



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Mikrolygmens veiksmų poveikio namų ūkių vartojimo išlaidoms vertinimas

Baigiamasis magistro studijų projektas

Judita Pilipauskaitė

Projekto autorė

Prof. dr. Alina Stundžienė

Vadovė

Doc. dr. Kristina Šutienė

Vadovė

Kaunas, 2025



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Mikrolygmens veiksmų poveikio namų ūkių vartojimo išlaidoms vertinimas

Baigiamasis magistro studijų projektas

Didžiųjų verslo duomenų analitika (6213AX001)

Judita Pilipauskaitė

Projekto autorė

Prof. dr. Alina Stundžienė

Vadovė

Doc. dr. Kristina Šutienė

Vadovė

Doc. dr. Daiva Laskienė

Recenzentė

Doc. dr. Mantas Landauskas

Recenzentas

Kaunas, 2025



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Judita Pilipauskaitė

Mikrolygmens veiksmų poveikio namų ūkių vartojimo išlaidoms vertinimas

Akademinių sąžiningumo deklaracija

Patvirtinu, kad:

1. baigiamąjį projektą parengiau savarankiškai ir sąžiningai, nepažeisdama(s) kitų asmenų autoriaus ar kitų teisių, laikydamasi(s) Lietuvos Respublikos autorių teisių ir gretutinių teisių įstatymo nuostatų, Kauno technologijos universiteto (toliau – Universitetas) intelektinės nuosavybės valdymo ir perdavimo nuostatų bei Universiteto akademinės etikos kodekse nustatytų etikos reikalavimų;
2. baigiamajame projekte visi pateikti duomenys ir tyrimų rezultatai yra teisingi ir gauti teisėtai, nei viena šio projekto dalis nėra plagijuota nuo jokių spausdintinių ar elektroninių šaltinių, visos baigiamojo projekto tekste pateiktos citatos ir nuorodos yra nurodytos literatūros sąrašė;
3. įstatymų nenumatytų piniginių sumų už baigiamąjį projektą ar jo dalis niekam nesu mokėjęs (-usi);
4. suprantu, kad išaiškėjus nesąžiningumo ar kitų asmenų teisių pažeidimo faktui, man bus taikomos akademinės nuobaudos pagal Universitete galiojančią tvarką ir būsiu pašalinta(s) iš Universiteto, o baigiamasis projektas gali būti pateiktas Akademinės etikos ir procedūrų kontrolieriaus tarnybai nagrinėjant galimą akademinės etikos pažeidimą.

Judita Pilipauskaitė

Patvirtinta elektroniniu būdu

Pilipauskaitė, Judita. Mikrolygmens veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimas. Magistro baigiamasis projektas / vadovės prof. dr. Alina Stundžienė ir doc. dr. Kristina Šutienė; Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas.

Studijų kryptis ir sritis (studijų krypties grupė): Taikomoji matematika (Matematikos mokslai).

Reikšminiai žodžiai: vartojimo išlaidos, požymių atranka, daugialypė tiesinė regresija, mašininis mokymasis, SHAP analizė.

Kaunas, 2025. 70 p.

Santrauka

Temos aktualumas. Namų ūkių vartojimas, kartu su investicijomis, valstybės išlaidomis ir grynuoju eksportu, yra naudojamas svarbiausio valstybės ekonominę gerovę atspindinčio rodiklio – bendrojo vidaus produkto – skaičiavime. Gyventojų vartojimo išlaidų ir jų struktūros analizė gali būti vienas iš rodiklių, padedančių vertinti ne tik šalies ekonominę padėtį, bet ir gyventojų gyvenimo kokybę, kadangi yra laikoma, jog kuo vartojimo lygis yra aukštesnis, tuo aukštesnis yra ir gyvenimo lygis šalyje. **Tyrimo objektas** – bendros namų ūkių vartojimo išlaidos ir jų struktūrinės dalys. **Tyrimo tikslas** – įvertinti mikrolygmens veiksnių poveikį bendroms namų ūkių vartojimo išlaidoms bei jų struktūrai. **Tyrimo metodai** – mokslinės literatūros analizė, statistiniai testai, mašininio mokymosi modeliai. **Tyrimo uždaviniai:**

1. apžvelgti vartojimo išlaidų pokyčius lemiančius veiksnius remiantis mokslinės literatūros analize;
2. aprašyti tyrime naudojamus duomenis, jų parengimo bei modeliavimo metodus;
3. atlikti požymių atranką, siekiant identifikuoti namų ūkių vartojimo išlaidas veikiančius kintamuosius;
4. sudaryti modelį, siekiant nustatyti kaip vartojimo išlaidos ir jų struktūra priklauso nuo įvairių namų ūkių charakteristikų Lietuvoje.

Projektas susideda iš trijų pagrindinių etapų. Mokslinės literatūros apžvalga, kurios metu buvo identifikuoti pagrindiniai su vartojimu susiję veiksniai ir jų poveikis, taip pat aptarti kitų tyrėjų atlikti darbai bei taikyti metodai. Metodologinėje dalyje buvo aprašyti tyrimui naudoti duomenys bei jų ypatumai ir pateikta taikytų metodų literatūrinė analizė. Tiriamojoje dalyje buvo pateikti tyrimo vykdymo sprendimai, geriausi gauti modeliai, jų rezultatai ir interpretacijos.

Remiantis SHAP analize, kuri buvo taikyta ekstremalaus gradientinio stiprinimo modeliui, buvo nustatyta, kad Lietuvos namų ūkių bendros vartojimo išlaidų augimas yra labiausiai susijęs su didesniu namų ūkio narių skaičiumi. Gyventojų amžius ir vartojimo išlaidos yra susieti netiesiniais ryšiais. Išlaidos didėja, jei gyventojai priklauso aukštesnei pajamų kvintilinei grupei, turi didesnio naudingo ploto būstus, augina vaikus, turi aukštąjį išsilavinimą ar dirba aukštesnės kvalifikacijos darbu. Panašios išvados buvo rastos ir tiriant maisto bei nealkoholinių gėrimų, informacijos ir ryšių paslaugų bei draudimo ir finansinių paslaugų išlaidų kategorijas. Be to, nustatyta, kad informacijos ir ryšių išlaidos didėja, jei asmuo negyvena kaime, yra samdomas darbuotojas ir negyvena vieno buto name, kai draudimo ir finansinių paslaugų išlaidos yra didinamos, jei asmens pagrindinės pajamos yra ne iš pensijos, jis negyvena vienas, yra būsto savininkas, turintis neišmokėtą paskolą.

Pilipauskaitė, Judita. Evaluation of Micro-Level Factors Impacting Household Consumption Expenditure. Master's Final Degree Project / supervisors Prof. Dr. Alina Stundžienė and Assoc. Dr. Kristina Šutienė; Faculty of Mathematics and Natural Sciences, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics (Mathematical Sciences).

Keywords: consumption expenditure, feature selection, multiple linear regression, machine learning, SHAP analysis.

Kaunas, 2025. 70.

Summary

Relevance of the topic. Household consumption, together with investment, government expenditure and net exports, is used in calculating the most important indicator reflecting a country's economic welfare – gross domestic product. Analysis of household consumption expenditure and its structure can be one of the indicators that helps to assess not only the country's economic situation but also the population's quality of life, since a higher level of consumption is considered to correspond to a higher standard of living. **Object of the study** – total household consumption expenditure and its structural components. **Aim of the study** – to assess the impact of micro-level factors on total household consumption expenditure and its structure. **Methods** – scientific literature review, statistical tests, machine learning models. **Tasks of the study:**

1. to review the factors determining changes in consumption expenditure on the basis of a scientific literature analysis;
2. describe the data used in the study, their preparation steps, and the modeling methods applied;
3. to perform feature selection in order to identify the variables affecting household consumption expenditure;
4. to build a model to determine how consumption expenditure and its structure depend on various household characteristics in Lithuania.

The project consists of three main stages. A scientific literature review identified the main consumption related factors and their impact and also discussed previous research and methods applied. The methodological part described the data used, their characteristics and provided a literature analysis of the methods applied. The empirical part presented the research decisions, the best models obtained, their results and interpretations.

Based on SHAP analysis applied to the extreme gradient boosting model, it was found that the growth of total consumption expenditure of Lithuanian households is most closely related to a larger household size. Age and consumption expenditure are linked by a non linear relationship. Expenditure increases if households belong to a higher income quintile, have larger usable floor space, raise children, have higher education or work in higher-skilled occupations. Similar conclusions were drawn when analysing the categories of food and non-alcoholic beverages, information and communication services, and insurance and financial services expenditure. In addition, it was established that expenditure on information and communication rises if a person does not live in the countryside, is an employee and does not live in a single-family house, while expenditure on insurance and financial services increases if the person's main income is not a pension, does not live alone, and is a homeowner with an outstanding mortgage.

Turinys

Lentelių sąrašas	7
Paveikslų sąrašas	8
Santrumpų ir terminų sąrašas	9
Įvadas.....	10
1. Gyventojų vartojimo išlaidų literatūros apžvalga.....	11
1.1. Gyventojų vartojimo išlaidų samprata ir svarba ekonomikai	11
1.2. Mikrolygmens veiksnių samprata.....	13
1.3. Gyventojų vartojimo išlaidų pokyčius lemiantys veiksniai.....	14
1.3.1. Mikrolygmens veiksnių įtaka vartojimo išlaidoms	15
1.3.2. Makrolygmens veiksnių įtaka vartojimo išlaidoms.....	19
1.4. Mikrolygmens ir egzogeninių veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo metodų apžvalga	21
1.5. Mikrolygmens veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo poreikis.....	27
2. Lietuvos namų ūkių duomenys ir tyrimo metodai	29
2.1. Lietuvos namų ūkių duomenų rinkinys ir jo paruošimas	29
2.2. Požymių atrankos metodai	32
2.3. Neįprastų reikšmių įtakos mažinimo metodikos	35
2.4. Regresijos uždavinio sprendimo modeliai.....	37
2.4.1. Atsitiktinių miškų regresijos algoritmas.....	38
2.4.2. Ekstremalaus gradientinio stiprinimo algoritmas regresijai	40
2.5. Modelių kokybės vertinimo ir interpretavimo metodai.....	42
3. Vartojimo išlaidų priklausomybės nuo namų ūkių charakteristikų vertinimas	44
3.1. Svarbiausių požymių atranka naudojant filtravimo metodus	44
3.2. Vartojimo išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimas	51
3.3. Mikrolygmens veiksnių poveikio vartojimo išlaidoms tyrimo apibendrinimas.....	59
Išvados	62
Literatūros sąrašas	64
Priedai.....	71

Lentelių sąrašas

1 lentelė. Individualaus vartojimo išlaidų pagal paskirtį klasifikavimas pagal COICOP 2018 klasifikatorių (šaltinis: oficialiosios statistikos portalas, 2024 ⁵)	12
2 lentelė. Namų ūkių išlaidų klasifikavimas reguliarumą, būtinumą ir naudojimo tikslą. Sudaryta pagal Gedminaitė [25 p. 11]	13
3 lentelė. Mikrolygmens veiksnių, lemiančių vartojimą, galimo poveikio gyventojų išlaidoms apibendrinimas	18
4 lentelė. Makrolygmens veiksnių, lemiančių vartojimą, galimo poveikio gyventojų išlaidoms aptarimas.....	21
5 lentelė. Mikrolygmens ir išorinių veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo metodų apibendrinimas, remiantis ankstesniais mokslinėje literatūroje aprašytais tyrimais	26
6 lentelė. Tyrime naudojamų nepriklausomų kintamųjų aprašymas.....	30
7 lentelė. Atsitiktinių miškų regresijos taikymo privalumai ir trūkumai (sudaryta pagal Gupta [102])	40
8 lentelė. Leveno testo rezultatai, kai priklausomas kintamasis yra bendros išlaidos	46
9 lentelė. K-imčių Andersono ir Darlingo testo rezultatai, siekiant įvertinti, ar bendrų vartojimo išlaidų skirstiniai atskirose grupėse yra panašūs	46
10 lentelė. Medianos testo rezultatai, skirti vartojimo išlaidų skirtumams tarp kategorinių grupių vertinti.....	47
11 lentelė. Porinių medianų testų rezultatai, kai pagal bendras vartojimo išlaidas nebuvo rasta skirtumų tarp kategorinių požymių grupių	48
12 lentelė. Daugialypės tiesinės regresijos modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos.....	52
13 lentelė. Atsitiktinių miškų regresijos modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos.....	53
14 lentelė. Ekstremalaus gradientinio stiprinimo algoritmo metodu sukurtų regresinių modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos.....	54

Paveikslų sąrašas

1 pav. Namų ūkių sprendimus veikiančių aspektų klasifikavimas. (Sudaryta pagal Korol'ą [13])..	13
2 pav. Tyrimo metodologijos schema	29
3 pav. Filtravimo požymių atrankos metodų veikimo principas (autoriai: Wang'as ir kt. [70])	32
4 pav. Pirmųjų keturių analizuojamų kiekybinių išlaidų kintamųjų histogramos.....	35
5 pav. ϕ_K koreliacijos koeficientų matrica tarp visų tyrimo priklausomų ir nepriklausomų kintamųjų	44
6 pav. Tarpusavio informacijos metodo panaudojimas požymių atrankai, kai vertinamos bendros išlaidos.....	49
7 pav. Moodo medianos testo panaudojimas požymių atrankai, kai vertinamos bendros vartojimo išlaidos.....	49
8 pav. Ekstremalaus gradientinio stiprinimo regresijos modelio diagnostika prognozuojant bendras išlaidas	55
9 pav. Požymių poveikis bendroms vartojimo išlaidoms, nustatytas naudojant SHAP metodą, pritaikytą pagal ekstremalaus gradientinio stiprinimo algoritmu sukurtą regresijos modelį	55
10 pav. SHAP priklausomybės diagramos, vertinančios kintamųjų tarpusavio sąveikas prognozuojamų bendrų vartojimo išlaidų atžvilgiu, kai vienas iš veiksnių yra 5-a pajamų kvintilinė grupė	57

Santrumpų ir terminų sąrašas

Santrumpos:

TI – tarpusavio informacijos matas.

VIF – dispersijos didinimo koeficientas (angl. *Variance Inflation Factor*).

RMSE – kvadratinės paklaidos šaknis.

MAE – vidutinė absoliutinė paklaida.

MedAE – medianinė absoliutinė paklaida.

Terminai:

Multikolinearumas – situacija, kai du ar daugiau nepriklausomųjų kintamųjų yra stipriai tarpusavyje susiję.

Post-hoc – papildoma analizė, atliekama po pagrindinio metodo taikymo.

Reikšmingumo lygmuo – tyrėjo pasirinkta riba, leidžianti klaidingai atmesti nulinę hipotezę.

Generalizacija – modelio gebėjimas teisingai prognozuoti naudojant anksčiau nematytus duomenis.

Ivadas

Tyrimo aktualumas. Gyventojų vartojimą dažniausiai apibrėžia namų ūkių galutinio vartojimo išlaidos. Nors kiekvieno gyventojų turimų pajamų paskirstymo taupymui ar vartojimui sprendimas yra individualus, tačiau šių sprendimų visuma turi įtakos makroekonominiai šalies veiklai. Gyventojų vartojimo išlaidos yra viena iš didžiausių bendrąją paklausą sudarančių komponentų rinką formuojančių pasiūlos ir paklausos jėgų kontekste, kai vartojimui skirtų prekių gamyba yra viena iš svarbiausių pasiūlą sudarančių dalių. Namų ūkių vartojimas, kartu su investicijomis, valstybės išlaidomis ir grynųjų eksportu, yra naudojamas svarbiausio valstybės ekonominę gerovę atspindinčio rodiklio – bendrojo vidaus produkto (BVP) – skaičiavime. Gyventojų vartojimo išlaidų ir jų struktūros analizė gali būti vienas iš rodiklių, padedančių vertinti ne tik šalies ekonominę padėtį, bet ir gyventojų gyvenimo kokybę, kadangi yra laikoma, jog kuo bendras vartojimo lygis yra aukštesnis, tuo aukštesnis yra ir gyvenimo lygis šalyje.

Tyrimo problema. Nepaisant vis dažniau atliekamų vartojimo analizės tyrimų, dauguma jų vis dar orientuojasi į makroekonominius veiksnius, o mikrolygmens kintamųjų, tokių kaip namų ūkio dydis, gyventojų amžius, išsilavinimas ar gyvenamoji vieta, poveikis vartojimo išlaidoms ir jų struktūrai lieka nepakankamai ištyrinėtas. Būtent mikrolygmuo yra susijęs su kasdieniais gyventojų vartojimo sprendimais ir atspindi realius socialinius bei ekonominius namų ūkių pasirinkimus, o tai yra svarbus aspektas siekiant suprasti kompleksinę vartojimo elgseną. Be to, net ir tie tyrimai, kurie į analizę įtraukia mikrolygmens veiksnius, dažnai pateikia nevienareikšmius rezultatus, o tai rodo šios srities tyrimų neišbaigtumą. Verta paminėti, kad tyrimų rezultatai gali skirtis dėl analizuojamos šalies ar regiono konteksto, o tai rodo, jog mikrolygmens veiksmų poveikis vartojimui nėra universalus ir priklauso nuo specifinių socialinių, ekonominių ar geografinių aplinkybių.

Tyrimo naujumas. Esami vartojimo išlaidų tyrimai dažniausiai apsiriboja daugialypės tiesinės regresijos modelių taikymu, tačiau pastaruoju metu vis labiau populiarėjantys pažangesni mašininio mokymosi metodai skatina tyrėjus juos taikyti ir šioje srityje, siekiant atskleisti galimai egzistuojančius sudėtingesnius ryšius tarp kintamųjų. Be to, daugumoje tyrimų yra analizuojamos bendros vartojimo išlaidos, o jų struktūrinės dalys lieka neištirtos. Pažymėtina ir tai, kad tokio pobūdžio tyrimas Lietuvos atveju iki šiol dar nebuvo atliktas.

Tyrimo tikslas – įvertinti mikrolygmens veiksmų poveikį bendroms namų ūkių vartojimo išlaidoms bei jų struktūrai.

Darbo uždaviniai:

1. apžvelgti vartojimo išlaidų pokyčius lemiančius veiksnius remiantis mokslinės literatūros analize;
2. aprašyti tyrime naudojamus duomenis, jų parengimo bei modeliavimo metodus;
3. atlikti požymių atranką, siekiant identifikuoti namų ūkių vartojimo išlaidas veikiančius kintamuosius;
4. sudaryti modelį, siekiant nustatyti kaip vartojimo išlaidos ir jų struktūra priklauso nuo įvairių namų ūkių charakteristikų Lietuvoje.

1. Gyventojų vartojimo išlaidų literatūros apžvalga

Šiame skyriuje apžvelgiama vartojimo išlaidų ir jų struktūros svarba bei aktualumas ekonomikai tiek mikro-, tiek makrolygmenimis. Identifikuojami labiausiai vartojimo išlaidų pokyčius lemiantys veiksniai, ir aptariamos šių veiksnių poveikio išlaidoms vertinimo metodologinės galimybės.

