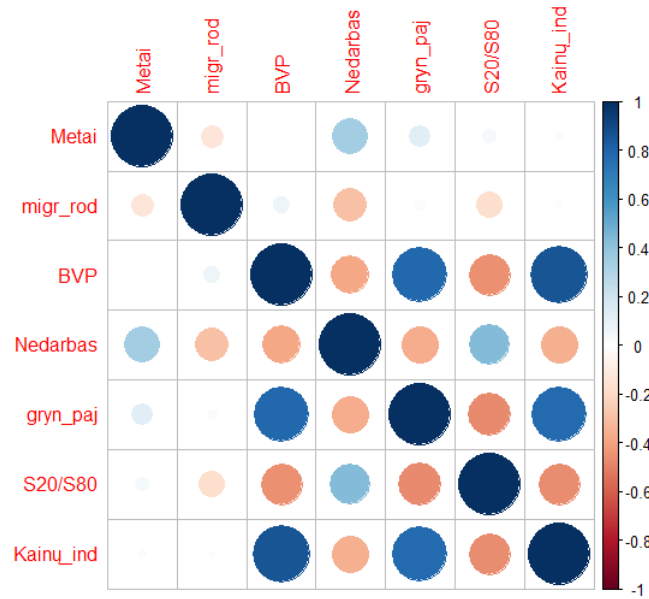


3.2.1 pav.: Migracijos rodiklis tūkstančiui gyventojų

Grafiko duomenys rodo, kad migracija Europoje yra pasiskirsčiusi labai netolygiai. Bendras ES migracijos rodiklio tūkstančiui gyventojų vidurkis yra teigiamas, 2009 - 2012 m. rodiklis buvo šiek tiek sumažėjęs, o 2013 m. šoktelėjo iki 6,3, Jungtinėje Karalystėje rodiklis visame laikotaryje yra stabilus, didelių pokyčių nevyko. Beveik visais metais viršija ES vidurkį, išskyrus 2013 metais. Tuo tarpu Lietuvos rodiklis visą laiką yra neigiamas. Migracijos situacija Lietuvoje plačiau aptariama 3.3 skyrelyje. Ekonominė krizė paveikė migracijos srautus, yra tokių valstybių, kurios iš valstybių į kurias emigruojama, pačios tapo valstybėmis, kuriose vyrauja emigracija. Tokių valstybių tarpe yra Ispanija. Grafike matomas didžiulis pokytis - 2007 m. migracijos rodiklis tūkstančiui gyventojų buvo net 17,2, o 2013 m. nusileido iki -5,4. Vokietijoje ekonominė krizė atnešė priešingų pokyčių, ten migracijos rodiklis stipriai išaugo ir Vokietija tapo populiari emigracijos kryptis. 2008 m. migracijos rodiklis joje buvo -0,7, rodiklis kasmet auga, o 2014 m. yra net 7,2.

Toliau nagrinėjama, kokią įtaką migracijai daro ekonominiai rodikliai, kurie iš jų turi reikšmingos įtakos migracijai, kurie - ne.

Apjungus duomenis, gaunami paneliniai duomenys. Jie sudaryti iš metų, valstybių, migracijos rodiklio tūkstančiui gyventojų, BVP, grynujų metinių namų ūkių pajamų, nedarbo lygio, kainų lygių indekso ir $S20/S80$ indekso. Nubraižomas Pirsono koreliacijos koeficientų grafikas (3.2.2 pav.) ir lentelė bei p reikšmės (3.2.3 pav.). Tarp migracijos ir kitų rodiklių silpna neigiama koreliacija yra tik su nedarbu, tačiau p reikšmė yra mažesnė, nei 0.001, todėl ji nėra reikšminga.



3.2.2 pav.: Koreliacijos koeficientų grafikas

```

Metai      Metai      migr_rod      BVP      Nedarbas      gryn_paj      S20/S80      Kainos
Metai      1.00000000  -0.13055047  0.01795184  0.3423117  0.11387804  0.04977841  0.02118378
migr_rod   -0.13055047  1.00000000  0.04319478 -0.2916419  0.02129152 -0.17619652 -0.01368926
BVP         0.01795184  0.04319478  1.00000000 -0.3439457  0.89802842 -0.53718662  0.86117204
Nedarbas   0.34231166 -0.29164192 -0.34394571  1.00000000 -0.31513498  0.43482108 -0.35423080
gryn_paj    0.11387804  0.02129152  0.89802842 -0.3151350  1.00000000 -0.50059790  0.76409138
S20/S80     0.04977841 -0.17619652 -0.53718662  0.4348211 -0.50059790  1.00000000 -0.46461038
Kainos      0.02118378 -0.01368926  0.86117204 -0.3542308  0.76409138 -0.46461038  1.00000000
> cor.test(du$migr_rod,du$BVP)$p.value
[1] 0.4613907
> cor.test(du$migr_rod,du$Nedarbas)$p.value
[1] 3.741444e-07
> cor.test(du$migr_rod,du$gryn_paj)$p.value
[1] 0.7166532
> cor.test(du$migr_rod,du$`S20/S80`)$p.value
[1] 0.002471397
> cor.test(du$migr_rod,du$Kainos)$p.value
[1] 0.8155039

```

3.2.3 pav.: Koreliacijos koeficientų lentelė ir p reikšmės

Išsamesniam migracijos rodiklio su kitais turimais rodikliais ryšio tyrimui, taikoma regresinė analizė. Pirmiausiai parenkamas jungusis modelis.

$$I_{it} = \alpha + \beta_1 BVP_{it} + \beta_2 GP_{it} + \beta_3 N_{it} + \beta_4 K_{it} + \beta_5 S_{it} + \epsilon_{it}, i = 1, \dots, 32, t = 1, \dots, 11$$

Čia i - valstybės, t - metai, α - laisvasis narys, BVP - bendrasis vidaus produktas, GP - grynosios pajamos, N - nedarbo lygis, K - kainų lygių indeksas, S - $S20/S80$, ϵ - atsitiktinės paklaidos.

Tariame, kad visi koeficientai nepriklauso nei nuo valstybės i , nei nuo metų t . Visus įrašus traktuojame kaip vieną vienalaikį duomenų rinkinį.

```
> jung=lm(migr_rod ~ `Kainos`+ Nedarbas + `S20/S80`+ BVP + `gryn_paj`, data=pduom)
> summary(jung)

Call:
lm(formula = migr_rod ~ Kainos + Nedarbas + `S20/S80` + BVP +
    gryn_paj, data = pduom)

Residuals:
    Min       1Q   Median       3Q      Max
-25.3723  -3.4694  -0.1874   2.8488  19.0478

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  45.31645    17.47550     2.593   0.0100 **
Kainos       -7.43736     3.03379    -2.452   0.0148 *
Nedarbas     -0.46544     0.09718    -4.790 2.68e-06 ***
`S20/S80`    -0.68968     0.42197    -1.634   0.1033
BVP           2.25310     1.64866     1.367   0.1728
gryn_paj     -2.58809     2.64156    -0.980   0.3280
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.455 on 287 degrees of freedom
(59 observations deleted due to missingness)
Multiple R-squared:  0.1182,    Adjusted R-squared:  0.1029
F-statistic: 7.698 on 5 and 287 DF,  p-value: 8.257e-07
```

3.2.4 pav.: Modelio taikymo rezultatai

Pasirinkus 0.05 reikšmingumo lygmenį, testas rodo, kad labai reikšmingos kovariantės yra laisvasis narys ir nedarbo lygis, šiek tiek mažiau reikšminga kovariantė - kainų lygio indeksas. Tai reiškia, kad augant kainoms ir didėjant nedarbui, migracijos rodiklis mažėja, vadinasi auga emigracija ir mažėja imigracija. Tam, kad būtų palikti tik reikšmingi regresoriai, pritaikoma pažingsninė regresija.

```
> step(lm(migr_rod ~ `Kainos`+ Nedarbas + `S20/S80`+ BVP + `gryn_paj`, data=pduomen),
+ direction = "backward")
Start: AIC=1098.77
migr_rod ~ Kainos + Nedarbas + `S20/S80` + BVP + gryn_paj

              Df Sum of Sq  RSS   AIC
- gryn_paj    1    40.00 11999 1097.8
- BVP         1    77.83 12037 1098.7
<none>                        11959 1098.8
- `S20/S80`   1    111.32 12071 1099.5
- Kainos      1    250.44 12210 1102.8
- Nedarbas    1    955.97 12915 1119.3

Step: AIC=1097.75
migr_rod ~ Kainos + Nedarbas + `S20/S80` + BVP

              Df Sum of Sq  RSS   AIC
- BVP         1    37.83 12037 1096.7
<none>                        11999 1097.8
- `S20/S80`   1    105.59 12105 1098.3
- Kainos      1    242.52 12242 1101.6
- Nedarbas    1    954.05 12954 1118.2

Step: AIC=1096.67
migr_rod ~ Kainos + Nedarbas + `S20/S80`

              Df Sum of Sq  RSS   AIC
<none>                        12037 1096.7
- `S20/S80`   1    160.08 12197 1098.5
- Kainos      1    331.59 12369 1102.6
- Nedarbas    1    947.01 12984 1116.9

Call:
lm(formula = migr_rod ~ Kainos + Nedarbas + `S20/S80`, data = pduomen)

Coefficients:
(Intercept)      Kainos      Nedarbas      `S20/S80`
   40.5350    -4.9599    -0.4632    -0.7895
```

3.2.5 pav.: Pažingsninė regresija

Atlikus pažingsninę regresiją, paliekami tik nedarbo, kainų lygio ir $S20/S80$ rodikliai.

Gaunamas toks įvertis:

$$\hat{I}_{it} = 40.54 - 0.46N_{it} - 5K_{it} - 0.79S_{it}, i = 1, \dots, 32, t = 1, \dots, 11$$

Tam, kad patikrinti ar tarp liekamųjų paklaidų nėra autokoreliacijos, naudojamas Durbino ir Vatsono testas.

$$H_0: \rho = 0$$

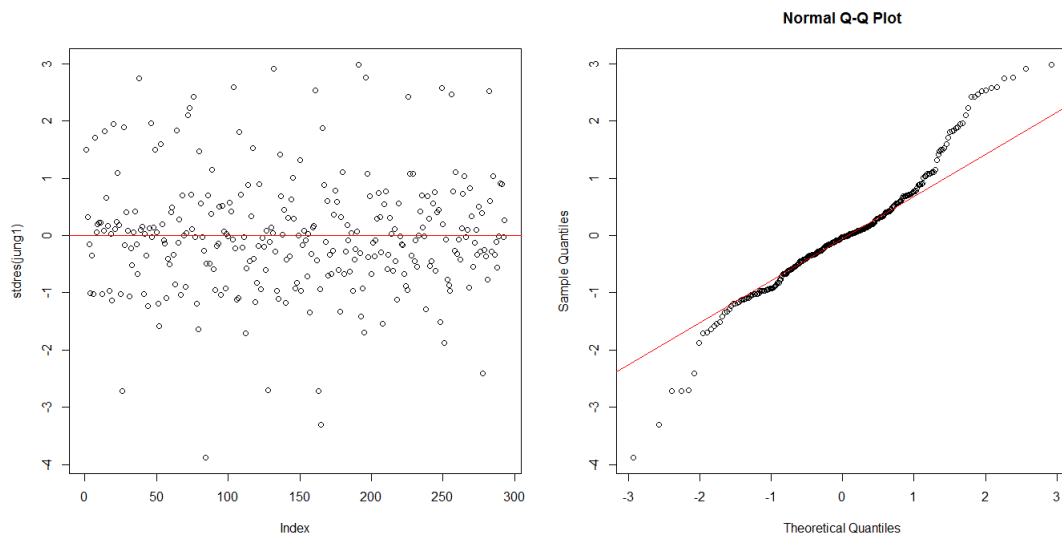
$$H_1: \rho \neq 0$$

Gauta testo reikšmė apytiksliai lygi 1.99, tai reiškia, kad autokoreliacijos nėra ir H_0 atmetama, nes reikšmė labai artima 2.

```
> durbinwatsonTest(jung$residuals)
[1] 1.958257
```

3.2.6 pav.: Durbino ir Vatsono testas

Taip pat reikia patikrinti, ar paklaidos nėra heteroskedastiškos. Tam nubraižomas taškinis standartizuotų liekanų grafikas ir pavaizduojamas jų vidurkis (3.2.7 1. pav.). Kadangi vidurkis yra 0, o liekanos išsidėsčiusios tolygiai aplink jį, galima daryti išvadą, kad liekanos yra homoskedastiškos.



3.2.7 pav.: 1. Standartizuotų liekamųjų paklaidų grafikas. 2. Kvantilių - kvantilių grafikas

Visi kintamieji turi tenkinti normalumo sąlygą. Antrame grafike (3.2.7 pav.) pavaizduotas kvantilių grafikas, kur ant vienos ašies atidedami teoriniai, ant kitos - empiriniai kvantiliai. Esant normaliajam pasiskirstymui taškai turi išsidėstyti arti linijos. Šiuo atveju dalis taškų nutolę nuo linijos. Toliau taikomas Andersono ir Darlingo testas normalumui tikrinti.

$$H_0: \text{Liekamosios paklaidos pasiskirsčiusios normaliai}$$

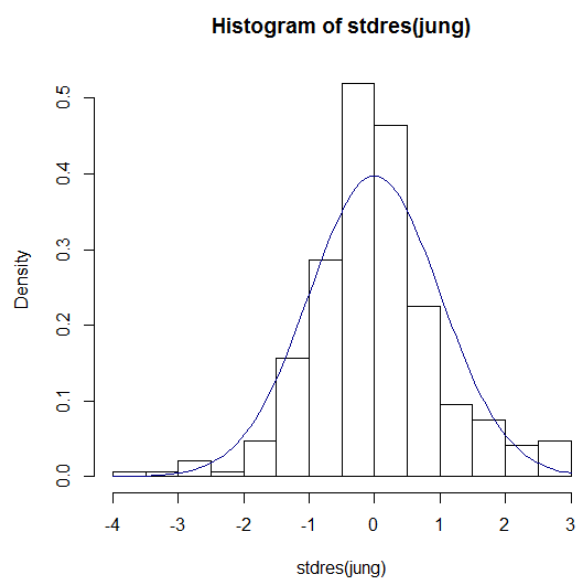
$$H_1: \text{Liekamosios paklaidos nėra pasiskirsčiusios normaliai}$$

```
> ad.test(stdres(jung))

Anderson-Darling normality test

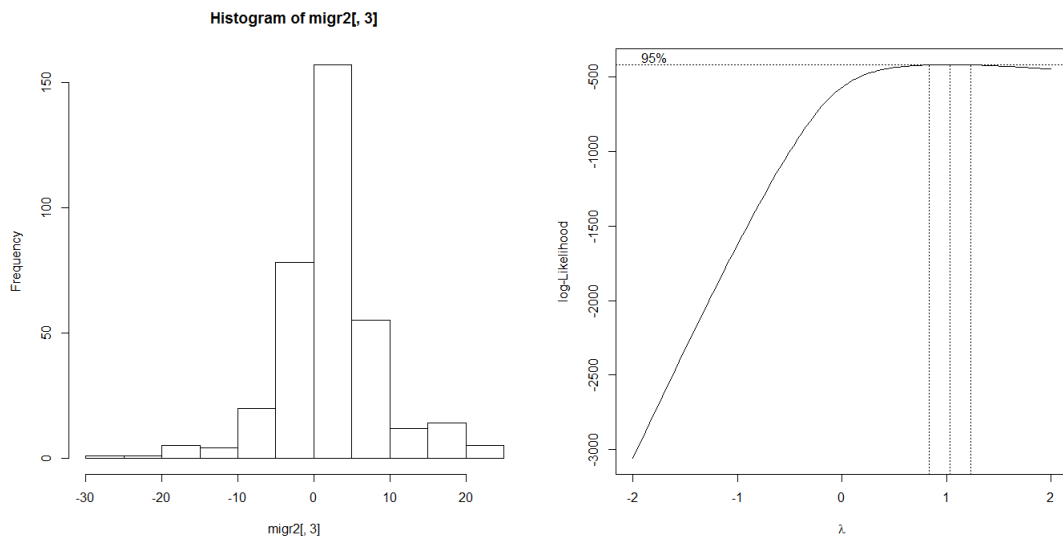
data:  stdres(jung)
A = 3.7272, p-value = 2.568e-09
```

