



VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS
FUNDAMENTINIŲ MOKSLŲ FAKULTETAS
MATEMATINĖS STATISTIKOS KATEDRA

Kristina Galiautdinovaitė

**EUROPOS ŠALIŲ PAJAMŲ NELYGYBĖS RODIKLIŲ
VERTINIMAS
ESTIMATION OF INCOME INEQUALITY INDICATORS OF
EUROPEAN COUNTRIES**

Baigiamasis magistro darbas

Taikomosios statistikos studijų programa, valstybinis kodas 621G31001
Statistinių metodų finansuose ir ekonomikoje specializacija
Statistikos studijų kryptis

Vilnius, 2016

Vilniaus Gedimino technikos universitetas

Fundamentinių mokslų fakultetas

Matematinės statistikos katedra

ISBN ISSN

Egz. sk.

Data-....-....

Taikomosios statistikos studijų programos baigiamasis magistro darbas

Pavadinimas **Europos šalių pajamų nelygybės rodiklių vertinimas**

Autorius **Kristina Galiautdinovaitė**

Vadovas doc. dr. **Danutė Krapavickaitė**

Kalba

☒ lietuvių

☐ užsienio

Anotacija

Baigiamajame magistro darbe supažindinama su statistiniais rodikliais, naudojamais gyventojų pajamų nelygybei vertinti. Atsižvelgiant į imties planą, pasirinkti rodiklių įvertiniai ir jų tikslumo matų įvertiniai. Darbui naudojami Europos socialinio tyrimo 2012 m. pradiniai statistiniai duomenys. Pateikiama informacija apie duomenų gavimą bei pertvarkymą skaičiavimams palengvinti, apžvelgiami imčių planai. Iš minėtų duomenų vertinami Lietuvos pajamų nelygybės rodikliai bei jų tikslumo matai priklausomai nuo imties plano, o gauti rezultatai lyginami su Statistikos departamento interneto puslapyje talpinamais rezultatais. Pasirinkus keletą kaimyninių šalių, joms taip pat vertinami pajamų nelygybės rodikliai. Gauti rezultatai lyginami su oficialiąja statistika ir tarpusavyje.

Darbą sudaro 6 dalys: įvadas, pajamų nelygybės rodikliai, Europos socialinis tyrimas, pajamų nelygybės rodiklių vertinimas, pabaiga, literatūros sąrašas.

Darbo apimtis – 39 p. teksto be priedų, 8 iliustr., 6 lent., 18 bibliografinių šaltinių.

Atskirai pridedami darbo priedai.

Prasminiai žodžiai: pajamų nelygybės rodikliai, Europos socialinis tyrimas, Gini koeficientas, Theil indeksai, Atkinsono indeksas, visraktis

Vilnius Gediminas Technical University
Faculty of Fundamental Science
Department of Mathematical Statistics

ISBN ISSN
Copies No.
Date ..-.-.-.-

Applied Statistics study programme master thesis.

Title: Estimation of income inequality indicators of European countries

Author **Kristina Galiautdinovaitė**

doc. dr. **Danutė Krapavickaitė**

Thesis language



Lithuanian



Foreign (English)

Annotation

In this master thesis main statistical indicators used to estimate income inequality are overviewed. The purpose of this thesis is to estimate indicators' estimators and measures of accuracy taking into account specific sampling design. Data of European Social Survey 2012 is used in this thesis. The information about data acquisition and restructuring is introduced and also sample designs are reviewed. From the mentioned data according to sample design Lithuanian income inequality indicators and their measures of accuracy are estimated and compared with the officially published statistics. Other European countries are chosen and same calculations are made. The results obtained are compared with the officially published statistics and with each other.

Structure: introduction, income inequality indicators, European Social Survey, estimation of income inequality indicators, conclusions, references.

Thesis consist of: 39 p. text without appendixes, 8 pictures, 6 tables, 18 bibliographical entries.

Appendixes included.

Keywords: income inequality indicators, European Social Survey, Gini coefficient, Theil indexes, Atkinson index, jackknife

Turinys

Įvadas	6
1. Pajamų nelygybės rodikliai	8
1.1. Gini koeficientas	9
1.2. Apibendrintieji entropijos matai ir Theilo indeksai	10
1.3. Atkinsono koeficientas	11
2. Europos socialinis tyrimas	12
2.1. Europos socialinis tyrimas	12
2.2. ESS Lietuvoje	12
2.3. Tyrimo temos	13
2.4. Tyrimo metodai	14
2.5. Šalių imčių planų apžvalga	14
3. Pajamų nelygybės rodiklių vertinimas	16
3.1. Gini koeficientas	16
3.1.1. Gini koeficientas ir Lorenzo kreivė	16
3.1.2. Gini koeficiento įvertis	19
3.1.3. Gini koeficiento įvertinys ESS duomenims	20
3.1.4. Dispersijos vertinimas visrakčio metodu	21
3.1.5. Palyginimas su kitomis šalimis	22
3.2. Apibendrintieji entropijos matai ir Theilo indeksai	26
3.2.1. Theilo indeksų įverčiai	26
3.2.2. Theilo indeksų įvertiniai ESS duomenims	28
3.2.3. Dispersijos vertinimas	28
3.2.4. Palyginimas su kitomis šalimis	29
3.3. Atkinsono koeficientas	30
3.3.1. Atkinsono koeficiento įverčiai	30
3.3.2. Atkinsono koeficiento įvertiniai ESS duomenims	31
3.3.3. Dispersijos vertinimas	31
3.3.4. Palyginimas su kitomis šalimis	32
Apibendrinimas	33
Priedai	35
Literatūra	47

Lentelių sąrašas

1.	Šalių imčių planai	14
2.	Gini koeficiento įvertinių palyginimas su oficialiąja statistika	23
3.	Theilo L koeficiento įvertinių palyginimas su oficialiąja statistika	29
4.	Theilo T koeficiento įvertinių palyginimas su oficialiąja statistika	29
5.	Atkinsono koeficiento ($\varepsilon = 0,5$) įvertinių palyginimas su oficialiąja statistika . .	32
6.	Atkinsono koeficiento ($\varepsilon = 1$) įvertinių palyginimas su oficialiąja statistika . . .	32

Paveikslų sąrašas

1.	Lorenzo kreivė	9
2.	Pajamų pasiskirstymo priklausomybė nuo populiacijos pasiskirstymo	17
3.	Gini koeficientas iš Lorenzo kreivės	18
4.	Lorenzo kreivė ESS duomenims	21
5.	Lietuvos Lorenzo kreivė ESS duomenims	23
6.	Estijos Lorenzo kreivė ESS duomenims	24
7.	Suomijos Lorenzo kreivė ESS duomenims	24
8.	Švedijos Lorenzo kreivė ESS duomenims	25

Ivadas

Nelygybė gali atsirasti dėl užimtumo modelių, pajamų šaltinių, individualių charakteristikų (išsilavinimo, amžiaus, lyties ir t.t.) ar namų ūkio charakteristikų (dirbančiųjų skaičiaus namų ūkyje, šeimos dydžio ir pan.). Kaip teigiama E. Di Falco straipsnyje [2], pajamų nelygybė yra kompleksinis fenomenas, kuris atsiranda dėl įvairių faktorių sąveikos. Natūralu, kad žmonės nuo seno domisi, kaip įvertinti ir palyginti nelygybę.

Iki šiol viena iš plačiausiai naudojamų priemonių pajamų nelygybei vertinti – Lorenzo kreivė – autoriaus M. Lorenzo pasiūlyta dar 1905 metais [12], o remiantis ja 1912 metais C. Gini [9] pristatė vieną iš šiuo metu populiariausių pajamų nelygybės vertinimo rodiklių – Gini koeficientą.

Visgi Gini koeficientas turėjo trūkumų ir 1970 metais A. B. Atkinsonas straipsnyje apie nelygybės matavimą [1] pristatė naują rodiklių klasę su laisvai parenkamu parametru, skirtu stebėti, kuri pajamų dalis turi didesnę įtaką nelygybei.

Kita rodiklių klasė apibrėžta 1980 metais A. F. Shorrocks [15] ir išvesta iš pertekliaus rodiklio informacijos teorijoje. Informacijos teorijoje pertekliaus rodiklis gali būti interpretuojamas kaip atsitiktinumo nebuvimas duomenyse. Panašiai gali būti interpretuojamas ir pajamų nelygybės vertinime.

2000 metų gruodį Europos Vadovų Taryba patvirtino savo sprendimą pradėti kovą su skurdu ir socialine atskirtimi, todėl Belgijoje, Briuselio priemiestyje Laekene 2001 metais vykusiam susirinkime buvo patvirtinti „Laeken“ rodikliai [17] – bendras Europos skurdo ir socialinės atskirties rodiklių rinkinys, kurį sudarė 18 rodiklių, tarp kurių yra ir Gini koeficientas. Siekiama, kad šių rodiklių visuma atvazduotų ES socialines problemas, todėl orientuojamasi į keturias svarbiausias socialines įtrauktis – finansinį skurdą, užimtumą, sveikatą ir išsilavinimą.

Kitas būdas tirti pajamų nelygybę Europoje – nuo 2001 metų vykdomas Europos socialinis tyrimas (angl. European Social Survey (ESS)) [5]. Jo tikslas – ne tik stebėti Europos socialinę padėtį, politines ir vertybines nuostatas, bet ir skatinti mokslinius tyrimus šiose srityse.

Europos Sąjungos pajamų ir gyvenimo sąlygų statistika (angl. The EU statistics on income and living conditions (EU-SILC)) – šaltinis nuorodų į palyginamąją pajamų pasiskirstymo ir socialinės įtraukties statistiką Europos Sąjungoje. Tam pradžią davė 2003 metų susitarimas tarp Eurostat, šešių ES šalių (Austrijos, Belgijos, Danijos, Graikijos, Airijos ir Liuksemburgo) ir Norvegijos. EU-SILC renka ir teikia informaciją apie pajamas, skurdą, socialinę atskirtį ir kitas gyvenimo sąlygas, taip pat apie šių duomenų kitimą ilguoju laikotarpiu.

2005 metais Pasaulio banko instituto parengtame skurdo analizės vadove [18] pristatomas įžanginis skurdo analizės kursas, apimantis pagrindinius skurdo vertinimo metodus ir jų taikymą imties duomenims. Vadove aptariami aktualūs pajamų nelygybės rodikliai, taip pat gerovės koncepcija, kodėl reikia vertinti skurdą, kaip tai padaryti, skurdo ribų nustatymas, indikatoriai ir jų palyginimas bei kitos temos.

Šiame darbe supažindinama su minėtais statistiniais rodikliais, kurie yra naudojami gyventojų pajamų nelygybei vertinti. Pasirinktais būdais užrašomos rodiklių įverčių formos ir

vertinami tikslumo matai, atsižvelgiant į imties planą. Kadangi pajamų nelygybės rodiklių įverčiai yra netiesiniai sumų atžvilgiu, jų dispersijas vertinti yra sunkiau, todėl darbe taikomas visrakčio metodas.

Baigiamajam darbui naudojami Europos socialinio tyrimo 2012 m. pradiniai statistiniai duomenys. Pateikiama informacija apie duomenų gavimą bei pertvarkymą skaičiavimams palengvinti, apžvelgiami imčių planai. Visi skaičiavimai atliekami programa *R*, papildomai naudojamas paketas *ineq*.

Iš minėtų duomenų vertinami Lietuvos pajamų nelygybės rodikliai bei jų paklaidos priklauso nuo imties plano, o gauti rezultatai lyginami su Statistikos departamento puslapyje talpinamais rezultatais. Pasirinkus keletą kaimyninių šalių, joms taip pat vertinami pajamų nelygybės rodikliai. Gauti rezultatai lyginami su oficialiąja statistika (Lietuvos Statistikos departamento [16] ir Eurostat [4] teikiamais duomenimis) ir tarpusavyje.

1. Pajamų nelygybės rodikliai

Pajamų nelygybė – vienas iš ekonominės nelygybės elementų. Pajamų nelygybė vadinamas namų ūkių ar individų pajamų netolygus pasiskirstymas, kuris paprastai pateikiamas kaip procentinė visų pajamų dalis, atitenkanti procentinei populiacijos daliai, pavyzdžiui, 70 proc. visų šalies pajamų priklauso 20 proc. jos rezidentų. Pajamų nelygybei gali daryti įtaką regionas, lytis, išsilavinimas, religinės pažiūros, socialinis statusas ir kiti veiksniai.

Dažnai ši sąvoka siejama su „teisingumu“. Manoma, kad egzistuoja „neteisybė“, jei visi turtingieji, kurių pasaulyje yra mažuma, turi neproporcingai dideles pajamas, lyginant su jų dalimi populiacijoje.

Pajamų nelygybės sąvoka yra bendresnė nei skurdo, nes yra apibrėžiama visai populiacijai, o ne jos daliai. Pasaulio banko parengtame vadove [18] pateikiamas kriterijų sąrašas, pagal kuriuos galime vertinti pajamų nelygybės rodiklių tinkamumą:

1. *Nepriklausomybė nuo vidurkio*, t. y., jei visų populiacijos elementų pajamas padidintume du kartus, rodiklio reikšmė neturėtų pasikeisti.
2. *Nepriklausomybė nuo populiacijos dydžio*.
3. *Simetriškumas*, t. y., jei keli žmonės pasikeičia pajamomis, nuo to rodiklio reikšmė nesikeičia.
4. *Pigou-Daltono jautrumo perdavimui kriterijus*. Remiantis šiuo kriterijumi, jei dalis pajamų būtų atimta iš daugiausiai uždirbančių asmenų ir perduota mažiausiai uždirbantiems, nelygybė turėtų sumažėti, o kartu pasikeisti ir tą nelygybę matuojanti rodiklio reikšmė.
5. *Išskaidomumas*. Tai reiškia, kad jei populiaciją suskaidytume į grupes, jose apskaičiuotume nelygybės rodiklių reikšmes, tai iš jų galėtume gauti bendrą populiacijos rodiklio reikšmę.
6. *Statistinis testavimas*, t. y., įmanoma patikrinti rodiklio kitimo laike reikšmingumą.