1.1. Gyventojų vartojimo išlaidų samprata ir svarba ekonomikai

Vartojimas yra einamiesiems žmonių poreikiams tenkinti skirtas galutinių prekių ar paslaugų naudojimas. Vartojimas yra skirstomas į asmeninį ir visuomeninį vartojimą¹. Šalies gyventojų vartojimą dažniausiai apibrėžia namų ūkių galutinio vartojimo išlaidos, kurios yra asmens arba žmonių grupės, gyvenančių kartu bendrame būste, patiriamos bendros namų ūkio išlaidos. Namų ūkių išlaidas apima tam tikros valstybės teritorijoje patirtos išlaidos, susijusios ne tik su prekių ir paslaugų pirkimu, bet ir su nuosavos produkcijos vartojimu². Nacionalinėse sąskaitose namų ūkių vartojimo išlaidos dažniausiai yra nurodomos kaip faktinis individualus suvartojimas, kurį apima namų ūkių tiesiogiai perkamos vartojimo prekės ir paslaugos bei paslaugos, kurias teikia ne pelno institucijos ir vyriausybė (pvz., sveikatos priežiūros ir švietimo paslaugos). Bakri'is ir kt. [1] priduria, kad vartojimo išlaidas apima ir visos prekės, paslaugos bei įrenginiai, kurie buvo gauti natūra, t. y., nemokamai arba su nuolaida. Faktinis individualus vartojimas, palyginti su siauresne namų ūkių vartojimo sąvoka, labiau tinka tarptautiniuose palyginimuose dėl įvertintos ne pelno institucijų ir valstybės sektoriaus įtakos³.

Magrabi'is ir kt. [2] pastebi, kad gerovė yra sveikatos, komforto ar laimės būseną, kurią lemia prekių ir paslaugų vartojimas. Jų teigimu, išmatuoti tikrąjį ryšį tarp vartojimo ir gerovės yra sunku, tačiau dažniausiai jis apibrėžiamas paprasčiausia teorija, t. y., kuo daugiau vartojama, tuo aukštesnis yra gerovės lygis. Ramanausko ir Jakaitienės [3] teigimu, namų ūkių išlaidos yra ypač svarbus ekonominis rodiklis. Namų ūkių sprendimas dėl vartojimui ar taupymui skirtos pajamų dalies rodo ne tik bendrą šalies ekonominę gerovę, bet ir kapitalo kaupimo ir ekonominio potencialo didinimo procesų priklausomybę nuo minėto sprendimo. Bendras šalies gerovės lygis gali būti nustatomas pagal tai, kad namų ūkiai didina išlaidas vartojimo tikslams ir mažiausia tų išlaidų dalis yra skiriama maisto prekėms įsigyti, kas ir parodo aukštesnį gyvenimo lygį [4]. Ismal'as-Abu'us ir kt. [5] teigimu, namų ūkiai gali būti laikomi skurstančiais tuomet, kai daugiau nei 60 % jų visų vartojimo išlaidų tenka maistui. Namų ūkių vartojimo ekonominį svarbumą atspindi tai, kad vartojimo išlaidų lyginamasis svoris bendrojo šalies produkto (toliau – BVP) apimtyje sudaro apie 60 % [6], kai iš formulės (žr. (1) formulę), matyti, kad vartojimas yra viena iš BVP skaičiavimui naudojamų dalių:

$$BVP = C + I + G + NX; \quad (1)$$

čia C – vartojimas, I – investicijos, G – valstybės išlaidos, NX – grynas eksportas.

Statistikos portalo, *Eurostat*⁴ duomenimis, namų ūkių galutinio vartojimo išlaidos 2022 m. visoje Europos Sąjungoje sudarė 52,1 % jos BVP. Gyventojų vartojimo išlaidos yra viena iš didžiausių bendrąją paklausą sudarančių komponentų rinką formuojančių pasiūlos ir paklausos jėgų kontekste, kai vartojimui skirtų prekių gamyba yra viena iš svarbiausių pasiūlos dalių [7]. Visgi vartojimo išlaidų

¹ <https://zodinas.vz.lt/Vartojimas>

² <https://data.europa.eu/data/datasets/qe6lqlgopcbjeac6su3dq?locale=lt>

³ [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Actual_individual_consumption_\(AIC\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Actual_individual_consumption_(AIC))

⁴ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Household_consumption_by_purpose

svarbumas ekonomikos augimo teorijoje yra dviprasmiškas. Keinso (*John Maynard Keynes*) ekonomikos teorijoje, kuria remiasi daugelis šių laikų ekonomistų, yra teigiama, kad vartotojų išlaidos trumpuoju laikotarpiu yra svarbiausias ekonominio aktyvumo veiksnys bei pagrindinė visuminės paklausos sudedamoji dalis, kai būtent visuminės paklausos svyravimai lemia produkcijos apimtį ir užimtumo pasikeitimus. Teorijoje taip pat pabrėžiama, kad vartojimo išlaidos turi „multiplikatyvų“ poveikį ekonomikai ir vartojimo išlaidų padidėjimas lemia didesnę nei proporcingą bendros ekonominės veiklos padidėjimą, tačiau ekonomikai skatinti ir reguliuoti šiuo būdu yra reikalingi kuo greitesni veiksmai, t. y., veikti trumpuoju laikotarpiu [8]. Ekonominėje teorijoje gerai žinomos Keinso „fiskalinė ir monetarinė“ politikos yra glaudžiai susijusios su vartotojų išlaidomis, nes fiskalinės politikos idėjos pagrindas yra tai, kad vyriausybė nuosmukio metu turėtų didinti išlaidas ir taip didinti bendrąją paklausą, o monetarinės politikos pagrindas – idėja mažinti palūkanų normas tikintis didesnio vartotojų ir įmonių skolinimosi su tikslu daugiau išleisti arba investuoti [9]. Kiti ekonomistai pripažįsta labiau klasikinę Žano Batisto Sė (*Jean-Baptiste Say*) ekonominę teoriją, teigiančią, kad „pasiūla sukuria sau paklausą“ ir ekonomikos augimo varomosios jėgos yra gamyba, taupymas ir investicijos, o per didelės vartojimui skiriamos išlaidos dabar, gali kelti pavojų būsimam ekonominės veiklos aktyvumo augimui dėl per mažų santaupų ir investicijų [10].

Remiantis oficialiosios statistikos portalu⁵, gyventojų vartojimo išlaidos pagal paskirtį Lietuvoje yra klasifikuojamos taikant 2018 m. versijos tarptautinį individualaus vartojimo išlaidų pagal paskirtį klasifikatorių (toliau – COICOP). Pagrindiniai vartojimo išlaidų klasifikavimo rezultatai, pateikti iki 1 kodo lygmens, yra apžvelgti 1 lentelėje.

1 lentelė. Individualaus vartojimo išlaidų pagal paskirtį klasifikavimas pagal COICOP 2018 klasifikatorių (šaltinis: oficialiosios statistikos portalas, 2024⁵)

COICOP 2018 kodas	Pavadinimas
01	Maistas ir nealkoholiniai gėrimai
02	Alkoholiniai gėrimai, tabakas ir narkotikai
03	Apranga ir avalynė
04	Būstas, vanduo, elektra, dujos ir kitas kuras
05	Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra
06	Sveikata
07	Transportas
08	Informacija ir ryšiai
09	Rekreacija, sportas ir kultūra
10	Švietimo paslaugos
11	Restoranai ir apgyvendinimo paslaugos
12	Draudimas ir finansinės paslaugos
13	Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos
14	Ne pelno institucijų, teikiančių paslaugas namų ūkiams, individualaus vartojimo išlaidos

Gedminaitė [11] teigia, kad namų ūkių išlaidas galima klasifikuoti ir pagal kitokius požymius, kurie buvo nurodyti 2 lentelėje. Pabrėžtina, kad dažniausiai naudojamas išlaidų skirstymas yra susijęs su jų būtinumu, kai pastarųjų sudėtis visoje išlaidų struktūroje ir gali būti vertinama kaip gyventojų gyvenimo kokybės rodiklis.

⁵ <https://osp.stat.gov.lt/individualaus-vartojimo-islaidu-pagal-paskirti-klasifikatorius-coicop>

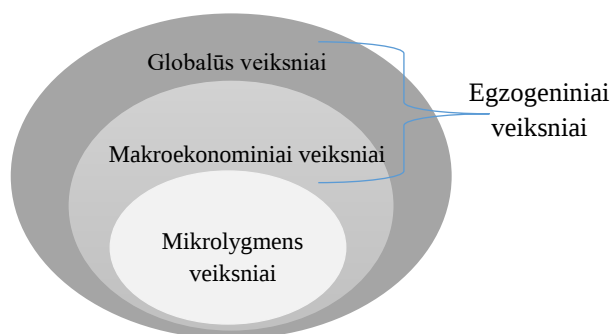
2 lentelė. Namų ūkių išlaidų klasifikavimas reguliarumą, būtinumą ir naudojimo tikslą. Sudaryta pagal Gedminaitė [25 p. 11]

Požymis	Išlaidų rūšys
Reguliarumas	• Pastovios (pvz., maistas, išlaidos būsto išlaikymui) • Reguliariosios (pvz., avalynė, drabužiai) • Vienkartinės (pvz., automobilio pirkimas)
Būtinumas	• Būtiniosios (pvz., maitinimasis, būsto išlaikymas, sveikatos priežiūros paslaugos) • Nebūtiniosios (pvz., draudimas, akcijų įsigijimas)
Naudojimo tikslas	• Vartojimui (pvz., paslaugų ir prekių pirkimas) • Mokesčiams ir kitoms privalomoms ar savanoriškoms įmokoms • Užsienio valiutos įsigijimui

Piekt [12] teigia, kad vartojimas yra vienas iš svarbiausių makroekonominės aplinkos veiksnių, kurio aktualumas išlieka aukštas nepaisant pokyčių, įvykusių makroekonominių teorijų formavimuose. Daugelis veiksnių turi įtakos namų ūkio išlaidų dydžiui ir struktūrai, tačiau pagrindinis įtakos turintis faktorius yra individualūs žmonių sprendimai, priklausantys nuo asmenų gyvenimo aplinkybių.

1.2. Mikrolygmens veiksnių samprata

Korol'as [13] teigia, kad namų ūkių sprendimus galima suskirstyti į dvi pagrindines grupes – egzogenines ir endogenines. Egzogeninei grupei priskiriami išoriniai veiksniai, kuriuos apima makroekonominė aplinka ir įvairūs globalūs reiškiniai, pavyzdžiui, globalizacija, klimato kaita, pandemijos. Makroekonominiai veiksniai formuoja regiono ekonomines sąlygas, kurios daro įtaką namų ūkių būsimoms pajamoms, turimam turtui bei paskolų prieinamumui, kai patys namų ūkiai neturi jokios tiesioginės kontrolės šių veiksnių atžvilgiu, dėl ko prie galiojančių ekonominių sąlygų jiems tenka prisitaikyti. Endogeniniai veiksniai apima tiesioginius namų ūkių sprendimus, kuriuos jie gali valdyti. Šie veiksniai apibrėžia mikrolygmens aplinką, kurią formuoja sociodemografiniai, psichologiniai ar mikroekonominiai požymiai. Pabrėžtina, kad veiksnių grupės tarpusavyje persidengia (žr. 1 pav.), nes namų ūkių sprendimai yra veikiami išorinės aplinkos [13].



1 pav. Namų ūkių sprendimus veikiančių aspektų klasifikavimas. (Sudaryta pagal Korol'ą [13])

Jajic'as, Herceg'as ir Bach'as [14] teigia, kad mikroekonominėje teorijoje viena optimizavimo problemos pusė apima vartotojų preferencijas, atspindinčias jų subjektyvias nuostatas ir įpročius, o kita – uždirdamas pajamas bei prekių ir paslaugų kainas, kurios formuoja asmenų biudžeto apribojimus. Neoklasikinė mikroekonominė teorija numato, kad namų ūkiai siekia maksimizuoti naudą arba minimizuoti išlaidas, kai vartotojų elgsena yra subjektyvi, o jų galimybės – objektyvios. Magrabi'is ir kt. [2] pritaria ir pateikia naudingumo funkciją:

$$U = u(A); \quad (2)$$

čia U – bendra individo nauda, gaunama tam tikru laikotarpiu, A – alternatyvų rinkinys.

Naudingumo funkcija gali būti apibrėžiama kaip alternatyvų rinkinio vertinimas, kai kiekviena alternatyva turi po kelis požymius, kurie veikia bendrą iš tos alternatyvos gaunamą naudą. Hauser'is ir Urban'as [15], taikydami adityviąją naudingumo funkciją, tyrė, kaip vartotojai priima sprendimus, kai turi biudžeto apribojimą ir geba įvertinti kiekvienos alternatyvos (automobilio ir atostogų) suteikiamą naudingumo kiekį. Bandymo metu buvo keltos dvi hipotezės: naudos už dolerį ir gryniosios naudos. Naudos už dolerį hipotezė teigė, kad vartotojas pasirinks tą alternatyvą, kuri suteiks daugiausiai naudos už kiekvieną išleistą dolerį, kai gryniosios naudos hipoteze buvo teigiama, jog vartotojas pasirinks tą alternatyvą, kuri suteiks daugiausiai gryniosios naudos, kai yra atsižvelgiama į galimą prarastą naudą pasirinkus kitą alternatyvą. Rezultatai iš apklausos, kurioje buvo teiraujamasi apie asmenų turimą turtą, pirkimo planus bei prioritetus, patvirtino, kad abi hipotezės galioja. Jajic'as, Herceg'as ir Bach'as [14] paaiškina, kad vartotojai skiriasi savo racionalumu, o tai ir daro reikšmingą poveikį galutiniam pirkimo sprendimui, net kai biudžetiniai apribojimai yra vienodi. Nors vartotojas ir siekia geriausio kainos ir vertės santykio, tačiau realios gyvenimo sąlygos ar nepilna informacija apie visas egzistuojančias alternatyvas dažnai lemia namų ūkių priimamus subracionalius sprendimus.

Mikroekonomika tiria smulkius ekonomikos vienetų bei sąveikas tarp jų, kurios gali būti analizuojamos tiek namų ūkių, tiek įmonių lygmenyje. Korol'as [13] pastebi, namų ūkių sprendimai, susiję su pajamų uždirbimu, darbo stažu, banko paskolų naudojimu, investavimu, turto įsigijimu ar pensijų planų sudarymu, yra smulkūs ekonominiai pasirinkimai. Ciarli'is ir kt. [16] teigia, kad mikroekonominiai veiksnius apima įmonių vidiniai sprendimai, tokie kaip gamybos ir organizacijos struktūriniai pakeitimai, pelno pasiskirstymas. Nuosavi būstai dažniau susiję su investicijomis į energijos taupymo priemones nei nuomojami būstai.

Sociodemografiniai veiksniai apima su socialiniais santykiais ir namų ūkių ypatumais susijusius veiksnius. Liu, Shryane'ė, Elliot'as [17] teigia, kad kasdienis gyvenimo būdas, t. y., įpročiai, šildymo ar vėsinimo dažnumas, įjungtų prietaisų palikimas, daro tiesioginę įtaką namų ūkių energijos vartojimui. Kaip socialinius veiksnius Korol'as [13] išskiria vartotojo dėmesį finansams, aukštojo mokslo siekimą, savo sveikatos priežiūros lygį. Pastaruoju metu visuomenėje populiarėjantis tvarumo požiūris ir didėjantis susirūpinimas aplinkos apsauga taip pat gali keisti vartotojų elgseną, susijusią su prekių pirkimu, naudojimu ir perdirbimu [18]. Adrian'as ir Daniel'ius [19] tyrinėdami maisto vartojimą, kaip svarbius socialinius veiksnius išskyrė namų ūkio šeimos gyvenimo ciklo etapą ir šeiminių statusą. Religijos įtaka taip pat gali būti klasifikuojama kaip socialinis veiksnys [13]. Sociodemografiniai veiksniai apima ir namų ūkio narių lytį, amžių, išsilavinimą, socialinį-ekonominį statusą, bei namų ūkio dydį [20].

Kaip psichologinis aspektas gali būti išskiriamas nekantrumas materialistinėje kultūroje, kadangi materialistai tiki, jog materialių vertybių įsigijimas yra pagrindinis sėkmės rodiklis, matuojantis asmenų laimę bei gyvenimo tikslų pasiekimą. Bianica [21] pabrėžia, kad vartotojų elgsena priklauso nuo namų ūkio narių motyvacijos, įsitikinimų, emocijų, charakterio.

1.3. Gyventojų vartojimo išlaidų pokyčius lemiantys veiksniai

Gyventojų vartojimo išlaidos yra kompleksinės ir priklauso nuo įvairių tarpusavyje susijusių veiksnių, kurie gali būti skirstomi į mikrolygmens, makrolygmens ar kitus veiksnius.

1.3.1. Mikrolygmens veiksmų įtaka vartojimo išlaidoms

Pagrindinis vartojimo išlaidų lygį lemiantis veiksnys yra namų ūkių pajamos, kurios tiesiogiai veikia vartojimo sprendimus, nes būtent namų ūkis priima sprendimą dėl savo pinigų paskirstymo, t. y., vartoti ar taupyti. Šių dviejų veiksmų sąveika ir yra dažniausiai aprašoma ekonominėse teorijose. Kaip teigia Gedminaitė [11], namų ūkių pajamas gali sudaryti skirtingi šaltiniai, t. y., pajamos iš samdomo darbo, individualios ne žemės ūkio ir individualios žemės ūkio veiklos, turto (palūkanos, gaunamos iš akcijų, vertybinių popierių ir pan.), rentos, socialinių išmokų ir kitų šaltinių (stipendijos, piniginės dovanos). Išskiriamosios grynosios (disponuojamosios) pajamos, kurios rodo pajamų dalį, kurią namų ūkis gali panaudoti vartojimui, nes tai yra pajamos, kurios liko asmeniui, atskaičius jam taikomus mokesčius. Svarbus ne tik pajamų dydis, bet ir jų patikimumas bei reguliarumas. Ramanauskas ir Jakaitienė [3] pastebi, kad 1936 m. Keinsio sukurtos absoliučiuųjų pajamų hipotezės pagrindas yra vartojimo priklausomybė nuo esamų pajamų lygio, kai kiti veiksniai – pinigų kiekis, palūkanų norma ar ekonominiai lūkesčiai – neturėtų turėti didelės reikšmės visuminiam vartojimui. Mokslininkai [3] taip pat pabrėžia, kad *Miltono Friedmano* 1957 m. iškelta nuolatinių pajamų hipotezė patikslina anksčiau sukurta teoriją ir teigia, jog vartojimą lemia ne tik esamos, bet ir būsimos asmens pajamos, nes būsimų pajamų lūkesčiai gali turėti įtakos taupymo ar dabartinio vartojimo skolinimosi sprendimams. Ding'as ir kt. [22] tyrimų metu nustatė, kad pajamų lygis Kinijoje yra pagrindinis požymis lemiantis kuro vartojimą namų ūkiuose, kadangi buvo rasta, jog 1 % padidėjus gyventojų pajamoms, energijos suvartojimas vienam gyventojui padidėja 0,75 %. Roshchina [23] savo vykdytame tyrime atrado, kad tarp vartojimo ir pajamų gali egzistuoti kvadratinė pajamų priklausomybė, t. y., nustatyta, kad alkoholio vartojimas auga iki tam tikro pajamų lygio, o vėliau pradeda mažėti. Ramanauskas ir Jakaitienė [3] pabrėžia, kad namų ūkių deklaruojamos pajamos gali būti mažesnės nei realiai gaunamos dėl namų ūkių šešėlinėje ekonomikoje vykdomos veiklos, t. y., vartojimo lygis gali būti aukštesnis už oficialiai skelbiamų pajamų lygį dėl neapskaitytų šešėlinių pajamų.