3.2.8 pav.: Andersono ir Darlingo testas



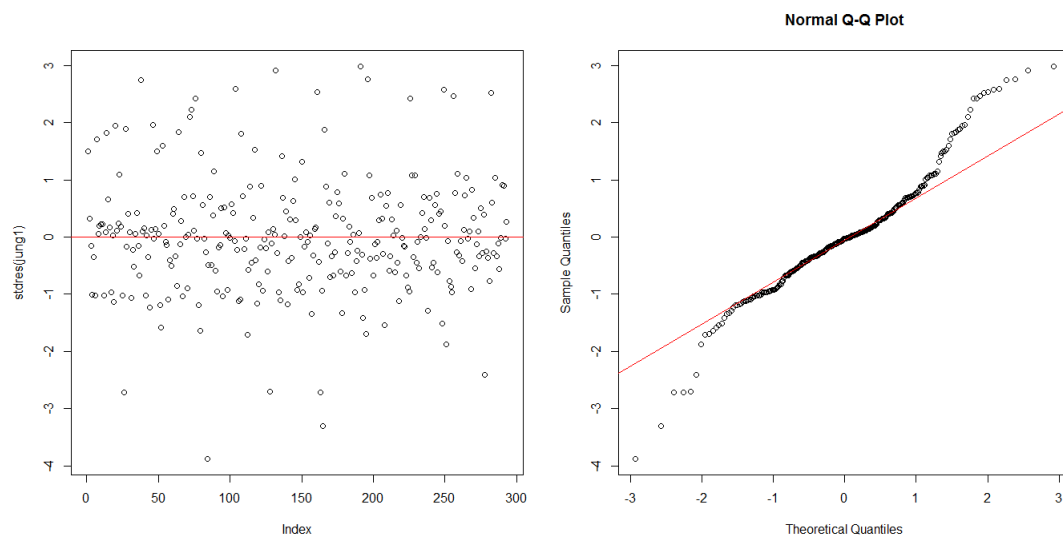
3.2.9 pav.: Liekamųjų paklaidų histograma

Iš gautų rezultatų matoma, kad hipotezės apie liekamųjų paklaidų normalųjį pasiskirstymą priimti negalime, nes 0.05 reikšmingumo lygmuo yra didesnis už gautą p reikšmę. Pavaizduojama paklaidų histograma (3.2.9 pav.). Gaunama histograma nėra simetriška. Tam, kad normalizuoti liekamąsias paklaidas, atliekama priklausomo kintamojo regresijoje - migracijos rodiklio Bokso ir Kokso transformacija.



3.2.10 pav.: 1. Migracijos rodiklio tūkstaniui gyventojų duomenų pasiskirstymas
2. Bokso ir Kokso transformacijos grafikas

Aukščiausias tikėtinas, kad duomenys bus pasiskirstę normaliai, įgyjamas, kai λ yra 1.03. Todėl dar kartą taikomas tas pats modelis, tik migracijos rodiklis pakeliamas 1.03 laipsniu. Regresijos rezultatų šį transformacija nepakeičia, gaunami tokie patys, kaip ir 3.2.4 paveikslėlyje. Andersono ir Darlingo testo reikšmė mažesnė už 0.05, todėl hipotezė apie duomenų normalųjį pasiskirstymą atmetama. Paveikslėlyje 3.2.11 matoma, kad ir transformacija nepadėjo normalizuoti duomenų.



3.2.11 pav.: 1. Standartizuotų liekamųjų paklaidų grafikas. 2. Kvantilių - kvantilių grafikas

Kita problema, galinti lemti standartizuotųjų liekamų normaliojo pasiskirstymo nebuvimą yra išskirtys duomenyse. Surandame šias išskirtis.

```
> is[is$`stdres(jung)`< -2.7 | is$`stdres(jung)`> 2.7,]
      stdres(jung) row.names(DF)
2005-Lietuva      -2.719708    2005-Lietuva
2006-Airija        2.745853    2006-Airija
2007-Rumunija     -3.878566    2007-Rumunija
2009-Islandija    -2.707802    2009-Islandija
2009-Kipras        2.912332    2009-Kipras
2010-Latvija      -2.721867    2010-Latvija
2010-Lietuva      -3.311370    2010-Lietuva
2011-Kipras        2.985254    2011-Kipras
2011-Liuksemburgas 2.753487    2011-Liuksemburgas
```

3.2.12 pav.: Duomenų išskirtys

Valstybės, kurios patenka tarp išskirčių yra pašalinamos. Dar kartą taikomas jungusis modelis. Gaunami šiek tiek kitokie rezultatai - nedarbas išlieka reikšminga kovariante, o vietoj kainų indekso, reikšminga kovariante tampa $S20/S80$ rodiklis.

```
Call:
lm(formula = migr_rod ~ Kainos + Nedarbas + `S20/S80` + BVP +
    gryn_paj, data = pduom2)

Residuals:
    Min       1Q   Median       3Q      Max
-7.5036 -2.4783 -0.0025  1.8318  8.0448

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 44.97202   10.83168   4.152 4.86e-05 ***
Kainos        1.15597    2.07343    0.558  0.5778
Nedarbas     -0.13063    0.05554   -2.352  0.0196 *
`S20/S80`    -0.49758    0.24585   -2.024  0.0443 *
BVP          -0.65115    1.33354   -0.488  0.6259
gryn_paj     -0.88542    1.97589   -0.448  0.6546
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.212 on 202 degrees of freedom
(44 observations deleted due to missingness)
Multiple R-squared:  0.07545,    Adjusted R-squared:  0.05256
F-statistic: 3.297 on 5 and 202 DF,  p-value: 0.006953
```

3.2.13 pav.: Jungusis modelis

Vėl taikoma pažingsninė regresija (3.2.16 pav.), nes yra nereikšmingų kovariančių. Pa-
liekami trys regresoriai: nedarbas, $S20/S80$ ir BVP.

```

> step(lm(migr_rod ~ `kainos` + Nedarbas + `S20/S80` + BVP + `gryn_paj`, data=na.omit(pduom2)),
+ direction = "backward")
Start: AIC=491.34
migr_rod ~ kainos + Nedarbas + `S20/S80` + BVP + gryn_paj

      Df Sum of Sq    RSS   AIC
- gryn_paj  1      2.072 2086.1 489.55
- BVP       1      2.460 2086.5 489.59
- kainos    1      3.207 2087.2 489.66
<none>                        2084.0 491.34
- `S20/S80` 1     42.262 2126.3 493.52
- Nedarbas  1     57.079 2141.1 494.96

Step: AIC=489.55
migr_rod ~ kainos + Nedarbas + `S20/S80` + BVP

      Df Sum of Sq    RSS   AIC
- kainos    1      5.645 2091.8 488.11
<none>                        2086.1 489.55
- BVP       1     23.883 2110.0 489.92
- `S20/S80` 1     45.934 2132.1 492.08
- Nedarbas  1     58.231 2144.3 493.27

Step: AIC=488.11
migr_rod ~ Nedarbas + `S20/S80` + BVP

      Df Sum of Sq    RSS   AIC
<none>                        2091.8 488.11
- BVP       1     30.674 2122.4 489.14
- `S20/S80` 1     42.476 2134.2 490.29
- Nedarbas  1     64.151 2155.9 492.39

Call:
lm(formula = migr_rod ~ Nedarbas + `S20/S80` + BVP, data = na.omit(pduom2))

Coefficients:
(Intercept)      Nedarbas      `S20/S80`          BVP
    37.7086      -0.1371      -0.4892      -0.6674

```

3.2.14 pav.: Pažingsninė regresija

Gaunamas toks įvertis:

$$\hat{I}_{it} = 37.71 - 0.14N_{it} - 0.49S_{it} - 0.67BVP_{it}, i = 1, \dots, 32, t = 1, \dots, 11$$

Dar kartą tikrinamos modelio prielaidos. Pirmiausia patikrinama ar nėra autokoreliuotumo, tam taikomas Durbinio ir Vatsonio testas.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Gauta testo reikšmė artima 2, tai reiškia, kad autokoreliacijos nėra ir H_0 atmetama.

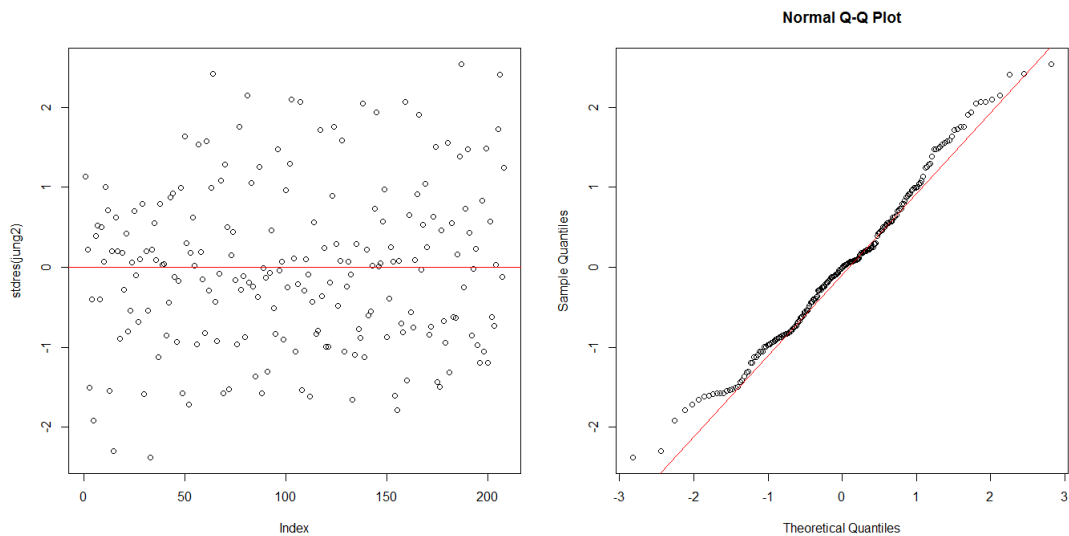
```

> durbinwatsonTest(stdres(jung2))
[1] 2.2075

```

3.2.15 pav.: Durbinio ir Vatsonio testas

Nubraižomas taškinis standartizuotų liekanų grafikas, pavaizduojamas jų vidurkis (3.2.14 pav.) ir tikrinama ar nėra heteroskedastiškumo.



3.2.16 pav.: 1. Standartizuotų liekamųjų paklaidų grafikas. 2. Kvantilių - kvantilių grafikas

Kadangi vidurkis yra apytiksliai 0, o liekamosios paklaidos išsidėsčiusios tolygiai aplink jį, galima daryti išvadą, kad liekanos yra homoskedastiškos. Toliau nagrinėjamas liekamųjų paklaidų pasiskirstymas. Antrame paveikslėlyje matoma, kad kvantilių reikšmės išsidėsčiusios arti linijos ir „uodegos“ nėra taip stipriai nutolusios nuo linijos, kaip 3.2.11 paveikslėlyje. Taikomas Andersono ir Darlingo testas paklaidų normalumui tikrinti.

H_0 : Standartizuotos liekamosios paklaidos pasiskirsčiusios normaliai

H_1 : Standartizuotos liekamosios paklaidos nėra pasiskirsčiusios normaliai

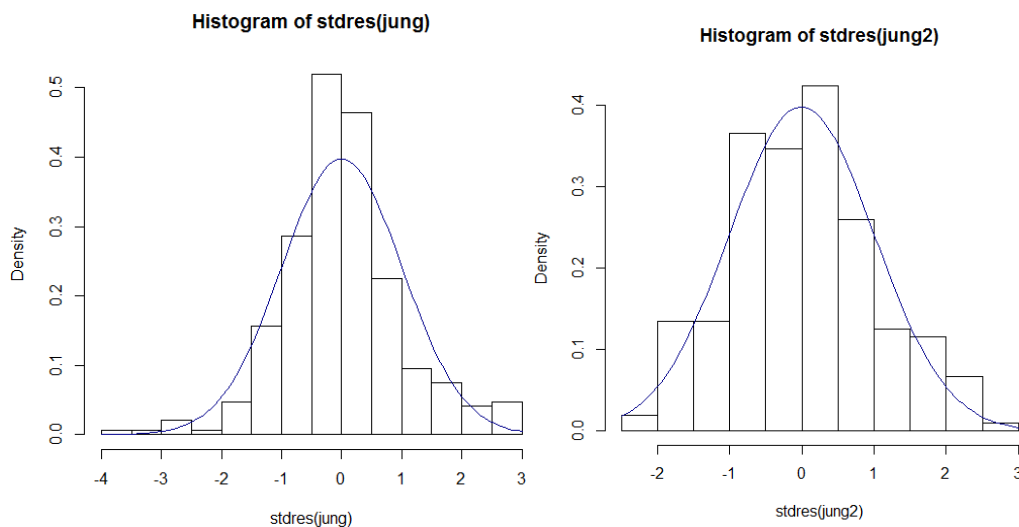
```
> ad.test(stdres(jung2))

Anderson-Darling normality test

data:  stdres(jung2)
A = 0.68167, p-value = 0.07401
```

3.2.17 pav.: Andersono ir Darlingo testas

Andersono ir Darlingo testo p reikšmė yra apytiksliai 0.07 ir su $\alpha = 0.05$, H_0 yra neatmetama. Tai reiškia, kad standartizuotos liekamosios paklaidos yra normaliai pasiskirsčiusios.



3.2.18 pav.: Liekamųjų paklaidų histogramų palyginimas

3.2.18 paveikslėlyje pavaizduotos dvi histogramos su normaliojo skirstinio tankio funkcijos kreivėmis. Pirmoji jau anksčiau pavaizduota - pirmojo jungiojo modelio standartizuotų liekamųjų paklaidų histograma, o antroji - jungiojo modelio taikyto duomenims, iš kurių buvo pašalintos išskirtys, standartizuotų liekamųjų paklaidų histograma. Pašalinus išskirtis, paklaidos tampa normaliai pasiskirsčiusiomis.