Paprasčiausias būdas įvertinti nelygybę yra išrikiuoti populiaciją pagal tiriamą kintamąjį didėjimo tvarka ir padalinti ją į penktadalius (kvintilius) arba kitokias dalis. Taip tirdami populiacijas jas galime nesunkiai lyginti tarpusavyje.

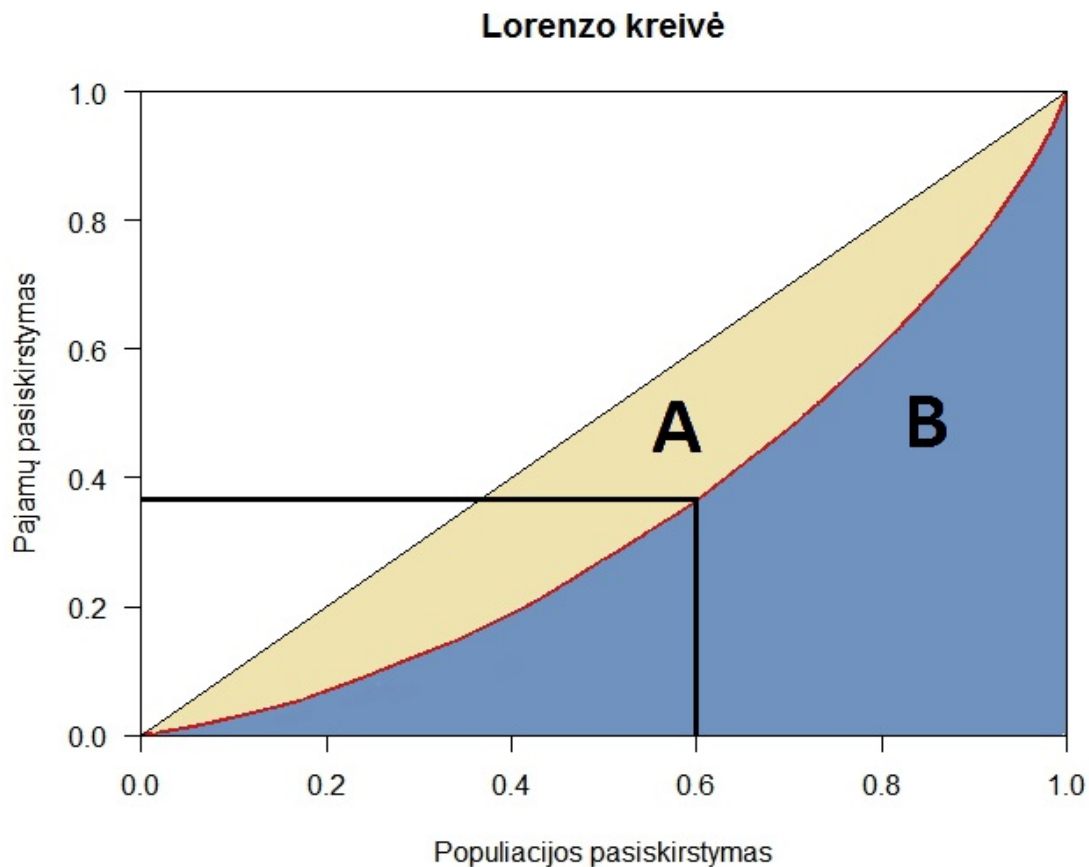
1.1. Gini koeficientas

Gini koeficientas skaičiuojamas pagal 1912 m. italų statistiko ir sociologo Corrado Gini sukurtą metodą [9], remiantis statistine matavimų dispersija, ir kinta skalėje nuo 0 iki 1. Pavyzdžiui, jei Gini koeficiento reikšmė yra 1, turime absoliučios, arba kitaip tobulos, pajamų nelygybės atvejį, o jei visi namų ūkiai turėtų vienodas pajamas – Gini koeficientas būtų lygus 0 – absoliuti lygybė.

Koeficientas yra apskaičiuojamas naudojantis Lorenzo kreive [12], parodančia gaunamų pajamų pasiskirstymą pagal namų ūkius.

Lorenzo kreivė – grafinis visuomenės pajamų pasiskirstymo atvaizdavimas. Kreivė parodo, kaip netolygiai pasiskirsto pajamos. Lorenzo kreivė apskaičiuojama turtą padalinus iš žmonių, kurie juo naudojasi, skaičiaus. Kuo kreivė tiesesnė, tuo pajamos pasiskirsto tolygiau ir atvirkščiai, kuo mažesnis kiekis turtingų žmonių ir kuo didesnį turtą ji valdo, tuo kreivė yra labiau įgaubta.

1 pav. pavaizduotame grafike juodomis tiesėmis žymimą atvejį reiktų interpretuoti taip: apytiksliai 37 proc. šalies namų ūkių gaunamų pajamų uždirba 60 proc. populiacijos. Tuo tarpu 45 laipsnių tiesė simbolizuoja visišką pajamų pasiskirstymo lygybę, kuomet visos 10 namų ūkių grupių, sudarančios po 10 proc. visos populiacijos, gautų po 10 proc. pajamų.



1 pav. Lorenzo kreivė

Gini koeficientas yra apskaičiuojamas kaip ploto, kurį riboja išgaubta Lorenzo kreivė iki 45 laipsnių tiesės (A plotas paveiksle) ir viso dešiniojo trikampio ploto (A+B) santykis.

Tarkime, turime baigtinę populiaciją $U = 1, \dots, N$, o tyrimo kintamasis Y yra pajamos su reikšmėmis $y_1, \dots, y_N$. Pažymėkime x_i taškus ant X ašies, kurie reiškia atitinkamą populiacijos dalį, o žyminčius pajamų dalį y_i – ant Y. Tada Gini koeficientą galime užrašyti:

$$G = 1 - \sum_{i=1}^N (x_i - x_{i-1}) (y_i + y_{i-1}). \quad (1)$$

Jeigu ant X ašies yra N vienodo ilgio intervalų, tuomet gauname paprastesnę išraišką:

$$G = 1 - \frac{1}{N} \sum_{i=1}^N (y_i + y_{i-1}). \quad (2)$$

Gini koeficientas yra pakankamai geras ir bene plačiausiai naudojamas dėl savo paprastumo, visgi kaip rodiklis jis nėra visiškai tinkamas, nes netenkina išskaidomumo kriterijaus.

1.2. Apibendrintieji entropijos matai ir Theilo indeksai

Pristatyti kaip pajamų nelygybės vertinimo rodikliai [15], šie matai tenkina visus šešis pageidaujamus kriterijus. Entropijos matas išreiškiamas viena bendrąja formule, o Theilo indeksas – vienas iš entropijos mato galimų variantų.

Bendroji entropijos matų formulė:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\mu} \right)^\alpha \right], \quad (3)$$

čia μ – pajamų vidurkis. Šio koeficiento reikšmės kinta nuo 0 iki ∞ . 0 reiškia lygų pajamų pasiskirstymą, o vis didesnė reikšmė – vis didėjančią nelygybę. Parametrą $\alpha \in \mathbb{R}$ nustatome patys, paprastai būna 0, 1 arba 2. Esant mažesnėms α reikšmėms $GE(\alpha)$ tampa jautresnis apatinėms pajamų reikšmėms, o pasirinkus didesnės α reikšmės daugiau įtakos indeksui turi aukščiausios pajamos.

Theilo T indeksas aprašomas formule:

$$GE(1) = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\mu} \ln \left(\frac{y_i}{\mu} \right). \quad (4)$$

Dar vienas entropijos variantas – Theilo L indeksas:

$$GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \left(\frac{\mu}{y_i} \right). \quad (5)$$

Theilo indeksai matuoja, kaip toli populiacija yra „nutolusi“ nuo idealios pusiausvyros, kai visi asmenys turi vienodas pajamas. Kuo šis indeksas yra didesnis, tuo didesnė pajamų nelygybė.

Pavyzdžiui, jei kiekvienas asmuo gauna vienodas pajamas, tuomet Theilo T koeficientas yra lygus 0. Jei populiacijoje visos esamos pajamos priklauso tik vienam asmeniui, tuomet $GE(1) = \ln N$.

1.3. Atkinsono koeficientas

A. B. Atkinsonas 1970 m. [1] pasiūlė kitą nelygybės rodiklių klasę. Joje taip pat yra laisvai keičiamas parametras ε . Visa šių koeficientų klasė apibrėžiama:

$$A_\varepsilon = 1 - \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\mu} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}, \quad \varepsilon \neq 1 \quad (6)$$

$$A_\varepsilon = 1 - \frac{\prod_{i=1}^N \left(y_i^{\frac{1}{N}} \right)}{\mu}, \quad \varepsilon = 1. \quad (7)$$

Atkinsono ε parametras parodo socialinio naudingumo dydį visiškai perskirsčius pajamas. Jei $\varepsilon = 0$, darome išvadą, kad bet kaip perskirsčius pajamas nelygybė nesumažės ir tokiu atveju Atkinsono indeksas yra lygus 0. Ir atvirkščiai, jei $\varepsilon = \infty$, tai gaunama nauda yra begalinė ir $A_\varepsilon = 1$. Taigi indeksas A_ε kinta tarp 0 ir 1.

Šis indeksas taip pat naudingas norint sužinoti, kuri pajamų skirstinio dalis turi didesnę reikšmę pajamų nelygybei. Kai Atkinsono koeficiento parametras ε artėja į 1, koeficientas tampa jautresnis mažesnėms pajamų reikšmėms ir, atvirkščiai, kai ε artėja į 0 – didesnėms pajamų reikšmėms.

Atkinsono indeksas taip pat tenkina visus šešis rodiklių kriterijus.

2. Europos socialinis tyrimas

Baigiamajam darbui naudojami Europos socialinio tyrimo 2012 m. pradiniai statistiniai duomenys. Informaciją apie socialinį tyrimą ir pačius duomenis galima rasti ESS internetiniame puslapyje [5].

2.1. Europos socialinis tyrimas

Europos socialinis tyrimas yra mokslininkų inicijuotas ir nuo 2001 m. Europos šalyse kas dvejus metus atliekamas tarptautinis tyrimas. Jį organizuoja Centrinė mokslinė grupė, kuriai vadovauja Rory Fitzgeraldas iš Londono City universiteto (Jungtinė Karalystė), kartu su šešiais instituciniais partneriais.

Europos socialiniame tyrime analizuojamos žmonių nuostatos, įsitikinimai ir elgsena daugiau nei trisdešimtyje skirtingų Europos šalių.

Pagrindiniai ESS tikslai:

- fiksuoti Europos socialinės struktūros, gyvenimo sąlygų ir nuostatų stabilumą bei kaitą, taip pat pateikti įžvalgas apie tai, kaip kinta Europos socialinis, politinis ir vertybinis paveikslas,
- įgyvendinti ir skleisti aukštus metodologinius tarptautinių tyrimų standartus socialiniuose moksluose tokiose srityse kaip klausimynų kūrimas ir testavimas, imčių sudarymas, duomenų rinkimas, paklaidų mažinimas ir klausimų patikimumo užtikrinimas,
- sukurti pagrįstus nacionalinės pažangos rodiklius, kurie remtųsi tuo, kaip piliečiai suvokia ir vertina pagrindinius savo visuomenių bruožus,
- skatinti ir įgyvendinti Europos socialinių mokslų tyrėjų kiekybinių lyginamųjų tyrimų ir duomenų analizės mokymus,
- didinti duomenų apie socialinius pokyčius matomumą ir prieinamumą mokslininkams, politikos formuotojams ir visuomenei.

2.2. ESS Lietuvoje

Lietuvai ESS mokslinėje patariamojoje taryboje atstovauja prof. Algis Krupavičius, Kauno technologijos universiteto (KTU) Politikos ir viešojo administravimo instituto (PVAI) direktorius.

ESS nacionalinis koordinatorius Lietuvoje yra dr. Vaidas Morkevičius, KTU PVAI vyresnysis mokslo darbuotojas. Nacionalinę ESS mokslinę grupę taip pat sudaro KTU PVAI darbuotojai Ligita Šarkutė, Audronė Telešienė ir Giedrius Žvaliauskas.

ESS įgyvendinimą Lietuvoje nuo 4-osios bangos rėmė Europos socialinis fondas, Lietuvos Respublikos švietimo ir mokslo ministerija bei Lietuvos mokslo taryba.

6-osios bangos ESS duomenų surinkimą (gyventojų apklausą) Lietuvoje atliko UAB „Baltijos tyrimai“.

Lietuva įsitraukė į ESS nuo 4-osios bangos ir 2015 m. Lietuvoje atlikta jau ketvirtoji tyrimo banga.

2.3. Tyrimo temos

ESS tyrime analizuojamos žmonių nuostatos, įsitikinimai ir elgsena skirtingose Europos šalyse. Tyrimo klausimynas sudarytas iš nuolatinių klausimų modulio, kartojamo vykdant kiekvienos bangos tyrimą, kelių kintančių klausimų modulių, tyrinėjančių tam tikras temas ir papildomo klausimyno, kuriame yra žmogiškųjų vertybių skalė, bei klausimų, skirtų metodologiniams eksperimentams. Nuolatiniai klausimai apima:

- žiniasklaidos vartojimą;
- politiką;
- pasitikėjimą institucijomis;
- imigraciją;
- nusikaltimų baime;
- sveikatą;
- religiją;
- socialinį gerbūvį;
- žmogiškąsias vertybes;
- socialines demografines respondentų ir jų šeimos narių charakteristikas, tokias kaip gyvenamoji vieta, išsilavinimas, šeimyninė padėtis, darbo užmokestis ir kt. Šiame darbe tai ir bus daugiausiai naudojami ir nagrinėjami duomenys.