Vartojimo išlaidų pokyčius lemia ir namų ūkių užimtumo lygis, nes vartojimui skirtos pajamos dažniausiai kyla iš asmens atliekamo darbo. Staigus pajamų sumažėjimas dėl darbo praradimo ar sutrumpėjusių darbo valandų mažina namų ūkių išlaidas nebūtiniems pirkiniams ir verčia atsakyti daugelio poreikių patenkinimo [12]. Ganong'as ir Noel'is [24] teigia, kad nedarbas sukelia ryškų pajamų sumažėjimą, tačiau šis poveikis yra trumpalaikis. Naudojant Jungtinių Amerikos Valstijų namų ūkių duomenis, minėti mokslininkai apskaičiavo, kad padidėjus nedarbui, išlaidos ilgalaikio vartojimo prekėms sumažėja 6%, tačiau taip pat kartu sumažėja maždaug ketvirtadalis patiriamų visų išlaidų, kurias paaiškina su darbu susijusių išlaidų sumažėjimas. Pologeogris [25] pabrėžia, kad ryšys tarp vartojimo ir nedarbo yra abipusis, nes jei namų ūkiai pastebimai sumažintų savo vartojimą, verslai prarastų tikslą plėstis ir investuoti į kapitalą bei darbo jėgą, nes nebūtų vartotojų paklausos, kas įtakotų sulėtėjimą ekonomikoje ir didėjantį nedarbą. Namazkhan'as, Albers'as ir Steg'as [26] teigia, kad namų ūkių užimtumo statusas yra aktualus ir vienai iš vartojimo struktūrą sudarančių dalių, nes buvo atrasta, kad pilnu etatu dirbantys ar bedarbiai vartoja daugiau dujų nei dirbantys ne pilnu etatu.

Asmenų vartojimo sprendimus lemia ir demografiniai veiksniai. Ramanausko ir Jakaitienės [3] teigimu 1954 m. F. Modigliani ir R. Brumberg iškelta paprasčiausia gyvenimo ciklo teorija teigia, kad gyventojai taupo ir kaupia turtą iki pensinio amžiaus, o po to jį eikvoja vartojimo tikslais. Visgi kuo pensinis amžius ilgesnis, tuo ilgiau nėra dirbama, dėl ko mažesnė pajamų dalis yra vartojama einamuoju laikotarpiu. Augant ekonomikai, gyventojai uždirba daugiau, todėl ir dabartinės kartos

dirbantieji gali sukaupti daugiau, o toks taupymas atsveria dabartinių pensininkų sukauptas santaupas vartojimo reikmėms. Bugheanu'us ir Strachinaru'us [27] pastebi, kad jauni gyventojai, neturintys nei turtinių nei šeimyninių išpareigojimų, yra labiau linkę išlaidauti nei tie gyventojai, kurie turi atlikti sudėtingesnius finansinius sprendimus ir atsižvelgti į svarbių žmonių poreikius. Mokslininkai pabrėžia, kad tyrimų metu buvo atrastas teigiamas ryšys tarp amžiaus ir polinkio vartoti, t. y., kuo asmuo yra vyresnis, tuo labiau tikėtina, jog augs jo pajamos, dėl ko ir polinkis vartoti didės. Arapova [28] su Azijos šalių 1991 – 2015 m. duomenimis atlikto tyrimo metu nustatė, jog gyventojų skaičiaus augimas yra tiesiogiai susijęs su didesniu vartojimo potencialu. York'as [29] pastebi, kad gyventojų struktūra veikia viena iš bendrų išlaidų sudedamųjų dalių – energijos vartojimo išlaidas, nes atliktų tyrimų su Europos Sąjungos šalimis metu buvo atrasta, kad didesni namų ūkiai yra linkę suvartoti daugiau energijos nei mažesni, nors ir vienam asmeniui tenkantis energijos kiekis dažnai būna mažesnis. Be to, vyresnio amžiaus žmonės yra linkę vartoti daugiau energijos nei jaunesni. Verta paminėti, kad atliktų tyrimų rezultatai rodo, jog nors didesnis asmenų skaičius namų ūkyje yra susijęs su didesnėmis vartojimo išlaidomis, tačiau išlaidos, tenkančios vienam asmeniui, augant gyventojų skaičiui, turi tendenciją mažėti [30]. Lund'as ir Derry [31] remiantis Nacionalinės maisto apklausos, vykdytos Didžiojoje Britanijoje, duomenimis, nustatė, kad analizuojant namų ūkių maisto vartojimo modelius, svarbūs požymiai yra namų ūkių gyvenamoji vieta, jų sudėtis, amžius bei būsto nuosavybės forma. Pabrėžtina, kad vertinant ryšį tarp lyties ir energijos vartojimo, atrastas ryšys yra silpnas arba statistiškai nereikšmingas [20].

Socialiniai požymiai, tokie kaip aukštesnis išsilavinimo lygis, lemia, kad gyventojai dažniau skiria savo pajamas vartojimo tikslams [3]. Varlamovos ir Larionavos [32] atlikta su Ekonominio bendradarbiavimo ir plėtros organizacijos (angl. *Organisation for Economic Co-operation and Development*, OECD) narėmis analizė, taikant daugialypę tiesinę regresiją, parodė teigiamą ryšį tarp gyventojų aukštojo išsilavinimo turėjimo ir jų didesnio polinkio vartoti. York'as [29] prideda, kad gyvenimo būdas taip pat turi poveikį energijos vartojimui, nes yra randama skirtumų tarp miesto ir kaimo gyventojų kuro rūšių pasirinkimų. Miestuose gyvenančių asmenų energijos vartojimas dažniausiai yra mažesnis, nes didžioji dauguma gyvena daugiabučiuose su prieiga prie švarių, efektyvių energijos šaltinių per elektros tinklą, gamtinių dujų sistemą ar centralizuotą šildymo tinklą. Tuo tarpu kaimuose gyvenantys asmenys renkasi anglis ar biomasę, kas didina energijos vartojimo poreikį. Ding'as ir kt. [22] pritaria ir pastebi, kad namų ūkių vartojimo elgsena priklauso ne tik nuo geografinės padėties, bet ir nuo klimato sąlygų. Jų tyrimų metu buvo atrasta, kad padidėjus metinei vidutinei temperatūrai 1 %, energijos vartojimas vienam gyventojui sumažėtų iki 0,823 %, kai tiek pat išaugęs urbanizacijos lygis energijos vartojimą vienam gyventojui sumažintų apie 0,54 %. Arapova [28] pabrėžia, kad plečiantis vidurinajai klasei, vartojimo potencialas jų gyvenamajame regione auga. Frederiks'as, Stenner'is ir Hobman'as [20] prideda, kad namų ūkių būsto pasirinkimas teigiamai veikia energijos suvartojimą, t. y., didesni būstai paprastai sunaudoja daugiau energijos, kai asmenys, gyvenantys individualiuose namuose ar kotedžuose, taip pat sunaudoja daugiau energijos nei tie, kurie gyvena daugiabučiuose. Pavlisos [33] atlikta kiekybinė lyginamoji analizė pagal Prancūzijos namų ūkių duomenis atskleidė, kad vartojimo prioritetai skiriasi priklausomai nuo namų ūkio pasirinktos profesijos, pavyzdžiui, verslo profesionalai daugiau išleidžia išvaizdą gerinančioms prekėms ir paslaugoms, o švietimo specialistai – kultūrinėms prekėms, tokioms kaip knygos ar filmai. Edinak'as ir Shirov'as [34] remiantis atlikta Rusijos namų ūkių analize, teigia, kad mažėjanti žemos kvalifikacijos darbo dalis užimtumo struktūroje teigiamai veikia namų ūkių vartojimo apimtį augimą. Khamis ir kt. [35] tirdami Rusijos namų ūkių alkoholio vartojimą nustatė, kad asmens

religingumas neigiamai koreliuoja su alkoholio vartojimu, kai tuo tarpu tėvų auklėjimo stilius turi prieštarinę poveikį.

Magrabi'is ir kt. [2] teigia, kad namų ūkiai vartoja ne būtinai pagal poreikį, bet ir vadovaudamiesi kitais kriterijais, tokiais kaip visuomenės ar mokslo nuomonė arba išoriniai standartai. Namazkhan'as, Albers'as ir Steg'as [26] pritaria, kad asmenų psichologiniai veiksniai yra svarbūs analizuojant vartojimo struktūrą. Tyrimo metu mokslininkai psichologinius veiksnius įvertino apklausos būdu, kurioje buvo analizuojama namų ūkių nuomonė apie 4 vertybes – biosferinės, egoistinės, hedonistinės ir altruistinės. Biosferinių vertybių identifikavimą apėmė respondentų požiūris į aplinkos apsaugą ir išteklių tausojimą, egoistinių – klausimai susiję su materialinių dalykų, pinigų svarba, hedonistinių – mėgavimosi gyvenimu, laisvalaikio praleidimo svarba ir altruistinių – požiūris į lygybę, taiką ar socialinį teisingumą. Jų tyrimo rezultatai parodė, kad gyvenantys didesniuose būstuose ir pasižymintys hedonistinės vertybėmis yra linkę vartoti mažiau dujų [26]. Tuo tarpu mažesniuose būstuose gyvenantys namų ūkiai ir pasižymintys egoistinėmis vertybėmis yra taip pat labiau linkę vartoti mažiau dujų. Pavlisos [33] tyrime nustatytas teigiamas ryšys tarp verslo profesionalų ir jų polinkio daugiau išleisti su išvaizda susijusioms prekėms gali būti aiškinamas psichologiniu aspektu, nes tokiu būdu yra siekiama socialinį statusą bei atitikti profesinės aplinkos išvaizdai keliamus reikalavimus. Šiuolaikiniame technologijų pasaulyje vis svarbesniu aspektu tampa tinklaraštininkų poveikis vartotojų pirkimo elgsenai [33]. Mokslininkų Serman'o ir Sims'o [36] atliktos apklausos, kurioje dalyvavo 202 Londone gyvenantys asmenys, rezultatai atskleidė, kad tokie veiksniai kaip pasitikėjimas, ankstesnė patirtis, suvokiama nauda ar socialinis patrauklumas reikšmingai veikia vartotojų ketinimus pirkti pagal tinklaraštininkų pateikiamas rekomendacijas.

Vartojimo galimybes riboja ne tik asmens pajamos, bet ir jo turimas turtas [37]. Ščupakovaitė [6] teigia, kad turint daugiau turto, asmeninis vartojimas yra didesnis. Magrabi'is ir kt. [2] pritaria, pabrėždami, kad turtas ir vartojimo kreditai gyventojams suteikia galimybę vartoti nenaudojant pajamų, nors ir pastebi, jog šie veiksniai gali turėti įtakos būsimam vartojimui sumažėjimui dėl finansinių aktyvų išsekimo, ilgalaikių prekių nusidėvėjimo ar skolų iš būsimų pajamų grąžinimo. May, Nodari'is ir Rees'as [38] aiškina, kad išaugus turtui namų ūkiai turi daugiau išteklių palaikyti vartojimo įpročius per visą savo gyvenimą. Verta paminėti, kad vartojimo reakcija į pokyčius turte gali pasireikšti dėl daugelio priežasčių. Viena iš priežasčių gali būti, kad būsto savininkai, dabartinį savo vartojimą pakoreguos daugiau nei būsimi būsto pirkėjai, nes jų likvidus turtas už būsto pardavimą padidėja iš karto, kai būsimų pirkėjų išlaidos slypi ateityje. Didesnis namų ūkių turtas gali būti priežastis didesniam skolinimuisi dėl siekio sušvelninti gaunamų pajamų pokyčius. Akcijų, ar kito finansinio turto vertės kilimas taip pat gali paskatinti didesnę namų ūkių vartojimą, jei tikimasi, kad šio turto vertė ir toliau augs. Kaip teigia May, Nodari'is ir Rees'as [38], ištirti tikslią turto ir vartojimo sąveiką yra sudėtinga, bet ryšys tarp vartojimo ir asmens turimo turto yra tvirtai pagrįstas.

Ščupakovaitė [6] teigia, kad padidėjęs namų ūkių taupymas trumpuoju laikotarpiu mažina poreikį vartoti. Visgi ilguoju laikotarpiu santaupos atstoja investicijas, kurios gali būti priežastis padidėjusiam vartojimui ateityje. Verta atsižvelgti į tai, kad labiau pasiturintys namų ūkiai gali taupyti daugiau, kai mažiau pasiturintys – gali visai netaupyti ir visas savo pajamas skirti dabartiniam vartojimui [11].

Kredito prieinamumą labiausiai lemia kredito rinkos liberalizavimas, kas reiškia finansų sektoriui ir kredito rinkoms taikomų apribojimų mažinimą. Kuo didesnis yra prieinamumas prie kredito, tuo asmenys yra labiau linkę imti paskolas ar kredito korteles [39]. Soman'o ir Cheema'o [40] atlikto

tyrimo apie paskolų limitus metu buvo išsiaiškinta, kad kuo namų ūkių kredito limitas didesnis, tuo labiau didinamos ir išlaidos vartojimui. Galimybės pasiimti paskolą vartojimo tikslais poveikis ypač stiprus tarp mažiau patyrusių vartotojų, kurie kredito prieinamumą linkę laikyti patikimu būsimų pajamų šaltiniu [40]. Remiantis 2017 m. atlikto Kinijos namų ūkių finansiniu tyrimu buvo nustatyta, kad prieiga prie internetinio vartojimo kredito reikšmingai teigiamai veikia namų ūkių vartojimo lygį [41]. Pažymėtina, kad stipresnis teigiamas poveikis buvo pastebėtas tarp namų ūkių, kurie gauna mažesnes pajamas ar gyvena mažiau išsivysčiusiuose regionuose. Verta paminėti, kad lėšos, gautos iš vartojimo kredito, daugiausiai buvo panaudotos ne kasdienio vartojimo prekių įsigijimui.

3 lentelė. Mikrolygmens veiksnių, lemiančių vartojimą, galimo poveikio gyventojų išlaidoms apibendrinimas

Veiksnių pobūdis	Veiksnys	Poveikis	Autoriai
Ekonominiai	Pajamos	Tiesioginis poveikis. Didesnės namų ūkių pajamos lemia didesnį jų vartojimo lygį.	Ramanauskas, Jakaitienė [3]; Gedminaitė [11]; Ding ir kt. [22]; Piekut [12]
	Užimtumas	Nedarbas mažina pajamų lygį, dėl ko mažėja ir asmenų poreikis vartoti. Pilnu etatu dirbantys asmenys yra linkę vartojimui išleisti daugiau, nei dirbantys ne pilnu etatu.	Ganong, Noel [24]; Namazkhan, Albers ir Stegas [26]; Piekut [12]; Pologeogris [25]
	Taupymas	Asmenų turimų pajamų atidėjimas taupymui tiesiogiai mažina pajamų kiekį skirtą vartojimui.	Gedminaitė [11]; Ščupakovaitė [6]
	Kreditų naudojimas	Pasiskolintos lėšos padidina asmenų disponuojamas pajamas trumpuoju laikotarpiu, dėl ko vartojimo lygis didėja, ypač ne kasdinių prekių vartojimui.	Soman ir Cheema [40]; Pettinger [39]; Li, Song, Wu, Huang [41]
	Namų ūkių turtas	Turint daugiau turto, asmeninis vartojimas yra didesnis.	Sun ir Lv [37]; Ščupakovaitė [6]; May, Nodari ir Rees [38]
Demografiniai	Amžius	Teigiamas ryšys tarp amžiaus ir polinkio vartoti. Gali egzistuoti apverstos U formos ryšys – vidutinio amžiaus žmonės vartoja daugiau nei jaunesni ar vyresni.	Ramanauskas ir Jakaitienė [3]; York [29]; Bugheanu ir Strachinaru [27]; Lee, Ong ir Lee [30]
	Namų ūkio sudėtis	Nevienareikšmis poveikis. Nors dauguma tyrimų rodo, kad didesni namų ūkiai išleidžia daugiau, tačiau atrandamos ir priešingos tendencijos, t. y., neturinčių šeimyninių įsipareigojimų asmenų vartojimo išlaidos didesnės.	York [29]; Varlamova ir Lariopnava [32]; Bugheanu ir Strachinaru [27]
	Lytis	Silpnas arba statistiškai nereikšmingas ryšys	Frederiks, Stenner ir Hobman [20]
Socialiniai	Išsilavinimo lygis	Aukštesnis išsilavinimo lygis yra siejamas su didesniu vartojimu.	Ramanauskas, Jakaitienė [3]; Varlamova ir Lariopnava [32]
	Gyvenimo būdas	Gyvenimas mieste yra susijęs su didesniu bendro vartojimo potencialu, tačiau mažesniu poreikiu vartoti energiją.	York [29]; Ding ir kt. [22]; Arapova [28]
	Profesijos pasirinkimas	Teigiamas ryšys tarp aukštesnės kvalifikacijos profesijų ir vartojimo išlaidų dydžio.	Edinak ir Shirov [34]
	Finansinis raštingumas	Teigiamas poveikis tarp mažiau patyrusių vartotojų ir polinkio naudotis paskolomis	Soman ir Cheema [40]; Li, Song, Wu, Huang [41]
Psichologiniai	Vertybės	Hedonistinės ir egoistinės vertybės yra siejamos su polinkiu mažesnę išlaidų dalį skirti būsto išlaikymo išlaidoms.	Namazkhan, Albers ir Stegas [26]
	Visuomenės standartai	Socialinio statuso palaikymas yra siejamas su didesnėmis išlaidomis įvaizdžio formavimui.	Magrabi [2]; Pavlisa [33]
	Nuomonė	Asmenų polinkis pirkti yra siejamas su mėgstamų tinklaraštininkų rekomendacijomis.	Serman ir Sims [36]

Mikrolygmeniui priskiriamos namų ūkio charakteristikos buvo apibendrintos (žr. 3 lent.) ir suskirstytos pagal jų ekonominį, demografinį, socialinį bei psichologinį pobūdį. Galima daryti išvadą, kad mikrolygmens veiksniai yra susiję su namų ūkių individualumu ir jų ypatybėmis.

1.3.2. Makrolygmens veiksnių įtaka vartojimo išlaidoms

Kozlov'as [42] pastebi, kad namų ūkių išlaidos dažnai yra finansuojamos skolintais pinigais, kas anksčiau buvo ypač aktualu dideliems pirkiniams įsigyti, tačiau pastaraisiais dešimtmečiais mažėjanti palūkanų norma skatino pigesnę skolinimąsi, o kartu ir prieinamesnės kredito pasiėmimo galimybės lėmė asmenų pinigų skolinimąsi ir kasdieniškiesiems vartojimui finansuoti. Kozlov'o [42] atliktų tyrimų metu buvo atrasta, kad 50% sumažėję būsto paskolų mokėjimai lėmė 35% padidėjimą perkant automobilius. Manasseh'as ir kt. [43] pastebi, kad mažesnė palūkanų norma skatina namų ūkių skolinimąsi didesnio vartojimo ar investavimo tikslais, kas sukelia padidėjusių išlaidų poveikį visos ekonomikos kontekste. Didesnė palūkanų norma, reiškia vartojimo išlaidų mažėjimą, nes dėl aukštesnio skolinimosi standarto vartotojai gali nebeturėti pakankamai disponuojamų pajamų vartojimo lygio didinimui, todėl palūkanų normos kitimas gali veikti ne tik vartotojus, bet ir įmones bei gamybos tempus. Palūkanų norma gali būti realioji ir nominalioji. Nominalioji palūkanų norma yra faktinė palūkanų norma, dėl kurios buvo susitarta ir kuri yra mokama, kai realioji palūkanų norma yra pakoreguota pagal infliaciją nominalioji palūkanų norma (Europos centrinis bankas⁶).