Toliau paneliniai duomenys nagrinėjami naudojant R paketą plm. Pažymima, kad panelinius duomenis sudaro stebimi objektai - valstybės ir laikas - metai. Kadangi tolesnei analizei naudojamos plm funkcijos, tai iš pradžių vėl taikomas jungusis modelis (3.2.19 pav.). Rezultatai gaunami tokie patys, kaip ir anksčiau taikytame jungiajame modelyje (3.2.13 pav.).

```
> pooling <- plm(Y~X, data=paneldata, model="pooling")
> summary(pooling)
oneway (individual) effect Pooling Model

Call:
plm(formula = Y ~ X, data = paneldata, model = "pooling")

Unbalanced Panel: n=21, T=5-11, N=208

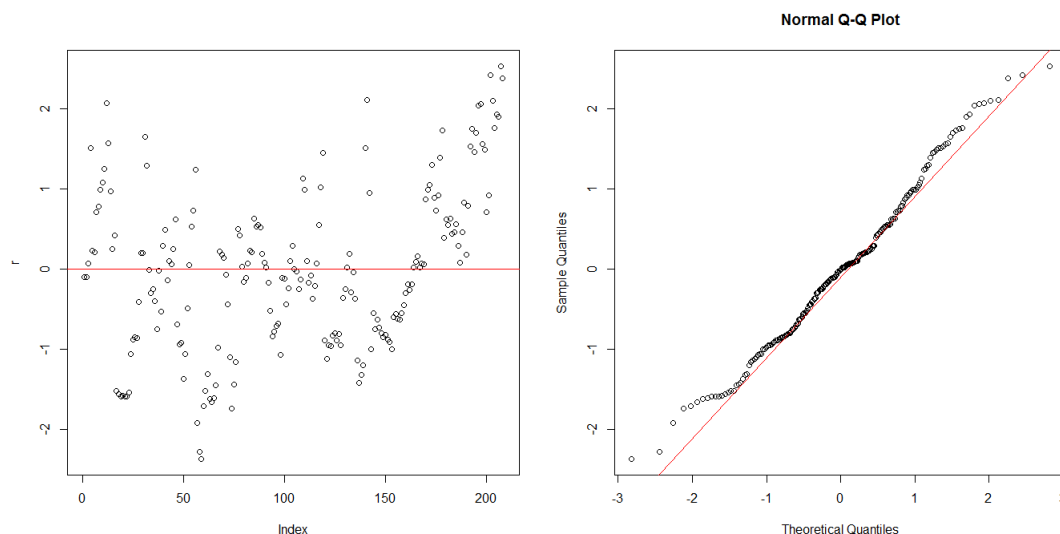
Residuals:
    Min.   1st Qu.   Median   3rd Qu.    Max.
-7.50000 -2.48000 -0.00247  1.83000  8.04000

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  44.972019  10.831679  4.1519 4.861e-05 ***
XBVP         -0.651147   1.333543 -0.4883  0.62588
XNedarbas    -0.130629   0.055536 -2.3521  0.01963 *
Xgryn_paj    -0.885419   1.975889 -0.4481  0.65455
XS20/S80     -0.497580   0.245847 -2.0239  0.04429 *
XKainos       1.155971   2.073434  0.5575  0.57779
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2254.1
Residual Sum of Squares: 2084
R-Squared:                0.075446
Adj. R-Squared:           0.07327
F-statistic: 3.29675 on 5 and 202 DF, p-value: 0.0069528
```

3.2.19 pav.: Jungusis modelis

Nubraižomi jungiojo modelio standartizuotų liekamųjų paklaidų ir kvantilių grafikai (2.2.20 pav.).



3.2.20 pav.: 1. Standartizuotų liekamųjų paklaidų grafikas. 2. Kvantilių - kvantilių grafikas

Iš standartizuotų liekamųjų paklaidų grafiko matosi atskirų grupių, kurios gali sukelti autokoreliaciją. Tai galima pamatyti iš netolygaus liekamųjų paklaidų išsidėstymo aplink vidurkį.

Taikant Durbino ir Watsono testą paneliniams duomenims, tikrinamas paklaidų autokoreliuotumas nagrinėjant atskirus objektus.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

```
> pdwtest(pooling)

Durbin-watson test for serial correlation in panel models

data: Y ~ X
DW = 0.34603, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

3.2.21 pav.: Durbino ir Watsono testas paklaidų autokoreliuotumui tikrinti

Testas parodo, kad H_0 yra atmetama, paklaidos yra autokoreliuotos.

Jungiojo modelio įvertinimui naudojamas „Wooldrige“ testas, kuris parodo ar buvo laike nekintančių individualių efektų, šiuo atveju, ar buvo valstybių, kurios iškreiptų modelio rezultatus.

H_0 : Nulinė koreliacija tarp tos pačios grupės (valstybės) liekamųjų paklaidų

H_1 : Koreliacija tarp tos pačios grupės (valstybės) liekamųjų paklaidų nėra lygi nuliui

```
> pwtest(pooling)

Wooldridge's test for unobserved individual effects

data: formula
z = 2.5911, p-value = 0.009566
alternative hypothesis: unobserved effect
```

3.2.22 pav.: „Wooldrige“ testas individualiems efektams

Testas parodo, kad buvo individualių efektų, į kuriuos jungiajame modelyje nėra atsižvelgiama. Todėl toliau taikomas fiksuotų efektų modelis (2.2.23 pav.), kuris turėtų atsižvelgti į atskirų valstybių efektą modeliui ir jį panaikinti. Be to fiksuotų efektų modelyje daroma prielaida, kad nestebimi individualūs efektai yra koreliuoti su regresoriais.

```
> fixed <- plm(migr_rod~BVP+Nedarbas+gryn_paj+S20.S80`+Kainos, data=paneldata, model="within"
, index = c("valstybė", "Metai"))
> summary(fixed)
Oneway (individual) effect within Model

Call:
plm(formula = migr_rod ~ BVP + Nedarbas + gryn_paj + S20.S80 +
    Kainos, data = paneldata, model = "within", index = c("valstybė",
    "Metai"))

Unbalanced Panel: n=21, T=5-11, N=208

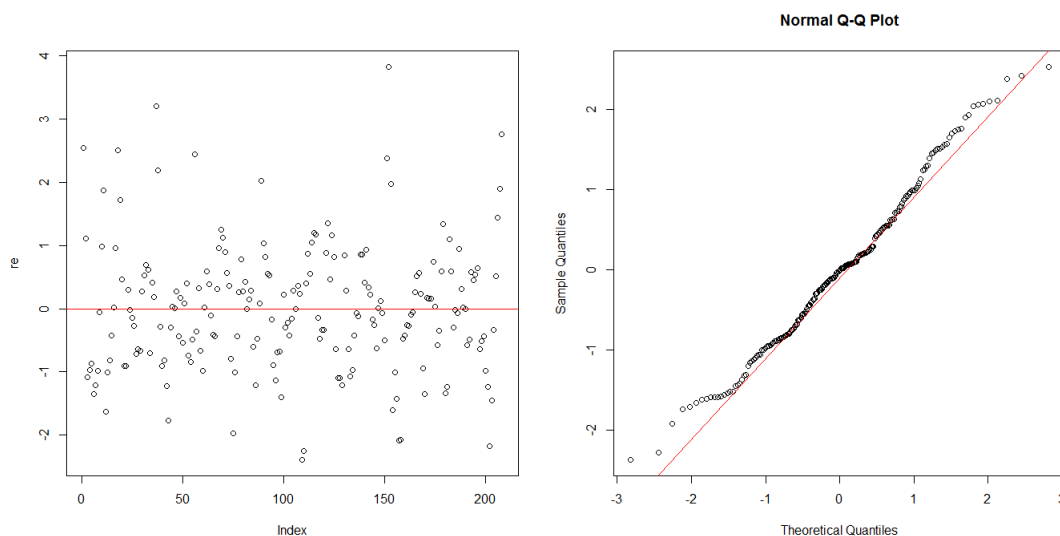
Residuals :
    Min. 1st Qu.  Median 3rd Qu.    Max.
-3.7400 -1.0000 -0.0234  0.8380  6.0000

Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
BVP          -1.282016   1.020114  -1.2567  0.21046
Nedarbas     -0.229520   0.039786  -5.7689 3.372e-08 ***
gryn_paj      3.333021   1.984127   1.6798  0.09470 .
S20.S80       0.602783   0.352382   1.7106  0.08886 .
Kainos        4.532788   2.564263   1.7677  0.07879 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    624.05
Residual Sum of Squares: 507.57
R-Squared:              0.18666
Adj. R-Squared:         0.16333
F-statistic: 8.35363 on 5 and 182 DF, p-value: 3.9418e-07
```

3.2.23 pav.: Fiksuotų efektų modelis

Fiksuotų efektų modelio rezultatai skiriasi nuo jungiojo modelio rezultatų. Esant reikšmingumo lygmeniui 0.05, reikšmingas regresorius tik vienas - nedarbas. Toliau pavaizduojamos fiksuotų efektų modelio liekamosios paklaidos (3.2.24 pav.).



3.2.24 pav.: Fiksuotų efektų modelio liekamųjų paklaidų grafikas

Palyginus su jungiojo modelio liekamųjų paklaidų grafiku (3.2.20 pav.), čia paklaidos daug tolygiau išsidėstę aplink vidurkį ir nesigrupuoja į atskirus klasterius. Patikrinamas autokoreliuotumas grupėse Durbino ir Vatsono testą paklaidų autokoreliuotumui tikrinti.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

```
> pdwtest(fixed)

Durbin-watson test for serial correlation in panel models

data: migr_rod ~ BVP + Nedarbas + gryn_paj + s20.s80 + kainos
DW = 0.9124, p-value = 8.545e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

3.2.25 pav.: Durbino ir Vatsono testas paklaidų autokoreliuotumui tikrinti

Ir fiksuotų efektų modelyje paklaidos yra autokoreliuotos atskirose grupėse.

F testas tikrina ar lyginant fiksuotų efektų ir jungųjų modelius, yra individualių arba laiko efektų.

```
> pFtest(fixed, pooling)

F test for individual effects

data: Y ~ X
F = 10.656, df1 = 29, df2 = 258, p-value < 2.2e-16
alternative hypothesis: significant effects
```

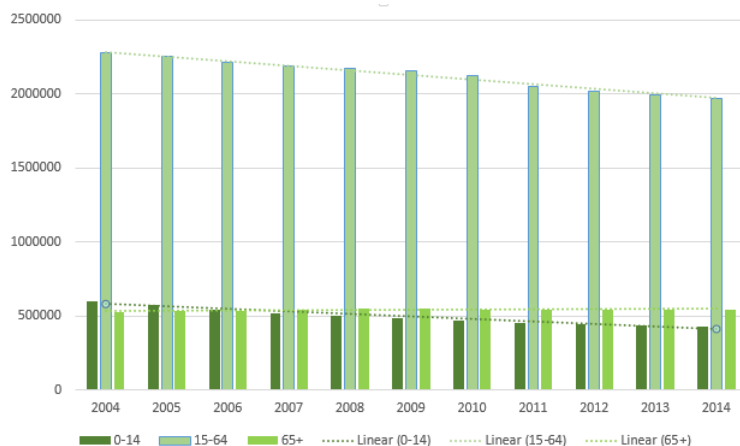
3.2.26 pav.: F testas individualių efektų tyrimui

Maža p reikšmė reiškia, kad viename iš modelių yra randami individualūs arba laiko efektai, tokiu atveju reikia rinktis fiksuotų efektų modelį, kuris atsižvelgia į šiuos efektus.

Pritaikius fiksuotų efektų ir jungųjų modelius, galima daryti išvadą, kad fiksuotų efektų yra tinkamesnis, nes atsižvelgia į individualius efektus. Nors abiem atvejais gaunamas autokoreliuotumas tarp atskirų grupių paklaidų, fiksuotų efektų modelio paklaidos yra gražiau pasiskirsčiusios. Visais atvejais gauta, kad nedarbas yra reikšminga kovariantė, tai reiškia, kad nedarbas tikrai įtakoja migraciją.

3.3 Lietuvos piliečių emigracijos kryptys

Lietuva yra viena iš Europos valstybių, kuriose vyrauja emigracija ir migracijos rodiklis tūkstančiui gyventojų yra neigiamas. Nuo nepriklausomybės atkūrimo iki 2015 metų iš Lietuvos emigravo beveik 870 tūkstančių žmonių. Tai yra milžiniški skaičiai tokiai mažai valstybei. Mūsų šalyje dėl migracijos atsiranda įvairių problemų: visuomenės senėjimas dėl darbingo amžiaus žmonių emigracijos, pensinio amžiaus žmonių dalies didėjimas visoje populiacijoje, kvalifikuotos darbo jėgos trūkumas, gimstamumo mažėjimas. Šias tendencijas galime matyti 3.3.1 pav., grafike pavaizduota, kaip nagrinėjamame laikotarpyje kito žmonių skaičius 0-14, 15-64 ir 65+ amžiaus grupėse.



3.3.1 pav.: Gyventojų skaičius skirtingose amžiaus grupėse

Grafike matomas žymus darbingo amžiaus žmonių (15-64 m.) ir vaikų (0-14 m.) mažėjimas, o pensinio amžiaus grupėje žmonių skaičius auga, tačiau ne tokiu sparčiu tempu. Galima daryti prielaidą, kad ateityje šios problemos tik gilės, nes mažėjant gimstamumui, darbingo amžiaus žmonių skaičius dar sparčiau mažės. Pažiūrėkime kaip keitėsi migracijos mastai Lietuvoje 2004-2014 m.(3.3.2 pav.).



3.3.2 pav.: Migracijos rodiklis Lietuvoje 2004-2014 m.

Visame laikotarpyje migracijos rodiklis tūkstančiui gyventojų yra neigiamas. Daug gyventojų emigravo po įstojimo į Europos Sąjungą - 2005 m. migracijos rodiklis buvo pasiekęs -15,4, tada šiek tiek pakilo, prasidėjus 2008-2010 m. pasaulinei ekonomikos krizei, dar stipriau nusileido 2010 m. net iki -25,2. Pasibaigus krizei, migracijos rodiklis kyla ir

iš tendencijos tiesės matome, kad bendra tendencija yra rodiklio augimas ir 2014 m., jis buvo didžiausias per visą nagrinėjamą laikotarpį - 4,2. Tai reiškia, kad po truputėlį mažėja emigracija, tačiau ji vis dar yra labai žema, lyginant su ES vidurkiu 2014 m. - 2,2.

Toliau panagrinėsime, kur Lietuvos piliečiai išvyksta gyventi ir dirbti. Tam, kad nustatyti kokios valstybės yra populiariausios emigracijos kryptys, taikoma klasterinė analizė. Valstybės klasterizuojamos pagal emigrantų skaičių 2014 metais. Klasterių skaičiaus nustatymui, taikomi šie kriterijai: pseudo t^2 , kubinio klasterizavimo ir Huberto. Visiems kriterijams nustatomas minimalus dviejų klasterių skaičius ir maksimalus - dešimties. Atstumų skaičiavimui pasirenkamas Euklido atstumo matas.

```
> NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+ method = "ward.D", index = "pseudot2", alphaBeale = 0.1)
$All.index
      2      3      4      5      6      7      8      9     10
198.1037 150.7090 19.8702 74.3217  3.2630 11.1514  3.9251 18.1999  0.0000

$All.CriticalValues
      2      3      4      5      6      7      8      9     10
 39.1821 38.1140 29.8538 34.1652 -4.0703 246.4030 -50.8045 27.7018  0.0000

$Best.nc
Number_clusters  Value_Index
         4.0000         19.8702
```

3.3.3 pav.: Pseudo t^2 kriterijaus rezultatai

Pseudo t^2 kriterijaus siūlomas klasterių skaičius yra keturi.

```
> NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+ method = "ward.D", index = "ccc", alphaBeale = 0.1)
$All.index
      2      3      4      5      6      7      8      9     10
 7.4586 10.4110  6.4321  9.5964  8.6139 11.1540 13.5653 15.1861 14.7506

$Best.nc
Number_clusters  Value_Index
         9.0000         15.1861
```

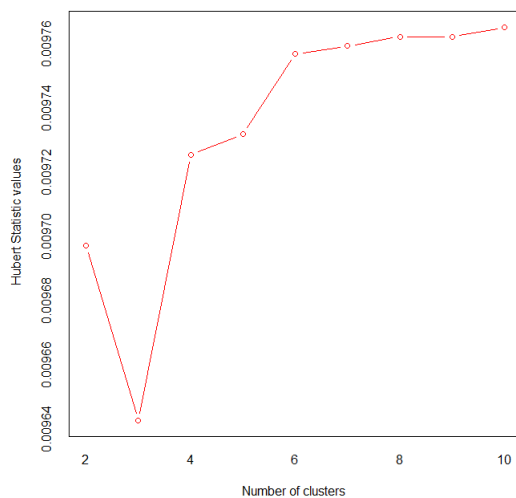
3.3.4 pav.: Kubinio klasterizavimo kriterijaus rezultatai