Kintančių klausimų moduliai skirti nagrinėti:

- imigraciją, pilietybę ir demokratiją (1 banga, 2002)
- ekonominę moralę, šeimą, darbą ir gerbūvį bei sveikatą ir sveikatos priežiūrą (2 banga, 2004)
- gyvenimo tarpsnius, asmeninį ir socialinį gerbūvį (3 banga, 2006)
- požiūrius į diskriminaciją dėl amžiaus ir požiūrius į socialinę gerovę (4 banga, 2008)
- šeimą, darbą ir gerbūvį po ekonominio nuosmukio ir pasitikėjimą policija bei teismais (5 banga, 2010)

- demokratijos supratimą ir vertinimą bei asmeninį ir socialinį gerbūvį po ekonominio nuosmukio (6 banga, 2012)
- požiūrius į imigraciją ir sveikatos nelygybę (7 banga, 2014)

2.4. Tyrimo metodai

ESS apklausa Lietuvoje įgyvendinama tiesioginės apklausos būdu, o papildomą klausimyną pildo pats apklausiamas asmuo. Klausimynai respondentams pateikiami lietuvių arba rusų kalbomis. Klausimynų pagrindas yra baziniai klausimynai anglų kalba, tačiau juose yra ir atskiroms šalims būdingų klausimų, pavyzdžiui, į skirtingų šalių klausimynus įtraukti skirtingi (arba adaptuoti, kad atspindėtų nacionalinį kontekstą) partijų, išpažįtamų religijų, išsimokslinimo ir šeimyninės padėties atsakymų variantai.

ESS imtys apima visų tyrime dalyvaujančių šalių namų ūkių gyventojus nuo 15 metų ir vyresnius nepriklausomai nuo jų tautybės, pilietybės ar kalbos. Imties elementai (asmenys, adresai arba namų ūkiai) išrenkami griežtai laikantis tikimybinų principų kiekvienoje imties pakopoje, o minimalus efektyvus siekiamas imties dydis (įvertinus tyrimo imties plano poveikį) yra 1500 arba 800 atsakytųjų (Lietuvoje – 1500 atsakytųjų).

Lietuvoje vykdomos ESS apklausos imties planas yra tikimybinis, o imčiai sudaryti naudojamas Lietuvos Respublikos rinkėjų sąrašas. Lietuvos teritorija yra sluoksniuojama pagal apskritis ir gyvenamosios vietos dydį. Sluoksniuose tikimybiskai išrenkamas tam tikras kiekis lizdų (5-ojoje bangoje lizdus sudarė rinkimų apylinkės, o 6-ojoje – seniūnijos) arba adresų proporcingai 15 metų ir vyresnių gyventojų skaičiui. Kiekviename lizde išrenkamas tam tikras kiekis adresų, o kiekviename adrese – vienas asmuo, vadovaujantis „paskutinio gimtadienio“ taisykle.

2.5. Šalių imčių planų apžvalga

Norint palyginti Lietuvos pajamų nelygybės rodiklį įvertius su kitomis Europos šalimis, pasirinktos Estija, Suomija ir Švedija. Visų šių šalių imčių planai pateikiami 1 lentelėje.

1 lentelė. Šalių imčių planai

Šalis	Imties planas	Sluoksniai
Lietuva	Sluoksninė imtis	Apskritys (10)
Estija	Sluoksninė imtis	Apskričių grupės (5)
Suomija	Paprastoji atsitiktinė imtis	–
Švedija	Paprastoji atsitiktinė imtis	–

Kaip teigiama D. Krapavickaitės ir A. Plikuso vadovėlyje [11], atsitiktinė imtis (angl. random sample) – tai imtis, sudaryta taip, kad visi individai turi tam tikrą tikimybę, nelygią nuliui, patekti į šią imtį. Jei išrinkimo į imtį tikimybės yra lygios, turime paprastąją atsitiktinę imtį (angl. simple random sample). Tokiu atveju bet kokia n dydžio imtis iš N dydžio populiacijos turi vienodą tikimybę būti išrinkta. Pavyzdžiui, tokia imtis gali būti sudaryta iš visų gyventojų

sąrašo, kai tiriamieji atrenkami, taikant atsitiktinių skaičių arba kitą tikimybių metodą. Toks imties planas buvo taikytas atliekant Europos socialinį tyrimą Suomijoje ir Švedijoje.

Kartais nėra galimybės rinkti PAI (nėra visos populiacijos elementų sąrašo ar dėl ribotų tyrimo išlaidų) arba tiesiog patogiau N dydžio populiaciją skirstyti į H nesikertančių sluoksnių, kurių dydžiai yra N_h , $h = 1, \dots, H$, atsižvelgiant į tam tikrą svarbų požymį, pavyzdžiui lytį, tautybę, gyvenamąją vietą ar kt. Tuomet iš kiekvieno sluoksnio atskirai, nepriklausomai nuo kitų, renkama n_h dydžio paprastoji atsitiktinė imtis, o visos tokios imtys sudaro sluoksninę atsitiktinę imtį (angl. stratified random sample), t.y. $n = \sum_{h=1}^H n_h$. Tyrimo kintamojo y i -ojo elemento reikšmė h -ajame sluoksnyje žymima y_{hi} .

Jei tyrimo kintamojo y reikšmių suma h -ajame sluoksnyje yra $t_h = \sum_{i=1}^{N_h} y_{hi}$, o sluoksnio vidurkis – $\mu_h = \frac{t_h}{N_h}$, tai populiacijos sumos įvertinys yra

$$\hat{t}_{sl} = \hat{t}_1 + \dots + \hat{t}_H,$$

o populiacijos vidurkio įvertinys:

$$\hat{\mu}_{sl} = \sum_{h=1}^H \frac{N_h}{N} \hat{\mu}_h.$$

Lietuvoje sluoksniai yra apskritys, jų iš viso yra $H = 10$, o Estijoje – apskričių grupės ($H = 5$).

3. Pajamų nelygybės rodiklių vertinimas

3.1. Gini koeficientas

Tarkime, kad turime U populiaciją, kurioje yra N asmenų, gaunančių teigiamas pajamas y su reišmėmis y_i , $i \in U$. Pažymėkime:

$t_y = \sum_{i \in U} y_i$ – visos populiacijos pajamų suma;

$\mu = \frac{t_y}{N}$ – populiacijos vidurkis.

Pirmas žingsnis norint apskaičiuoti Gini koeficientą yra variacinės eilutės užrašymas. Laikykime, kad pajamos y_k , $k \in [1, N]$ yra išrikiuotos nuo mažiausių iki didžiausių:

$$y_1 \leq y_2 \leq \dots \leq y_N.$$

Antras žingsnis – apskaičiuoti visų įmanomų pajamų porų skirtumų absoliutinių dydžių vidurkį Δ :

$$\Delta = \frac{\sum_{i=1}^N \sum_{j>i}^N |y_i - y_j|}{\frac{N(N-1)}{2}}. \quad (8)$$

Kadangi populiacija buvo išrikiuota didėjančia tvarka, šią išraišką galime supaprastinti:

$$\Delta = -2\mu \frac{N+1}{N-1} + \frac{4}{N(N-1)} \sum_{i=1}^N i y_i.$$

Apibrėžkime Gini koeficientą kaip funkciją nuo dviejų argumentų: Δ ir μ :

$$G = \frac{\Delta}{2\mu} \quad (9)$$

Akivaizdu, kad Δ yra teigiamas, be to

$$\begin{aligned} \Delta &\leq -2\mu \frac{N+1}{N-1} + \frac{4}{N(N-1)} \sum_{i=1}^N N y_i \\ &= -2\mu \frac{N+1}{N-1} + \frac{4}{N-1} N\mu \\ &= -2\mu \left(\frac{N+1}{N-1} - \frac{2N}{N-1} \right) = 2\mu. \end{aligned}$$

Taigi, jei populiacijoje visi gauna vienodas pajamas, gauname $\Delta = 0$ ir $G = 0$. Ir atvirkščiai, kai $\Delta = 2\mu$, turime visišką pajamų nelygybę ir $G = 1$.

3.1.1. Gini koeficientas ir Lorenzo kreivė

Tarkime, kad Q_k yra populiacijos pajamų dalis, kurią uždirba pirmieji k išrikiuotos pajamų sekos nariai, o P_k – populiacijos dalis, kurią sudaro tie patys pirmieji k narių.

Kai k kinta nuo 1 iki N , tai pažymėkime pirmųjų k populiacijos elementų gaunamą pajamų dalį

$$Q_k = \frac{\sum_{i=1}^k y_i}{\sum_{i=1}^N y_i} = \frac{\sum_{i=1}^k y_i}{t_y},$$

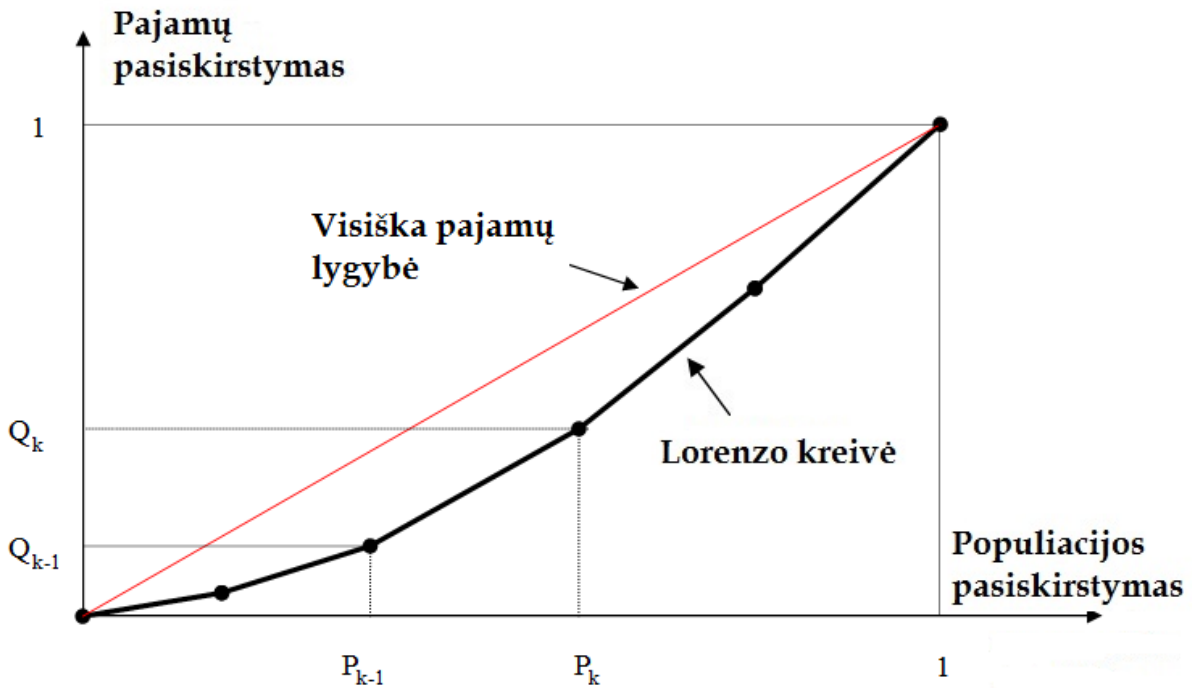
o pirmųjų k populiacijos elementų sudaromą dalį

$$P_k = \frac{k}{N}.$$

Laikykime, kad

$$P_0 = Q_0 = 0.$$

Lorenzo kreivė – grafinis visuomenės pajamų pasiskirstymo atvaizdavimas. Tai kreivė, tiesėmis jungianti taškus (P_k, Q_k) , ir esanti žemiau visiškos pajamų lygybės linijos.

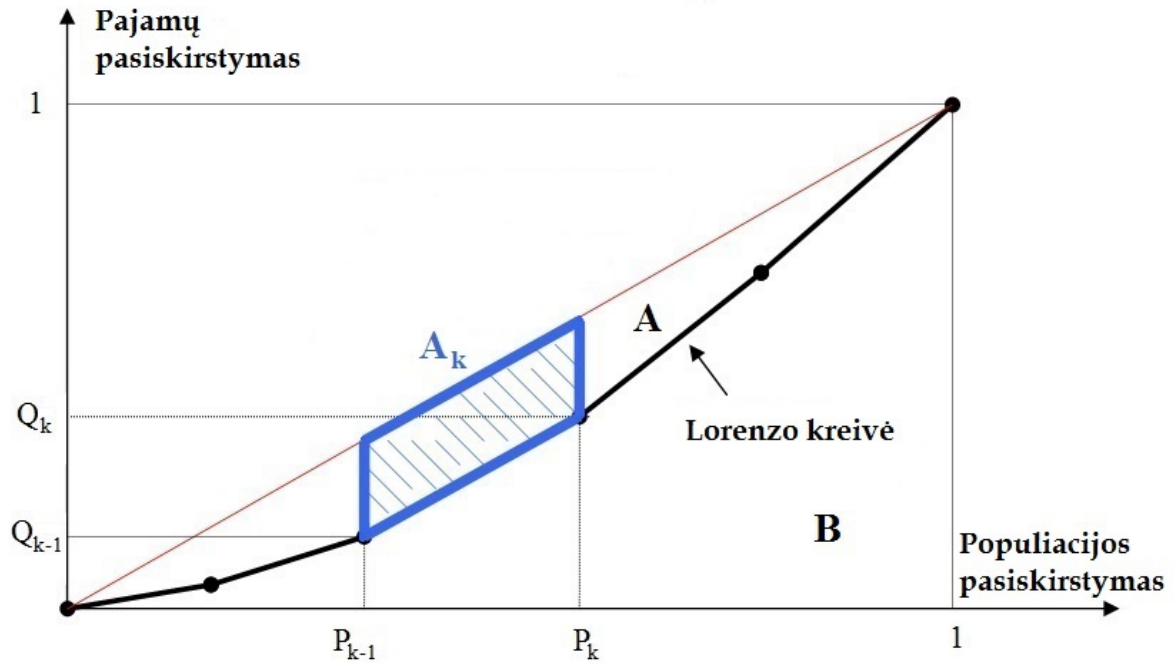


2 pav. Pajamų pasiskirstymo priklausomybė nuo populiacijos pasiskirstymo

Pažymėkime sritį tarp idealaus pasiskirstymo linijos ir Lorenzo kreivės raide A , o visą sritį po idealaus pasiskirstymo kreivės – B . Tada išreiškiamą šių dviejų dalių plotų santykį pažymėkime $R = \frac{S_A}{S_B}$.