Kainų lygiui kylant, didėja prekių ir paslaugų savikaina, dėl ko mažėja pinigų perkamoji galia, o to rezultatas – sumažėjęs bendras vartojimas [44]. Bendrą kainų lygio kilimą, kuris turi įtakos piniginio vieneto perkamosios galios kritimui, atspindi infliacijos veiksnys. Pinigų perkamoji galia yra suma, už kurią namų ūkiai gali faktiškai nusipirkti prekių ar paslaugų (Europos centrinis bankas⁶). Namų ūkių vartojimui įtakos turi ir infliacijos lūkesčiai, dėl kurių namų ūkiai gali padidinti savo vartojimą, jei tikisi, kad kainos ateityje toliau augs arba jį sumažinti, jei tikisi, kad ateityje kainos stabilizuosis arba mažės, dėl ko būtų atidėtas sprendimas pirkti dabartiniu momentu. Infliacijos lūkesčiai taip pat mažėja, jei namų ūkiai tampa optimistiškesni dėl savo realiųjų pajamų, o tai nutinka, kai infliacijos tempai yra suderinti su gaunamų pajamų (pvz., darbo užmokesčio) didėjimu [45]. Mandeya ir Sin'o -Yu'so [46] atliktų tyrimu metu buvo atrasta, kad infliacija kenkia ekonomikos augimui, nes ekonominio neapibrėžtumo sąlygomis skatina namų ūkius taupyti, kas sumažina paklausą, dėl ko sulėtėja ir ekonominis augimas. Radu'us, Kenny ir Reute'ė [44] taip pat pastebi, kad esant pastoviai nominaliajai palūkanų normai, realioji palūkanų norma gali sumažėti dėl numato infliacijos padidėjimo, kas skatintų dabartinį vartojimą bei bendrą paklausą, dėl sumažėjusių gyventojų paskatų taupyti ateičiai. Charalampakis ir kt. [47] teigia, kad infliacija veikia vartojimo išlaidų struktūrą, nes mažas pajamas gaunantys namų ūkiai yra mažiau apsaugoti nuo infliacijos poveikio ir pakilus kainų lygiui, turės skirti dar didesnę savo pajamų dalį būtiniausioms prekėms įsigyti. Ding'as ir kt. [22] pateikia konkretų pavyzdį, t. y., bendrosiose vartojimo išlaidose būsto energijos išlaidos didėja kartu su elektros tarifų augimu. Be to, buvo atrastas ryšys ir tarp namų ūkių būsto išlaikymo išlaidų bei namų ūkių priežiūros išlaidų, kadangi didžiausią įtaką energijos vartojimo augimui turėjo namų ūkių elektrinių prietaisų pirkimas, kurių kainos taip pat darė didesnę įtaką energijos vartojimui nei pati elektros kaina.

Ghosh'as ir kt. [48] Brazilijoje atliktų tyrimų metu buvo išsiaiškinta, kad vartotojų pasitikėjimas ateitimi yra svarbus vartotojų išlaidas paaiškinantis veiksnys, nes jis, kartu su kitais

⁶ https://www.ecb.europa.eu/ecb-and-you/explainers/tell-me/html/nominal_and_real_interest_rates.lt.html

makroekonominiais veiksniais, suteikia papildomos informacijos apie vartojimo elgseną. Vartotojų pasitikėjimo rodiklis vartotojų elgsenai turi didesnę nuspėjamąją galią ekonominio nuosmukio metu. Vartotojų pasitikėjimo rodiklis nurodo namų ūkių pasitikėjimą šalies ekonomine ateitimi ir yra apskaičiuojamas remiantis namų ūkių atsakymais į konkrečius su jų finansine padėtimi šiuo metu ir ateityje bei ekonominiu stabilumu susijusiais klausimais (Oficialios statistikos portalas⁷). Remiantis nuolatinių pajamų hipoteze, namų ūkiai vartojimui naudoja pastovias pajamas ir visada apgalvotai formuoja su jomis susijusius lūkesčius, todėl remiantis vartotojo turima atitinkama informacija, galima pastebėti, jog $(t+1)$ laikotarpio sprendimas vartoti jau yra numatytas t laikotarpio vartojime (C_t), kas rodo, jog $(C_{t+1}-C_t)$ yra vartojimo elgsenos pasireiškimas $(t+1)$ periodu, kuris buvo pagrįstas papildoma informacija apie turimas pastovias pajamas $(t+1)$ periode [49].

Pettinger [3937] teigia, kad vartotojo turimas būstas yra svarbiausia turto forma, o bendras būstų kainų kilimas rodo vartotojų pasitikėjimą ir turto vertės didėjimą, kas skatina didinti vartojimo lygį. Tą patvirtina ir Berger'io bei kt. [50] atlikti tyrimai, kurių metu buvo išvystyta vartojimo kitimo reakcijos į būstų kainų šokus skaičiavimui skirta formulė, t. y., ribinis polinkis vartoti laikinąsias pajamas yra padauginamas iš būsto vertės. Skaičiavimai parodė, kad vartojimo atsakas į būstų kainų pokyčius yra labai stiprus. Andersen'as ir Petersen'as [51] pastebi, kad būstų savininkai, ima naujas hipotekos paskolas ir didina savo vartojimą, kai būstų vertė kyla. Poveikis vartojimui dar labiau išauga tarp fiksuotų palūkanų hipotekos skolintojų, kurie siekia paskolų refinansavimo su tikslu fiksuoti mažesnę rinkos palūkanų normą, todėl pinigų politika gali sustiprinti ryšį tarp būsto kainų ir vartotojų išlaidų, paveikdama hipotekos paskolų palūkanų normas. Pettinger [39] pabrėžia, kad brangesnė nuoma ar didesnės įsigyjamų būstų kainos gali sumažinti jaunesnių gyventojų bendras vartojimo išlaidas, dėl jų siekio taupyti būsto užstatui ir didesnės pajamų dalies išleidžiamos nuomai.

Mokesčių našta, mažina gyventojų disponuojamąsias pajamas, todėl mažina ir vartojimo lygį. Alm'as ir Ganainy'is [52] teigia, kad vienas iš pagrindinių motyvų vartojimo mokesčio įvedimui buvo tai, kad tikėtasi, jog toks mokestis gyventojus paskatins mažiau vartoti ir labiau taupyti, kas turėtų generuoti didesnę ekonomikos augimą, tačiau empiriniai tyrimai rodo, kad vartojimo mokesčiai labai ribotai veikia gyventojų polinkį taupyti. Šių mokslininkų atlikti tyrimai parodė, kad Europos Sąjungos šalyse, pridėtinės vertės mokesčio tarifą padidinus vienu procentiniu punktu trumpuoju laikotarpiu bendras vartojimo lygis sumažėja apie vieną procentą, o ilguoju laikotarpiu – truputį daugiau nei vieną procentą. Sun'as ir Lv'as [37] Kinijoje atlikto tyrimo metu buvo nustatyta, kad progresyvinis gyventojų pajamų ir nekilnojamojo turto mokesčiai iš tikrųjų teigiamai veikia privatų vartojimą, dėl geresnio mokesčių naštos paskirstymo tarp skirtingas pajamas uždirbančių gyventojų. Visgi apyvartos mokestis (įskaitant vartojimo mokesčius) stabdo privataus vartojimo lygį, nes mokesčių našta, tenkanti didesnes pajamas gaunantiems namų ūkiams, yra mažesnė. Verta paminėti, kad atliktas tyrimas parodė, kad mokesčių poveikis labiau lemia kaimo gyventojų vartojimo išlaidų pasikeitimą nei mieste gyvenančių namų ūkių pasikeitimus vartojimo išlaidose. Arapova [28] pabrėžia, kad mokesčiais surinktos lėšos vėliau gali tapti vyriausybės išlaidomis, kas yra veiksminga priemonė šalies plėtros augimui, grindžiamam vartojimu.

Palūkanų normą, infliaciją, vartotojų pasitikėjimo rodiklį, turimo turto vertę ir vyriausybės politiką galima priskirti makroekonominiam lygmeniui, nes šie veiksniai priklauso nuo bendrų šalyje vykstančių ekonominių procesų. Susisteminta informacija (žr. 4 lent.) apibendrina mokslinėje literatūroje apžvelgtų vartojimą lemiančių veiksnių poveikį.

⁷ <https://osp.stat.gov.lt/covid19-statistika/itaka-ekonomikai/vartotoju-nuomoniui-tyrimai>

4 lentelė. Makrolygmens veiksnių, lemiančių vartojimą, galimo poveikio gyventojų išlaidoms aptarimas

Veiksny	Poveikis	Autoriai
Palūkanų norma	Mažėjanti realioji palūkanų norma skatina skolinimąsi, dėl to didėja vartojimas.	Kozlov [42]; Manasseh ir kt. [43]
Infliacija	Tiesioginis ir netiesioginis poveikis. Kylant kainoms, mažėja pinigų perkamoji galia, t. y., mažėja disponuojamos pajamos, kurios turi tiesioginį poveikį bendro vartojimo mažėjimui	Radu, Kenny, Reute [44]; Mandeya ir Sin-Yu [46]
Infliacijos lūkesčiai	Infliacijos lūkesčiai tiesiogiai skatina vartojimą, jeigu yra tikimasi, kad kainos ateityje augs. Ekonominio neapibrėžtumo sąlygomis inflacijos lūkesčiai skatina gyventojus labiau taupyti nei vartoti. Dėl numato inflacijos padidėjimo gali pradėti mažėti realioji palūkanų norma, kuri paskatintų skolinimąsi, dėl kurio didėtų namų ūkių pajamos.	Coibion ir kt. [45]
Vartotojų pasitikėjimo rodiklis	Augant vartotojų pasitikėjimui ateitimi, išlaidos vartojimui taip pat auga. Didesnė nuspėjamoji galia pasireiškia ekonominio nuosmukio metu.	Hall [49]; Ghosh ir kt. [48]
Būstų kainos	Būstų kainų kilimas rodo vartotojų pasitikėjimą ir turto vertės didėjimą, tai skatina didinti vartojimo lygio augimą.	Pettinger [39]; Berger ir kt. [50]; Andersen, Petersen [51].
Vyriausybės politika	Abipusis poveikis. Taikomi mokesčiai mažina gyventojų disponuojamąsias pajamas, todėl mažina ir galimybę daugiau vartoti, ypač tai pastebima vartojimo mokesčių atveju. Mokesčiais surinktos lėšos vėliau gali būti naudojamos vyriausybės išlaidoms, skatinančioms namų ūkių vartojimą.	Alm, Ganainy [52], Arapova [28]; Sun, Lv [37]

Vartojimas yra sudėtingas reiškinys, kurį gali veikti daug ir įvairių veiksnių: vyriausybės sprendimai, ekonominė šalies būklė, kultūra, technologijos, aplinkos sąlygos, tvarumo siekiai, asmens psichologija ar individualumas ir kita [53]. Visgi buvo nustatyta, kad didžiausią poveikį namų ūkių vartojimui turi jų uždirbamos pajamos. Vartojimo išlaidų didinimui turi poveikį ir kiti mikrolygmens požymiai: namų ūkių sprendimas dirbti pilnu etatu, netaupyti, naudotis paskolomis, turėti turtinę nuosavybę, siekti aukštesnio išsilavinimo lygio, gyventi mieste, pasirinkti aukštesnės kvalifikacijos profesiją, taip pat aktualu ir jų vyresnis amžius, turimos vertybės, visuomenėje egzistuojantys standartai bei teigiama nuomonė apie mėgstamų tinklaraštininkų rekomendacijas. Makrolygmens požymiai, turintys įtakos vartojimui, apima palūkanų normą, inflacijos lygį ir jo lūkesčius, vartotojų pasitikėjimą šalies ekonomine padėtimi ateityje, vyraujančias būsto kainas bei taikomą vyriausybės politiką.

1.4. Mikrolygmens ir egzogeninių veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo metodų apžvalga

Dėl savo paprastumo ir aiškos interpretacijos dažniausiai yra taikomi **daugialypės tiesinės regresijos modeliai**, kurių parametrai yra vertinami mažiausių kvadratų (angl. *Ordinary Least Squares*, OLS) metodu. Paik [54] pateikė regresijos modelio lygtį:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_K X_{iK} + \epsilon_i ; \quad (3)$$

čia y_i – priklausomas kintamasis (vartojimo išlaidos), X_i – nepriklausomi kintamieji, β_i – koeficientai, β_0 – laisvasis narys, ϵ_i – paklaida.

Paik [54] taikytu modeliu siekė paaiškinti svarbiausio JAV ekonomikos augimą skatinančio veiksnio – vartojimo – pasikeitimų priežastis. Darbe buvo aiškinamasi vartojimo priklausomybė nuo namų ūkių disponuojamųjų pajamų, naftos kainos, palūkanų normos ir recesijos per 1950 – 2013 metų laikotarpį. Kintamųjų reikšmingumas buvo įvertintas t-testo p-reikšme, pagal kurią buvo nustatyta ar nepriklausomas kintamasis turi statistiškai reikšmingą ryšį su priklausomu kintamuoju. Atrasta, kad

1 % padidėjus namų ūkių disponuojamosioms pajamoms, vartojimas padidėja 0,67 %, kai 1 procentiniu punktu padidėjus palūkanų normai, vartojimas padidėja 0,03 %, o naftos kainai padidėjus 1 JAV doleriu, vartojimas sumažėja per 0,01 %, kai tuo tarpu recesijos laikotarpiu išlaidos vartojimui sumažėja net 1,24 %. Modelio kokybės vertinimas parodė, kad pakoreguotas determinacijos koeficientas (R^2) siekė 0,72, Durbino-Vatsono statistikos testas patvirtino teigiamos serijinės koreliacijos nebuvimą, Vaito statistika neparodė heteroskedastiškumo, o žemos VIF reikšmės patvirtino multikolinearumo problemos nebuvimą, dėl to modelio rezultatai buvo pripažinti statistiškai patikimais ir tinkamai pritaikytais. Verta pažymėti, kad neįtraukus recesijos laikotarpio kaip paaiškinamojo kintamojo, pakoreguoto R^2 reikšmė sumažėjo iki 0,64. Tiesinės regresijos modelis buvo pritaikytas ir vertinant Malaizijos namų ūkių vartojimo elgseną 1970 – 2014 m. laikotarpiu pagal vidutines namų ūkių pajamas ir vartotojų kainų indeksą [1]. Nustatyta, kad 1 % padidėjus namų ūkių pajamoms, Malaizijos namų ūkių vartojimo išlaidos padidėja per 0,98 %, kai atitinkamai padidėjus infliacijai, kuri buvo išreikšta vartotojų kainų indeksu, namų ūkių vartojimo išlaidos Malaizijoje sumažėja per 0,292 %. Abu nepriklausomi kintamieji gali paaiškinti 86 % priklausomo kintamojo variacijos, nes modelio tikslumo vertinimo metrika (R^2) siekė net 0,86. Visgi, remiantis individualiais t-testo rezultatais, buvo nustatyta, kad vartotojų kainų indeksas nebuvo statistiškai reikšmingas kintamasis aiškinant namų ūkių vartojimo išlaidas. Be to, nors Durbino ir Watsono, Parto ir Glejsero statistikų testais nebuvo nustatyta atitinkamai autokoreliacijos ar heteroskedastiškumo problemų, tačiau modelyje egzistavo multikolinearumo problema, nes buvo nustatytas itin stiprus tiesinis ryšys tarp nepriklausomųjų kintamųjų, t. y., Pirsono koreliacijos koeficientas tarp pajamų ir infliacijos siekė 0,981. Verta paminėti, kad siekiant pagerinti regresijos modelių kokybę, tiek JAV, tiek Malaizijos duomenų analizėje buvo naudojamos logaritmuotos vartojimo išlaidos. Varlamova ir Larionova [32] atliko tyrimą, kuriame analizavo OECD šalių bendrųjų vartojimo išlaidų bei vienos iš jų struktūrinių dalių – sveikatos išlaidų – priklausomybę nuo įvairių makroekonominių ir sociodemografinių veiksnių. Naudojant tiesinės regresijos modelį buvo atrasta, kad importas, trumpalaikės palūkanų normos, infliacija, mokesčiai, pensininkų išlaikymo santykis, valdžios išlaidos, aukštasis išsilavinimas ir namų ūkių pajamos statistiškai reikšmingai veikė vartojimo išlaidas. Šalies importo padidėjimas turėjo neigiamą poveikį namų ūkių bendrosioms vartojimo išlaidoms. Tuo tarpu atskiro sveikatos išlaidų tiesinės regresijos modelio rezultatai parodė, jog šias išlaidas statistiškai reikšmingai veikė namų ūkių pajamos, bendrosios valdžios išlaidos, trumpalaikės palūkanų normos ir aukštojo išsilavinimo turėjimas. Verta paminėti, kad bendrosios vartojimo išlaidos buvo logaritmuotos, kai sveikatos išlaidų duomenims tokia transformacija nebuvo taikoma. Tiesinių daugialypių regresijos modelių, sudarytų bendrojo vartojimo ir sveikatos išlaidų pokyčiams įvertinti, pakoreguoti determinacijos koeficientai atitinkamai siekė 0,80 ir 0,75, kas iš visų kurtų modelių buvo didžiausia aiškinamoji galia. Verta paminėti, kad abiejų išlaidų atvejais didžiausius koeficientus turėjo namų ūkių pajamos.

Mokslininkai nagrinėjo ir namų ūkių išlaidų struktūrines dalis atskirai. Singh'as ir Kumar'as [55] tyrė kaip namų ūkių energijos, t. y., būsto priežiūrai ir transportui, suvartojimas priklauso nuo jų išsilavinimo lygio, vidutinio amžiaus, gyvenamosios erdvės ploto, turimos žemės ploto ir uždirbamų pajamų kiekio. Taikant pažingsninės tiesinės regresijos modelį (angl. *Stepwise Regression*), kuris užtikrina tik statistiškai reikšmingų kintamųjų patekimą į modelį ir kartu tvarkosi su galima multikolinearumo problema, buvo sukurtas bendras modelis, apimantis visas Indijos valstijas. Bendro modelio aiškinamoji galia (R^2) siekė 0,74 ir atskleidė, kad namų ūkių išsilavinimas turi didžiausią teigiamą poveikį energijos suvartojimui, kuriam teigiamą poveikį taip pat turi gyventojų gyvenamosios erdvės, žemės plotai ir pajamos, kai tuo tarpu atvirkštinį poveikį – amžius. Didesni

modelių tikslumai (pakoreguotas R^2 siekė 0,87 – 0,93) buvo gauti, kai pažingsninės tiesinės regresijos modeliai buvo sukurti atskirai trims teritorijų tipams – lygumai, pakrantėms ir kalvotoms vietovėms. Buvo nustatyta, kad, nepaisant geografinių padalijimų, gyventojų pajamos daro teigiamą įtaką energijos vartojimui, kai tuo tarpu lygumose ir kalnuotose vietovėse gyvenančiųjų energijos suvartojimą teigiamai veikia gyvenamosios erdvės plotas, o pakrantėje gyvenančiųjų energijos vartojimui jis neturėjo statistiškai reikšmingos įtakos. Be to, pakrantėse ir kalnuotose vietovėse gyvenančiųjų energijos suvartojimui didžiausią poveikį turėjo išsilavinimo lygis, kai lygumose gyvenančiųjų energijos vartojimui šis kintamasis pasirodė statistiškai nereikšmingas ir pagrindinį poveikį turėjo gyventojų turimas žemės plotas. Verta paminėti, kad šiame tyrime nebuvo patikrintos patikimos tiesinės regresijos taikymui reikalingos homoskedastiškumo ir liekanų nepriklausomumo prielaidos. Roshchina [23] alkoholio vartojimo kiekiui įvertinti, remdamasi Rusijos ilgalaikio stebėjimo tyrimo (RLMS) duomenimis, surinktais 1994 – 2011 m. laikotarpiu, taikė Tobit regresijos modelį. Toks modelis gali būti laikomas tiesinės regresijos atmaina, t. y., cenzūruotu tiesinės regresijos modeliu, nes yra taikomas tais atvejais, kai daugelis priklausomojo kintamojo reikšmių yra lygios nuliui. Atlikto tyrimo kontekste nulinės reikšmės atspindėjo itin mažą arba visišką alkoholio nevartojimą tarp apklaustųjų, dėl to ir buvo pasirinktas Tobit regresijos metodas, kuriuo pirmiausia yra apskaičiuojama tikimybė, kad stebiny bus lygus nuliui, o po to yra modeliuojamas latentinis (paslėptas) priklausomasis kintamasis, jei stebimoji reikšmė yra didesnė už nulį. Verta pastebėti, kad Tobit metodu nulinės reikšmės nėra visiškai ignoruojamos, nes jos yra įtraukiamos per tikimybės komponentą ir tokiu būdu visi stebiniai yra įtraukiami į skaičiavimus. Pagrindiniai rezultatai parodė, kad vyrai vartoja daugiau alkoholio nei moterys, o amžiaus poveikis alkoholio vartojimui yra apverstos U formos, t. y., vidutinio amžiaus žmonės vartoja daugiau nei jaunesni ar vyresni. Ypač svarbūs teigiamą įtaką turintys buvo pajamų, šeimos dydžio, gyvenančių mieste, nedarbo lygio kintamieji, kai atvirkštinę įtaką – aukštesnis išsilavinimo lygis, religingumas ir kainų lygis. Kadangi alkoholio vartojimo vertinimo tyrimo kontekste buvo taikytas cenzūruotas regresijos modelis, t. y., Tobit, tradiciniai modelio kokybės vertinimo metodai negalėjo būti taikomi dėl sumažintos stebėjimų dalies, dėl kurios pilnai nėra įvertinamas priklausomojo kintamojo pasiskirstymas, kas lemia rezultatų nepatikimumą. Dėl šios priežasties modelio tinkamumas buvo įvertintas kokybiniu būdu, t. y., jo rezultatus lyginant su kitų mokslininkų gautais ir publikuotais rezultatais mokslinėje literatūroje.