Pagal kubinio klasterizavimo kriterijų tinkamiausias klasterių skaičius yra, kai kriterijus įgyja didžiausią reikšmę. Šiuo atveju, didžiausia reikšmė įgyjama ties devyniais klasteriais.

```
> NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+ method = "ward.D", index = "hubert", alphaBeale = 0.1)
*** : The Hubert index is a graphical method of determining the number of clusters.
      In the plot of Hubert index, we seek a significant knee that corresponds to a
      significant increase of the value of the measure i.e the significant peak in Hu
bert
      index second differences plot.

$All.index
      2      3      4      5      6      7      8      9     10
0.0097 0.0096 0.0097 0.0097 0.0098 0.0098 0.0098 0.0098 0.0098
```

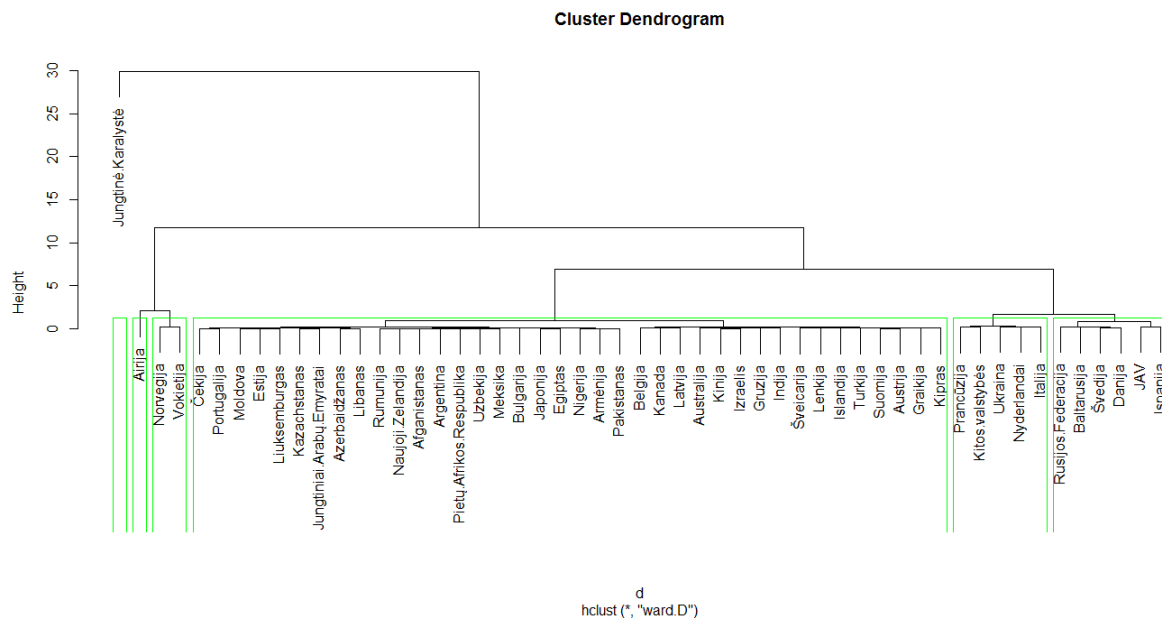
3.3.5 pav.: Huberto kriterijaus rezultatai



3.3.6 pav.

Tinkamiausias klasterių skaičius yra tas, prie kurio Huberto kriterijus įgyja mažiausią reikšmę. Iš rezultato ir grafikų matyti, kad šiuo atveju geriausias siūlomas klasterių skaičius yra trys.

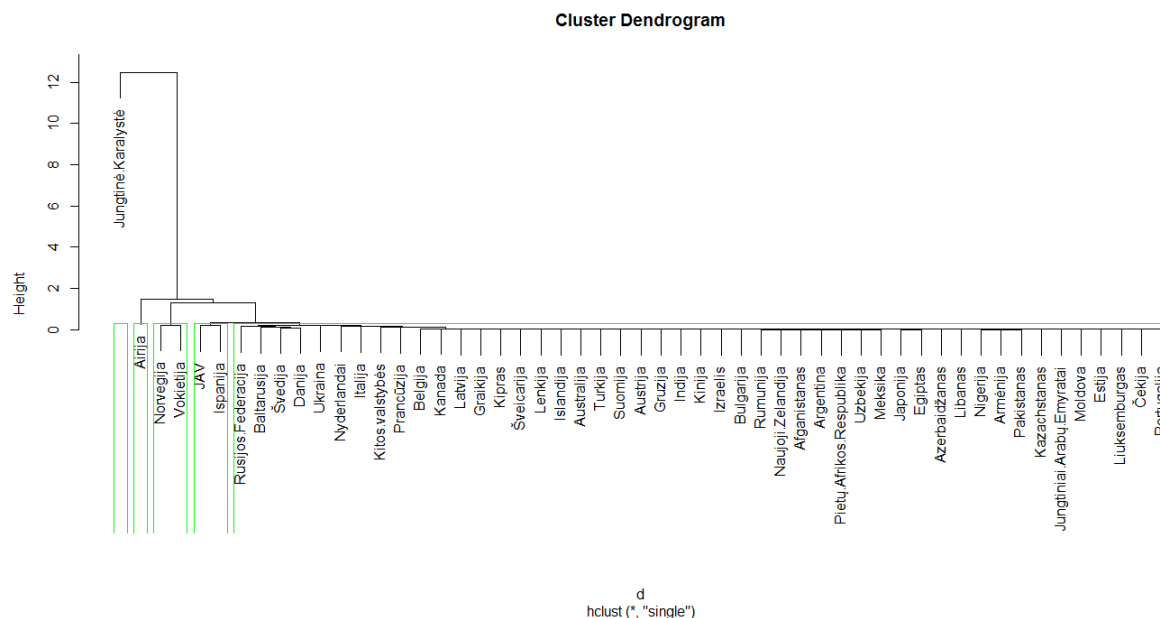
Visi tikrinti kriterijai siūlo skirtingus grupių skaičius: tris, keturias arba devynias. Kadangi visi rezultatai skirtingi, klasterių skaičius parenkamas išsinagrinėjus dendrogramas, kurios braižomas naudojant skirtingus klasterizavimo metodus. Pasirenkami trys hierarchinio klasterizavimo metodai: Vardo, vienutinės jungties ir vidutinės jungties.



3.3.7 pav.: Dendrograma. Vardo metodas

Šioje dendrogramoje (2.3.7 pav.) valstybės suskirstomos į šešis klasterius. Ryškiausiai iš visų išsiskirianti grupė sudaryta iš vienos valstybės - Jungtinės Karalystės. Tai yra pati populiariausia emigracijos kryptis. Antroje grupėje taip pat yra tik viena valstybė -

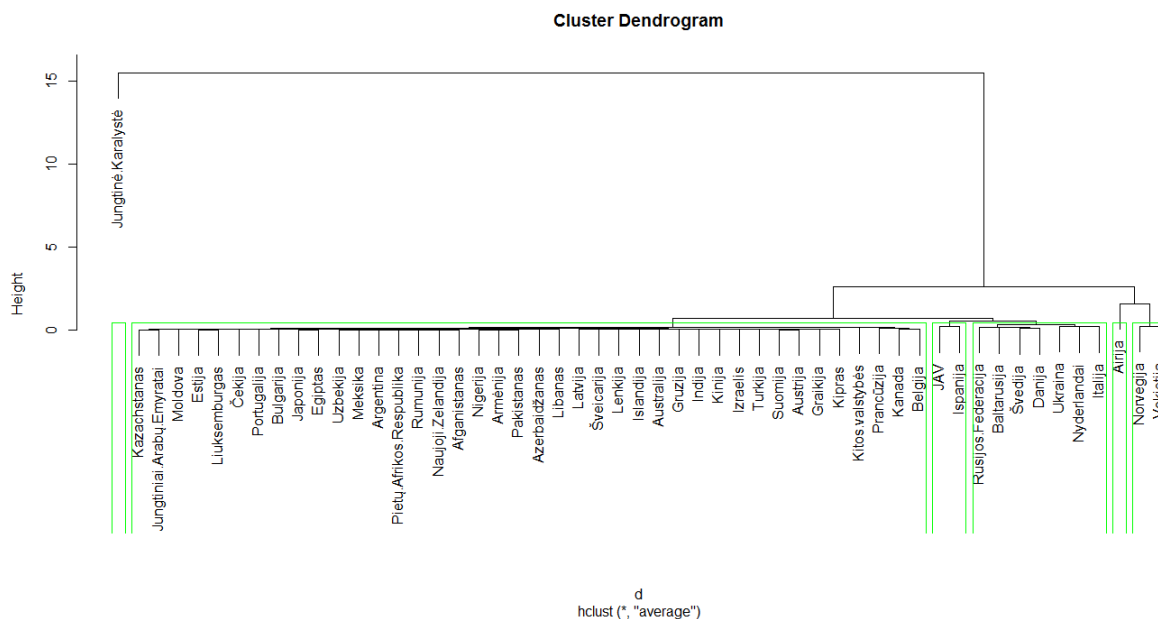
Airija, į trečiąją patenka Norvegija ir Vokietija. Paskutinis klasteris sudarytas iš Rusijos Federacijos, Baltarusijos, Švedijos, Danijos, JAV ir Ispanijos. Į dar vieną mažesnę grupę patenka Prancūzija, Ukraina, Nyderlandai, Italija ir kitos valstybės, o didžiausia grupė sudaryta iš valstybių, kurios rečiausiai pasirenkamos emigracijai, į ją patenka tokios šalys kaip Čekija, Naujoji Zelandija, Kinija ir t.t.



3.3.8 pav.: Dendrograma. Vienetinės jungties metodas

Taikant vienetinės jungties metodą, rezultatai gaunami šiek tiek kitokie, didžioji dalis valstybių apjungiamos panašiam aukštyje, todėl ši dendrograma (2.3.8 pav.) skirstoma į penkis klasterius. Pirmieji trys gaunami tokie patys, kaip ir pirmoje dendrogramoje: labiausiai išsiskiriantis yra Jungtinė Karalystė, toliau yra Airija, o trečiasis klasteris sudarytas iš Vokietijos ir Norvegijos. Ketvirtas klasteris irgi nedidelis, į jį patenka JAV bei Ispanija, o penktajame yra visos likusios valstybės, tarp kurių ir Rusijos Federacija, Švedija ir kitos valstybės.

Pirmosios dvi dendrogramos ganėtinai skiriasi, todėl tam, kad tiksliau suskirstyti turimas valstybes į grupes, toliau nagrinėjamas dar vienas hierarchinio klasterizavimo metodas - vidutinės jungties. Dendrograma, nubraižyta naudojant šį metodą vaizduojama 5 paveikslėlyje. Išnagrinėjus dendrogramą, psirenkama valstybes grupuoti į šešis klasterius, kaip ir naudojant Vardo metodą.

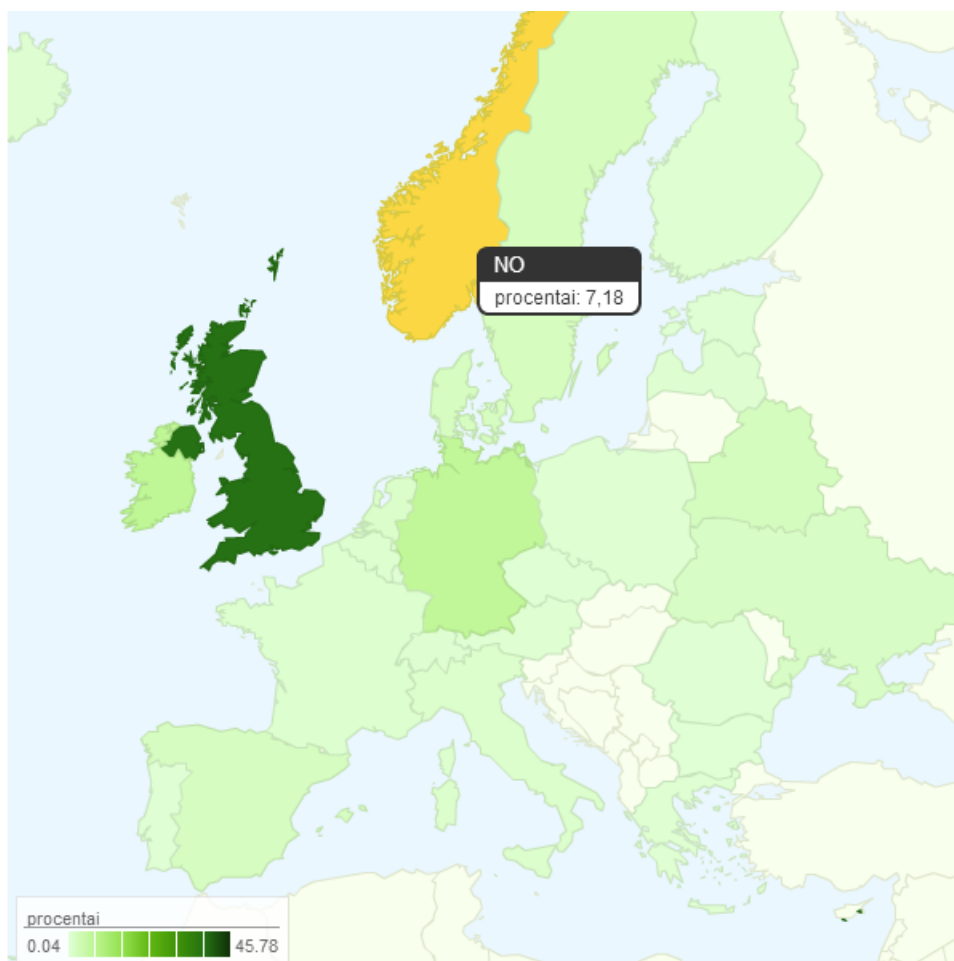


3.3.9 pav.: Dendrograma. Vidutinės jungties metodas

Šioje dendrogramoje keturios grupės sutampa su antrosios rezultatais: Jungtinė Karalystė; Airija; Norvegija ir Vokietija; JAV ir Ispanija vėl patenka į tuos pačius klasterius. Penktasis klasteris sudarytas iš Rusijos Federacijos, Baltarusijos, Švedijos, Danijos, Ukrainos, Nyderlandų ir Italijos. Į paskutiniąją grupę patenka visos kitos valstybės.

Klasterizavime yra gana daug subjektyvumo, tyrėjas, remdamasis įvairiais klasterių skaičiaus nustatymui skirtais indeksais, pats pasirenka į kiek grupių skirstys savo duomenis. Yra įvairių klasterizavimo metodų, kuriuos pritaikius, gaunami skirtingi rezultatai, todėl nėra vieno atsakymo į išsikeltus klausimus. Apibendrinus rezultatus, gautus taikant Vardo, vietinės jungties ir vidutinės jungties metodus, galima teigti, kad labiausiai iš kitų valstybių išsiskiria: Jungtinė Karalystė, Airija, Norvegija ir Vokietija, tai lemia dideli emigrantų, pasirinkusių šias valstybes, skaičiai.

3.3.10 paveikslėlyje pavaizduotas Europos žemėlapis, kuriame pažymėtos valstybės, į kurias 2014 m. emigravo Lietuvos piliečiai. Žemėlapis nubraižytas naudojant R paketą googleVis (kodas 1 pr). Paleidus kodą, žemėlapis atidaromas internetinėje naršyklėje, užvedus pelytės kursorių ant žalia atspalvio valstybės, galima sužinoti, kokia dalis visų 2014 m. emigravusių Lietuvos piliečių, emigravo į pasirinktą valstybę. Žemėlapyje (3.3.10 pav.) parodyta informacija apie Norvegiją - 2014 m. 7.18% emigrantų persikėlė gyventi į šią valstybę.