Parodysime, kad Gini koeficientas yra išreiškiamas

$$G = \frac{N}{N-1} R. \quad (10)$$



3 pav. Gini koeficientas iš Lorenzo kreivės

Dviejų kreivių ribojamą trapecijos plotą S_A galime apytiksliai apskaičiuoti, sudėdami sričių A_k plotus S_{A_k} . Kiekvienos iš tų dalių plotas užrašomas

$$S_{A_k} = \frac{(P_{k-1} - Q_{k-1} + P_k - Q_k) \cdot (P_k - P_{k-1})}{2}.$$

Tada bendras srities A plotas yra

$$S_A = \sum_{k=1}^N S_{A_k}.$$

Iš čia apskaičiuojame plotų santykį R :

$$\begin{aligned} R &= \frac{S_A}{S_B} = \frac{\sum_{k=1}^N \frac{(P_{k-1} - Q_{k-1} + P_k - Q_k) (P_k - P_{k-1})}{2}}{\frac{1}{2}} \\ &= \sum_{k=1}^N (P_{k-1} - Q_{k-1} + P_k - Q_k) (P_k - P_{k-1}) \\ &= \sum_{k=1}^N \left(\frac{k}{N} - \frac{\sum_{i=1}^k y_i}{t_y} + \frac{k-1}{N} - \frac{\sum_{i=1}^{k-1} y_i}{t_y} \right) \left(\frac{k}{N} - \frac{k-1}{N} \right) \\ &= \frac{1}{N^2} \sum_{k=1}^N (2k-1) - \frac{1}{N t_y} \sum_{k=1}^N \left(2 \sum_{i=1}^k y_i - y_k \right) \\ &= \dots = \frac{2 \sum_{i=1}^N i y_i - (N+1) t_y}{N t_y} \end{aligned}$$

Kadangi

$$R = \frac{N-1}{N} \cdot \left(\frac{2 \sum_{i=1}^N i \cdot y_i - (N+1) t_y}{(N-1) t_y} \right) = \frac{N-1}{N} G,$$

tai gauname Gini koeficiento išraišką, apibūžtą 10 formulėje. Kai N pakankamai didelis, $R \approx G$.

3.1.2. Gini koeficiento įvertis

Ankstesniame skyrelyje gavome plotų santykio R išraišką:

$$R = \frac{2 \sum_{i=1}^N i y_i - (N+1) t_y}{N t_y}$$

Atliekame pertvarkymus:

$$\begin{aligned} R &= \frac{2 \sum_{i=1}^N i y_i - t_y - N t_y}{N t_y} = \frac{2 \sum_{i=1}^N i y_i - t_y}{N t_y} - 1 \\ &= \frac{2 \sum_{i=1}^N i y_i - t_y}{N t_y} - 1 \\ &= \frac{2 \sum_{i \in U} r_i y_i - t_y}{N t_y} - 1, \end{aligned}$$

čia r_i yra i -ojo populiacijos elemento rangas, o t_y – kintamojo y populiacijos reikšmių suma.

Tarkime, kad turime n dydžio imtį u iš populiacijos U . Elemento imties plano svoris ω_i – tai populiacijos elementų skaičius, kuriam jis atstovauja, būdamas išrinktas į imtį. Pažymėkime suminius svorius $W_i = \sum_{j \in s, j \leq i} \omega_j$.

Norėdami įvertinti santykį R , vertinkime jo išraiškos skaitiklyje esančią sumą $S_1 = \sum_{i \in U} r_i y_i$ (šios dalies įrodymas pateikiamas EU-SILC Teoriniame Gini koeficiento tyrime [3]):

$$\hat{S}_1 = \sum_{i \in u} y_i \left(\omega_i W_i - \frac{1}{2} \omega_i^2 + \frac{1}{2} \omega_i \right),$$

o t_y ir N :

$$\hat{t}_y = \sum_{i \in u} \omega_i y_i$$

$$\hat{N} = \sum_{i \in u} \omega_i$$

Atliekame pertvarkymus:

$$\begin{aligned}\hat{G} &= \frac{2 \sum_{i \in u} y_i \cdot (\omega_i W_i - \frac{1}{2} \omega_i^2 + \frac{1}{2} \omega_i) - \sum_{i \in u} \omega_i \cdot y_i}{(\sum_{i \in u} \omega_i) \cdot (\sum_{i \in u} \omega_i \cdot y_i)} - 1 \\ &= \frac{2 \sum_{i \in u} y_i \cdot \omega_i \cdot W_i - \sum_{i \in u} y_i \cdot \omega_i^2 + \sum_{i \in u} y_i \cdot \omega_i - \sum_{i \in u} \omega_i \cdot y_i}{(\sum_{i \in u} \omega_i) \cdot (\sum_{i \in u} \omega_i \cdot y_i)} - 1\end{aligned}$$

Taigi galutinis Gini koeficiento įvertinys yra:

$$\hat{G} = \frac{2 \sum_{i \in u} y_i \cdot \omega_i \cdot W_i - \sum_{i \in u} y_i \cdot \omega_i^2}{(\sum_{i \in u} \omega_i) \cdot (\sum_{i \in u} \omega_i \cdot y_i)} - 1. \quad (11)$$

Ši išraiška tinka ir PAI, ir sluoksninei imčiai.

3.1.3. Gini koeficiento įvertinys ESS duomenims

Pasiėmus lietuvišką statistinio tyrimo klausimyną [7] ir anglišką kintamųjų vardų sąrašą [6], buvo išrinkti baigiamojo darbo tyrimui reikalingi kintamieji: namų ūkio dydis, narių skaičius, bendros pajamos, apskritis. Iš Statistikos departamento internetinio puslapio paimta informacija apie vyresnių nei 15 metų amžiaus gyventojų skaičių kiekvienoje apskrityje.

Kadangi duomenys apie pajamas pateikiami užkoduoti, t.y. pateikiamos raidės, žyminčios pajamų intervalą, į kurį patenka kiekvienas namų ūkis, o tyrimui reikalingi tikslūs skaičiai, buvo pasirinkti intervalų vidurio taškai, o kraštinius intervalus atitinkančios reikšmės pasirinktos laisvai. Taip pat dalis respondentų nepateikė duomenų apie namų ūkio pajamas, todėl skaičiavimuose naudoti tik pilni duomenys.

Nors pagal tyrimo Lietuvoje aprašymą buvo naudojamas sudėtingesnis daugiapakopės imties planas, iš duomenų galima atkurti tik sluoksninę imtį.

Visi skaičiavimai buvo atliekami su statistine programa *R*. Prieduose pateikiami du būdai Gini koeficiento įvertinio skaičiavimui, viename iš jų formulė skaičiuojama paprastai, kitame – pasinaudota *R* paketu *ineq*.

Abiem atvejais rezultatai sutampa, kadangi paketuose naudojama ta pati formulė:

$$\hat{G} = \frac{2 \sum_{i \in u} y_i \omega_i W_i - \sum_{i \in u} y_i \omega_i^2}{(\sum_{i \in u} \omega_i) (\sum_{i \in u} \omega_i y_i)} - 1,$$

čia y_i yra sąlyginės namų ūkio pajamos, apskaičiuotos namų ūkio pajamas padalinus iš sąlyginio namų ūkio narių skaičiaus. Pastarasis apskaičiuojamas taip: pirmam namų ūkio nariui priskiriamas koeficientas 1, antram namų ūkio nariui priskiriamas koeficientas 0,7, likusiems namų ūkio nariams priskiriami koeficientai 0,5. Sąlyginis namų ūkio narių skaičius lygus šių koeficientų sumai.

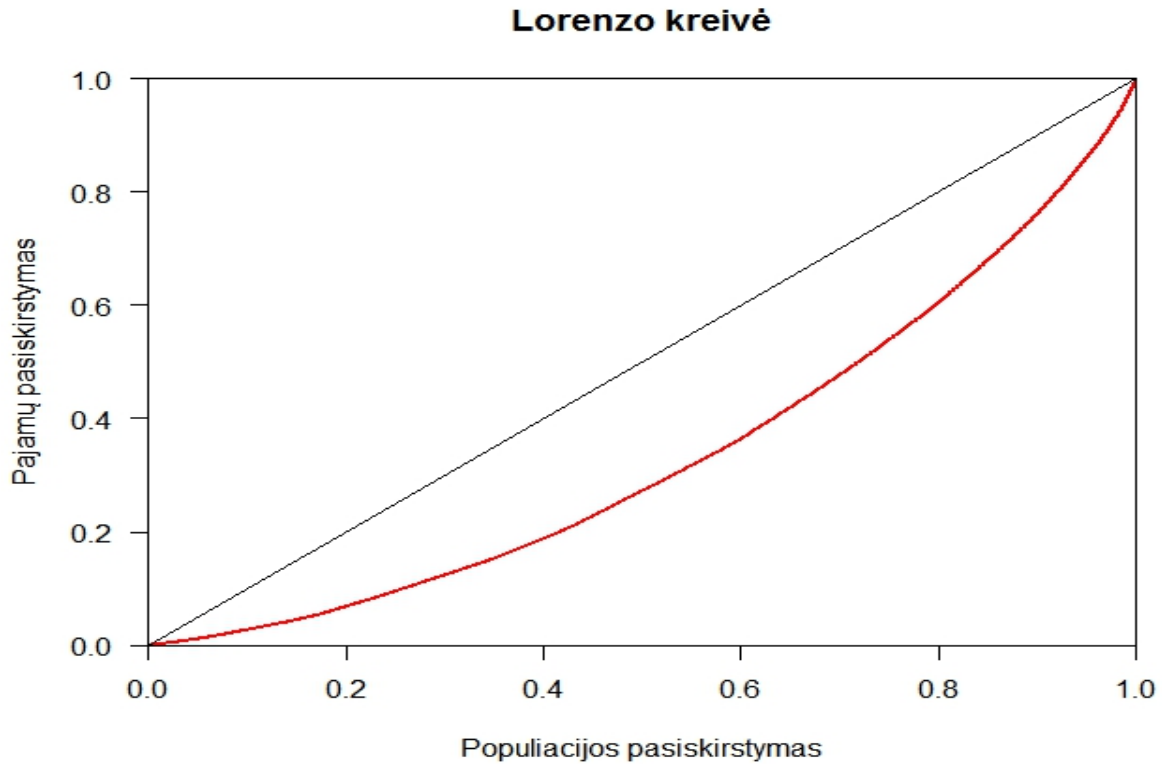
Imties plano svoriai ω_i apskaičiuojami kiekvienam namų ūkiui atskirai apskrities, kuriai priklauso elementas, dydį padalinus iš atitinkamo imties sluoksnio dydžio.

Taigi atlikę visus žingsnius, gauname, kad:

$$\hat{G} = 0,332$$

Pagal paskutinius Statistikos departamento duomenis [16] 2012 metais Lietuvoje Gini koeficientas buvo 0,326, o pagal Eurostat duomenis [4] - 0,343, taigi gauname pakankamai tikslų įvertinį iš ESS duomenų.

Nagrinėjamų duomenų Lorenzo kreivė:



4 pav. Lorenzo kreivė ESS duomenims

3.1.4. Dispersijos vertinimas visrakčio metodu

Särndalas, Swenssonas ir Wretmanas savo knygoje [14] apžvelgia įvertinio $\hat{\theta}$ dispersijos $\mathbf{D}\hat{\theta}$ vertinimą visrakčio metodu.

Kuomet imtis yra paprastoji atsitiktinė, tuomet a -oji pseudoreikšmė:

$$\hat{\theta}_a = n\hat{\theta} - (n-1)\hat{\theta}_{(a)},$$

čia $\hat{\theta}_{(a)}$ – įvertis, gaunamas iš $\hat{\theta}$, pašalinus iš imties a -ąjį elementą.

Tada alternatyvus visrakčio įvertinys:

$$\hat{\theta}_{vis} = \frac{1}{n} \sum_{a=1}^n \hat{\theta}_a.$$

Pseudoreikšmių $\hat{\theta}_a$, $a = 1, \dots, n$, imties dispersija

$$\hat{s}_{vis}^2 = \frac{1}{n-1} \sum_{a=1}^n \left(\hat{\theta}_a - \hat{\theta}_{vis} \right)^2.$$

Tuomet tiek įvertinio $\hat{\theta}_{vis}$, tiek $\hat{\theta}$ dispersijos įvertiniu laikysime

$$\hat{\mathbf{D}}\hat{\theta} = \left(1 - \frac{\lambda n}{N} \right) \frac{\hat{s}_{vis}^2}{n},$$

čia λ apibrėžiamas:

$$\lambda = \begin{cases} 0, & \text{jei imtis gražintinė} \\ 1, & \text{jei imtis negražintinė} \end{cases}$$

Sluoksninės imties atveju įvertinio $\hat{\theta}$ dispersijos $\hat{\mathbf{D}}\hat{\theta}$ įvertinys yra:

$$\hat{\mathbf{D}}\hat{\theta} = \sum_{h=1}^H \left(1 - \frac{\lambda_h n_h}{N_h} \right) \frac{n_h - 1}{n_h} \sum_{a=1}^{n_h} \left(\hat{\theta}_{(ha)} - \hat{\theta} \right)^2,$$

čia $\hat{\theta}_{(ha)}$ – įvertis, gaunamas iš $\hat{\theta}$, pašalinus iš imties h -ojo sluoksnio a -ąjį elementą.