Kitas poveikiui vertinti dažnai taikomas ekonometrinis metodas yra **kvantilinė regresija**. Mignouna ir kt. [57] atliko tyrimą, kuriame analizavo 2012 m. duomenis apie 1 400 atsitiktinai atrinktų namų ūkių, gyvenusių Nigerijos ir Ganos kaimo vietovėse, bendrųjų vartojimo išlaidų priklausomybę nuo socioekonominių veiksnių. Tyrime buvo lyginami du ekonometriniai metodai, t. y., tiesinė daugialypė regresija, paremta paprasčiausių mažiausių kvadratų metodu, ir kvantilinė regresija, kuriai buvo taikytas mažiausių absoliučių nukrypimų nuo pasirinkto kvantilio įverčio metodas. Mignouna ir kt. [57] pateikia kvantilinės regresijos išraiškos lygtį:

$$Kvant_q(Y_i|X_i) = X_i a_q + \varepsilon_{qi} ; \quad (4)$$

čia X_i – aiškinamųjų kintamųjų vektorius (nuo 1 iki n), a_q – parametų vektorius, $Kvant_q(Y/X)$ – sąlyginis priklausomojo kintamojo y q -asis kvantilis, pagal aiškinamąjį kintamąjį X .

Kvantilinės regresijos metodas pateikia įvairius įverčius, apibūdinančius priklausomojo ir aiškinamųjų kintamųjų ryšį skirtinguose priklausomojo kintamojo pasiskirstymo taškuose, t. y., kvantiliuose, kas leidžia geriau suprasti aiškinamųjų kintamųjų ne tik vidutinį poveikį. Taikant kvantilinę regresiją, yra daroma prielaida, kad priklausomojo kintamojo sąlyginis kvantilis su

aiškinamaisiais kintamaisiais yra susijęs tiesiniu ryšiu, tačiau, lyginant su tiesine regresija, šis metodas pasižymi didesniu atsparumu heteroskedastiškumui ir ekstremalioms reikšmėms. Verta paminėti, kad 50-asis kvantilis atitinka sąlyginės medianos regresijos atvejį. Tyrimo metu buvo atsižvelgta į multikolinearumo problemą naudojant VIF [57]. Gauti kvantilinės regresijos įverčiai, apskaičiuoti pagal 25-ąją, 50-ąją ir 75-ąją sąlyginių vartojimo išlaidų pasiskirstymo kvantilius, atskleidė kiekvieno nepriklausomojo kintamojo poveikį skirtingiems išlaidų lygiams, t. y., namų ūkių vartojimui pagal žemesnes ar aukštesnes jų išlaidų lygio grupes. Kvantilinės regresijos rezultatų tinkamumui įvertinti buvo taikytas pseudo R^2 matas, kuris, skirtingai nei klasikinis determinacijos koeficientas, nėra grįstas dispersijos paaiškinimu ir matuoja, kiek sumažėja liekanų suma, kai į modelį yra įtraukiami aiškinamieji kintamieji, palyginti su baziniu modeliu, kuriame aiškinamieji kintamieji nėra įtraukti. Nors teoriškai pseudo R^2 reikšmė gali siekti vienetą, jos interpretacija skiriasi nuo klasikinio R^2 [58]. Mignouna ir kt. [57] tyrimo kontekste pseudo R^2 reikšmės tarp skirtingų kvantilių svyravo nuo 0,04 iki 0,07, kai pseudo R^2 reikšmės didėjo priklausomai nuo analizuoto kvantilio, kas rodė, jog modelis geriau paaiškino vartojimo išlaidas tiems namų ūkiams, kurie pateko į 75-ąją kvantilį pagal išlaidas, nei tiems, kurie pagal vartojimo išlaidas pateko į žemesnius kvantilius (25-ąją ir 50-ąją). Bendrai tyrimo rezultatai parodė, kad aiškinamųjų kintamųjų įtaka nėra vienoda visame vartojimo išlaidų pasiskirstyme, nes Nigerijoje, remiantis tiesinės regresijos modeliu, tik amžius, išsilavinimas ir namų ūkio dydis buvo reikšmingi veiksniai, kai kvantilinės regresijos modeliai, priklausomai nuo vartojimo išlaidų kvantilio, papildomai reikšmingais veiksniais nustatė namų ūkių pagrindinę veiklą ir turimos žemės plotą. Pabrėžtina, kad išsilavinimo lygis buvo reikšmingas abiejų modelių atveju tiek Nigerijoje, tiek Ganoje. Caglayan'as ir Astar'as [59] aiškinosi Turkijos namų ūkių vartojimo išlaidas lemiančius veiksnius taikant kvantilinę regresiją su kvantilėmis nuo 0,1 iki 0,9. Šiame tyrime įverčiai buvo apskaičiuoti naudojant pusiau logaritminę modelio formą, kas reiškė, kad vartojimo išlaidoms buvo pritaikytas logaritmas. Tyrimu nustatyta, kad namų ūkio galvos amžiui padidėjus vienu vienetu vartojimo išlaidos 10-ajame, 50-ajame ir 90-ajame kvantiliuose atitinkamai išauga per 0,49 %, 0,25 % ir 0,33 %, kai miesto gyventojų vartojimo išlaidos visuose kvantiliuose buvo didesnės nei kaimo gyventojų ir namų ūkio dydžio išaugimas vienu asmeniu vartojimo išlaidas visuose kvantiliuose padidina per 16,9 % – 10-ajame kvantilyje ir per 10,69 % – 90-ajame kvantilyje. Modelio tinkamumo mato – pseudo R^2 – reikšmės visuose kvantiliuose siekė 0,30 – 0,31.

Poveikio vertinimui naudojami ir **mašininio mokymosi modeliai**, kurie gali būti pranašesni, nes jų veikimas nėra apribotas tiesinei regresijai taikomoms prielaidomis. Šie modeliai gali tiksliau aptikti sudėtingus ir netiesinius ryšius tarp aiškinamųjų kintamųjų ir priklausomojo kintamojo bei įvertinti stiprias sąveikas tarp nepriklausomų kintamųjų, kurių tradiciniai tiesiniai modeliai nesugebėtų tinkamai atskleisti. Lee'ė, Ong'as ir Lee'ė [30] tyrė vienam asmeniui tenkančių vartojimo išlaidų priklausomybę nuo socioekonominių veiksnių, naudodami 2019 m. atliktos Malaizijos namų ūkių vartojimo apklausos duomenis, apėmusios 16 tūkst. namų ūkių. Tyrimo metu buvo lyginami du modeliai – tiesinė daugialypė regresija ir atsitiktinių miškų regresija (angl. *Random forest*). Puttanapong'as ir Lim'as [60] teigia, kad atsitiktinių miškų regresija yra vienas iš dažniausiai taikomų mašininio mokymosi algoritmų, kuris regresijos užduotims spręsti pasitelkia ansamblinę logiką, t. y., naudojami keli sprendimų medžių modeliai, duodantys atskiras prognozes, iš kurių galutiniame rezultate yra išvedamas vidurkis, kas padeda išvengti persimokymo problemos. Medžių įvairumui užtikrinti, kiekvienas iš jų yra apmokomas su skirtinga ir atsitiktinai atrinkta duomenų dalimi. Be to, atsitiktiniais miškais paremtas modelis gali apskaičiuoti nepriklausomų kintamųjų indėlį į bendrą netvarkos (angl. *impurity*) sumažėjimą, kas leidžia įvertinti kiekvieno iš požymių svarbą. Mosca ir kt. [61] priduria, kad priklausomybių tarp aiškinamųjų kintamųjų ir priklausomojo kintamojo

ištyrimui galima atlikti SHAP analizę. Lee'ė, Ong'as ir Lee'ė [30] teigimu, remiantis tiesinės regresijos rezultatais, svarbiausi namų ūkių vartojimo išlaidas lemiantys veiksniai yra pajamos, išsilavinimo lygis ir namų ūkio dydis, kai pajamos ir išsilavinimas daro teigiamą poveikį logaritmuotomis vienam gyventojui tenkančioms vartojimo išlaidoms, o didesnis namų ūkio dydis yra susijęs su mažesnėmis išlaidomis. Remiantis daliniu R^2 buvo nustatyta, kad vien tik pajamos tenkančios vienam gyventojui paaiškina 54,39 % visos variacijos, kai tuo tarpu namų ūkio dydis paaiškina 9,55 %, o išsilavinimo lygis – 8,72 %. Atsitiktinių miškų modelis buvo derinamas parenkant optimalius hiperparametrus, o po modelio apmokymo testavimo duomenų rinkinys buvo naudojamas tiek modelio veikimo kokybės vertinimui, tiek SHAP analizei. Pastaroji parodė, kad pajamos yra svarbiausias veiksnys, aiškinantis vartojimo išlaidas, kai namų ūkio dydis – antras pagal svarbumą. Be to, taikant atsitiktinių miškų regresijos modelį, buvo atrasta, kad tarp namų ūkio galvos amžiaus ir vienam asmeniui tenkančių vartojimo išlaidų egzistuoja netiesinis, apverstos U formos ryšys, kas galėjo būti priežastis, dėl kurios tiesinėje regresijoje amžiaus kintamasis buvo pripažintas nereikšmingu. Abiejų modelių prognozavimo kokybė buvo įvertinta taikant determinacijos koeficientą (R^2) ir vidutinę kvadratinę paklaidą (toliau – MSE). Nors atsitiktinių miškų modelis pagal testavimo duomenis pasiekė aukščiausią R^2 ir žemiausią MSE vertes, tačiau skirtumas nuo tiesinio modelio buvo nedidelis, t. y., tiesinio ir atsitiktinių miškų modelių R^2 ir MSE atitinkamai buvo 0,847 ir 0,848 bei 0,0454 ir 0,0453, kas rodo, jog dauguma svarbiausių socioekonominių veiksnių (ypač pajamos ir išsilavinimas) su vartojimu yra susiję tiesiniais ryšiais, dėl ko tiesinis modelis pasiekė panašų tikslumą kaip ir atsitiktinių miškų modelis.

Puttanapong'as ir Lim'as [60] tyrė Kambodžos namų ūkių ne maisto vartojimo išlaidų priklausomybę nuo ekonominių (įvairaus turto nuosavybė, būsto plotas, užimtumo statusas) ir geoduomenų (gyvenamojo ploto dydis, gyventojų tankumas, urbanizacijos lygis, žemės ūkio paskirties plotai) veiksnių, naudodami Pasaulio banko 2019 m. gyvenimo lygio matavimo tyrimo ir GEE (angl. *Google Earth Engine*) platformos duomenis. Tyrime, be atsitiktinių miškų, buvo taikyti ir atraminių vektorių regresijos (angl. *Support Vector Regression*, SVR), dirbtinių neuroninių tinklų (angl. *neural networks*, NN) bei apibendrintos mažiausių kvadratų regresijos (angl. *Generalized Least Squares*, GLS) metodai. Rivas'as-Perea'as ir kt. [62] teigia, kad SVR buvo sukurtas remiantis atraminių vektorių mašinos (SVM) principais, juos pritaikant regresijos uždaviniams. Šio metodo pagrindinis tikslas yra rasti optimalią hiperplokštumą, kuri leistų duomenų taškams tilpti į ϵ ribas, t. y., paklaidos intervalą. SVR metodu yra naudojama bauda, kuri taikoma, jei paklaidos yra didesnės nei leistina. Matematiškai galima teigti, kad SVR sprendžia kvadratinio optimizavimo uždavinį su apribojimais, kai galutinis sprendinys priklauso nuo atraminių vektorių, t. y., tų duomenų taškų, kurie yra ant ϵ ribos arba už jos. Ciaburro'as ir Venkateswarn'as [63] teigia, kad dirbtiniai neuroniniai tinklai yra skaičiavimo modelis, sukurtas pagal žmogaus smegenų architektūrą, nes tinklus sudaro tarpusavyje susijungę mazgai, vadinami neuronais, kurių paskirtis yra atskleisti duomenyse paslėptus dėsningumus. NN gali būti naudojamas ne tik regresijos, bet ir klasifikavimo bei klasterizavimo uždavinių sprendime. Neuroniniai tinklai yra sudaryti iš sluoksnių hierarchijos, t. y., turi įėjimo sluoksnį, kuris priima aiškinamuosius kintamuosius, išėjimo sluoksnį, kuriuo yra pateikiamos prognozės, bei paslėptuosius sluoksnius, kurie randasi tarp įėjimo ir išėjimo sluoksnių ir kurie padeda apdoroti sudėtingus ir netiesinius ryšius tarp duomenų. Neuronai taip pat naudoja aktyvacijos funkciją, kuri yra skirta padėti tinklui geriau prisitaikyti prie sudėtingų duomenų. Tarp neuronų egzistuojantys ryšiai turi tam tikrą svorį, kurie yra koreguojami NN mokymosi metu siekiant sumažinti pasirinktą nuostolių funkciją, t. y., svoriai yra taisomi iki kol sumažėja prognozavimo klaida. Šiame procese dažniausiai yra taikoma atgalinės sklaidos (angl. *backpropagation*) technika,

kuri naudoja optimizacijos algoritmus, tokius kaip gradientinis nusileidimas (angl. *gradient descent*). Ergun'as, Owusu'us ir Rivas'as [64] teigia, kad GLS yra statistinis regresijos metodas ir yra patobulinta klasikinės tiesinės regresijos metodo versija, nes GLS geba spręsti tokias problemas kaip heteroskedastiškumas ir liekanų koreliacija (autokoreliacija). Matematiškai taikant GLS metodą yra naudojamas variacijos – kovariacijos matricos *Cholesky* skaidymas tam, kad būtų suformuota svorių matrica, kuri leidžia transformuoti pradinį regresijos modelį ir gauti tikslesnius bei nešališkus įverčius. Puttanapong'as ir Lim'as [60] visų keturių metodų palyginimui naudojo šaknies vidutinę kvadratinę prognozės paklaidą (toliau – RMSE) ir determinacijos koeficientą (R^2) kaip modelio tinkamumo kriterijus. Geriausią prognozavimo tikslumą pasiekė atsitiktinių miškų algoritmas, kai po jo sekė neuroninis tinklas, atraminių vektorių regresija ir apibendrinta mažiausių kvadratų regresija. Atsitiktinių miškų metodu buvo pasiekta mažiausia RMSE reikšmė, t. y., 0,09, o R^2 siekė 0,41, kai prasčiausi rezultatai buvo gauti taikant GLS metodą, kurio R^2 buvo tik 0,26, o RMSE – 0,09. Verta paminėti, kad tyrimo rezultatai parodė, jog didžiausią ne maisto vartojimo išlaidų aiškinamąją galią turėjo socioekonominiai rodikliai (automobilio, mobiliojo telefono, motociklo turėjimas bei gyvenamojo ploto dydis), kai tuo tarpu geoduomenys, nors ir buvo statistiškai reikšmingi, tačiau buvo mažiau svarbūs.

5 lentelė. Mikrolygmens ir išorinių veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo metodų apibendrinimas, remiantis ankstesniais mokslinėje literatūroje aprašytais tyrimais

Metodai	Modelių tinkamumo vertinimo kriterijai	Privalumai	Trūkumai
Tiesinė regresija (OLS, pažingsninė, Tobit, GLS)	pakoreguotas R^2 , RMSE, MSE, multikolinearumo vertinimas, autokoreliacijos testai, heteroskedastiškumo testai	Paprasti, lengvai interpretuojami. Tobit – tinka duomenimis su nulinėmis reikšmėmis, pažingsninė valdo multikolinearumo problemą, GLS sprendžia heteroskedastiškumo ir autokoreliacijos problemas.	Reikalauja prielaidų laikymosi (normalumas, homoskedastiškumas, nepriklausomumas), didesnis jautrumas multikolinearumui. Tobit regresijos modelio vertinimui klasikinis R^2 yra netinkamas.
Kvantilinė regresija	Pseudo R^2 , multikolinearumo vertinimas	Vertina aiškinamųjų kintamųjų poveikį ne vidurkio prasme, bet skirtinguose priklausomojo kintamojo kvantiliuose, atspari heteroskedastiškumui ir ekstremalioms reikšmėms.	Pseudo R^2 interpretacija skiriasi nuo klasikinio R^2 , reikalaujama tiesinio ryšio tarp kintamųjų ir priklausomo kintamojo, aktualus galimas koeficientų nestabilumas dėl multikolinearumo problemos, lyginant su tiesine regresija, rezultatų interpretacija yra sudėtingesnė.
Mašininio mokymosi algoritmai (atsitiktiniai miškai, atraminių vektorių regresija, dirbtiniai neuroniniai tinklai)	R^2 , RMSE, MSE	Aptinka sudėtingus netiesinius ryšius, automatiškai identifikuoja stiprias sąveikas tarp kintamųjų, nereikalauja klasikinės tiesinės regresijos prielaidų tenkinimo, tinkamesni dideliems duomenų rinkiniams apdoroti, RF yra atsparus persimokymui.	Sudėtingesnis rezultatų interpretavimas (reikia papildomų metodų, pvz., SHAP analizės), reikia optimizuoti hiperparametrus, SVR ir NN yra jautrūs kintamųjų matavimo skalių skirtumams, modelių apmokymui reikia didesnių kompiuterinių resursų.