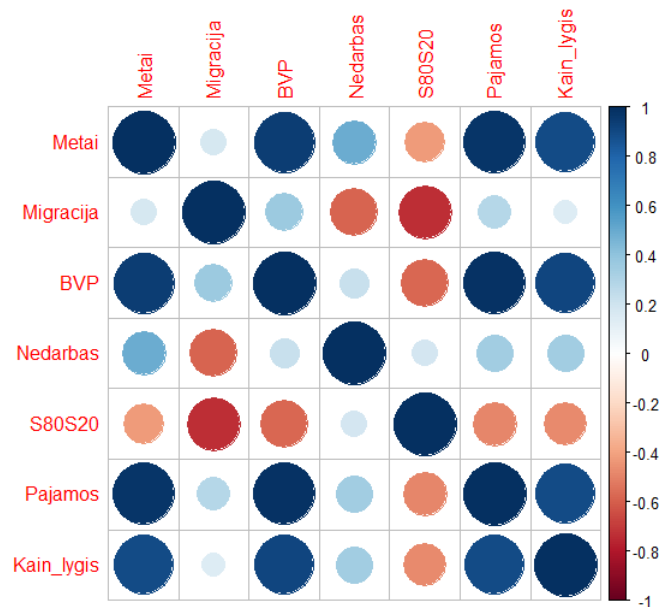


3.3.10 pav.: Europos valstybės, į kurias 2014 m. emigravo Lietuvos piliečiai

Jungtinė Karalystė yra labiausiai išsiskirianti ir tamsiausio atspalvio valstybė, nes šią emigracijos kryptį pasirinko net 45.78% emigravusių Lietuvos piliečių. Žemėlapyje matyti, kad pati populiariausia valstybė tarp Lietuvos emigrantų yra Jungtinė Karalystė, kitos ryškesnės valstybės: Norvegija, Vokietija ir Airija, į jas išvyksta po 7 – 8% visų emigrantų.

3.4 Veiksniai darantys įtaką migracijai Lietuvoje

Tam, kad ištirti, kurie iš nagrinėjamų indeksų daro įtaką migracijai Lietuvoje, pirmiausiai apskaičiuojami Pirsono koreliacijos koeficientai ir pavaizduojamas jų grafikas.



3.4.1 pav.: Koreliacijos koeficientų grafikas

```

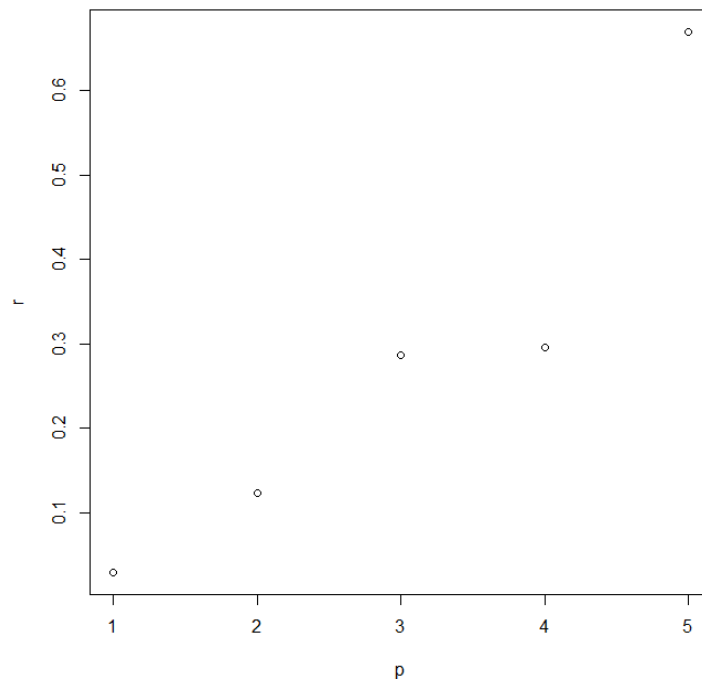
      Metai Migracija      BVP      Nedarbas      S80S20      Pajamos Kain_lygis
Metai      1.0000000  0.1717762  0.9418375  0.4972408 -0.4262022  0.9782698  0.8969302
Migracija  0.1717762  1.0000000  0.3670148 -0.5825586 -0.7308771  0.2847104  0.1405875
BVP        0.9418375  0.3670148  1.0000000  0.2260017 -0.5701297  0.9811845  0.9118784
Nedarbas   0.4972408 -0.5825586  0.2260017  1.0000000  0.1809911  0.3422506  0.3483261
S80S20     -0.4262022 -0.7308771 -0.5701297  0.1809911  1.0000000 -0.4814442 -0.4795406
Pajamos     0.9782698  0.2847104  0.9811845  0.3422506 -0.4814442  1.0000000  0.8948909
Kain_lygis  0.8969302  0.1405875  0.9118784  0.3483261 -0.4795406  0.8948909  1.0000000
> cor.test(LT$Migracija,LT$Metai)$p.value
[1] 0.6135352
> cor.test(LT$Migracija,LT$BVP)$p.value
[1] 0.2668733
> cor.test(LT$Migracija,LT$Nedarbas)$p.value
[1] 0.06001498
> cor.test(LT$Migracija,LT$Pajamos)$p.value
[1] 0.3961201
> cor.test(LT$Migracija,LT$S80S20)$p.value
[1] 0.01061451
> cor.test(LT$Migracija,LT$Kain_lygis)$p.value
[1] 0.6801193

```

3.4.2 pav.: Koreliacijos koeficientų lentelė ir p reikšmės

Iš gautų rezultatų matoma, kad stipri neigiama koreliacija yra tarp migracijos ir $S20/S80$ indekso, tačiau ji nereikšminga, nes p reikšmė mažesnė už 0.05 reikšmingumo lygmenį. Reikšminga neigiama koreliacija yra tarp migracijos bei namų ūkių grynųjų metinių pajamų. Vidutinė neigiama koreliacija tarp migracijos ir kainų lygio yra reikšminga, nes p reikšmė didesnė už 0.05. Tarp metų ir BVP bei migracijos koreliacija reikšminga, tačiau labai maža.

Toliau, tam, kad patikrinti ar migracijos rodiklis priklauso nuo metų, parenkamas polinominės regresijos modelis. Nubraižomas R kvadratų reikšmių grafikas. Pagal šį grafiką, pasirenkama trečia polinomo eilė.



3.4.3 pav.: R kvadrat reikšmių grafikas

Pritaikomas trečios eilės polinominės regresijos modelis:

$$M = \alpha + \beta t + \beta t^2 + \beta t^3 + \epsilon$$

čia M - migracijos rodiklis, t - metai, α - laisvasis narys, β_1 ir β_2 - koeficientai, ϵ - liekamosios paklaidos.

Rezultatai parodo, kad reikšmingų kovariančių nėra.

```
> migr <- lm(Migracija ~ Metai + I(Metai^2) + I(Metai^3), data=LT)
> summary(migr)

Call:
lm(formula = Migracija ~ Metai + I(Metai^2) + I(Metai^3), data = LT)

Residuals:
    Min       1Q   Median       3Q      Max
-13.792  -1.562   1.757   2.819   6.937

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.085e+05  8.762e+05   0.923   0.383
Metai        -8.052e+02  8.722e+02  -0.923   0.383
I(Metai^2)    2.005e-01  2.171e-01   0.923   0.383
I(Metai^3)           NA           NA      NA      NA

Residual standard error: 6.359 on 8 degrees of freedom
Multiple R-squared:  0.123,    Adjusted R-squared:  -0.09626
F-statistic: 0.561 on 2 and 8 DF,  p-value: 0.5916
```

3.4.4 pav.: Trečio laipsnio polinominė regresija

Kadangi tarp migracijos rodiklio ir nedarbo lygio bei BVP yra reikšminga koreliacija, parenkamas daugialypės regresijos modelis:

$$M = \alpha + \beta_1 N + \beta_2 BVP + \epsilon$$

čia M - migracijos rodiklis, N - nedarbo lygio rodiklis, α - laisvasis narys, β_1 ir β_2 - koeficientai, ϵ - liekamosios paklaidos.


```

> Migr2 <- lm(Migracija ~ Nedarbas+BVP, data=LT)
> summary(Migr2)

Call:
lm(formula = Migracija ~ Nedarbas + BVP, data = LT)

Residuals:
    Min       1Q   Median       3Q      Max
-7.9031 -2.1259  0.4353  2.1657  5.8508

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.243e+01  6.113e+00  -2.033  0.0765 .
Nedarbas    -9.893e-01  3.232e-01  -3.061  0.0156 *
BVP         1.429e-03  6.229e-04   2.293  0.0510 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.287 on 8 degrees of freedom
Multiple R-squared:  0.6014,    Adjusted R-squared:  0.5018
F-statistic: 6.036 on 2 and 8 DF,  p-value: 0.02523

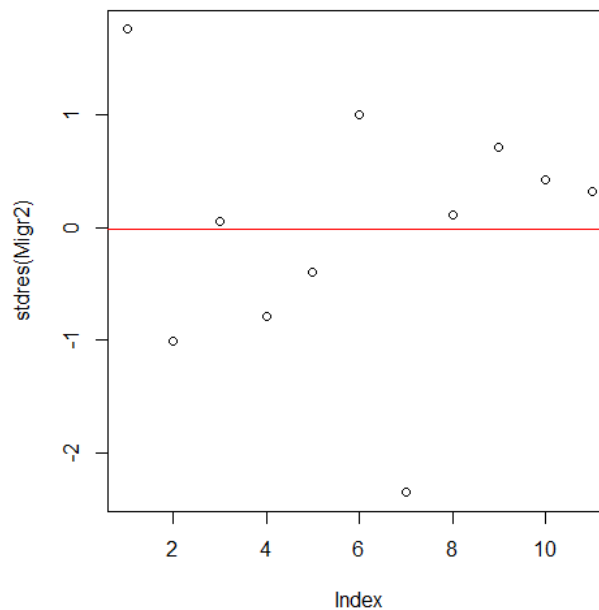
```

3.4.5 pav.: Tiesinė regresija

Gaunama, kad su 0.1 reikšmingumo lygmeniu reikšmingos abi kovariantės. Gaunama, kad augant nedarbo lygiui, migracijos rodiklis mažėja, tai reiškia, kad emigracija didėja, imigracija mažėja. Esant BVP augimui, migracijos rodiklis taip pat auga.

Gaunamas toks įvertis: $M = -12.43 - 0.9893N + 0.00143BVP$.

Toliau tiriamos liekamosios paklaidos. Pirmiausia nubraižomas standartizuotų liekamųjų paklaidų grafikas. Gaunama, kad paklaidų visdurkis yra labai artimas 0, paklaidų reikšmės priklauso intervalui $[-3; 3]$, todėl paklaidos yra homoskedastiškos.



3.4.6 pav.: Standartizuotų liekamųjų paklaidų grafikas

Tam, kad patikrinti ar tarp liekamųjų paklaidų nėra autokoreliacijos, naudojamas Durbino ir Vatsonio testas.

$$H_0: \rho = 0$$

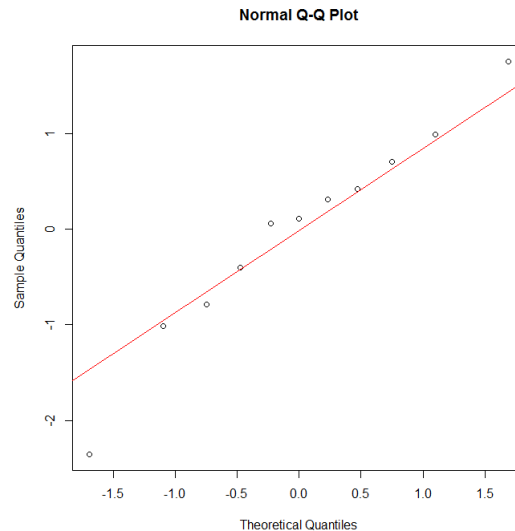
$$H_1: \rho \neq 0$$

```
> durbinwatsonTest(stdres(Migr2))
[1] 2.404066
```

3.4.7 pav.: Durbino ir Vatsono testas

Gauta testo reikšmė yra 2.4, tai reiškia, kad autokoreliacijos nėra.

Liekamųjų paklaidų pasiskirstymui ištirti, nubraižomas teorinių kvantilių ir empirinių kvantilių grafikas. Kvantilių grafike tik pirmasis taškas yra labai nutolęs nuo tiesės, visi kiti - išsidėstę aplink ją.



3.4.8 pav.: Kvantilių - kvantilių grafikas

Tam, kad patikrinti ar paklaidos yra normaliai pasiskirsčiusios, taikomas Šapiro ir Vilko testas.

H_0 : Standartizuotos liekamosios paklaidos pasiskirsčiusios normaliai

H_1 : Standartizuotos liekamosios paklaidos nėra pasiskirsčiusios normaliai

```
> shapiro.test(stdres(Migr2))

Shapiro-Wilk normality test

data:  stdres(Migr2)
W = 0.96951, p-value = 0.8815
```

3.4.9 pav.: Šapiro ir Vilko testas

Gauta testo p reikšmė yra didesnė už pasirenkamą 0.05 reikšmingumo lygmenį, todėl H_0 nėra atmetama ir paklaidos yra pasiskirsčiusios normaliai.

Atlikus analizę nustatyta, kad migraciją rodiklį Lietuvoje įtakoja BVP ir nedarbo lygis, tačiau šie rezultatai nėra labai patikimi, kadangi duomenų laiko eilutę sudaro vos vienuolika metų.

3.5 Lietuvos savivaldybių klasterizacija pagal nedarbą ir migraciją

Ankstesniame skyrelyje tarp migracijos rodiklio ir nedarbo buvo stipri neigiama koreliacija, todėl pažiūrėkime kaip pagal šiuos du rodiklius galima suskirstyti Lietuvos savivaldybes. Savivaldybių suskirstymui vėl taikoma klasterinė analizė. Klasterių skaičiaus

nustatymui, taikomi šie kriterijai: pseudo t^2 , kubinio klasterizavimo ir Huberto. Visiems kriterijams nustatomas minimalus dviejų klasterių skaičius ir maksimalus - dešimties. Atstumų skaičiavimui pasirenkamas Euklido atstumo matas.

```
> NbClust(mydata2[,-1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+ #Pseudo t squared
+ method = "ward.D", index = "pseudot2", alphaBeale = 0.1)
$All.index
      2      3      4      5      6      7      8      9     10
26.0197  9.1260 23.2640  1.0422 15.0756 16.0093 14.4106 11.1958  8.3905

$All.CriticalValues
      2      3      4      5      6      7      8      9
 92.6281 130.1288 130.1288 -24.9172 108.5678 -335.8023 -24.9172 -56.6808
      10
-4016.9908

$Best.nc
Number_clusters  value_index
      2.0000      26.0197

$Best.partition
[1] 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 2 1 2 2 2 1 1 2 1 1 1 1 1 1
[45] 1 1 2 2 2 1 1 2 1 2 1 2 1 1 1 2
```

3.5.1 pav.: Pseudo t^2 kriterijaus rezultatai

Pseudo t^2 kriterijaus siūlomas klasterių skaičius yra du.

```
> NbClust(mydata2[,-1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+ method = "ward.D", index = "pseudot2", alphaBeale = 0.1)
$All.index
      2      3      4      5      6      7      8      9     10
26.0197  9.1260 23.2640  1.0422 15.0756 16.0093 14.4106 11.1958  8.3905

$All.CriticalValues
      2      3      4      5      6      7      8      9
 92.6281 130.1288 130.1288 -24.9172 108.5678 -335.8023 -24.9172 -56.6808
      10
-4016.9908

$Best.nc
Number_clusters  value_index
      2.0000      26.0197

$Best.partition
[1] 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 2 1 2 2 2 1 1 2 1 1 1 1 1 1
[45] 1 1 2 2 2 1 1 2 1 2 1 2 1 1 1 2
```