Naudojantis *R* programa Gini koeficiento įvertinio dispersijos visrakčio įvertinį apskaičiuojame pagal nurodytas formules, t.y. iš imties u pašaliname vieną elementą ir likusiems perskaičiuojame imties plano svorius. Naujai imčiai su perskaičiuotais svoriais skaičiuojame Gini koeficiento įvertinį. Kartojame žingsnius n kartų ir gauname n skirtingų Gini koeficiento įverčių. Įrašome į $\hat{\mathbf{D}}\hat{\theta}$ formulę ir gauname

$$\hat{\mathbf{D}}\hat{G} = 4,65 \cdot 10^{-5}.$$

Variacijos koeficiento įvertis šiuo atveju lygus

$$\hat{cv}(\hat{G}) = \frac{\sqrt{\hat{\mathbf{D}}\hat{G}}}{\hat{G}} = 0,02.$$

3.1.5. Palyginimas su kitomis šalimis

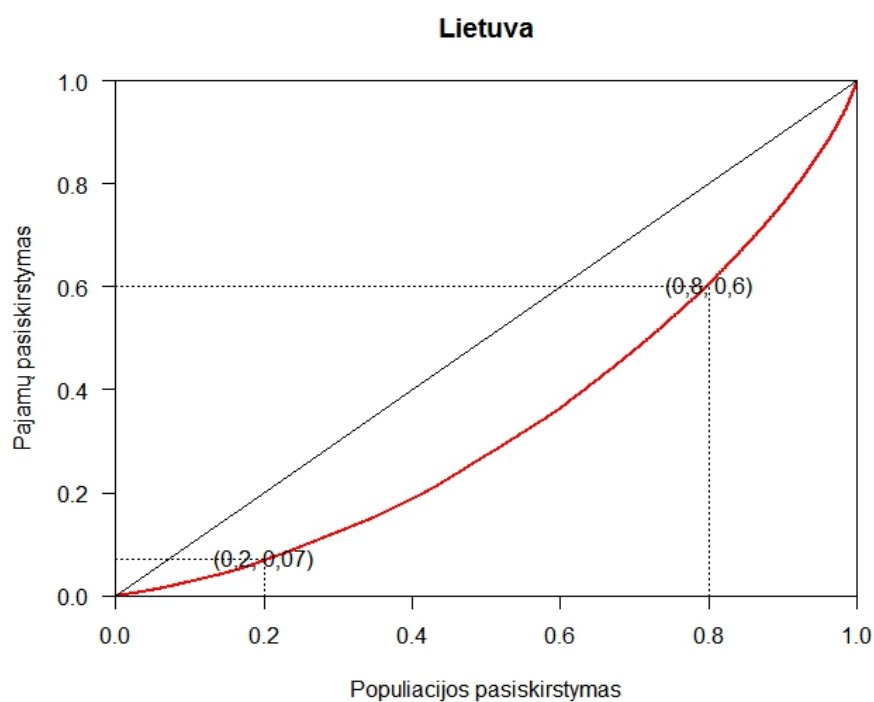
Norint palyginti Lietuvos rezultatus su kitomis šalimis, iš ESS svetainės gauti tiriamųjų šalių duomenys apie jų gyventojų pajamas.

Visoms šalims paskaičiuoti svoriai (PAI atveju jie visi lygūs $\frac{N}{n}$), sąlyginės namų pajamos. Tuomet naudojantis anksčiau pateiktomis formulėmis, kiekvienai šaliai įvertinti Gini koeficientai bei variacijos koeficientai. Rezultatai kartu su oficialiosios statistikos duomenimis [4] pateikiami 2 lentelėje:

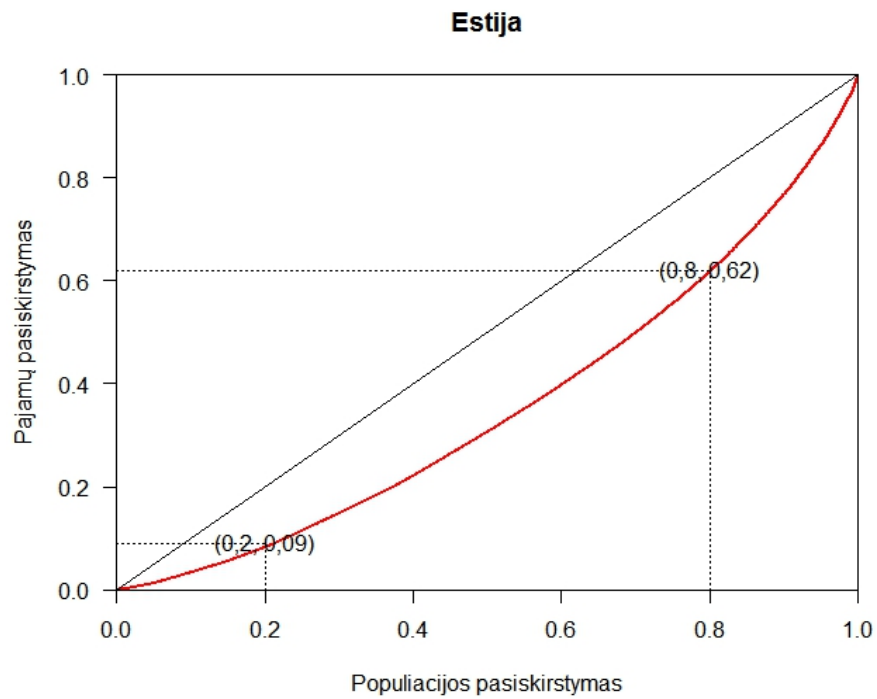
2 lentelė. Gini koeficiento įvertinių palyginimas su oficialiaja statistika

Šalis	$\widehat{G}$	Oficialioji statistika	$\widehat{cv}(\widehat{G})$	Imties dydis	Populiacijos dydis (mln.)
Lietuva	0,332	0,343	0,021	1317	2,5
Estija	0,288	0,326	0,019	1970	1,3
Suomija	0,264	0,252	0,018	2061	4,1
Švedija	0,289	0,245	0,016	1675	7,2

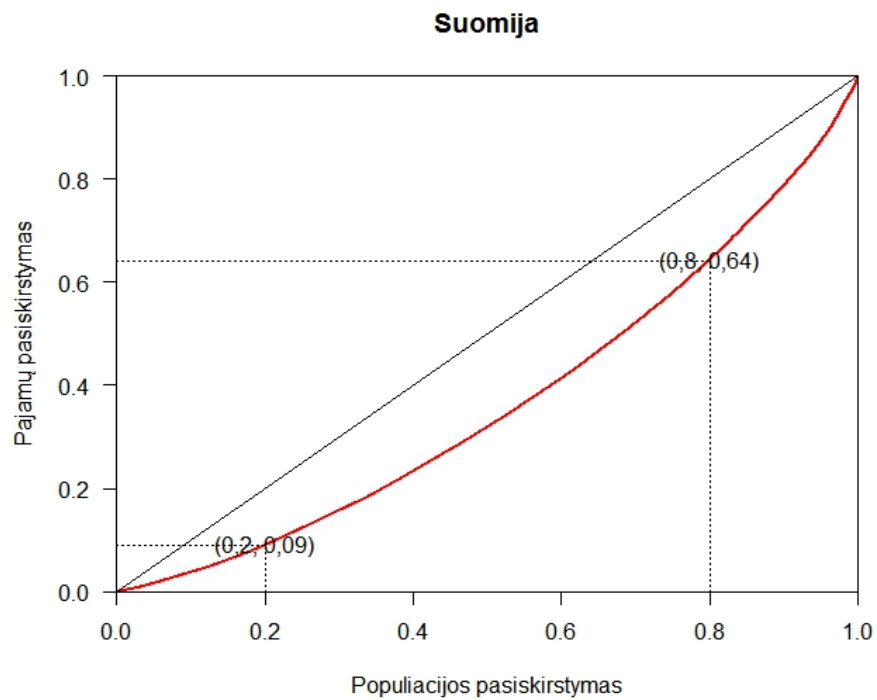
Taip pat pateikiamos šių šalių Lorenzo kreivės su išskirtomis reikšmėmis, kai populiacijos dalis yra 20 % ir 80 %.



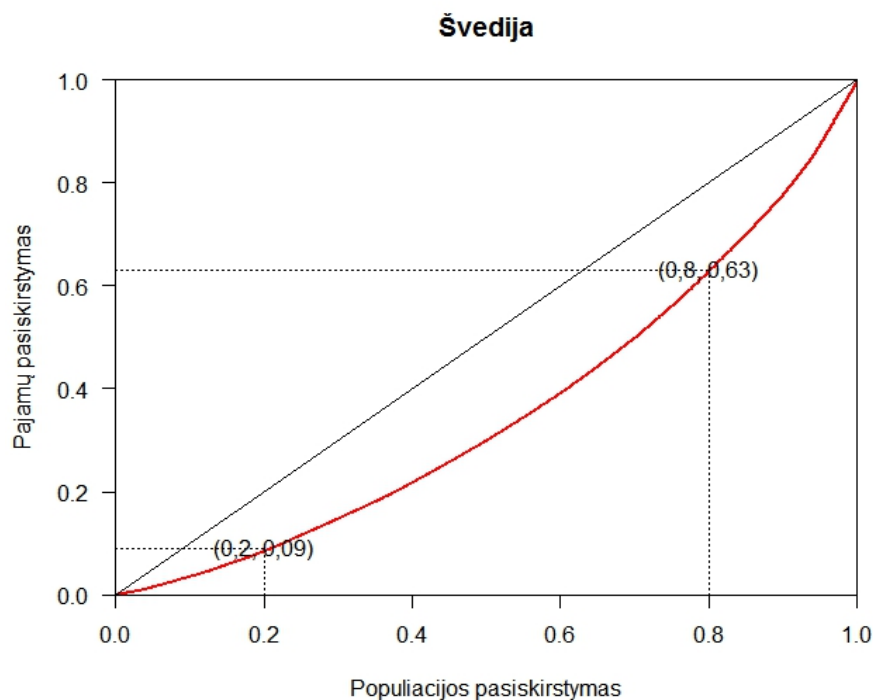
5 pav. Lietuvos Lorenzo kreivė ESS duomenims



6 pav. Estijos Lorenzo kreivė ESS duomenimis



7 pav. Suomijos Lorenzo kreivė ESS duomenimis



8 pav. Švedijos Lorenzo kreivė ESS duomenims

Iš 2 lentelės matome, kad nors koeficientų įvertiniai nedaug skiriasi vienas nuo kito, visgi didžiausia pajamų nelygybė yra Lietuvoje, tada Švedijoje, Estijoje ir mažiausia – Suomijoje.

Kadangi visų jų Gini koeficientai panašūs, kreivės taip pat nedaug skiriasi. Lietuvos atveju plotas tarp idealaus pajamų pasiskirstymo linijos ir Lorenzo kreivės didžiausias, todėl ir Gini koeficientas didžiausias.

Galima matyti, kad visų šalių 20 % gyventojų gauna apytiksliai tokią pat dalį šalies pajamų, išskyrus Lietuvą, ten tai pačiai daliai individų tenkančių pajamų dalis yra nežymiai mažesnė nei kitose tiriamose šalyse. 80 % šalies gyventojų didžiausia pajamų dalis gaunama Suomijoje, o mažiausia – Lietuvoje.

Taip pat galima pastebėti, kad mažesni variacijos koeficientai gaunami PAI atvejais, tačiau ne visada yra galimybė naudoti būtent šį imties planą, nes tam reikia turėti visų populiacijos elementų sąrašą. Beje tikslumui įtaką daro ne tik imties planas, bet ir duomenų surinkimas bei tai, kad ESS duomenys apie pajamas pateikiami intervalais.

3.2. Apibendrintieji entropijos matai ir Theilo indeksai

Kaip jau buvo minėta 2 skyriuje, bendroji apibendrintųjų entropijos matų formulė yra:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\mu} \right)^\alpha \right],$$

čia μ – populiacijos pajamų y vidurkis. Parametrą $\alpha \in \mathbb{R}$ nustatome patys.

Ši matų grupė išvesta iš informacijos teorijos, konkrečiai iš pertekliaus, daugžodžiavimo teorijos (angl. redundancy theory).

Išskiriami du atskiri Entropijos matai – Theilo T ir Theilo L indeksai. Abu jie matuoja, kaip toli populiacija yra „nutolusi“ nuo idealios pusiausvyros, kai visi asmenys turi vienodas pajamas.

Theilo L indeksas aprašomas formule:

$$GE(0) = \frac{1}{N} \sum_{i=1}^N \log \left(\frac{\mu}{y_i} \right).$$

Dar vienas entropijos variantas – Theilo T indeksas:

$$GE(1) = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\mu} \log \left(\frac{y_i}{\mu} \right).$$

Kadangi skaičiavimuose yra populiacijos pajamų vidurkio bei atskiro elemento pajamų santykio logaritmas, indeksų įverčiai bus paslinktieji, reikės vertinti poslinkio reikšmingumą.

3.2.1. Theilo indeksų įverčiai

Pradėkime nuo Theilo L indekso vertinimo. Atliekame pertvarkymus:

$$\begin{aligned} GE(0) &= \frac{1}{N} \sum_{i=1}^N \log \left(\frac{\mu}{y_i} \right) = \frac{1}{N} \sum_{i=1}^N (\log \mu - \log y_i) \\ &= \frac{1}{N} \left(\sum_{i=1}^N \log \mu - \sum_{i=1}^N \log y_i \right) = \log \mu - \frac{1}{N} \sum_{i=1}^N \log y_i. \end{aligned}$$

Pažymėkime:

$$z_i := \log y_i$$

Tada galime užrašyti indekso $GE(0)$ įvertį paprastosios atsitiktinės imties atveju:

$$\widehat{GE}(0) = \log \hat{\mu} - \hat{\mu}_z. \quad (12)$$

Įvertis yra nepaslinktasis, kai $\mathbf{E}(\widehat{GE}(0)) = GE(0)$.