Mokslinėje literatūroje (žr. 5 lent.), tyrinėjant namų ūkių vartojimo išlaidų priklausomybę nuo mikrolygmens ir išorinių veiksnių, yra plačiai taikomi ekonometriniai tiesiniai modeliai ir kartais – kvantilinės regresijos modeliai. Pastaruoju metu vis dažniau yra eksperimentuojama ir su

pažangesniais mašininio mokymosi algoritmais, tokiais kaip atsitiktinių miškų regresija, dirbtiniai neuroniniai tinklai ar atraminių vektorių regresija, kurie taikomi siekiant geriau atskleisti netiesines priklausomybes tarp aiškinamųjų ir priklausomųjų kintamųjų. Daugelio darbų tyrimų rezultatai patvirtino, kad pagrindiniai vartojimo išlaidas veikiantys aiškinamieji kintamieji yra disponuojamosios pajamos, namų ūkio dydis ir išsilavinimo lygis. Nors tiesinės regresijos modeliai dažniausiai paaiškino bent 70 % visos išlaidų dispersijos ($R^2 \geq 0,70$), tačiau dalinio R^2 analizė papildomai parodė, kad vien tik kiekybinės disponuojamosios pajamos gali paaiškinti daugiau kaip 50 % visos vartojimo išlaidų variacijos, nes tyrimuose, kuriuose nebuvo skaitinio pajamų kintamojo, net ir taikant pažangius mašininio mokymosi algoritmus, modelių tikslumas siekė tik 40 %, kas parodo, jog net ir pažangių metodų taikymas negali kompensuoti svarbiausio aiškinamojo kintamojo nebuvimo. Pabrėžtina, kad mašininio mokymosi metodų taikymas vartojimo išlaidų vertinimui yra palyginti jauna sritis, todėl ir tyrimų rezultatai nebuvo vienareikšmiai, t. y., kai kuriuose tyrimuose atsitiktinių miškų modelių tikslumas buvo panašus į tiesinės regresijos, kai tuo tarpu kituose – atsitiktinių miškų regresija per 0,14 vienetų pagerino determinacijos koeficiento (R^2) reikšmę, ją palyginus su apibendrintos mažiausių kvadratų regresijos tikslumo vertinimo matu.

1.5. Mikrolygmens veiksnių poveikio namų ūkių vartojimo išlaidoms vertinimo poreikis

Singh'as ir Kumar'as [55] pastebi, kad didėjantys skirtumai tarp skirtingų socialinių ir ekonominių sluoksnių kelia susirūpinimą dėl vis didėjančio visuomenės susiskaidymo šalyje. Norit spręsti tokias problemas, yra svarbu planuoti veiksmus atsižvelgiant į konkrečių gyventojų poreikius, kurie gali būti atrasti taikant mikrolygmens planavimą, apimančią išsamią vietos gyventojų pagrindinių poreikių analizę ir tokių politikos sprendimų formavimą, kurios užtikrintų racionalų išteklių naudojimą. Lee'ė, Ong'as ir Lee'ė [30] teigia, kad vien namų ūkių pajamų lygis gali neatspindėti tikrosios namų ūkių padėties, ypač jei ji apima didėjančius namų ūkių įsiskolinimus, skirtingas pragyveno kaštus priklausomai nuo pasirinktos gyvenamosios vietos, bei gyventojų amžiaus ar gyvenimo ciklo skirtumus. Dėl to nenuostabu, kad yra akcentuojama būtinybė skurdo ir socialinės atskirties vertinime taikyti platesnius mikrolygmens rodiklius, tokius kaip vartojimo išlaidų struktūra pagal pagrindines sritis – maistą, sveikatą, būstą, švietimą, transportą ir kt., kas suteikia galimybę išsamiu įvertinti gyventojų materialinę padėtį ir atskleisti realius apribojimus vartojant būtiniausias prekes ir paslaugas. Verta paminėti, kad pagal pajamų duomenis taip pat dažnai yra nepakankamai tiksliai identifikuojamos pažeidžiamos gyventojų grupės, ypač urbanizuotose teritorijose, kuriuose pajamų lygis gali būti aukštesnis, tačiau kartu ir gyvenimo kaštai gerokai didesni. Atsižvelgiant į šias aplinkybes, galima teigti, kad yra būtina stiprinti mikrolygmens analizės taikymą politikos formavimo procesuose, su tikslu geriau suprasti skurdo dinamiką ir namų ūkių vartojimo elgseną, kas padėtų kurti tikslingesnes ir labiau bendruomenių kontekstui pritaikytas intervencijas, taip pat prognozuoti vartojimo tendencijas, atsižvelgiant į dabartinius ekonominius, socialinius ir demografinius veiksnius.

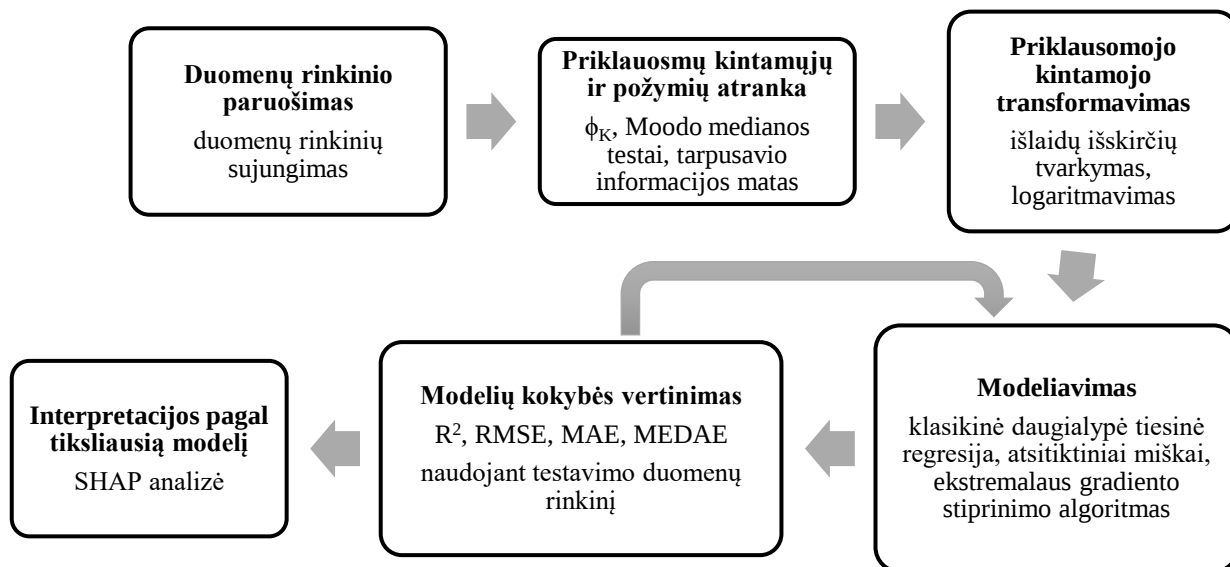
Singh'as ir Kumar'as [55] taip pat pabrėžia, kad, nepaisant plėtojamos literatūros ir atliekamų tyrimų, vis dėl to mikrolygmens kintamųjų poveikio vartojimui empiriniai įrodymai išlieka toli nuo nuoseklumo ir galutinumo, kas kelia iššūkius bandant padaryti apibendrintas išvadas bei pabrėžia vartotojų elgsenos analizės sudėtingumą. Lee'ė, Ong'as ir Lee'ė [30] siekdami geriau suprasti šį sudėtingumą ir sumažinti tyrimų spragų poveikį, pradėjo taikyti pažangesnes vertinimo metodikas, t. y., šiuolaikines analitines priemones, tokias kaip mašininio mokymosi metodai [30]. Tokių metodų taikymas, ne tik suteikia galimybę tiksliau įvertinti vartojimo išlaidas, bet ir geriau suprasti, kaip konkretūs mikrolygmens veiksniai (amžius, šeimos dydis, išsilavinimas, gyvenamoji vieta) veikia

vartojimo procesą, nes yra atsižvelgiama į sudėtingesnius bei netiesinius ryšius tarp kintamųjų, kurie dažnai lieka nepastebėti taikant tradicinius tiesinius metodus. Mašininio mokymosi metodų taikymas praplečia analitinių priemonių galimybes ir padidina poreikį tolimesniems tyrimams, kurie yra susiję su namų ūkių vartojimo analizės gilinimu.

Vartojimo išlaidų vertinimo priklausomai nuo mikrolygmens veiksmų tyrimų poreikį rodo ir atliktų analizių nevienareikšmiškos išvados, kurios, atrodo, priklauso nuo nagrinėjamos šalies konteksto. Lee'ė, Ong'as ir Lee'ė [30] tyrinėjo Malaizijos namų ūkių duomenis ir nustatė, kad tarp namų ūkio galvos amžiaus ir vienam asmeniui tenkančių vartojimo išlaidų egzistuoja netiesinis, apverstos U formos ryšys, kai tuo tarpu Bugheanu'us ir Strachinaru'us [27] nagrinėję Rumunijos atvejį, pastebėjo, kad jauni gyventojai yra labiau linkę išlaidauti, o Caglayan'as ir Astar'as [59], analizuodami Turkijos namų ūkius, nustatė, kad namų ūkio galvos amžiui padidėjus vienu vienetu, vartojimo išlaidos taip pat išauga. Singh'as ir Kumar'as [55] tyrė Indijos gyventojų vartojimą pagal regionus ir nustatė, kad namų ūkių, gyvenančių lygumose, vartojimą statistiškai reikšmingai veikia turimos žemės plotas, bet ne išsilavinimo lygis, kai pakrantėse gyvenantiems statistiškai reikšmingas buvo turimas išsilavinimo lygis, bet ne žemės plotas. Caglayan'as ir Astar'as [59] pastebėjo, kad Turkijoje miesto gyventojų vartojimo išlaidos yra didesnės nei kaimo gyventojų, kai York'as [29], tyrinėjęs bendras Europos Sąjungos šalių išlaidas, pastebėjo, jog miestų gyventojų energijos vartojimas yra mažesnis, juos lyginat su kaimo gyventojais, nors ir kaimo gyventojai dažniau naudoja kietąjį kurą, didinantį energijos vartojimo poreikį. Atsižvelgiant į tai, jog skirtingose šalyse ar regionuose mikrolygmens veiksmų poveikis vartojimo išlaidoms gali būti nevienodas, kyla poreikis toliau plėtoti šią temą. Ypač tai gali būti tai aktualu Lietuvos kontekste, kadangi iki šiol tokio pobūdžio tyrimai su šia valstybe nebuvo atlikti.

2. Lietuvos namų ūkių duomenys ir tyrimo metodai

Darbo procesą galima bendrai aprašyti keliais pagrindiniais etapais (žr. 2 pav.), kurie apima visą tyrimo eigą – nuo duomenų lentelių sujungimo iki SHAP analizės.



2 pav. Tyrimo metodologijos schema

Remiantis anksčiau atliktais tyrimais, aprašytais mokslinėje literatūroje (žr. 1.4 skyrelį), vartojimo išlaidų ir jų struktūrinių dalių priklausomybių nuo namų ūkio mikrolygmens charakteristikų vertinimui šiame darbe buvo nuspręsta taikyti tiek klasikinį daugialypės tiesinės regresijos modelį, tiek pažangesnius mašininio mokymosi metodus, t. y., atsitiktinių miškų regresijos modelį, kuris jau buvo panaudotas ankstesniuose tyrimuose, bei ekstremalaus gradiento stiprinimo algoritmą, kuriuo buvo siekiama ne tik geresnių rezultatų, bet ir praplėsti ankstesnių tyrimų metodologiją. Siekiant užtikrinti tinkamiausio modelio rezultatų interpretavimą ir įvertinti, kaip konkretūs mikrolygmens veiksniai prisidėjo prie vartojimo išlaidų prognozavimo, buvo nuspręsta taikyti jau anksčiau tyrimuose paminėtą SHAP metodą, kuris dažniausiai yra naudojamas interpretacijoms atlikti taikant mašininio mokymosi modelius.

Šio tyrimo pagrindinis tikslas buvo atlikti regresinę analizę, naudojant kiekybinį priklausomąjį kintamąjį (vartojimo išlaidas) ir kategorinius bei kiekybinius veiksnus (namų ūkių charakteristikas). Duomenys buvo apdoroti naudojant „Python“ programavimo kalbos 3.12.8 versiją, pasitelkiant šio įrankio platų matematinių ir statistinės analizės bibliotekų rinkinį. Rossum'as [65] pastebi, kad bibliotekos yra jau parašyto kodo rinkiniai, kuriuos iš karto galima panaudoti analizėje. Dažniausiai buvo naudojamos funkcijos, kurios, iškvietus, automatiškai atliko tam tikrus veiksmus ar skaičiavimus. Pagrindiniai naudoti įrankiai buvo duomenų analizės pandas biblioteka, vizualizacijų *matplotlib* biblioteka, statistinių metodų *scipy.stats* modulis, skaitinių skaičiavimų *numpy* biblioteka, modeliavimo *scikit-learn* biblioteka, Bajeso optimizavimo *scikit-optimize* biblioteka.

2.1. Lietuvos namų ūkių duomenų rinkinys ir jo paruošimas

Pasirinktas tirti objektas – Lietuvos namų ūkių mėnesinės išlaidos, skirtos vartojimui. Apžvelgti tyrimo kintamieji ir pateikta jų aprašomoji analizė.

Tyrimui atlikti buvo naudoti 2021 m. atlikto Lietuvos namų ūkių biudžetų statistinio tyrimo duomenys, gauti iš laisvai prieinamo **Valstybės duomenų agentūros**⁸ šaltinio. Verta paminėti, kad 2021 m. duomenys yra naujausi prieinami duomenys, nes minėta agentūra tokią valstybinio lygio apklausą atlieka tik kas ketverius metus, tai reiškia, kad kita apklausa turėtų būti organizuojama 2025 metais.

Apklausos duomenys buvo išskaidyti į tris duomenų rinkinius: vartojimo išlaidų duomenis, namų ūkių biudžetų statistinio tyrimo duomenis ir individualius namų ūkiuose gyvenančių asmenų duomenis. Kadangi pagrindinis duomenų rinkinys (vartojimo išlaidos) turėjo 4334 eilutes, kuriose buvo pateiktas tik namų ūkio galvos unikalus identifikatorius, visi trys rinkiniai buvo sujungti naudojant šį identifikatorių. Verta paminėti, kad šioje analizėje namų ūkio galva yra asmuo, didžiąją savo pajamų dalį skiriantis šeimos poreikiams tenkinti (Valstybės duomenų agentūra⁸). Iš vartojimo išlaidų duomenų lentelės buvo atrinktos tik 14 pagrindinių vartojimo išlaidų kategorijos, remiantis oficialiu individualaus vartojimo išlaidų paskirties pirmo lygio klasifikavimu (žr. 1 lent.). Į tyrimą nebuvo įtrauktos išlaidos, susijusios su ne pelno institucijų, teikiančių paslaugas namų ūkiams, veikla, todėl analizei buvo planuota naudoti 14 priklausomų kintamųjų:

- bendrosios vartojimo išlaidos, susidedančios iš 13 išlaidų tipų, eurai. Naudojamas kintamasis – „visos namų ūkio vartojimo išlaidos (mėnesinės)“;
- trylika kitų kintamųjų, atspindinčių vartojimo išlaidų struktūrą pagal kategorijas: maistas ir nealkoholiniai gėrimai, alkoholiniai gėrimai, tabakas ir narkotikai, apranga ir avalynė, būstas, vanduo, elektra, dujos ir kitas kuras, būsto apstatymas, namų ūkio įranga ir kasdienė namų priežiūra, sveikata, transportas, informacija ir ryšiai, švietimo paslaugos, poilsis, sportas ir kultūra, restoranai ir apgyvendinimo paslaugos, draudimas ir finansinės paslaugos, asmens priežiūra, socialinė apsauga, įvairios prekės ir paslaugos.

Kalbant apie tyrime naudotus nepriklausomus kintamuosius, iš galutinės duomenų lentelės buvo pašalinta namų ūkių duomenyse ir individualiuose namų ūkių duomenyse pasikartojusi informacija (namų ūkio išsilavinimas, socialinė grupė, amžiaus grupės, užimtumo statusas). Galutinai paruoštame duomenų rinkinyje liko 26 kategoriniai ir kiekybiniai kintamieji, atspindintys mikrolygmens namų ūkių charakteristikas. Šie kintamieji tolimesnėje analizėje gali būti vadinami požymiais, nepriklausomais ar aiškinamaisiais kintamaisiais, veiksniais. Tirti paruoštų nepriklausomų kintamųjų tipai ir aprašymai pateikti 6 lentelėje.

6 lentelė. Tyrime naudojamų nepriklausomų kintamųjų aprašymas

Kintamasis	Duomenų tipas	Aprašymas
hh_ident	Sveikasis skaičius	Unikalus namų ūkio identifikatorius
Amžius	Sveikasis skaičius	Namų ūkio galvos amžius
Gyvenamoji vieta	Kategorinis	5 didieji miestai, Kiti miestai, Kaimas
Regionas	Kategorinis	Sostinės regionas, Vidurio ir vakarų Lietuvos regionas
Namų ūkio tipas	Kategorinis	Vienas gyvenantis asmuo, Kiti namų ūkiai su išlaikomais vaikais, Pora be išlaikomų vaikų, Pora su išlaikomais vaikais, Vienas suaugęs asmuo su išlaikomais vaikais, Kiti namų ūkiai be išlaikomų vaikų

⁸[https://atviri-](https://atviri-duomenys.stat.gov.lt/search?categories=%252Fcategories%252Fstatistiniai%2520tyrimai%252Fnam%25C5%25B3%2520%25C5%25ABki%25C5%25B3%2520biud%25C5%25BEet%25C5%25B3)

[duomenys,stat.gov.lt/search?categories=%252Fcategories%252Fstatistiniai%2520tyrimai%252Fnam%25C5%25B3%2520%25C5%25ABki%25C5%25B3%2520biud%25C5%25BEet%25C5%25B3](https://atviri-duomenys.stat.gov.lt/search?categories=%252Fcategories%252Fstatistiniai%2520tyrimai%252Fnam%25C5%25B3%2520%25C5%25ABki%25C5%25B3%2520biud%25C5%25BEet%25C5%25B3)

Išlaikomų vaikų skaičius	Kategorinis	Namų ūkiai be išlaikomų vaikų, Namų ūkiai su 1 išlaikomu vaiku, Namų ūkiai su 2 išlaikomais vaikais, Namų ūkiai su 3 ir daugiau išlaikomų vaikų
Socialinė-ekonominė grupė	Kategorinis	Senatvės pensininkai, Samdomieji darbuotojai, Bedarbiai, Savarankiškai dirbantys asmenys
Pajamų šaltinis	Kategorinis	Pensija, Darbo užmokestis, Pašalpa, stipendija ir kitas pragyvenimo šaltinis, Pajamos iš savo ar šeimos verslo
Namų ūkio galvos išsilavinimas	Kategorinis	Aukštesniojo ir aukštojo nesuteikiantis povidurinis, Vidurinis, Aukštesnysis ir aukštasis, Pradinis, nebaigė pradinės mokyklos, Pagrindinis
Pajamų kvintilinė grupė	Kategorinis	1-a pajamų kvintilinė grupė, 2-a pajamų kvintilinė grupė, 3-a pajamų kvintilinė grupė, 4-a pajamų kvintilinė grupė, 5-a pajamų kvintilinė grupė
Būsto tipas	Kategorinis	Butas daugiabučiame name, Vieno buto namas, Butas dviejų butų ar sublokuotame name, Kita gyvenamoji patalpa
Būsto valdos statusas	Kategorinis	Būsto savininkas, neturintis išmokėti paskolos, Būsto savininkas, turintis neišmokėtą paskolą, Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina, Gyvena nemokėdamas nuomos, Nuomojasi šį būstą mažesne nei rinkos kaina
Centralizuotas šalto vandens tiekimas	Kategorinis	Taip, Ne
Centralizuotas karšto vandens tiekimas	Kategorinis	Taip, Ne
Pagrindinis būsto šildymo būdas	Kategorinis	Centrinis šildymas iš šilumos tinklų, Centrinis šildymas iš vietinio šilumos šaltinio, Nėra stacionaraus šilumos šaltinio, Krosnis, židiny
Asmenų skaičius namų ūkyje	Dešimtainis skaičius	Asmenų skaičius iš viso
Namų ūkio užimamas naudingasis plotas	Dešimtainis skaičius	Būsto plotas, m ²
Motinos gimimo šalis	Kategorinis	Lietuva, Kita šalis, Nenurodė
Gimimo šalis	Kategorinis	Lietuva, Kita šalis
Pagrindinės pilietybės šalis	Kategorinis	Lietuva, Kita šalis, Netaikoma
Tėvo gimimo šalis	Kategorinis	Lietuva, Kita šalis, Nenurodė
Lytis	Kategorinis	Moteris, Vyras
Teisinė santuokinė padėtis	Kategorinis	Išsituokęs (-usi), Nevedęs (netekėjusi), Vedęs (ištekęs), Našlys (-ė)
Gyvena su partneriu	Kategorinis	Ne, Taip, teisiniu pagrindu, Taip, be teisinio pagrindo
Pagrindinis darbas: profesija	Kategorinis	Netaikoma, Specialistai, Technikai ir jaunesnieji specialistai, Įstaigų tarnautojai, Nenurodė, Vadovai, Kvalifikuoti darbininkai ir amatininkai, Paslaugų sektoriaus darbuotojai ir pardavėjai, Įrenginių ir mašinų operatoriai ir surinkėjai, Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai
Šiuo metu mokosi	Kategorinis	Ne, Taip
Darbo sutarties tipas	Kategorinis	Netaikoma, Neterminuota sutartis, Terminuota sutartis