3.5.2 pav.: Kubinio klasterizavimo kriterijaus rezultatai

Pagal kubinio klasterizavimo kriterijų tinkamiausias klasterių skaičius yra, kai kriterijus įgyja didžiausią reikšmę. Šiuo atveju, didžiausia reikšmė įgyjama ties dviem klasteriais.

```

> NbClust(mydata2[,-1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
+         method = "ward.D", index = "ccc", alphaBeale = 0.1)
$All.index
      2      3      4      5      6      7      8      9     10
12.2050  8.2738  6.7329  8.5164 11.1076 11.0764 11.1394 10.8279 10.4977

$Best.nc
Number_clusters      value_index
          2.000          12.205

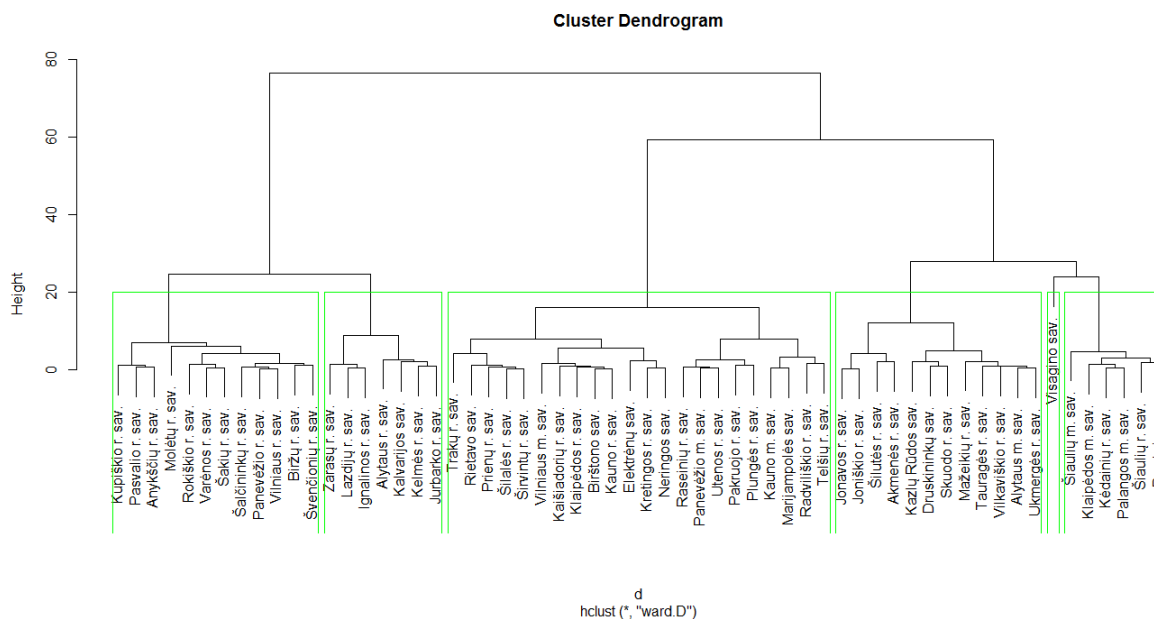
$Best.partition
[1] 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 2 1 2 2 2 1 1 2 1 1 1 1 2 1 1 1 1
[45] 1 1 2 2 2 1 1 2 1 2 1 2 1 1 1 2

```

3.5.3 pav.: Huberto kriterijaus rezultatai

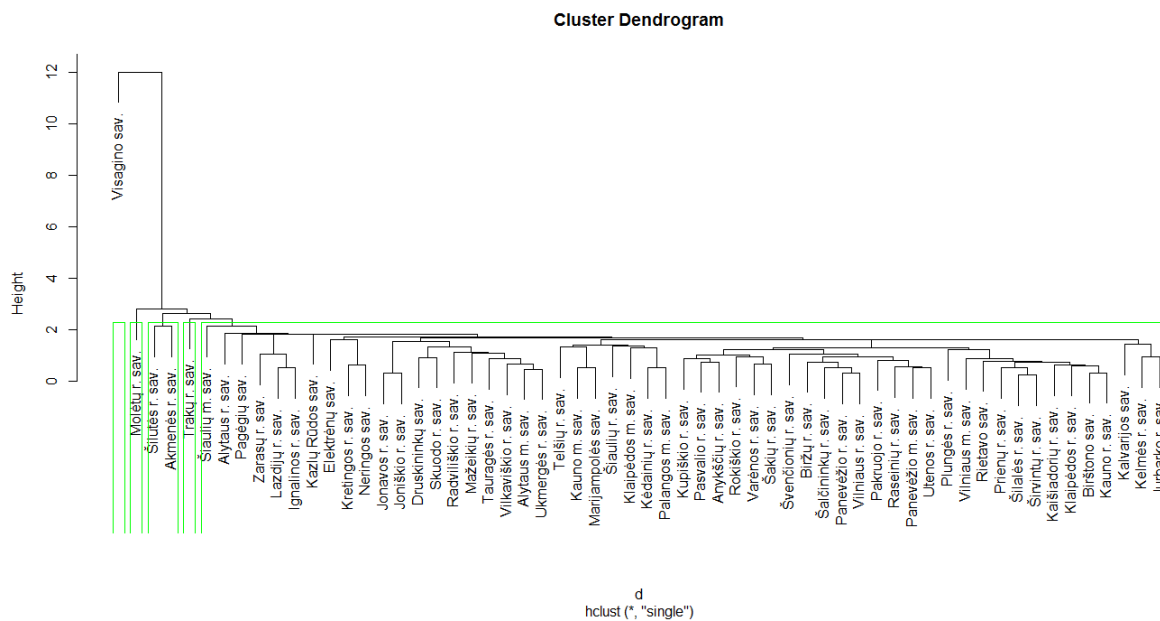
Tinkamiausiais klasterių skaičius yra tas, prie kurio Huberto kriterijus įgyja mažiausią reikšmę. Iš rezultato ir grafikų matyti, kad šiuo atveju geriausias siūlomas klasterių skaičius yra taip pat du.

Nors ir visi taikomi kriterijai siūlo rinktis du klasterius, klasterių skaičius parenkamas individualiai, kiekvienai dendrogramai, pagal tai, kiek grupių galima išskirti vizualiai.



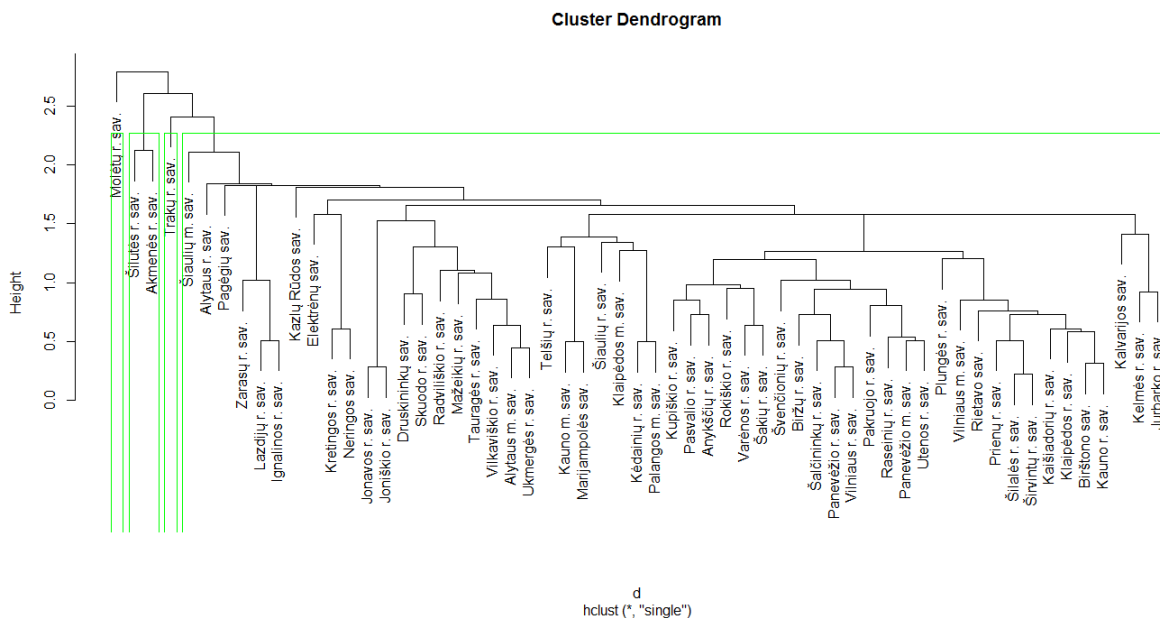
3.5.4 pav.: Dendrograma. Vardo metodas

3.5.4 paveikslėlyje vaizduojama Ward metodu nubraižyta dendrograma. Ji suskirstoma į šešis klasterius. Labiausiai išsiskiria klasteris, sudarytas tik iš Visagino savivaldybės. Yra dar 2 smulkesni klasteriai. Į vieną patenka Kėdainių r., Klaipėdos m., Palangos m., Šiaulių m., Šiaulių r. ir Pagėgių savivaldybės, į kitą: Alytaus r., Lazdijų r., Kalvarijos, Kelmės r., Jurbarko r., Ignalinos r. ir Zarasų r. savivaldybės. Likusios trys grupės gausesnės. Į pirmąją patenka tokios savivaldybės kaip Vilniaus r., Varėnos r., Biržų r. ir kitos savivaldybės. Antroje didesnėje grupėje Vilniaus m., Kauno m., Panevėžio m. ir kitos savivaldybės. Trečiąją sudaro Akmenės r., Druskininkų, Alytaus m. ir kitos.



3.5.5 pav.: Dendrograma. Vienetinės jungties metodas

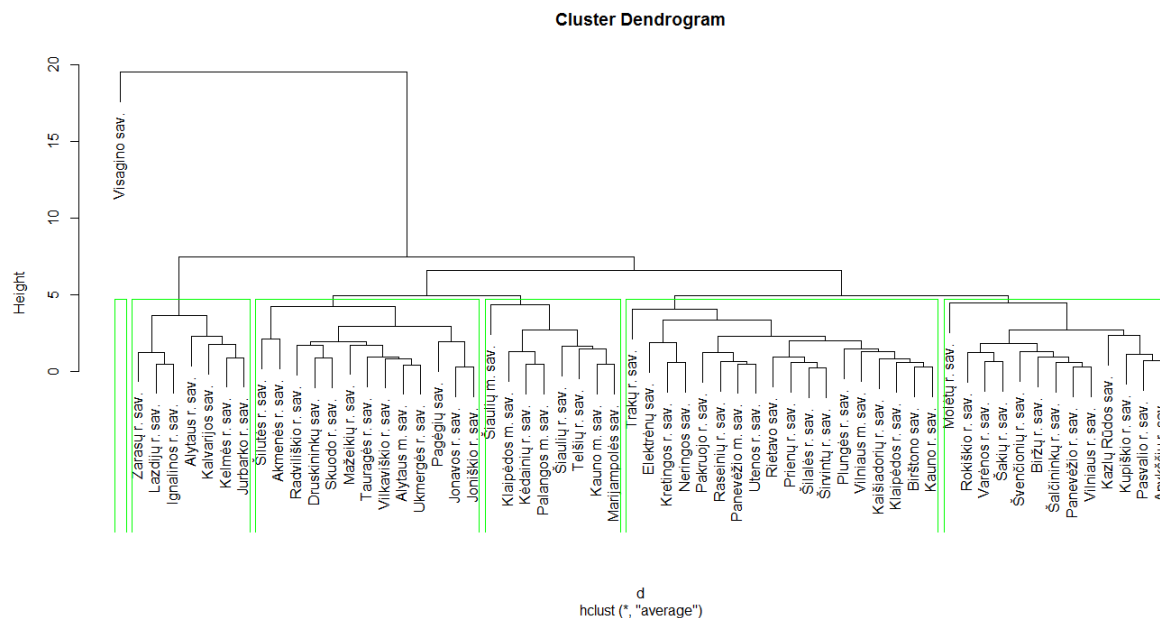
Dendrograma, sukurta naudojant vienetinės jungties metodą (3.5.5 pav.) skirstoma į penkias grupes. Ji sudaryta iš keturių mažų ir vienos didelės grupės. Kaip ir pirmoje dendrogramoje, Visagino sav. patenka į atskirą klasterį. Kadangi Visagino rajono savivaldybė labai išsiskiria nuo kitų, kad būtų aiškiau matyti kitus klasterius, ši savivaldybė panaikinama ir pažiūrima, kaip išsidėsto kitos savivaldybės.



3.5.6 pav.: Dendrograma. Vienetinės jungties metodas (be Visagino r. sav.)

Pirmasis, nors ir pašalintas, klasteris yra Visagino r. sav. Kitas klasteris irgi sudarytas iš

vienos - Molėtų rajono savivaldybės. Į trečiąją patenka Šilutės r. ir Akmenės r. savivaldybės. Trakų r. sav. sudaro ketvirtąjį klasterį, o į paskutinį, penktąjį, patenka visos likusios savivaldybės. Tarp jų didžiųjų miestų, kurortinių Lietuvos miestų ir kitos savivaldybės.



3.5.7 pav.: Dendrograma. Vidutinės jungties metodas

Paskutinė dendrograma sudaryta naudojant vidutinės jungties metodą, ji skirstoma į šešis klasterius. Pirmajame yra Visagino savivaldybė, į antrąjį patenka Zarasų r., Ignalinos r. ir kitos savivaldybės. Trečiasis sudarytas iš Druskininkų, Alytaus m. ir kitų savivaldybių. Toliau seka klasteris su Šiaulių m., Klaipėdos m., Kauno m. ir kitomis savivaldybėmis. Tarp penktojo klasterio savivaldybių patenka tokios kaip Kretingos r., Kauno r. ir Vilniaus m. sav. Paskutinajame yra Molėtų r., Vilniaus r., Anykščių r. savivaldybės ir kitos.

Gauti rezultatai labai skirtingi. Vienintėlis visose dendrogramose išskirtas klasteris yra Visagino sav. Tai reiškia, kad nedarbo ir migracijos rodikliai šioje savivaldybėje ypatingai skiriasi nuo kitų. Kitos savivaldybės grupuojamos labai įvairiai.

4 IŠVADOS

Atlikus analizę nustatyta, kad iš pasirinktų ekonominių trisdešimt dviejų Europos valstybių rodiklių, migracijos rodiklį reikšmingai įtakoja nedarbo rodiklis, šis rodiklis buvo reikšmingas regresorius visuose taikytuose modeliuose. Augant šiam rodikliui, migracijos rodiklis mažėja, tai reiškia, kad mažėja skirtumas tarp imigracijos ir emigracijos bei nedarbo augimas skatina emigraciją.

Ištyrus tuos pačius ekonominius rodiklius Lietuvoje, buvo gauta, kad migraciją Lietuvoje taip pat įtakoja nedarbas.

Iš grafinės nedarbo rodiklio analizės galima spręsti, kad didelius migracijos rodiklio pokyčius lėmė 2008-2010 m. pasaulinė ekonominė krizė. Lietuvoje ji sukėlė emigracijos išaugimą. Kitose Europos valstybėse buvo įvairių pokyčių, įtakotų šios krizės. Pavyzdžiui Ispanijoje, iki 2008 m. migracijos rodiklis buvo teigiamas - vyravo imigracija, o po 2008 m. rodiklis pradėjo leisti bei tapo neigiamas. Ekonominės situacijos pasikeitimas lėmė emigracijos išaugimą. Buvo ir priešingų pokyčių. Vokietijoje kaip tik krizės laikotarpyje migracijos rodiklis pradėjo sparčiai augti ir dabar į Vokietiją plūsta migrantai.