Patikrinsime įverčio poslinkio reikšmingumą, laikydami mūsų turimą imtį u populiacija U^* ir rinkdami iš jos paprastasias atsitiktines imtis u^* .

Dabar populiacijos dydis $N^* = n = 1317$, o renkamų imčių dydžiai $n^* = 300$.

Tada Theilo L indeksas populiacijai U^* yra lygus:

$$GE(0)^* = 0,1879956.$$

Išrinkę $J = 1000$ skirtingų paprastųjų atsitiktinių imčių u^* iš populiacijos U^* ir kiekvienai paskaičiavę $\widehat{GE}(0)^{(j)}$, gauname jų empirinį vidurkį:

$$\widehat{\widehat{GE}}(0)^* = 0,1872651.$$

Matome, kad empirinis vidurkis apytiksliai lygus populiacijos Theilo L indeksui, tad galime daryti prielaidą, kad PAI įvertis yra apytiksliai nepaslinktasis.

Vėlgi pažymėkime:

$$z_i := \log y_i.$$

Atliekame pertvarkymus

$$\begin{aligned} \widehat{GE}_{sl}(0) &= \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} \log \left(\frac{\hat{\mu}_{sl}}{y_i} \right) = \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} (\log \hat{\mu}_{sl} - \log y_i) \\ &= \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} \log \hat{\mu}_{sl} - \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} \log y_i \\ &= \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \log \hat{\mu}_{sl} \cdot n_h - \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} z_i. \end{aligned}$$

Sluoksnei imčiai Theilo L indekso įvertį galima užrašyti:

$$\widehat{GE}_{sl}(0) = \log \hat{\mu}_{sl} - \hat{\mu}_{zsl}. \quad (13)$$

Theilo T indekso formulę galima pertvarkyti:

$$\begin{aligned} GE(1) &= \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\mu} \log \left(\frac{y_i}{\mu} \right) = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\mu} (\log y_i - \log \mu) \\ &= \frac{1}{N} \frac{1}{\mu} \sum_{i=1}^N y_i \log y_i - \frac{1}{N} \frac{\log \mu}{\mu} \sum_{i=1}^N y_i \\ &= \frac{1}{N\mu} \sum_{i=1}^N y_i \log y_i - \log \mu. \end{aligned}$$

Pažymėkime:

$$a_i := y_i \log y_i.$$

Tada paprastosios atsitiktinės imties atveju taikome tokį Theilo T indekso įvertinį:

$$\widehat{GE}(1) = \frac{\widehat{\mu}_a}{\widehat{\mu}} - \log \widehat{\mu}. \quad (14)$$

Analogiškai Theilo L indekso atvejui laikydami mūsų imtį nauja populiacija gauname Theilo T indeksą ir empirinį įverčių vidurkį:

$$GE(1)^* = 0,1769364$$

$$\overline{\widehat{GE}}(1)^* = 0,1774264$$

Įvertis apytiksliai nepaslinktasis, o sluoksninės imties atveju jo forma:

$$\widehat{GE}_{sl}(1) = \frac{1}{N\widehat{\mu}_{sl}} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in i_h} y_i \log y_i - \log(\widehat{\mu}_{sl}).$$

Pažymėjus $a_i := y_i \log y_i$, galutinė Theilo T indekso įverčio forma yra:

$$\widehat{GE}_{sl}(1) = \frac{\widehat{\mu}_{a\,sl}}{\widehat{\mu}_{sl}} - \log(\widehat{\mu}_{sl}). \quad (15)$$

Ji panaši į paprastosios atsitiktinės imties įvertį, tik šiuo atveju vietoj PAI vidurkių skaičiuojame sluoksninės imties vidurkius.

3.2.2. Theilo indeksų įvertiniai ESS duomenims

Naudojami tie patys duomenys bei pažymėjimai kaip ir Gini koeficiento įvertinio skaičiavimo atveju.

$$\widehat{GE}_{sl}(0) = 0,192,$$

$$\widehat{GE}_{sl}(1) = 0,181.$$

Kai imtis yra paprastoji atsitiktinė, skaičiuojant abu indeksus su *R* paketu *ineq* gaunamos tokios pačios reikšmės kaip ir pagal išvestas įvertinių formules, todėl galima manyti, kad gavome teisingus įvertinius.

3.2.3. Dispersijos vertinimas

Visrakčio metodu įvertinę Theilo indeksų įvertinių dispersijas gauname

$$\widehat{DGE}_{sl}(0) = 5,9 \cdot 10^{-5},$$

$$\widehat{DGE}_{sl}(1) = 6,8 \cdot 10^{-5}.$$

Variacijos koeficientų įverčiai:

$$\widehat{cv}(\widehat{GE}_{sl}(0)) = 0,040,$$

$$\widehat{cv}(\widehat{GE}_{sl}(1)) = 0,046.$$

3.2.4. Palyginimas su kitomis šalimis

Kitoms šalims, atsižvelgiant į imties planą, apskaičiuoti Theilo L ir Theilo T indeksų įverčiai, taip pat variacijos koeficientų įverčiai ir rezultatai kartu su oficialiosios statistikos duomenimis [4] pateikiami 3 ir 4 lentelėse.

3 lentelė. Theilo L koeficiento įvertinių palyginimas su oficialiąja statistika

Šalis	$\widehat{GE}_{sl}(0)$	Oficialioji statistika	$\widehat{cv}(\widehat{GE}_{sl}(0))$	Imties dydis	Populiacijos dydis (mln.)
Lietuva	0,192	0,208	0,040	1317	2.5
Estija	0,147	0,188	0,038	1970	1.3
Suomija	0,118	0,108	0,036	2061	4.1
Švedija	0,140	0,114	0,034	1675	7.2

4 lentelė. Theilo T koeficiento įvertinių palyginimas su oficialiąja statistika

Šalis	$\widehat{GE}_{sl}(1)$	Oficialioji statistika	$\widehat{cv}(\widehat{GE}_{sl}(1))$	Imties dydis	Populiacijos dydis (mln.)
Lietuva	0,181	0,205	0,046	1317	2.5
Estija	0,142	0,179	0,042	1970	1.3
Suomija	0,117	0,113	0,040	2061	4.1
Švedija	0,136	0,105	0,035	1675	7.2

Iš gautų rezultatų matome, kad didžiausia pajamų nelygybė abiem atvejais stebima Lietuvoje (ten koeficientų įvertinių reikšmės yra didžiausios), tada Estijoje, Švedijoje, o mažiausia – Suomijoje. Pagal oficialiąją statistiką, vienu atveju mažiausia nelygybė yra Suomijoje, kitu – Švedijoje.

Galima pastebėti, kad Lietuvos ir Estijos rodiklių įvertiniai yra šiek tiek mažesni nei oficialiosios statistikos duomenys, Suomijos rodiklių įvertiniai pakankamai tikslūs, o Švedijos – didesni už Eurostat duomenis. Mažesnius variacijos koeficientų įverčius gauname paprastosios atsitiktinės imties atveju, t.y. Suomijos ir Švedijos.

3.3. Atkinsono koeficientas

Priklausomai nuo parametro ε , Atkinsono koeficientų A_ε klasė apibrėžiama:

$$A_\varepsilon = 1 - \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\mu} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}, \quad \varepsilon \neq 1$$

$$A_\varepsilon = 1 - \frac{\prod_{i=1}^N \left(y_i^{\frac{1}{N}} \right)}{\mu}, \quad \varepsilon = 1.$$

Atkinsono koeficiento įverčiai bus paslinktieji dėl kintamųjų kėlimo laipsniu.

3.3.1. Atkinsono koeficiento įverčiai

Pažymėkime:

$$x_i := y_i^{1-\varepsilon}.$$

Atkinsono koeficiento paprastosios atsitiktinės imties įvertį, kai $\varepsilon \neq 1$ apibrėžiame:

$$\hat{A}_\varepsilon = 1 - \left[\frac{1}{n} \sum_{k=1}^n \left(\frac{y_k}{\hat{\mu}} \right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} = 1 - \frac{\hat{\mu}_x^{\frac{1}{1-\varepsilon}}}{\hat{\mu}}. \quad (16)$$

Pasirenkame $\varepsilon = 0,5$ ir vertiname poslinkį, apskaičiavę Atkinsono koeficiento įverčių empirinį vidurkį:

$$A_{0,5}^* = 0,0871352,$$

$$\widehat{A}_{0,5}^* = 0,0866724.$$

Įvertis apytiksliai nepaslinktasis.

Kai

$$x_i := y_i^{1-\varepsilon},$$

gauname

$$\begin{aligned} \hat{A}_{\varepsilon,sl} &= 1 - \left(\sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \frac{1}{\hat{\mu}_{sl}^{1-\varepsilon}} \sum_{i \in i_h} x_i \right)^{\frac{1}{1-\varepsilon}} = 1 - \left(\frac{1}{\hat{\mu}_{sl}^{1-\varepsilon}} \sum_{h=1}^H \frac{N_h}{N} \frac{1}{n_h} \sum_{i \in i_h} x_i \right)^{\frac{1}{1-\varepsilon}} \\ &= 1 - \left(\frac{1}{\hat{\mu}_{sl}^{1-\varepsilon}} \right)^{\frac{1}{1-\varepsilon}} (\hat{\mu}_{xsl})^{\frac{1}{1-\varepsilon}}. \end{aligned}$$

Tada sluoksninės imties atveju indekso forma:

$$\hat{A}_{\varepsilon,sl} = 1 - \frac{\hat{\mu}_{xsl}^{\frac{1}{1-\varepsilon}}}{\hat{\mu}_{sl}}. \quad (17)$$

Paprastosios atsitiktinės imties atveju Atkinsono koeficiento įvertis, kai $\varepsilon = 1$:

$$\widehat{A}_1 = 1 - \frac{\prod_{i=1}^n \left(y_i^{\frac{1}{n}}\right)}{\bar{y}}.$$

Apibrėžkime kintamąjį x su reikšmėmis $x_i = \log y_i$, kur $i = 1, 2, \dots, N$. Tada $y_i = e^{x_i}$ ir

$$\prod_{i=1}^N y_i^{\frac{1}{N}} = \prod_{i=1}^N e^{\frac{x_i}{N}} = \exp \left\{ \sum_{i=1}^N \frac{x_i}{N} \right\} = e^{\mu_x}.$$

Iš čia gauname, kad

$$\widehat{A}_1 = 1 - \frac{e^{\widehat{\mu}_x}}{\widehat{\mu}}. \quad (18)$$

Analogiškai vertiname poslinkį:

$$\begin{aligned} A_1^* &= 0,1713817, \\ \widehat{A}_1^* &= 0,1724119. \end{aligned}$$

Įvertis apytiksliai nepaslinktasis.

Atkinsono koeficiento įvertis, kai $\varepsilon = 1$ ir turime sluoksнинę imtį,

$$\widehat{A}_{1sl} = 1 - \frac{e^{\widehat{\mu}_{xsl}}}{\widehat{\mu}_{sl}} \quad (19)$$

Taigi forma panaši į PAI iverčio, tik vidurkiai yra sluoksнинės imties

3.3.2. Atkinsono koeficiento įvertiniai ESS duomenims

Naudojami tie patys duomenys bei pažymėjimai, kaip ir Gini koeficiento įvertinio skaičiavimo atveju.

$$\widehat{A}_{0,5sl} = 0,090,$$

$$\widehat{A}_{1sl} = 0,175.$$

PAI atveju buvo gautos panašios reikšmės, todėl galima manyti, kad gavome teisingus įvertinius.

3.3.3. Dispersijos vertinimas

Visrakčio metodu įvertinę Atkinsono indekso įvertinio dispersiją gauname

$$\widehat{\mathbf{D}}\widehat{A}_{0,5sl} = 1,3 \cdot 10^{-5}$$

$$\widehat{\mathbf{D}}\widehat{A}_{1sl} = 4,0 \cdot 10^{-5}$$

Variacijos koeficientų įverčiai:

$$\hat{c}v\left(\hat{A}_{0,5\,sl}\right) = 0,040,$$

$$\hat{c}v\left(\hat{A}_{1\,sl}\right) = 0,036.$$

3.3.4. Palyginimas su kitomis šalimis

Kitoms šalims, atsižvelgiant į imties planą, apskaičiuoti Atkinsono indeksų įverčiai, kai $\varepsilon = 0,5$ ir $\varepsilon = 1$, taip pat variacijos koeficientų įverčiai ir rezultatai kartu su oficialiosios statistikos duomenimis [4] pateikiami 5 ir 6 lentelėse.

5 lentelė. Atkinsono koeficiento ($\varepsilon = 0,5$) įvertinių palyginimas su oficialiąja statistika

Šalis	$\hat{A}_{0,5\,sl}$	Oficialioji statistika	$\hat{c}v(\hat{A}_{0,5\,sl})$	Imties dydis	Populiacijos dydis (mln.)
Lietuva	0,090	0,097	0,040	1317	2.5
Estija	0,068	0,087	0,038	1970	1.3
Suomija	0,057	0,053	0,036	2061	4.1
Švedija	0,067	0,052	0,033	1675	7.2

6 lentelė. Atkinsono koeficiento ($\varepsilon = 1$) įvertinių palyginimas su oficialiąja statistika

Šalis	$\hat{A}_{1\,sl}$	Oficialioji statistika	$\hat{c}v(\hat{A}_{1\,sl})$	Imties dydis	Populiacijos dydis (mln.)
Lietuva	0,175	0,187	0,036	1317	2.5
Estija	0,133	0,171	0,035	1970	1.3
Suomija	0,111	0,102	0,034	2061	4.1
Švedija	0,130	0,107	0,032	1675	7.2

Pagal gautus rezultatus, didžiausia pajamų nelygybė stebima Lietuvoje (ten koeficiento įvertinio reikšmė yra didžiausia), tada Estijoje, Švedijoje, o mažiausia – Suomijoje. Pagal oficialiąją statistiką, vienu atveju mažiausia nelygybė yra Švedijoje (nors skirtumas nuo Suomijos nedidelis), kitu – Suomijoje.