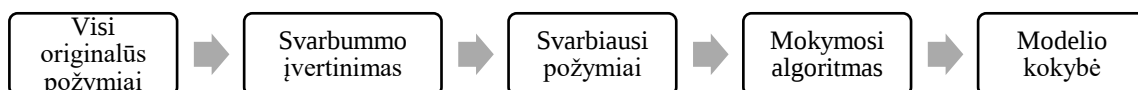
Penki didieji miestai yra Vilnius, Kaunas, Klaipėda, Šiauliai ir Panevėžys, kiti miestai – kiti miestai ir rajonų centrai, o kaimas apima kaimo vietas ir vienkiemius. Bendras analizuojamų duomenų (priklausomų ir nepriklausomų kartu) dydis yra 4334 eilučių ir 40 stulpelių (žr. 1 priedą). Panaudota vienoje iš svarbiausių duomenų analizės bibliotekų „pandas“ esanti funkcija *isna()* parodė (žr. 2 priedą), kad turimuose duomenyse taip pat nėra trūkstamų reikšmių. Vartojimo išlaidų ir jos struktūrinių dalių pagal namų ūkių kvintilinę pajamų ir išlaikomų vaikų skaičiaus kintamųjų grupes aprašomoji analizė pateikta 4 priede. Aprašomoji statistika pagal likusius kategorinius kintamuosius pateikta 29 priede. Verta paminėti, kad ypač mažas kategorinių kintamųjų grupių stebinių skaičius gali lemti prastą modelių apmokymo kokybę. Pastebėta (žr. 29 priedą), kad „Būsto tipas: kita gyvenamoji patalpa“ kategorijoje yra tik 3 stebiniai, o „Pagrindinis būsto šildymo būdas: nėra stacionaraus šilumos šaltinio“ kategorijoje – 16 stebinių, tai rodo grupių nepakankamą reprezentatyvumą, kuris gali lemti nepakankamą duomenų pasiskirstymą apmokymo ir testavimo imtyse.

2.2. Požymių atrankos metodai

Nors ir buvo planuojama taikyti mašininio mokymosi modelius, pasižyminčius savaiminiu atsparumu išskirtinėms reikšmėms, netiesiškumui ar multikolinearumui, tačiau dirbant su didelės apimties duomenų rinkiniais, neretai pasitaiko nereikšmingų arba mažai informatyvių nepriklausomų kintamųjų (požymių), kurių įtraukimas gali apsunkinti skaičiavimus, nepagerinti modelio tikslumo, ženkliai padidinti skaičiavimo išteklių (tiek laiko, tiek atminties) poreikį bei sulėtinti analizės procesus. Be to, mažai informatyvūs kintamieji gali tapti triukšmo šaltiniu ir padidinti modelio permokymo riziką [66]. Procesas, kurio metu yra pašalinami nereikalingi požymiai, neturintys ryšio su priklausomu kintamuoju vadinamas **požymių atranka** [67]. Požymių atranka apima iš esamo kintamųjų rinkinio svarbiausių elementų identifikavimą ir atrinkimą.

Bendrai naudojami požymių atrankos metodai, taikomi mašininio mokymosi modeliams, yra skirstomi prižiūrimus (angl. *supervised*) ir neprižiūrimus (angl. *unsupervised*). Skirtumas tarp prižiūrimų ir neprižiūrimų metodų priklauso nuo to, ar požymiai atrenkami atsižvelgiant į priklausomąjį kintamąjį, ar ne [68]. Prižiūrimi metodai remiasi ryšiu tarp požymių ir priklausomojo kintamojo ir siekia nustatyti, kurie požymiai yra svarbūs prognozavimui, o tai yra ypač naudinga analizuojant duomenis su etiketėmis (angl. *labeled data*), t. y., kai yra aiški reikšmė, kurią modelis turi prognozuoti naudojant klasifikavimo arba regresijos metodus.

Kadangi regresijos pobūdžio uždaviniai reikalauja prižiūrimų metodų taikymo, jie toliau ir buvo naudojami. Prižiūrimų metodų grupei priklauso požymių atrankos filtravimo (angl. *filter-based*), apvijos (angl. *wrapper*) ir įterptieji (angl. *embedded*) metodai. Pabrėžtina, kad filtravimo metodai yra vieninteliai, kuriuos galima taikyti prieš kuriant modelius ar naudojant algoritmus, todėl jie ir buvo naudojami tyrimo metu. Filtravimo pagrindu paremti požymių atrankos metodai (žr. 3 pav.) naudoja koreliaciją ar statistinius rodiklius, siekdami įvertinti kiekvieno požymio santykį su priklausomu kintamuoju. Šie metodai vertina kiekvieno veiksnio svarbumą individualiai arba poromis, neatsižvelgiant į konkretaus mokymosi algoritmo veikimą [69].



3 pav. Filtravimo požymių atrankos metodų veikimo principas (autoriai: Wang'as ir kt. [70])

Verta paminėti, kad atsižvelgiant į tai, jog duomenų rinkinyje kiekybinių požymių skaičius yra mažas, t. y., vos trys kintamieji, siekiant išlaikyti vientisą atrankos procesą, buvo pasirinkta taikyti tik kategoriniams kintamiesiems skirtus metodus. Dėl šios priežasties vieninteliai skaitinio tipo analizės kintamieji, t. y., asmenų skaičius namų ūkyje, namų ūkio užimamas naudingasis plotas bei amžius, požymių atrankos analizės metu buvo pakeisti į kategorinius kintamuosius. Santykinės namų ūkių charakteristikos į kategorinius kintamuosius buvo paverstos juos skaidant taip: asmenų skaičius padalintas į tris grupes: 1-2 asmenys, 3-4 asmenys ir 5 ar daugiau asmenų, kai naudingasis plotas – į keturis lygius: mažiau nei 50 m², 50-99 m², 100-199 m² ir daugiau nei 200 m², o amžius į 5 intervalus: iki 30 metų, 30-39 metų, 40-49 metų, 50-59 metų ir 60 metų bei daugiau.

Filtravimo pagrindu paremtų požymių atrankos metodų, skirtų kategorinių požymių atrankai, kai priklausomas kintamasis – kiekybinis, naudotinas technikas apima dispersijos analizę (angl. *ANOVA Fišerio Test*), Kruskalo ir Voliso testas (angl. *Kruskal-Wallis test*), tarpusavio informacijos metodas (angl. *mutual information*), Moodo medianos testas, *Phik* koreliacijos koeficientas. Pažymėtina, kad

dėl duomenų ypatumų tyrime buvo išbandyti, tačiau nepasiteisino kai kurie metodai, tokie kaip dispersinė analizė bei Kruskalo ir Voliso testas, todėl jų metodologiniai aprašai buvo pateikti 5 priede.

Tarpusavio informacijos (TI) metodas yra informacijos teorijos pagrindu sukurtas matas, naudojamas vertinti priklausomybę tarp požymio ir priklausomojo kintamojo [74]. Šis metodas nustato, kiek informacijos apie tikslinį kintamąjį suteikia konkretus požymis, kitaip tariant, kiek galima sumažinti netikrumą (angl. *uncertainty*) apie tikslinį kintamąjį, žinant požymio reikšmę. TI yra plačiai taikomas požymių atrankos užduotyse, nes geba aptikti tiek tiesinius, tiek netiesinius ryšius. Vis dėlto, šis metodas analizuoja kiekvieną savybę atskirai, todėl gali praleisti svarbias sąveikas tarp požymių. Mokslininkai Peng'as, Long'as ir Ding'as [75] dvejų atsitiktinių kintamųjų x ir y tarpusavio informaciją apibrėžia remiantis jų marginaliniais skirstiniais $p(x)$ $p(y)$ ir bendru pasiskirstymu $p(x,y)$:

$$I(x; y) = \sum_x \sum_y p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right); \quad (5)$$

čia $p(x,y)$ – jungtinė tikimybė tarp x ir y , $p(x)$ ir $p(y)$ atitinkamai x ir y marginalinės tikimybės.

Remiantis maksimalaus reikšmingumo (angl. *Max-Relevance*) atrankos metodu yra pasirenkami požymiai, kurie turi didžiausią reikšmę priklausomam kintamajam. Požymiai, turintys didžiausią TI $I(x;y)$ reikšmę, yra laikomi svarbiausiais, nes jie parodo stiprų ryšį tarp priklausomo (x) ir nepriklausomų (y) kintamųjų bei reikšmingą įtaką tiksliniam kintamajam. Jei $p(x,y) = p(x)p(y)$, tai rodo, kad x ir y yra nepriklausomi ir TI reikšmė tampa lygi nuliui. Verta paminėti, kad TI metodu gauti rezultatai dėl savilygiavertiškumo savybės išlieka aktualūs net ir tuo atveju, kai priklausomojo kintamojo reikšmės transformuojamos naudojant invertuojamą monotonišką funkciją [76]. Hernandez'as ir Samego'as [77] pastebi, kad TI gali pervertinti egzistuojančią priklausomybę, jei tiriami kintamieji pasižymi asimetriniu pasiskirstymu. TI reikšmėms gauti buvo naudota *mutual_info_regression()*⁹ funkcija iš *scikit-learn* bibliotekos.

Moodo medianos testas yra nparametrinis metodas, skirtas įvertinti medianų skirtumus tarp dviejų ar daugiau nepriklausomų imčių ar grupių, net jei jų dydžiai yra nevienodi [82]. Šis testas nereikalauja normalumo ir homogeniškumo prielaidų atitikimo, taip pat yra mažiau jautrus ekstremalioms išskirtims bei duomenų asimetrijai, todėl yra ypač tinkamas naudoti su netaisyklingais duomenų rinkiniais. Hussain'as ir kt. [83] pabrėžia, kad testas yra tinkamas naudoti, kai yra abejojama, ar grupių skirstiniai yra panašios formos. Mokslininkai Smida ir kt. [84] teigia, kad atliktų tyrimų metu buvo įrodyta, jog medianos testas yra galingesnis metodas nei Studento t-testas ir Mano ir Vitnio testas vienmačių duomenų atvejais, kai skirstiniai pasižymi sunkiomis uodegomis. Freidlin'as ir Gastwirth'as [85] įspėja, kad medianos testas ypač mažose imtyse gali turėti mažą statistinę galią, t. y., tikimybę tinkamai atmesti nulinę hipotezę. Nulinės hipotezės vertinimui Moodo medianos testas naudoja χ^2 testinę statistiką ir jei nulinė yra hipotezė teisinga, ši statistika seka χ^2 skirstinį su $k-1$ laisvės laipsniais. Oti'is, Olusola, Esemokumo'as [86] teigia, kad medianos testas veikia taip: yra atsitiktinai parenkama reikšmė X_i yra iš imties, kurios dydis yra m ir kuri nepriklausomai paimta iš populiacijos X , o Y_i – iš imties, kurios dydis yra n ir kuri nepriklausomai paimta iš populiacijos Y . Vėliau abi imtys (m ir n) yra sujungiamos į vieną bendrą imtį tam, kad būtų apskaičiuota bendroji mediana N . Po to kai bendroji imtis yra surūšiuojama didėjančia arba mažėjančia tvarka, vėliau yra nustatoma kiek kiekvienos imties stebėjimų yra virš arba žemiau bendros medianos ir šie dažniai yra

⁹ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html

surašomi į $2 \times n$ kontingencijos lentelę. Nulinė hipotezė teigia, kad tiriamos imtys yra paimitos iš populiacijų, kurios turi tą pačią medianą. Tad jei abi imtys yra paimitos iš populiacijų, kurių medianos yra vienodos, yra galima tikėtis, kad maždaug pusė stebėjimų kiekvienoje imtyje bus virš bendros medianos N , o kita pusė – žemiau. Reikiama statistika yra gaunama naudojant kontingencijos lentelę iš kurios yra paskaičiuojama χ^2 statistika pagal formulę:

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}; \quad (6)$$

čia n – bendras dažnis, n_i – bendras arba ribinis dažnis i -oje eilutėje, o n_j – bendras arba ribinis dažnis j -ame stulpelyje.

Jei apskaičiuota χ^2 reikšmė yra lygi arba didesnė už kritinę reikšmę χ^2 prie pasirinkto reikšmingumo lygmens, tada nulinė hipotezė yra atmetama. Maža pagal šią statistiką apskaičiuota p -reikšmė (pvz., mažesnė nei 0,05) leidžia atmesti nulinę hipotezę ir priimti alternatyviąją, t. y. bent vienos grupės mediana statistiškai reikšmingai skiriasi nuo kitų, kai yra naudojamas dviejų krypčių (angl. *two-tailed*) alternatyvios hipotezės variantas. Atrenkant požymius pagal jų svarbą atsako kintamajam, p -reikšmė gali būti naudojama siekiant atmesti nereikšmingas savybes. Kuo yra didesnė χ^2 statistikos reikšmė, o kartu ir mažesnė p -reikšmė, tuo labiau rodo, kad tam tikro požymio medianos reikšmingai skiriasi tarp grupių. Toks požymis gali būti svarbus atsako kintamojo elgesio paaiškinimui [87]. Verta paminėti, kad Moodo medianos testu gauti rezultatai galioja, net jei ir vėliau kintamiesiems yra taikomos duomenų transformacijos, kurios nekeičia reikšmių išsidėstymo tvarkos. Moodo medianos testas tyrime realizuotas naudojant *scipy.stats* bibliotekos *median_test()*¹⁰ funkciją.

Moodo medianos testu taip pat galima atlikti ir *post-hoc* analizę, taikant porinius medianos testus [88]. Jei po bendresnės analizės buvo nustatyti reikšmingi grupių skirtumai, vėliau yra galima atlikti daugybinius porinius palyginimus, siekiant nustatyti, kurios grupės tarpusavyje statistiškai reikšmingai skyrėsi. Šiuo atveju yra naudojama 2×2 kontingencijos lentelė. Verta pabrėžti, kad naudojamas 5 % reikšmingumo lygmuo, kuris reiškia, kad yra 5 % tikimybė klaidingai atmesti nulinę hipotezę, atliekant daug nepriklausomų testų tampa nepatikima, nes nors ir kiekvieno iš atliekamų testų klaidingų atradimų tikimybė vis dar išlieka 5%, tačiau bendra klaidingų atradimų tikimybė didėja. Lesack'as ir Naugler'is [89] teigia, kad klaidingai teigiamų rezultatų tikimybė didėja proporcingai atliekamų testų skaičiui. Norint sumažinti neteisingai atmestų nulinių hipotezių skaičių, atsirandantį dėl daugybinio testavimo, galima taikyti vieną iš sprendimo būdų – koreguoti p reikšmes naudojant konservatyvią Bonferroni korekciją, mažiau konservatyvią Bonferroni ir Holmo korekciją arba koreguoti klaidingų atradimų dažnį taikant Benjamino ir Hochbergo metodą. Naujesnis ir lankstesnis Benjamino ir Hochbergo korekcijos metodas orientuojasi į klaidingų atradimų dažnio (angl. *false discovery rate*) kontrolę eksperimentuose, dėl to padidėja šio metodo statistinė galia. Benjamino ir Hochbergo korekcijos formulė:

$$P_{pakoreguota} = \frac{P_{originali} \times n}{k}; \quad (7)$$

čia n – testuojamų hipotezių skaičius, o k – nepakoreguotų p -reikšmių eilės numeris (nuo mažiausios iki didžiausios). P -reikšmių korekcija buvo taikyta kartu su medianos testu naudojant *multipletests()*¹¹ funkciją iš *statsmodels.stats.multitest* bibliotekos, pasirinkus Benjamino ir Hochbergo metodą („*fdr_bh*“).

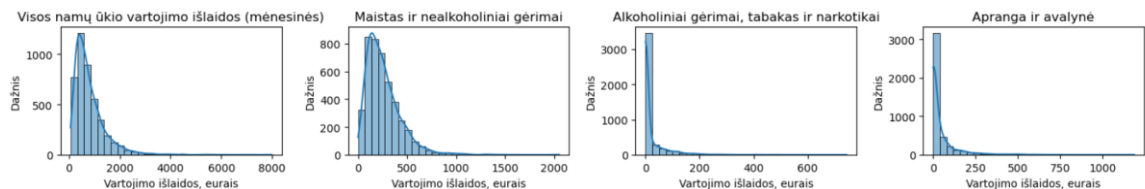
¹⁰ https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.median_test.html

¹¹ <https://www.statsmodels.org/dev/generated/statsmodels.stats.multitest.multipletests.html>

Phik koreliacijos koeficientas (ϕ_K) yra naujas ir praktiškas koreliacijos koeficientas, kuris yra tradicinio Pirsono koreliacijos koeficiento patobulinimas. Baak'as ir kt. [90] teigia, kad skirtingai nei tradiciškai naudojami koreliacijos koeficientai, ϕ_K vienu metu gali būti taikomas ne tik intervaliniams, bet ir ranginiams ar kategoriniams duomenims, nes kiekvienas kintamasis yra traktuojamas kaip kategorinis, t. y., tinkamas su mišraus tipo kintamaisiais. Naujasis koreliacijos koeficientas taip pat geba aptikti tiek tiesines, tiek netiesines priklausomybes. Esant dvimačiam normaliam pasiskirstymui ϕ_K sutampa su Pirsono koreliacijos koeficientu. Universalusis koreliacijos koeficientas yra išvestas iš Pirsono χ^2 kontingencijos testo, t. y., jis tikrina dviejų (ar daugiau) kintamųjų nepriklausomybę kontingencinėje lentelėje, kurioje eilutės ir stulpeliai atitinka skirtingų kintamųjų kategorijas, kai langeliuose yra laikomi įrašų, priklausančių abiejų kategorijų sankirtai, skaičius. ϕ_K koreliacijos koeficiento reikšmės yra apribotos intervale nuo 0 iki 1, kai 0 reiškia visišką nepriklausomumą, o 1 – visišką priklausomybę. Tyrime koreliacijos koeficiento apskaičiavimui buvo panaudota *phik*¹² bibliotekos *phik_matrix()* funkcija. Baak'as ir kt. [90] taip pat pateikia ir kintamųjų sąryšio statistinio reikšmingumo įvertinimo metodiką, kuri yra pagrįsta hibridiniu metodu. Hibridinis metodas yra naujas ir susideda iš *Monte Karlo* imčių simuliacijos bei χ^2 pasiskirstymo koregavimo, kai naudojamas χ^2 testo statistinis kriterijus, taikymo. Jei p-reikšmė, apskaičiuota hibridiniu metodu, yra mažesnė už 5 % reikšmingumo lygį, tai yra laikoma, jog kintamieji vienas su kitu yra statistiškai reikšmingai susiję.