Valstybės į kurias emigruoja Lietuvos piliečiai suskirstytos į klasterius pagal emigrantų skaičių 2014 metais. Visais taikomais klasterizavimo metodais, Jungtinė Karalystė ypatin-gai skiriasi nuo kitų valstybių, tai lemia ypatingai aukštas migrantų skaičius - beveik pusė emigrantų išvyksta gyventi į šią valstybę.

SANTRAUKA

Šiame darbe atliekamas migracijos rodiklio ir kitų ekonominių rodiklių statistinis tyrimas. Naudojami 2004-2014 m. metiniai duomenys apie trisdešimt dvi Europos valstybes gauti iš Eurostat duomenų bazės. Taip pat naudoti Lietuvos statistikos bei Europos migracijos tinklo duomenų bazių duomenys.

Darbo tikslas yra ištirti ar nagrinėjami ekonominiai rodikliai turi įtakos migracijai ir jei turi, kokią įtaką jai daro.

Duomenų analizei taikomi šie metodai: polinominė regresija, pažingsninė regresija, Durbino ir Watsono, Andersono ir Darlingo testai, klasterinė analizė. Taip pat taikoma regresinė panelinių duomenų analizė.

Atlikus statistinę analizę, gauta, kad migracijos rodiklį tiek Europoje, tiek ir atskirai Lietuvoje įtakoja nedarbo lygio rodiklis. Šiam rodikliui didėjant, migracijos rodiklis mažėja, t.y. auga emigracija, o imigracija mažėja. Populiariausia Lietuvos piliečių emigracijos kryptis yra Jungtinė Karalystė, į ją išvyksta beveik penkiasdešimt procentų visų emigrantų. Pagal emigrantų skaičių ir metus suklasterizuotos valstybės, į kurias emigruoja Lietuvos piliečiai, o Lietuvos savivaldybės suskirstytos pagal nedarbo lygį ir migracijos rodiklį.

SUMMARY

The statistical analysis of migration index and other economical indexes is implemented in this work. A part of the data that is used in bachelor thesis is taken from Eurostat database. These data include thirty two countries and time series of 2004 - 2014 years. Another part of data is from Lithuanian Statistics Department' database and Europe Migration Net database.

The main goal of this work is to analyse whether there is a relation between migration index and other investigated economical indexes.

Methods applied for data analysis: polynomial regression, stepwise regression, Durbin - Watson test, Anderson - Darling test, cluster analysis. A regression analysis for panel data is also used in this work.

After statistical analysis revealed that migration index in Europe and individually in Lithuania is dependent on unemployment index. When unemployment rate is growing migration index is falling, that means emigration is growing and immigration is getting lower. Most popular Lithuanian citizens' emigration direction is United Kingdom. Almost half of all emigrants go to this country. Countries where departs our emigrants are clustered by emigrants' number and year and Lithuania's municipalities where grouped by unemployment rate and migration index.

Literatūra

- [1] V. Čekanavičius, G. Murauskas „Statistika ir jos taikymai“ II, Vilnius, TEV, 2008
- [2] V. Bagdonavičius, J.J. Kruopis, MATEMATINĖ STATISTIKA II, 2015, Vilniaus universiteto leidykla. URL: <http://www.statistika.mif.vu.lt/atsisiuntimui/statistika/>
- [3] R. Lapinskas „Sekinių kompiuterinė analizė“ URL: <http://ec.europa.eu/eurostat/data/database>
- [4] Eurostat duomenų bazė URL: <http://ec.europa.eu/eurostat/data/database>
- [5] Europos migracijos tinklo internetinis puslapis URL: <http://123.emn.lt/>
- [6] J. M. Wooldrige „Econometric analysis of cross section and panel data“ URL: https://jrvargas.files.wordpress.com/2011/01/wooldridge_j-_2002_econometric_analysis_of_cross_section_and_panel_data.pdf
- [7] V. Čekanavičius, G. Murauskas, Taikomoji regresinė analizė socialiniuose tyrimuose, 2014, Vilniaus universiteto leidykla, 561 p., ISBN 978-609-459-300-0. URL: <http://www.statistika.mif.vu.lt/atsisiuntimui/statistika/>
- [8] Lietuvos statistikos departamento duomenų bazė URL: <http://osp.stat.gov.lt/>

Priedai

1 priedas. R kodas

Žemėlapis R kodas:

```
install.packages("googlevis")
library(googlevis)

valstybe <- c("GB", "IE", "NO", "DE", "ES", "BY", "SE", "DK", "NL", "UA", "IT", "FR",
             "BE", "LV", "PL", "CH", "IS", "FI", "AT", "GR", "CY", "EE", "CZ", "LU",
             "PT", "BG", "RO")
procentai <- c(45.78, 8.40, 7.18, 7.93, 2.56, 2.84, 2.61, 2.57, 1.52, 2.05, 1.02, 0.58, 0.64, 0.39, 0.29,
              0.38, 0.38, 0.23, 0.25, 0.24, 0.24, 0.16, 0.13, 0.13, 0.10, 0.04, 0.05)
duom <- data.frame(valstybe, procentai)

Geo <- gvisGeoMap(duom, locationvar="valstybe", numvar="procentai",
                 options=list(height=600, width=600, dataMode='regions', region="150", showLegend=TRUE))
plot(Geo)
```

Klasterinės analizės R kodas:

```
library(cluster)
library(NbClust)
mydata <- read.table("C:/Users/Egle/Desktop/Duomenys_R/Emigracijos_kryptys.csv",
                    header=TRUE, sep=";", dec = ",")
data <- scale(t(mydata[c(-1)]))
NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
        method = "ward.D", index = "all", alphaBeale = 0.1)
#Pseudo t kvadrat
NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
        method = "ward.D", index = "pseudot2", alphaBeale = 0.1)
#Kubinio klasterizavimo kriterijus
NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
        method = "ward.D", index = "ccc", alphaBeale = 0.1)
#Huberto
NbClust(data[, -1], diss = NULL, distance = "euclidean", min.nc = 2, max.nc = 10,
        method = "ward.D", index = "hubert", alphaBeale = 0.1)
#Dendrogramos
par(mfrow=c(1,1))
d <- dist(data, method = "euclidean") #
fit <- hclust(d, method="ward")
plot(fit)
rect.hclust(fit, k=6, border="green")
groups <- cutree(fit, k=6)

fit2 <- hclust(d, method="single")
plot(fit2)
rect.hclust(fit2, k=5, border="green")
groups <- cutree(fit2, k=5)

fit3 <- hclust(d, method="average")
plot(fit3)
rect.hclust(fit3, k=6, border="green")
groups <- cutree(fit3, k=6)
```

Kitos taikytos formulės ir funkcijos:

```
#Koreliacijos koeficientai
kintamasis <- cor(duomenys)
#Koreliacijos koeficientų grafikas
corrplot(kintamasis, method="circle") #koreliacijos koeficientų grafikas

#Daugialytė regresija
modelis=lm(priklusomas_kint ~ nepr_kint_1 + ... + nepr_kint_N, data=duomenys)
summary(modelis)

#Pažingsninė regresija
step(regresijos_funkcija_lm,direction="backward")

#Durbino ir vatsono testas, paketas car
durbinwatsonTest(modelis)

#Standartizuotų liekamųjų paklaidų gavimas iš modelio
stdres(modelis)

#Standartizuotų liekamųjų paklaidų kvantilių grafikas
qqnorm(stdres(modelis))
qqline(stdres(modelis),col="red")
```

2 priedas. Lentelės

Valstybė	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Airija	12,2	15,2	22,2	16,9	3,7	-4,2	-5,6	-7,4	-7,6	-5,5	-3,2
Austrija	6,6	6,1	3	2,8	2,9	2,1	2,6	3,7	5,2	6,5	7,7
Belgija	3,2	4,7	4,9	5,5	5,9	5,9	7,9	6,3	4,3	2,3	3
Bulgarija	-2,1	-2,2	-2,3	-2,3	-2,4	-2,5	-2,4	-0,7	-0,3	-0,2	-0,3
Čekija	1,3	3	2,9	7,7	6,5	2,4	1,4	1,6	1	-0,1	2,1
Danija	0,9	1,2	1,9	3,7	4,6	2,8	3	2,4	3	3,8	4,8
ES (28 valst.)	3,4	3,1	2,8	3,1	2,4	1,4	1,5	1,4	1,8	6,3	2,2
Estija	-2,7	-3,8	-4	-2,1	-1,5	-1,6	-2,8	-2,9	-2,8	-2	-0,5
Griekija	2,6	2,9	2,2	2,1	2,1	1,3	-0,1	-2,9	-6	-5,4	-4,3
Islandija	2	13	17,3	16,6	3,3	-15	-6,5	-4,3	-0,9	5,1	3,4
Ispanija	15,5	14,5	14,9	17,2	9,5	3	1,6	1,4	-3	-5,4	-2
Italija	6,3	3,5	2,7	7,5	6,1	3,6	3,4	1,3	6,2	19,7	1,8
Jungtinė Karalystė	4,3	4,9	4,5	4,9	4,1	3,8	4,2	3,4	2,6	3,8	4,9
Kipras	9,7	11	13,7	19,9	21,1	22	19,2	21,3	-0,7	-14	-17,6
Kroatija	3,4	2,5	2,3	2,1	1,4	0,2	-1	-0,9	-0,9	-1,1	-2,4
Latvija	-6,8	-4,9	-4	-3,6	-10,3	-16,1	-17	-9,7	-5,8	-7,1	-4,3
Lenkija	-0,2	-0,3	-0,9	-0,5	-0,4	0	0,1	-0,3	-0,1	-0,7	-0,3
Lietuva	-9,5	-15,4	-7,5	-6,7	-5,1	-10,1	-25,2	-12,6	-7,1	-5,7	-4,2
Liuksemburgas	9,6	13,1	11,3	12,5	15,8	13,2	15,1	21,2	18,9	19	19,9
Malta	4,8	4	0,1	3,8	5,7	5,6	0,2	4	7,4	7,6	7,1
Norvegija	2,9	4	5,1	8,4	9,1	8	8,6	9,4	9,4	7,7	7,8
Olandija	-0,6	-1,4	-1,6	-0,1	1,9	2,3	2	1,8	0,8	1,2	2,1
Portugalija	1,4	1,5	1,6	2,1	0,9	1,5	0,4	-2,3	-3,6	-3,5	-2,9
Prancūzija	3,2	3	1,8	1,2	0,9	0,5	0,6	0,3	1,1	0,7	0,7
Rumunija	-4,5	-4	-4,1	-21,9	-8	-5,4	-2,4	-2,4	-1,1	-0,4	-0,8
Slovakija	-0,2	-0,1	-0,1	0,4	0,4	-0,1	-0,9	0,5	0,6	0,4	0,3
Slovėnija	0,9	3,2	3,1	7,1	9,2	5,6	-0,3	1	0,3	0,2	-0,2
Suomija	1,3	1,7	2	2,6	2,9	2,7	2,6	3,1	3,3	3,3	2,8
Švedija	2,8	3	5,6	5,9	6	6,7	5,3	4,8	5,4	6,9	7,9
Šveicarija	5,1	4,3	4,9	9,4	12,1	8,8	7,7	8,3	8,3	10,2	9,4
Vengrija	1,8	1,7	2,1	1,4	1,6	1,7	1,2	1,3	1,6	0,6	1,1
Vokietija	1	1	0,3	0,5	-0,7	-0,1	1,6	3,4	4,9	5,6	7,2

4.0.1 pav.: Migracijos rodiklis tūkstančiui gyventojų

ValstMetai	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ES (28 valst)	22382,1	23288,9	24540,2	25918,9	25976,1	24409,1	25428,2	26126,9	26649,9	26822,7	27534,3
Airija	38784,6	41340,3	43944,0	45402,8	42072,0	37473,9	36522,7	38053,9	38153,0	39086,1	41047,8
Austrija	29659,5	30849,7	32283,5	34087,8	35138,5	34335,7	35277,8	36850,7	37708,3	38202,0	38709,3
Belgija	28732,1	29818,6	31077,0	32567,6	33193,1	32435,5	33681,2	34462,2	34918,8	35182,9	35758,0
Bulgarija	2716,8	3121,7	3584,8	4319,2	4971,2	4987,9	5082,9	5557,4	5690,2	5753,5	5900,2
Čekija	9404,1	10726,1	12103,7	13458,2	15561,7	14229,9	14946,3	15599,1	15297,5	14923,0	14719,6
Danija	37482,6	39344,0	41564,9	42855,9	44027,8	41770,0	43636,6	44253,0	45321,1	45556,4	46307,2
Estija	7105,4	8288,1	10010,9	12097,8	12340,7	10590,3	11039,2	12535,2	13587,2	14403,3	15171,3
Graikija	17706,5	18162,6	19797,1	21085,0	21877,9	21409,6	20327,9	18612,0	17246,7	16393,6	16249,9
Islandija	37990,8	45932,1	45441,8	50590,2	33980,6	28822,2	31467,7	33081,0	34565,9	35947,0	39443,2
Ispanija	20246,1	21492,9	22903,3	24133,4	24441,3	23335,9	23252,1	22937,2	22274,9	22069,7	22384,7
Italija	25202,1	25752,3	26680,6	27657,2	27840,6	26671,9	27127,7	27606,6	27185,7	26922,8	26551,3
Jungtinė Karalystė	30929,5	32329,0	34036,6	35513,3	30994,3	26876,3	29008,6	29608,7	32342,8	31967,5	35031,2
Kipras	19156,4	20388,6	21694,1	23028,7	24244,8	23191,9	23338,7	23277,3	22585,4	20862,8	20272,4
Kroatija	7772,1	8468,9	9321,3	10183,3	11161,9	10462,4	10459,2	10421,9	10274,5	10203,1	10129,9
Latvija	5122,8	6094,3	7736,1	10249,5	11094,9	8660,5	8381,2	9709,9	10750,5	11268,4	11781,8
Lenkija	89983,0	108823,1	7165,6	8226,9	9541,8	8251,8	9513,9	9988,2	10226,9	10367,2	10806,6
Lietuva	5365,6	6259,6	7319,3	8935,8	10177,5	8459,8	8920,4	10241,5	11098,1	11784,2	12381,4
Liuksemburgas	60797,7	64465,7	71222,1	77209,4	77816,2	73491,8	78725,7	82500,2	83021,5	86662,4	88956,3
Malta	12172,0	12770,1	13299,0	14194,5	15027,5	14938,5	15939,8	16594,2	17313,5	18170,8	19056,0
Norvegija	46512,4	53910,7	59326,9	62619,0	66878,2	58006,1	66606,4	72810,2	79560,4	77880,8	73911,6
Olandija	32226,5	33461,6	35460,1	37491,2	38960,5	37456,9	38100,3	38600,9	38562,5	38788,6	39381,9
Portugalija	3989,8	4156,1	15815,2	16659,5	16949,4	16609,7	17017,1	16662,4	15973,4	16235,8	16633,9
Prancūzija	27463,5	28228,4	29310,1	30570,6	31181,7	30132,2	30908,1	31691,7	31970,4	32264,5	32364,2
Rumunija	2853,2	3752,0	4629,9	5934,7	6900,6	5890,8	6245,3	6599,6	6643,7	7205,4	7531,3
Slovakija	6460,0	7299,9	8449,1	10439,0	12246,8	11856,9	12501,3	13063,4	13400,4	13645,8	13951,5
Slovėnija	13893,1	14635,3	15754,1	17485,8	18878,7	17795,2	17710,2	17996,5	17508,3	17440,8	18098,8
Suomija	30361,1	31391,9	32843,9	35358,3	36545,9	33987,7	34962,6	36624,9	36990,0	37470,1	37655,1
Švedija	34251,9	34758,0	37012,1	39111,8	38366,5	33455,8	39512,8	43008,1	44642,7	45600,4	44849,9
Šveicarija	43045,0	44201,0	45866,4	46461,2	49559,1	50479,0	56402,7	63740,0	65077,1	64146,8	64963,6
Vengrija	8253,3	8966,8	9065,1	10093,8	10701,7	9338,1	9805,8	10084,8	9965,1	10220,5	10553,3
Vokietija	27512,1	27888,9	29030,9	30531,9	31158,0	30002,6	31540,2	33065,0	34295,2	35030,9	36099,3