Matome, kad vėlgi Lietuvos ir Estijos rodiklių įvertiniai yra mažesni nei oficialiosios statistikos duomenys, Suomijos rodiklių įvertiniai pakankamai tikslūs, o Švedijos - didesni už Eurostat duomenis. Variacijos koeficientų įverčiai vėlgi mažesni paprastosios atsitiktinės imties atveju, t.y. Suomijos ir Švedijos.

Apibendrinimas

- Šiame darbe buvo apžvelgti pagrindiniai pajamų nelygybės rodikliai: Gini koeficientas kartu su Lorenzo kreive, apibendrintieji entropijos matai ir Theilo indeksai, Atkinsono koeficientas, trumpai aptartos jų savybės.
- Pristatytas Europos socialinis tyrimas ir jo metu gauti duomenys, pateikta informacija apie ESS Lietuvoje, jo temas, tyrimo metodus. Skyriaus pabaigoje apžvelgti tiriamųjų šalių (Lietuvos, Estijos, Suomijos ir Švedijos) imčių planai.
- Prieš vertinant pajamų nelygybės rodiklius, buvo paruošti skaičiavimams iš ESS internetinio puslapio [5] gauti tiriamųjų šalių duomenys – palikti duomenys be praleistų reikšmių, pajamų intervalus žyminčios raidės pakeistos į tikslias reikšmes, apskaičiuoti sąlyginiai namų ūkių narių skaičiai bei sąlyginės namų ūkio nario pajamos. Taip pat iš Statistikos departamento internetinio [16] puslapio gauta informacija apie gyventojų skaičių kiekvienoje apskrityje atsižvelgiant į ESS duomenų amžiaus ribas. Apskaičiuoti imties planų svoriai.
- Pristatyti Gini koeficiento algebrinis ir geometrinis išvedimai, užrašyta Gini koeficiento įverčio forma, tinkanti tiek paprastajai atsitiktinei imčiai, tiek sluoksninei imčiai. Pagal šią formulę, naudojantis R programa, apskaičiuotas Gini koeficiento įvertinys tiriamosioms šalims. Gini koeficiento įvertinio dispersija vertinta visrakčio būdu. Šiame skyriuje taip pat trumpai aptartas visrakčio taikymas turint paprastąją atsitiktinę imtį bei sluoksninę imtį.
- 2015 m. rugpjūtį pranešimas apie Gini koeficiento vertinimą Baltijos ir Skandinavijos šalyse pristatytas BaNoCoss konferencijoje [8], vykusioje Helsinkyje, Suomijoje.
- Supaprastintos pagrindinių apibendrintosios entropijos matų – Theilo L ir Theilo T indeksų formos, abiems išvesti įvertiniai paprastosios atsitiktinės imties ir sluoksninės imties atvejams. Patikrintas šių įvertinių poslinkio reikšmingumas. Įsitikinus, kad įvertiniai apytiksliai nepaslinktieji, apskaičiuoti įvertiniai tiriamosioms šalims. Dispersija, kaip ir Gini koeficiento atveju, vertinta visrakčio būdu.
- Analogiškai Theilo indeksams, supaprastinta Atkinsono koeficiento forma, priklausanti nuo parametro ε , išvesti įvertiniai paprastosios atsitiktinės imties ir sluoksninės imties atvejams. Patikrintas šių įvertinių poslinkio reikšmingumas. Įsitikinus, kad įvertiniai apytiksliai nepaslinktieji, apskaičiuoti įvertiniai tiriamosioms šalims. Dispersija vertinta visrakčio būdu.
- Kiekvienas aptartas pajamų nelygybės rodiklis apskaičiuotas Lietuvai, Estijai, Suomijai ir Švedijai, atsižvelgiant į šių šalių imčių planus. Rezultatai pristatomi lentelėse kartu su Eurostato informacija [4], variacijos koeficientų įverčiais, imties ir populiacijos dydžiais.

Pagal pasirinktų šalių duomenis nubrėžtos Lorenzo kreivės. Gauta, kad iš Europos socialinio tyrimo duomenų gautieji įvertiniai nors ir šiek tiek skiriasi nuo oficialiosios statistikos pateikiamų duomenų, tačiau yra pakankamai tikslūs.

- Įvairiuose Europos šalių reitinguose Suomija iš Švedija užima gana aukštas vietas kaip vienos iš geresnių šalių, pvz. pagal BVP Švedija yra geriausiųjų dešimtuose, Suomija – 16. Pagal vidutinį darbo užmokestį abi šalys yra pirmame dešimtuose. Pagal World Happiness Report 2010-2012 metų duomenis [10] Švedija yra 5 geriausia šalis, Suomija – 6. Šiame tyrime jos buvo laikytos etalonais, tačiau įdomu tai, kad gauti rezultatai ne itin daug skyrėsi nuo Lietuvos ir Estijos.
- Taip pat visuomet buvo įdomu lyginti Lietuvą su Estija. BVP didesnis Lietuvoje, tačiau nedarbo lygis, minimalus ir vidutinis darbo užmokestis – Estijoje, o pagal World Happiness Report Lietuva yra 29 vietoje, Estija – 30. Galima daryti išvadą, kad nuo kaimynų atsilieka, bet visgi nedaug. Pajamų nelygė taip pat Estijoje mažesnė.

Išvados

- Lietuvoje pagal aprašą, esantį Europos socialinio tyrimo internetiniame puslapyje, buvo renkama daugiapakopė imtis, tačiau šiame darbe ji buvo traktuota kaip paprastoji sluoksninė imtis. Iš duomenų neįmanoma atkurti, kaip buvo renkamos kitos pakopos (nėra nurodytų seniūnijų, adresų grupių), kas yra būtina dispersijos vertinimui. Su kitų šalių duomenimis tokių problemų nekilo.
- Lyginant pagal visus įvertintus pajamų nelygės rodiklius gavome, kad Lietuvoje pajamų nelygė yra didžiausia iš šiame darbe tirtų šalių. Antroje vietoje visais atvejais, išskyrus Gini koeficientą ESS duomenims, yra Estija. Pagal iš ESS duomenų išskaičiuotą Gini koeficiento įvertinį Estija būtų reitinguojama trečioje vietoje.
- Suomijos ir Švedijos rodikliai reitinguojami taip: pagal iš ESS duomenų paskaičiuotus pajamų nelygės rodiklius įvertinius visais atvejais mažiausia pajamų nelygė yra Suomijoje, o Švedija yra trečioje vietoje visais atvejais, išskyrus pagal Gini koeficientą. Pagal oficialiosios statistikos duomenis šių šalių reitingavimas skiriasi – Suomija dviem atvejais iš penkių užima ketvirtąją vietą, o likusiais – trečią. Švedija atvirkščiai – dviem atvejais trečią ir likusiais ketvirtąją vietą.

Priedai

```
#duomenų pertvarkymas
df = read.table("C:/Users/Kristina/Desktop/MAGISTRINIS/duom/R.txt")
library(laeken)
require(ineq)

n <- 1317
N <- 2535329
n_h<-c(96,295,111,64,190,119,56,75,43,268)
N_h<-c(131815,505915,280652,132498,208974,248649,90477,
124468,127928,683953)

#priskiriame ekvivalentinį NŪ dydį kiekvienam stebiniui
for (i in 1:n){
  if(df[i,1] == 0) df[i,6] <- 0
  if(df[i,1] == 1) df[i,6] <- 1
  if(df[i,1] == 2) df[i,6] <- 1.7
  if(df[i,1] >= 3) df[i,6] <- 1.7+(df[i,1]-2)*0.5
}

#pagal užkoduotas reikšmes priskiriame NŪ pajamas
#(pasirinkti intervalų vidurio taškai, o kraštinės reikšmės - laisvai)
for (i in 1:n){
  if(df[i,2] == 1) df[i,7] <- 500
  if(df[i,2] == 2) df[i,7] <- 775
  if(df[i,2] == 3) df[i,7] <- 1025
  if(df[i,2] == 4) df[i,7] <- 1350
  if(df[i,2] == 5) df[i,7] <- 1650
  if(df[i,2] == 6) df[i,7] <- 1900
  if(df[i,2] == 7) df[i,7] <- 2150
  if(df[i,2] == 8) df[i,7] <- 2550
  if(df[i,2] == 9) df[i,7] <- 2900
  if(df[i,2] == 10) df[i,7] <- 5000
}

#ekvivalentinės vieno asmens pajamos
for (i in 1:n){
  df[i,8] <- df[i,7]/df[i,6]
```

```

}

#imties plano svoriai
for (i in 1:n){
  df[i,9] <- df[i,4]/df[i,5]
}

vid_sl <- function(d,K){
  rezultatas<-1/K*sum(d[,8]*d[,9])
  return(rezultatas)
}

#Gini koeficientą apskaičiuojanti funkcija
Gini <- function(d){
  duom<-arrange(d,d[,8])
  wi<-duom[,9]
  yi<-duom[,8]
  Wi<-cumsum(wi)
  rezultatas<-100*((2*sum(Wi*wi*yi)-sum((wi^2)*yi))
  /(sum(wi)*sum(wi*yi))-1)
  return(rezultatas)
}

#Lorenzo kreivė
require(pracma)

Yi<-yi*wi
Q_k<-cumsum(Yi)/sum(Yi)
P_k<-cumsum(wi)/N
plot(P_k, Q_k, type="l", main="Lorenz curve",
xlab="Proportion of the population",
ylab="Proportion of the total income")
lines(P_k, P_k, col="red")

#visraktis sluoksn. GINI
duom<-arrange(df,df[,8])
duom_i<-0
g_a<-0
H<-c(0,0,0,0,0,0,0,0,0,0,0)
G<-Gini/100

```

```

g_a<- Gini(duom)

for (i in 1:n){
  duom_i<-duom[-i,]
  for (j in 1:(n-1)){
    if (duom_i[j,3]==duom[i,3]) duom_i[,9]<-duom_i[,4]/(duom_i[,5]-1)
  }
  g_a<- Gini(duom_i)
  if (duom[i,3]=="LT0013") H[1]<-H[1]+(g_a-G)^2
  if (duom[i,3]=="LT0023") H[2]<-H[2]+(g_a-G)^2
  if (duom[i,3]=="LT0033") H[3]<-H[3]+(g_a-G)^2
  if (duom[i,3]=="LT0043") H[4]<-H[4]+(g_a-G)^2
  if (duom[i,3]=="LT0053") H[5]<-H[5]+(g_a-G)^2
  if (duom[i,3]=="LT0063") H[6]<-H[6]+(g_a-G)^2
  if (duom[i,3]=="LT0073") H[7]<-H[7]+(g_a-G)^2
  if (duom[i,3]=="LT0083") H[8]<-H[8]+(g_a-G)^2
  if (duom[i,3]=="LT0093") H[9]<-H[9]+(g_a-G)^2
  if (duom[i,3]=="LT00A3") H[10]<-H[10]+(g_a-G)^2
  duom_i<-0
}

D_vis <- sum((1-n_h/N_h)*(n_h-1)/n_h*H)
cv <- sqrt(D_vis)/G

#THEIL L PAI
GE_0 <- function(d,K){
  z<-d
  z[,8]<-log(d[,8])
  rezultatas<-log(mean(d[,8]))-mean(z[,8])
  return(rezultatas)
}
GE_0(df,N)

#THEIL L sluoksninei
GE_0_sl <- function(d,K){
  z<-d
  z[,8]<-log(d[,8])
  rezultatas<-log(vid_sl(d,K))-vid_sl(z,K)
  return(rezultatas)
}

```

```

}
GE_0_sl(df,N)

#iš R paketo
ineq(df[,8], parameter=0, type="entropy")

#THEIL T PAI
GE_1 <- function(d,K){
  z<-d
  z[,8]<-d[,8]*log(d[,8])
  rezultatas<-mean(z[,8])/mean(d[,8])-log(mean(d[,8]))
  return(rezultatas)
}
GE_1(df,N)

#THEIL T sluoksninei
GE_1_sl <- function(d,K){
  z<-d
  z[,8]<-d[,8]*log(d[,8])
  rezultatas<-vid_sl(z,K)/vid_sl(d,K)-log(vid_sl(d,K))
  return(rezultatas)
}
GE_1_sl(df,N)

#iš paketo
ineq(df[,8], parameter=1, type="entropy")

#Paslinktumo vertinimas
paslinktasGE0 <- function(d){
  n0 <- 300
  kart <- 1000
  pop <- d[,8]
  GE0 <- data.frame(matrix(ncol = 1, nrow = kart))
  for (i in 1:kart) {
    imt=sample(pop,size = n0)
    GE0[i,1]<-log(mean(imt))-mean(log(imt))
  }
  rezultatas<-mean(GE0[,1])
  return(rezultatas)
}

```

```

paslinktasGE1 <- function(d){
  n0 <- 300
  kart <- 1000
  pop <- d[,8]
  GE1 <- data.frame(matrix(ncol = 1, nrow = kart))
  for (i in 1:kart) {
    imt=sample(pop,size = n0)
    GE1[i,1]<-mean(imt*log(imt))/mean(imt)-log(mean(imt))
  }
  rezultatas<-mean(GE1[,1])
  return(rezultatas)
}

paslinktasGE0(df)
paslinktasGE1(df)

#formulės populiacijos rodikliui paskaičiuoti
GE_0_pop<-log(mean(df[,8]))-1/n*sum(log(df[,8]))
GE_1_pop<-1/(n*mean(df[,8]))*sum(df[,8]*log(df[,8]))-log(mean(df[,8]))

#visr. GE_0_sl
GE_0_slu<-GE_0_sl(df,N)
GE_0_sl_a<-0
duom<-df
M<-N-1
H<-c(0,0,0,0,0,0,0,0,0,0,0)

for (i in 1:n){
  duom_i<-duom[-i,]
  for (j in 1:(n-1)){
    if (identical(duom_i[j,3],duom[i,3]))
      duom_i[j,9]<-duom_i[j,4]/(duom_i[j,5]-1)
  }
  GE_0_sl_a<-GE_0_sl(duom_i,M)
  if (duom[i,3]=="LT0013") H[1]<-H[1]+(GE_0_sl_a-GE_0_slu)^2
  if (duom[i,3]=="LT0023") H[2]<-H[2]+(GE_0_sl_a-GE_0_slu)^2
  if (duom[i,3]=="LT0033") H[3]<-H[3]+(GE_0_sl_a-GE_0_slu)^2
  if (duom[i,3]=="LT0043") H[4]<-H[4]+(GE_0_sl_a-GE_0_slu)^2
  if (duom[i,3]=="LT0053") H[5]<-H[5]+(GE_0_sl_a-GE_0_slu)^2

```

```

    if (duom[i,3]=="LT0063") H[6]<-H[6]+(GE_0_sl_a-GE_0_slu)^2
    if (duom[i,3]=="LT0073") H[7]<-H[7]+(GE_0_sl_a-GE_0_slu)^2
    if (duom[i,3]=="LT0083") H[8]<-H[8]+(GE_0_sl_a-GE_0_slu)^2
    if (duom[i,3]=="LT0093") H[9]<-H[9]+(GE_0_sl_a-GE_0_slu)^2
    if (duom[i,3]=="LT00A3") H[10]<-H[10]+(GE_0_sl_a-GE_0_slu)^2
    duom_i<-0
  }

D_vis <- sum((1-n_h/N_h)*(n_h-1)/n_h*H)
cv <- sqrt(D_vis)/GE_0_slu

#visr. GE_1_sl
GE_1_slu<-GE_1_sl(df,N)
GE_1_sl_a<-0
duom<-df
M<-N-1
H<-c(0,0,0,0,0,0,0,0,0,0)
for (i in 1:n){
  duom_i<-duom[-i,]
  for (j in 1:(n-1)){
    if (identical(duom_i[j,3],duom[i,3]))
      duom_i[j,9]<-duom_i[j,4]/(duom_i[j,5]-1)
  }
  GE_1_sl_a<-GE_1_sl(duom_i,M)
  if (duom[i,3]=="LT0013") H[1]<-H[1]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0023") H[2]<-H[2]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0033") H[3]<-H[3]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0043") H[4]<-H[4]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0053") H[5]<-H[5]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0063") H[6]<-H[6]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0073") H[7]<-H[7]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0083") H[8]<-H[8]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT0093") H[9]<-H[9]+(GE_1_sl_a-GE_1_slu)^2
  if (duom[i,3]=="LT00A3") H[10]<-H[10]+(GE_1_sl_a-GE_1_slu)^2
  duom_i<-0
}

D_vis <- sum((1-n_h/N_h)*(n_h-1)/n_h*H)
cv <- sqrt(D_vis)/GE_1_slu

```

```
#Atkinsono <>1 PAI
Atk <- function(d,eps,K){
  z<-d
  z[,8]<-d[,8]^(1-eps)
  rezultatas<-1-(mean(z[,8])^(1/(1-eps)))/mean(d[,8])
  return(rezultatas)
}
Atk(df,0.5,N)
```

```
#Atkinsono <>1 sluoksninei
Atk_sl <- function(d,eps,K){
  z<-d
  z[,8]<-d[,8]^(1-eps)
  rezultatas<-1-(vid_sl(z,K)^(1/(1-eps)))/vid_sl(d,K)
  return(rezultatas)
}
Atk_sl(df,0.5,N)
```

```
#iš paketo
Atkinson(df[,8], parameter = 0.5, na.rm = TRUE)
```

```
#Atkinsono =1 PAI
Atk_1 <- function(d,K){
  z<-d
  z[,8]<-d[,8]^(1/length(d[,8]))
  rezultatas<-1-prod(z[,8])/mean(d[,8])
  return(rezultatas)
}
Atk_1(df,N)
```

```
#Atkinsono =1 sluoksninei
Atk_1_sl <- function(d,K){
  z<-d
  z[,8]<-log(d[,8])
  rezultatas<-1-exp(vid_sl(z,K))/vid_sl(d,K)
  return(rezultatas)
}
Atk_1_sl(df,N)
```

```
#iš paketo
```

```

Atkinson(df[,8], parameter = 1, na.rm = TRUE)

#Paslinktumo vertinimas
paslinktasATK <- function(d){
  n0 <- 300
  kart <- 1000
  pop <- d[,8]
  ATK <- data.frame(matrix(ncol = 1, nrow = kart))
  for (i in 1:kart) {
    imt=sample(pop,size = n0)
    ATK[i,1]<-1-(mean(imt^(1-eps))^(1/(1-eps)))/mean(imt)
  }
  rezultatas<-mean(ATK[,1])
  return(rezultatas)
}

paslinktasATK1 <- function(d){
  n0 <- 300
  kart <- 1000
  pop <- d[,8]
  ATK1 <- data.frame(matrix(ncol = 1, nrow = kart))
  for (i in 1:kart) {
    imt=sample(pop,size = n0)
    ATK1[i,1]<-1-prod(imt^(1/length(imt)))/mean(d[,8])
  }
  rezultatas<-mean(ATK1[,1])
  return(rezultatas)
}

paslinktasATK(df)
paslinktasATK1(df)

#formulės populiacijos rodikliui paskaičiuoti
eps<-0.5
Atk_pop<-1-(1/n*sum ((d[,8]/mean(d[,8]))^(1-eps)))^(1/(1-eps))
Atk_1_pop<-1-prod(d[,8]^(1/n))/mean(d[,8])

#visr. Atk_sl
Atk_slu<-Atk_sl(df,0.5,N)
Atk_sl_a<-0

```

```

duom<-df
M<-N-1
H<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:n){
  duom_i<-duom[-i,]
  for (j in 1:(n-1)){
    if (identical(duom_i[j,3],duom[i,3]))
      duom_i[j,9]<-duom_i[j,4]/(duom_i[j,5]-1)
  }
  Atk_sl_a<-Atk_sl(duom_i,0.5,M)
  if (duom[i,3]=="LT0013") H[1]<-H[1]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0023") H[2]<-H[2]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0033") H[3]<-H[3]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0043") H[4]<-H[4]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0053") H[5]<-H[5]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0063") H[6]<-H[6]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0073") H[7]<-H[7]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0083") H[8]<-H[8]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT0093") H[9]<-H[9]+(Atk_sl_a-Atk_slu)^2
  if (duom[i,3]=="LT00A3") H[10]<-H[10]+(Atk_sl_a-Atk_slu)^2
  duom_i<-0
}

D_vis <- sum((1-n_h/N_h)*(n_h-1)/n_h*H)
cv <- sqrt(D_vis)/Atk_slu

#visr. Atk_1_sl
Atk_1_slu<-Atk_1_sl(df,N)
Atk_1_sl_a<-0
duom<-df
M<-N-1
H<-c(0,0,0,0,0,0,0,0,0,0)

for (i in 1:n){
  duom_i<-duom[-i,]
  for (j in 1:(n-1)){
    if (identical(duom_i[j,3],duom[i,3]))
      duom_i[j,9]<-duom_i[j,4]/(duom_i[j,5]-1)
  }
}

```

```

Atk_1_sl_a<-Atk_1_sl(duom_i,M)
if (duom[i,3]=="LT0013") H[1]<-H[1]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0023") H[2]<-H[2]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0033") H[3]<-H[3]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0043") H[4]<-H[4]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0053") H[5]<-H[5]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0063") H[6]<-H[6]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0073") H[7]<-H[7]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0083") H[8]<-H[8]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT0093") H[9]<-H[9]+(Atk_1_sl_a-Atk_1_slu)^2
if (duom[i,3]=="LT00A3") H[10]<-H[10]+(Atk_1_sl_a-Atk_1_slu)^2
duom_i<-0
}

```

```

D_vis <- sum((1-n_h/N_h)*(n_h-1)/n_h*H)
cv <- sqrt(D_vis)/Atk_1_slu

```

```

#visraktis PAI - visi rodikliai iškart

```

```

g_sum<-0
tl_sum<-0
tt_sum<-0
a_sum<-0
al_sum<-0

```

```

g_vis<-0
tl_vis<-0
tt_vis<-0
a_vis<-0
al_vis<-0

```

```

duom_i<-0

```

```

M<-N-1

```

```

g<-Gini(df)
tl<-GE_0(df,N)
tt<-GE_1(df,N)
a<-Atk(df,0.5,N)
al<-Atk_1(df,N)

```

```

for (i in 1:n){
  duom_i<-df[-i,]
  g_aa<-Gini(duom_i)
  tl_aa<-GE_0(duom_i,M)
  tt_aa<-GE_1(duom_i,M)
  a_aa<-Atk(duom_i,0.5,M)
  al_aa<-Atk_1(duom_i,M)
  g_a<-n*g-(n-1)*g_aa
  tl_a<-n*tl-(n-1)*tl_aa
  tt_a<-n*tt-(n-1)*tt_aa
  a_a<-n*a-(n-1)*a_aa
  al_a<-n*al-(n-1)*al_aa
  g_vis_<-g_vis_+g_a
  tl_vis_<-tl_vis_+tl_a
  tt_vis_<-tt_vis_+tt_a
  a_vis_<-a_vis_+a_a
  al_vis_<-al_vis_+al_a
  duom_i<-0
}

```

```

g_vis<-g_vis_/n
tl_vis<-tl_vis_/n
tt_vis<-tt_vis_/n
a_vis<-a_vis_/n
al_vis<-al_vis_/n

```

```

for (i in 1:n){
  duom_i<-df[-i,]
  g_aa<-Gini(duom_i)
  tl_aa<-GE_0(duom_i,M)
  tt_aa<-GE_1(duom_i,M)
  a_aa<-Atk(duom_i,0.5,M)
  al_aa<-Atk_1(duom_i,M)
  g_a<-n*g-(n-1)*g_aa
  tl_a<-n*tl-(n-1)*tl_aa
  tt_a<-n*tt-(n-1)*tt_aa
  a_a<-n*a-(n-1)*a_aa
  al_a<-n*al-(n-1)*al_aa
  g_sum<-g_sum+(g_a-g_vis)^2
  tl_sum<- tl_sum+(tl_a-tl_vis)^2

```

```

tt_sum<-tt_sum+(tt_a-tt_vis)^2
a_sum<-a_sum+(a_a-a_vis)^2
al_sum<-al_sum+(al_a-al_vis)^2
duom_i<-0
}

```

```

g_s_2<-g_sum/(n-1)
tl_s_2<-tl_sum/(n-1)
tt_s_2<-tt_sum/(n-1)
a_s_2<-a_sum/(n-1)
al_s_2<-al_sum/(n-1)

```

```

g_D_vis <- (1-n/N)*g_s_2/n
tl_D_vis <- (1-n/N)*tl_s_2/n
tt_D_vis <- (1-n/N)*tt_s_2/n
a_D_vis <- (1-n/N)*a_s_2/n
al_D_vis <- (1-n/N)*al_s_2/n

```

```

g_cv <- sqrt(g_D_vis)/g
tl_cv <- sqrt(tl_D_vis)/tl
tt_cv <- sqrt(tt_D_vis)/tt
a_cv <- sqrt(a_D_vis)/a
al_cv <- sqrt(al_D_vis)/al

```

Literatūra

1. Atkinson, A. B., On the measurement of inequality, *Journal of Economic Theory*, 1970, 2 (3).
2. Di Falco, E., Income inequality statistics. *Statistics in focus*, 2014, 12/2014.
3. European Commission, Eurostat. *Theoretical study of the Gini index*. EU-SILC 131-A/04, 2004.
4. European Commission's DG for Employment, *Social Affairs & Inclusion* [interaktyvus]. Brussels [žiūrėta 2016 m. gegužės 22d.]. Prieiga per internetą: <<http://ec.europa.eu>>.
5. ESS ERIC. *European Social Survey* [interaktyvus]. London [žiūrėta 2016 m. gegužės 22 d.]. Prieiga per internetą: <www.europeansocialsurvey.org>.
6. The ESS Core Scientific Team (CST), *ESS1-6, European Social Survey Cumulative File Rounds 1-6, Study Documentation*.
7. Europos socialinis tyrimas, *Pagrindinis klausimynas (5 banga, 2010/11)*. Vilnius ir Kaunas, 2011.
8. Galiautdinovaitė, K., *Income inequality measures for Baltic and Nordic countries* [interaktyvus]. Helsinki [žiūrėta 2016 m. gegužės 22d.]. Prieiga per internetą: <<https://wiki.helsinki.fi/display/banocoss2015/Presentations>>.
9. Gini, C., *Variabilità e mutabilità (angl. Variability and Mutability)*, Bologna: 1912.
10. Helliwell, J., Layard, R., Sachs, J., (September 9, 2013). *World Happiness Report 2013*. United Nations Sustainable Development Solutions Network, 2013.
11. Krapavickaitė, D.; Plikusas, A., *Imčių teorijos pagrindai*. Vilnius: Technika, 2005.
12. Lorenz, M. O., Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 1905, 9 (70).
13. Rudzkienė, V. *Socialinė statistika*. Vilnius: Mykolo Romerio universiteto Leidybos centras, 2005.
14. Särndal, C.-E.; Swensson, B.; and Wretman, J. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.
15. Shorrocks, A. F., The Class of Additively Decomposable Inequality Measures. *Econometrica*, 1980, 48 (3).
16. Statistikos departamentas. *Statistikos departamento rodiklių duomenų bazė* [interaktyvus]. Vilnius [žiūrėta 2016 m. gegužės 22d.]. Prieiga per internetą: <<http://osp.stat.gov.lt/>>.

17. Working group „Statistics on income, poverty & social exclusion“, „*Laeken*“ *indicators – detailed calculation methodology*. Luxembourg, 2003.
18. World Bank. *Introduction to poverty analysis*. Washington, DC: World Bank Group, 2005.