4.0.2 pav.: BVP gyventojui (€)

Valst/Metai	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ES (28 valst.)							4,9	5,0	5,0	5,0	5,2
Ainija	4,9	5,0	4,9	4,8	4,4	4,2	4,7	4,6	4,7	4,5	4,7
Austrija	3,8	3,8	3,7	3,8	4,2	4,2	4,3	4,1	4,2	4,1	4,1
Belgija	3,9	4,0	4,2	3,9	4,1	3,9	3,9	3,9	4,0	3,8	3,8
Bulgarija			5,1	7,0	6,5	5,9	5,9	6,5	6,1	6,6	6,8
Čekija		3,7	3,5	3,5	3,4	3,5	3,5	3,5	3,5	3,4	3,5
Danija	3,4	3,5	3,4	3,7	3,6	4,6	4,4	4,4	4,5	4,0	4,1
Estija	7,2	5,9	5,5	5,5	5,0	5,0	5,0	5,3	5,4	5,5	6,5
Graikija	5,9	5,8	6,1	6,0	5,9	5,8	5,6	6,0	6,6	6,6	6,5
Islandija	3,4	3,5	3,7	3,9	3,8	4,2	3,6	3,3	3,4	3,4	3,1
Ispanija	5,2	5,5	5,5	5,5	5,6	5,9	6,2	6,3	6,5	6,3	6,8
Italija	5,6	5,6	5,4	5,4	5,2	5,3	5,4	5,7	5,6	5,8	5,8
Jungtinė Karalystė		5,9	5,4	5,3	5,6	5,3	5,4	5,3	5,0	4,6	5,1
Kipras		4,3	4,3	4,4	4,3	4,4	4,5	4,3	4,7	4,9	5,4
Kroatija							5,5	5,6	5,4	5,3	5,1
Latvija		6,7	7,8	6,4	7,3	7,4	6,8	6,5	6,5	6,3	6,5
Lenkija		6,6	5,6	5,3	5,1	5,0	5,0	5,0	4,9	4,9	4,9
Lietuva	6,6	6,9	6,3	5,9	6,1	6,4	7,3	5,8	5,3	6,1	6,1
Liuksemburgas	3,9	3,9	4,2	4,0	4,1	4,3	4,1	4,0	4,1	4,6	4,4
Malta		3,9	4,0	3,9	4,3	4,0	4,3	4,0	3,9	4,1	4,0
Norvegija	3,6	4,1	4,8	3,5	3,7	3,5	3,4	3,3	3,2	3,3	3,4
Olandija		4,0	3,8	4,0	4,0	4,0	3,7	3,8	3,6	3,6	3,8
Portugalija	7,0	7,0	6,7	6,5	6,1	6,0	5,6	5,7	5,8	6,0	6,2
Prancūzija	4,2	4,0	4,0	3,9	4,4	4,4	4,4	4,6	4,5	4,5	4,3
Rumunija				7,8	7,0	6,7	6,0	6,2	6,3	6,6	7,2
Slovakija		3,9	4,1	3,5	3,4	3,6	3,8	3,8	3,7	3,6	3,9
Slovėnija		3,4	3,4	3,3	3,4	3,2	3,4	3,5	3,4	3,6	3,7
Suomija	3,5	3,6	3,6	3,7	3,8	3,7	3,6	3,7	3,7	3,6	3,6
Švedija	3,3	3,3	3,6	3,3	3,5	3,7	3,5	3,6	3,7	3,7	3,9
Šveicarija				4,7	4,9	4,8	4,5	4,5	4,4	4,2	
Vengrija		4,0	5,5	3,7	3,8	3,5	3,4	3,9	4,0	4,3	4,2
Vokietija		3,8	4,1	4,9	4,8	4,5	4,5	4,5	4,3	4,6	5,1

4.0.3 pav.: 20% didžiausias pajamas gaunančių asmenų pajamos, padalintos 80% mažiausias pajamas gaunančių asmenų pajamų

Valsts/Metals	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ES (28 valst.)	9,2	8,9	8,2	7,1	7,0	8,9	9,5	9,8	10,4	10,8	10,2
Airija	4,5	4,4	4,4	4,7	6,4	12,0	13,9	14,7	14,7	13,1	11,3
Austrija	5,8	5,6	5,3	4,9	4,1	5,3	4,8	4,8	4,9	5,4	5,6
Beļģija	7,4	8,5	8,3	7,5	7,0	7,9	8,3	7,2	7,6	8,4	8,5
Bulgārija	12,1	10,1	9,0	6,9	5,6	6,8	10,3	11,3	12,3	13,0	11,4
Čekija	8,2	7,9	7,2	5,3	4,4	6,7	7,3	6,7	7,0	7,0	6,1
Dānija	5,2	4,8	3,9	3,8	3,4	6,0	7,5	7,6	7,5	7,0	6,6
Estija	10,2	8,0	5,9	4,6	5,5	13,5	16,7	12,3	10,0	8,6	7,4
Griekija	10,3	10,0	9,0	8,4	7,8	9,8	12,7	17,9	24,5	27,5	26,5
Islandija	4,0	2,5	2,8	2,3	2,9	7,2	7,6	7,0	6,0	5,4	4,9
Ispanija	11,1	9,2	8,5	8,2	11,3	17,9	19,9	21,4	24,8	26,1	24,5
Itālija	7,9	7,7	6,8	6,1	6,7	7,8	8,4	8,4	10,7	12,2	12,7
Jungtinā Karaliste	4,6	4,8	5,4	5,3	5,6	7,6	7,8	8,1	7,9	7,6	6,1
Kipras	4,4	5,3	4,6	3,9	3,7	5,4	6,3	7,9	11,9	15,9	16,1
Krievija	13,7	12,7	11,2	9,9	8,6	9,2	11,7	13,7	16,0	17,3	17,3
Latvija	11,7	10,0	7,0	6,1	7,7	17,5	19,5	16,2	15,0	11,9	10,8
Lenkija	19,1	17,8	13,9	9,6	7,1	8,2	9,7	9,7	10,1	10,3	9,0
Lietuva	10,7	8,3	5,8	4,3	5,8	13,8	17,8	15,4	13,4	11,8	10,7
Liuksemburgas	5,1	4,5	4,7	4,1	5,1	5,1	4,4	4,9	5,1	5,9	5,9
Malta	7,3	6,9	6,8	6,5	6,0	6,9	6,9	6,4	6,3	6,4	5,8
Norvēģija	4,3	4,4	3,4	2,5	2,5	3,1	3,5	3,2	3,1	3,4	3,5
Olandija	4,7	4,7	3,9	3,2	2,8	3,4	4,5	5,0	5,8	7,3	7,4
Portugālija	6,4	7,7	7,8	8,1	7,7	9,8	11,0	12,9	15,8	16,4	14,1
Prancūzija	8,9	8,5	8,5	7,7	7,1	8,7	8,9	8,8	9,4	9,9	10,3
Rumunija	7,7	7,2	7,3	6,4	5,8	6,9	7,0	7,2	6,8	7,1	6,8
Slovākija	18,8	16,3	13,4	11,1	9,5	12,0	14,4	13,6	14,0	14,2	13,2
Slovēnija	6,0	6,5	6,0	4,9	4,4	5,9	7,3	8,2	8,9	10,1	9,7
Suomija	10,4	8,4	7,7	6,9	6,4	8,2	8,4	7,8	7,7	8,2	8,7
Švedija	6,7	7,8	7,1	6,2	6,2	8,4	8,6	7,8	8,0	8,1	8,0
Šveicarija	4,3	4,5	4,0	3,7	3,4	4,1	4,6	4,1	4,2	4,4	4,6
Vengrija	5,8	7,2	7,5	7,4	7,8	10,0	11,2	11,0	11,0	10,2	7,7
Vācija	10,7	11,2	10,3	8,7	7,5	7,8	7,0	5,8	5,4	5,2	5,0

4.0.4 pav.: Nedarbo lygis (proc.)

Valst/Metai	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ES (28 valst.)	17004	17630	18378	19147	19401	19132	19596	19918	20281	20345	20709
Airija	17326	18663	19192	20159	19945	19232	19400	18783	18760	18608	18847
Austrija	21024	21834	23136	23722	23884	23384	24085	24456	25323	24930	25850
Belgija	19301	19827	20434	21019	21610	21487	22220	22606	23137	22994	23924
Bulgarija	5309	5484	6002	6519	7352	7153	7512	8155	8475	8921	
Čekija	11946	12579	13263	14070	13497	13850	14191	14672	14892	14958	15567
Danija	17014	17076	18168	18871	19067	19288	20446	21042	21641	21407	22216
Estija	8085	8681	9668	10821	11791	10948	11138	11910	12249	12993	13074
Graikija			17629	18494	19330	19167	17914	16412	15498	14823	14703
Islandija	17527	18229	19172	21519	20973	18541	17197	18139	19179	19633	
Ispanija	16833	17307	17998	18165	18488	18374	18092	18133	18010	18113	18340
Italija	18746	19178	20061	20972	21385	20402	21131	21201	20592	20618	20733
Jungtinė Karalystė	21405	22048	22883	23161	22415	22338	21788	21439	22338	22045	22086
Kipras	15557	16477	17694	18842	20894	19420	19545	18628	18094	18305	16252
Kroatija	9489	9696	9805	10514	11230	10952	11179	11701	12259	12252	12339
Latvija	7731	8548	10112	11352	11838	9910	10022	9932	10600	11496	11802
Lenkija	8637	8815	9358	10456	10828	11351	12257	12796	13550	13864	14062
Lietuva	9399	9963	11048	11430	12530	11710	12515	13232	13777	14770	15074
Liuksemburgas											
Malta											
Norvegija	20768	22077	21510	23230	23546	23732	24569	25208	26391	26905	27813
Olandija	20247	20625	21975	23135	23267	22725	22405	22725	22813	22436	
Portugalija	13495	14793	15337	15814	16050	15924	16489	16212	15995	16704	16822
Prancūzija	19747	20316	20939	22015	22083	21800	22651	23085	23336	23268	24283
Rumunija	5233	5521	6240	7308	8518	8039	8500	8698	8813	11460	9152
Slovakija	8918	9726	10361	11902	12782	12832	13867	13936	14206	14758	15653
Slovėnija	13617	14307	14975	15680	16205	15345	15720	16116	16073	15948	16149
Suomija	16857	17301	18191	19640	20520	20461	21320	22084	22837	23049	23057
Švedija	18090	18151	19119	20611	21194	20944	21258	22231	23109	23164	23579
Šveicarija	21239	21841	22821	24675	25227	25059	25371	26453	27680	28151	
Vengrija	10298	10829	11198	11122	11132	11033	11501	12247	12299	12603	13102
Vokietija	20232	21434	22111	22864	23164	22495	23996	24910	25766	25826	26736

4.0.5 pav.: Namų ūkių grynosios (atskaičius mokesčius) metinės pajamos (Eur)

ValstMetai	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ES(28 valst.)	100,0	100	100	100	100	100	100	100	100	100	100
Airija	119,5	126	125	124,6	131,2	128,7	120,8	121,8	119,7	118,7	122,9
Austrija	103,9	110,1	109,4	110,2	113,9	116,2	114,4	120,5	118,9	121	120,4
Belgija	108,5	109,1	110,2	111	112,9	115,3	115	110,1	110,4	111,9	107,9
Bulgarija	35,1	54	56,2	59,9	65,6	66,6	65,3	69,6	69,6	70,3	70,5
Čekija	53,2	66,8	68,9	70,9	80,7	74,9	77,4	83,1	84	83,7	80,1
Danija	134,2	141,7	141,9	143,3	136,1	138,2	135,6	143	142,5	139,8	144,8
Estija	57,7	73,1	75,5	79,6	82,8	80,6	81,4	87	86,3	88,4	89,3
Graikija	82,8	96	96,6	95,9	100,4	104,1	101,9	105	104,5	102,3	102,9
Islandija	128,5	173,4	163,6	157,3	106,9	104,4	113,1	113,7	116,6	119,2	119,1
Ispanija	90,2	91,2	92,3	92,8	95,6	96,6	94	94,5	92,8	93,1	92,6
Italija	103,7	116,2	115,2	114,4	104,3	108,1	106,4	112,7	111,6	111,6	110,2
Jungtinė Karalystė	110,7	113,5	113,7	114,9	101,3	96,9	101,2	97,3	103,8	100,7	104,5
Kipras	88,1	104,5	106,8	105,3	102,7	108,9	107,7	112,4	110,3	108,5	108,2
Kroatija	81,2	86,4	88,8	88,8	93,3	94,7	92,4	93,8	91,5	92,6	91,6
Latvija	48,9	65,1	69,5	75,8	85,3	85,9	83,9	87,9	87,4	86,6	89,8
Lenkija	48,8	64,6	66,3	68,8	74,1	63,5	69,8	62,4	61,8	61,6	64,2
Lietuva	48,4	61,5	64,2	69,3	72,1	73,6	72,2	78,7	77,7	77,6	79,1
Liuksemburgas	109,8	114,9	115,4	116,1	113	116,5	115,5	117,1	116,7	119,1	116,4
Malta	87,5	82,5	83,4	83,5	85,1	93,5	92,8	96,2	98,1	100	103,4
Norvegija	127,8	160	158,3	157,9	154,7	154,5	165,8	182,9	183,9	175,2	165,7
Olandija	108,0	88,5	88,5	86,6	95	97,7	95,9	98	95,6	96,1	98,7
Portugalija	85,1	88,3	87,7	86,4	93,4	92,9	91,4	90,6	91	91,1	94,1
Prancūzija	111,7	105,6	104,8	103	108,5	110,2	109	109,6	108,5	107,7	109,5
Rumunija	38,2	68,5	70,7	74,9	71,7	66	66,8	72,6	68	69,4	67,5
Slovakija	51,2	64,4	66,7	73,5	80	81,4	81,1	86,9	88,2	89,7	89,6
Slovėnija	72,7	87,3	87,6	90,8	93,7	96,1	95,3	96,6	97,1	98,7	96
Suomija	115,8	121,5	120,8	119,1	114,8	119,2	112,5	117	120,2	123,9	123,1
Švedija	118,5	120,2	119,2	117,7	109,8	104,1	115,4	122,3	123,4	124,6	125,4
Šveicarija	135,0	147	141,8	132,1	131,5	140,9	149,1	157,7	155,9	152	154,9
Vengrija	59,8	71,7	70,7	79,3	83,4	78,8	81,2	82,4	81,3	79,8	79,9
Vokietija	106,5	105,7	105,6	105,4	110	110,5	109,9	106,2	106,9	109	103,8

4.0.6 pav.: Kainų lygio indeksai

Metai	Migracija	BVP	Nedarbas	S80 S20	Pajamos	Kain_lygis
2004	-9,5	5365,6	10,7	6,6	9399	48,4
2005	-15,4	6259,6	8,3	6,9	9963	61,5
2006	-7,5	7319,3	5,8	6,3	11048	64,2
2007	-6,7	8935,6	4,3	5,9	11430	69,3
2008	-5,1	10177,5	5,8	6,1	12530	72,1
2009	-10,1	8459,8	13,8	6,4	11710	73,6
2010	-25,2	8920,4	17,8	7,3	12515	72,2
2011	-12,6	10241,5	15,4	5,8	13232	78,7
2012	-7,1	11098,1	13,4	5,3	13777	77,7
2013	-5,7	11764,2	11,8	6,1	14770	77,6
2014	-4,2	12381,4	10,7	6,1	15074	79,1

4.0.7 pav.: Duomenys apie Lietuvą