2.3. Neįprastų reikšmių įtakos mažinimo metodikos

Iš kiekybinių kintamųjų aprašomosios statistikos (žr. 4 priedą bei išlaidų histogramų (žr. 4 pav.) matyti, kad duomenyse yra namų ūkių, kurie bendram vartojimui skiria neįprastai dideles sumas. Histogramos buvo gautos naudojant „seaborn“ bibliotekos *histplot()* funkciją.



4 pav. Pirmųjų keturių analizuojamų kiekybinių išlaidų kintamųjų histogramos

Nors buvo pateikti tik bendrų vartojimo išlaidų ir jų trijų pagrindinių kategorijų skirstiniai (žr. 4 pav.), tačiau yra aiški bendra tendencija ir kitų kintamųjų (žr. 3 priedą) pasiskirstymuose, t. y., nei vieno iš skaitinių kintamųjų, susijusių su išlaidomis, pasiskirstymas nėra panašus į teorinį, t. y., varpo formos normalųjį pasiskirstymą. Skirstiniai yra stipriai asimetriški, nes vartojimo išlaidų reikšmės yra artimos nuliui, taip pat matyti ir ilgos „uodegos“, rodančios ekstremalias reikšmes dešinėje pusėje.

Nyitrai ir Virag'as [91] teigia, kad yra būtina spręsti išskirčių (angl. *outliers*) problemą, nors ir nėra aišku, kaip galima tiksliausiai jas apibrėžti, siekiant užtikrinti didžiausią modelių prognozavimo tikslumą. Bendrai yra išskiriamos dvi pagrindinės metodų grupės, skirtos sumažinti išskirčių daromą įtaką – išskirčių turinčių stebėjimų pašalinimas ir kraštinių reikšminių apribojimas. Gussenbauer'is [92] analizuodamas namų ūkių vartojimo išlaidų duomenis pagal penkių šalių namų ūkių apklausas, nustatė, kad ne visada yra galima atskirti ar didelės išlaidos yra tikri atvejai, ar klaidingi (dėl matavimo klaidų) įrašai, todėl pabrėžė, kad automatiškai šalinti dideles išlaidų reikšmes be tvirtų

¹² https://phik.readthedocs.io/_/downloads/en/stable/pdf/

įrodymų yra klaida. Daugiausiai išleidžiančių namų ūkių pašalinimas gali iškreipti analizę, ypač kai yra analizuojamas pajamų ar išlaidų pasiskirstymas, kuriame natūraliai egzistuoja ekstremalios reikšmės, todėl išskirčių apribojimo būdai tokiais atvejais yra tinkamesni, nes leidžia sumažinti ekstremalių reikšmių įtaką neišmetant svarbių duomenų.

Vienas iš tokių metodų yra **vinzorizacija (angl. winsorization)**, skirta ekstremalių reikšmių įtakai sumažinti, apribojant tam tikrą procentą didžiausių ir mažiausių reikšmių iki nustatytų ribų. Vinsorizacijos naudojimo tikslas yra, neprarandant duomenų, sumažinti kraštutinių taškų įtaką vidurkiui ir dispersijai bei pagerinti modelių stabilumą. Taikant šį metodą stebinių kiekis nesumažėja, bet yra pakoreguojamos pačios kraštutinės reikšmės, t. y., reikšmės, kurios viršija nustatytas ribas, yra pakeičiamos pačių ribų reikšmėmis. Nyitrai ir Virag'as [91] pataria taikant vinsorizaciją naudoti skirtingo griežtumo ekstremalių reikšmių apribojimo ribas, pavyzdžiui, 1 %, 3 % arba 5 % procentilius. Naudojant 1 % vinsorizaciją, nuo kiekvieno duomenų galo (apatinių ir viršutinių reikšmių) būtų pakoreguotos po 1 %, t. y., iš viso 2 % visų reikšmių. Kraštutinės reikšmės, patekusios į 2 %, būtų pakeičiamos atitinkamai 1-ojo ir 99-ojo procentilio reikšmėmis. Naudojant 3 % vinsorizaciją, būtų koreguojamos po 3 % ekstremalių reikšmių nuo kiekvieno duomenų galo, t. y., iš viso 6 % visų reikšmių. Tuo tarpu 5 % vinsorizacijos taikymo atveju būtų koreguojamos po 5 % reikšmių nuo kiekvieno galo (iš viso 10 % visų reikšmių), tai reiškia, kad kraštutinės mažiausios ir didžiausios (išsiskiriančios) reikšmės būtų pakeičiamos į atitinkamai 5-ojo ir 95-ojo procentilio reikšmes. Pasirinktos per didelės vinsorizacijos ribos lemia reikšmingos informacijos apie duomenų variaciją praradimą bei analizės rezultatų iškraipymą. Vinsorizacijos atlikimui yra naudojama *winsorize()*¹³ funkcija iš *scipy.stats* bibliotekos.

Kitas plačiai naudojamas sprendimas – išlaidų kintamiesiems **natūrinio logaritmo**, žymimo $\ln(x)$ su pagrindu $e \approx 2,718$, **transformacijos pritaikymas**. Išsiskiriantys stebėjimai gali pernelyg paveikti tiek tiesinius, tiek medžiais grįstus modelius [30]. Bellégo'as, Benatta, Pape'ė [93] pritaria ir teigia, jog taikant logaritmavimą, didelės reikšmės yra suspaudžiamos labiau nei mažos, todėl kintamojo reikšmių pasiskirstymas priartėja prie simetrijos bei homoskedastiškumo prielaidos tenkinimo. Verta paminėti, kad jei logaritmė transformacija nebūtų taikoma, tai vienas labai didelis stebėjimas galėtų smarkiai iškreipti regresijos modeliu gautus įverčius, todėl transformacija prisideda prie parametrų stabilumo užtikrinimo. Bellégo'as, Benatta, Pape'ė [93] pastebi, kad yra susiduriama su iššūkiu, kai priklausomas kintamasis Y turi nulines reikšmes, nes paprasta $\ln(Y)$ transformacija yra galima tik kai priklausomas kintamasis yra griežtai teigiamas. Kadangi išlaidų duomenyse gali būti namų ūkių, kurie tam tikroms išlaidų kategorijoms neskiria lėšų visiškai, tiesioginės logaritmo transformacijos nebūtų galima taikyti, nes $\ln(0) = -\infty$. Dėl šios priežasties dažnai yra taikomas mokslininkų vadinamas „populiarus sprendimas“, t. y., prie kiekvienos stebėjimo reikšmės Y_i yra pridėjama teigiama konstanta Δ , gaunant $\tilde{Y}_i = Y_i + \Delta > 0$ ir atliekant logaritmę transformaciją $\ln(\tilde{Y}_i)$. Formulė:

$$\ln(Y_i + \Delta) = X_i' \beta + \omega_i; \quad (8)$$

čia Δ – tyrėjo pasirenkama konstanta, $\omega_i = \log\left(U_i + \frac{\Delta}{\exp(X_i' \beta)}\right)$ – transformuotas klaidos terminas.

Klaidos terminas pagal savo prigimtį koreliuoja su X_i , lemdamas klaidingą ryšį tarp kintamųjų [93]. Verta paminėti, kad nors ir konstantos parinkimas yra svarbus, tačiau $e^{X_i' \beta}$ gali tapti artima nuliui, dėl to gali atsirasti itin dideli šališkumai. Tyrime buvo naudota *NumPy* bibliotekos $\log1p$ ¹⁴ funkcija, tai

¹³ <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mstats.winsorize.html>

¹⁴ <https://numpy.org/doc/2.1/reference/generated/numpy.log1p.html>

pasirinkta konstanta buvo lygi 1. Siekiant atstatyti taip transformuotų duomenų prognozes į originalią skalę be sisteminio šališkumo, reikėtų naudoti retransformacijos korekcijos koeficientą (angl. *smearing factor*), t. y., jei $z_i = \ln(1 + Y_i)$, kai prognozės yra $\hat{z}_i = X_i' \beta$, o prognozės liekanos būtų $\hat{\epsilon}_i = z_i - \hat{z}_i$, tai korekcijos koeficiento gavimo formulė [93]:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n \exp(\hat{\epsilon}_i); \quad (9)$$

čia n – stebinių mokymo imtyje skaičius.

Gautos prognozės yra pataisomos apskaičiuotu koeficientu ir grąžinamos į savo originalią skalę. Rezultate yra gaunamos nešališkos prognozės originalioje kintamojo matavimo skalėje, kurias galima palyginti su tikromis reikšmėmis neiškreipiant modelio kokybės metrikų skaičiavimo. Prognozuotų reikšmių grąžinimui į pradinę skalę tyrime kartu su apskaičiuotu korekcijos koeficientu buvo naudojama *NumPy* bibliotekos *expm1*¹⁵ funkcija.

2.4. Regresijos uždavinio sprendimo modeliai

Norint atlikti kiekybinę regresijos analizę, kokybinius kintamuosius yra būtina perkoduoti į fiktyvius (angl. *dummy*) kintamuosius, kurie gali priimti tik dvi reikšmes: 1, jei stebėjimas priklauso tam tikrai grupei, ir 0 – jei nepriklauso, dėl to padidėja stulpelių skaičius duomenyse. Toks kodavimo būdas pasižymi tuo, jog yra standartinis ir nešališkas visiems kategoriniams kintamiesiems, t. y., tiek turintiems natūralią eilės tvarką (ranginiai), tiek jos neturintiems (nominalieji). Wooldridge'ė [94] teigia, kad kuriant fiktyvius kintamuosius yra būtina nurodyti bazinę grupę, kurios visi atitinkami fiktyvieji kintamieji yra lygūs nuliui, o kitų grupių poveikis yra vertinamas jos atžvilgiu. Jei bazinė grupė nėra nurodoma, t. y., yra sukuriama fiktyvieji kintamieji visiems kategorijos lygiams, susidaro situacija, kad vienas iš kintamųjų tampa visiškai priklausomas nuo kitų, t. y., egzistuoja tobulas kolinearumas, dar kitaip mokslininkų pavadintas „fiktyviųjų kintamųjų spąstais“. Dėl šios priežasties klasikinis tiesinės regresijos modelis nebepajėgia išspręsti lygčių sistemos ir gauti koeficientus. Atliekant tyrimą fiktyvių kintamųjų kūrimui buvo naudota *get_dummies()*¹⁶ funkcija iš *pandas* bibliotekos nurodant pašalinti pirmąją grupę iš kiekvieno kategorinio kintamojo, taip ją paverčiant referencine (bazine) grupe.

Visai turimai duomenų aibei buvo pritaikytas atsitiktinis skaidymas į mokymo ir testavimo imtis. Salazar'as ir kt. [95] teigia, kad mokymo imtį sudaro atsitiktinai atrinkta stebėjimų dalis, naudojama modelio parametrus išmokyti, kai testavimo imtis yra likusi atsitiktinė duomenų dalis, kuri yra laikoma „nematyta“ ir todėl naudojama modelių prognozavimo su naujais duomenimis objektyvumui vertinti. Duomenų padalijimas yra standartinis duomenų paruošimo modeliavimui žingsnis, kuris užtikrina nepriklausomą modelio kokybės vertinimą, saugo nuo galimo permokymo, t. y., per didelio modelio prisitaikymo prie duomenų išmokstant duomenyse esantį „triukšmą“, ir leidžia palyginti skirtingų metodų kokybę naudojant vienodą testavimo imtį. Tyrime duomenų padalijimui buvo naudota *train_test_split()* iš *sklearn* bibliotekos. Mokymo imčiai buvo skirta 80 % duomenų, o testavimo – 20 % visų stebinių.

Mašininio mokymosi algoritmų mokymo procese yra naudojami dvejų tipų parametrai: modelio parametrai, kurie yra gaunami iteratyviai modelio mokymosi metu, ir hiperparametrai, kurie negali būti nustatyti iteratyviai ir turi būti parenkami rankiniu būdu arba taikant optimizavimo metodus.

¹⁵ <https://numpy.org/doc/2.1/reference/generated/numpy.expm1.html>

¹⁶ https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html

Hiperparametų derinimas leidžia valdyti modelių mokymosi proceso eigą, jo sudėtingumą, padeda pagerinti tikslumą, naujų duomenų apibendrinamumą. Yan'as ir kt. [96] pabrėžia, kad tobulėjant mašininio mokymosi algoritmams, mokslininkai, remdamiesi empiriniais eksperimentais su įvairiais duomenų rinkiniais, jau yra pateikę kai kurių hiperparametrų numatytas reikšmes, tačiau jos kol kas yra grindžiamos tik praktine patirtimi, o ne teorija. Siekiant kurti tikslius prognozavimo modelius, yra taikomos optimizavimo ir derinimo strategijos, kurios padeda parinkti konkrečiam duomenų rinkiniui tinkamiausias hiperparametrų reikšmes. Dažniausiai naudojami optimizavimo algoritmai yra tinklelio paieška (angl. *grid search*) ar atsitiktinė paieška (angl. *random search*), kurie yra paprasti ir lengvai suprantami [30], tačiau Yan'as ir kt. [96] teigia, kad jie yra imlūs skaičiavimo ištekliams ir neefektyvūs, todėl yra verta rinktis metodą, gerinantį paieškos efektyvumą, t. y., naudoti Bajeso optimizacijos algoritmą. Šis metodas yra globalios paieškos metodas, kuris geba panaudoti ankstesnę informaciją, t. y., naudodamas praeities bandymų informaciją koncentruojasi į perspektyviausius hiperparametrų derinius. Bajeso optimizacijos algoritmo esmė yra pagal vertinimo funkciją sukurti pradinį kandidatinių sprendimų rinkinį, numatyti *priori* pasiskirstymo prielaidą ir nuolat papildyti rinkinį naujais kandidatų sprendimais, kurie yra vertinami pagal nustatytą vertinimo funkciją [96]. Po pakartotinių iteracijų yra randamas globalusis maksimumas. Verta paminėti, kad algoritmas pasižymi mažu skaičiavimo sudėtingumu ir efektyvumu, nes nereikia bandyti visų hiperparametrų kombinacijų kaip, pavyzdžiui, tinklelio paieškos atveju. Tyrime buvo naudojamas *BayesSearchCV*¹⁷ paieškos algoritmas iš *scikit-optimize* paketo.

Daugialypė tiesinė regresija yra plačiai taikoma analizuojant vartojimo išlaidų priklausomybę nuo mikrolygmens namų ūkių charakteristikų, todėl šis metodas tyrime buvo naudotas kaip orientacinis atskaitos taškas palyginimui, o jo metodologijos aprašas pateiktas 6 priede. Pagrindinis dėmesys buvo skiriamas pažangesnių metodų – atsitiktinių miškų ir ekstremalaus gradiento stiprinimo algoritmų – taikymui, siekiant įvertinti galimas netiesines priklausomybes. Ekstremalaus gradiento stiprinimo algoritmas, kiek yra žinoma, iki šiol nebuvo taikytas bendro vartojimo išlaidų ir jo struktūrinių dalių vertinimo kontekste, todėl jo įtraukimas gali būti vertinamas kaip metodologijos tokioje srityje praplėtimas.

2.4.1. Atsitiktinių miškų regresijos algoritmas

Šis metodas yra prižiūrimo mokymosi algoritmas, taikomas tiek klasifikavimo, tiek regresijos uždavinių sprendimui. Atsitiktinių miškų regresija jau buvo išbandyta atliekant su išlaidomis susijusius tyrimus (žr. 1.4. skyrių). Puttanapong'as ir Lim'as [60] teigia, kad atsitiktiniai miškai sudaryti iš daugybės sprendimo medžių, kurių kiekvienas yra treniruojamas naudojant atsitiktinės imties su gražinimu (angl. *bootstrap*) metodą, kuris reiškia, jog iš visos turimos duomenų imties yra atsitiktinai traukiami stebėjimai su gražinimu, todėl tas pats stebėjimas gali patekti į vieną imtį kelis kartus, kai kiti stebėjimai gali nepatekti į jokią imtį. Atsitiktinės imties su gražinimu metodo taikymas užtikrina įvairovę tarp skirtingų medžių, o dėl to sumažėja ir persimokymo rizika. Verta pabrėžti, kad sprendimų medis yra sudarytas iš mazgų (angl. *nodes*), kuriuose kiekvieną kartą ir yra priimamas sprendimas apie duomenų pagal tam tikrą kintamąjį skaidymą. Priklausomai nuo sprendimo, stebėjimai toliau keliauja į kitą mazgą šakomis, kurios jungia visus mazgus ir bendrai žymi duomenų padalijimo kelią. Procesas tęsiasi iki kol yra pasiekiami lapai (angl. *leaf nodes*), kuriuose ir yra pateikiama galutinė modelio prognozė. Skaidymo kokybė regresinio uždavinio atveju kiekviename mazge yra vertinama pagal maksimalų dispersijos (angl. *impurity*), t. y., vidinė tikslo kintamojo

¹⁷ <https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html>

reikšmių variacija mazge, sumažėjimą. Kuo padalinus duomenis labiau sumažėja dispersija, tuo geresniu yra laikomas ir atliktas skaidymas. Dispersijos sumažėjimas skaičiuojamas pagal formulę:

$$\Delta i(s_t, t) = i(t) - \left(\frac{n_L}{n} i(t_L) + \frac{n_R}{n} i(t_R) \right); \quad (10)$$

čia $i(t)$ – dispersija prieš skaidymą mazge t , $i(t_L)$, $i(t_R)$ – dispersija po skaidymo (kairėje ir dešinėje šakose), n_L, n_R – stebėjimų skaičius šakose, n – visų stebėjimų skaičius mazge t .

Taikant atsitiktinius miškus, galutinė prognozė yra gaunama apskaičiuojant visų sprendimų medžių prognozių vidurkį:

$$\hat{y} = \frac{1}{N_T} \sum_{i=1}^{N_T} M_i(x); \quad (11)$$

čia N_T – medžių skaičius ansamblyje, $M_i(x)$ – i -tojo medžio prognozė duotam stebėjimui X .

Bendras procesas, apimantis atsitiktinės imties su grąžinimu naudojimą ir medžių prognozių agregavimą, yra vadinamas „bagging“ metodu, kurio naudojimas padeda mažinti variaciją, neprarandant prognozavimo tikslumo.

Tyrime atsitiktinių miškų regresijos modelis buvo realizuotas naudojant *RandomForestRegressor*¹⁸ klasę iš *scikit-learn* bibliotekos. Lee'ė, Ong'as ir Lee'ė [30] bei Yan'as ir kt. [96] teigia, kad nors atsitiktinių miškų regresijai pritaikytas algoritmas turi nemažai hiperparametrų pasirinkimų, tačiau galima išskirti svarbiausius:

- Medžių skaičius (angl. *number of estimators/trees*). Nurodoma kiek medžių turi būti sukuriami modelyje. Didesnis medžių skaičius padeda sumažinti atsitiktinių klaidų įtaką, todėl modelis yra stabilesnis. Visgi nauda gali būti iki tam tikros ribos, vėliau didinant medžių skaičių modelio tikslumas gali nebesikeisti, o skaičiavimų ir atminties sąnaudos – išaugti. Standartiškai pasirenkamas medžių skaičius nuo 10 iki 500.
- Minimalus stebinių skaičius lapo mazge (angl. *minimum number of samples in a leaf*). Apribotas stebinių skaičius kiekviename lapo mazge, padeda išvengti per smulkių šakų susidarymo ir permokymo problemos. Minimalų stebinių skaičių rekomenduojama rinktis intervale nuo 2 iki 10 stebinių.
- Didžiausias medžio gylis (angl. *maximum depth*). Gylis yra didžiausias atstumas nuo šaknies mazgo iki lapo mazgo, t. y., medžio duomenų padalijimo kartai, nuo pradžios iki galutinio sprendimo. Per mažas medžio gylis gali lemti prastą modelio tikslumą, nes sukuriami per paprasti modeliai, kurie nesugeba atpažinti sudėtingų duomenų ryšių bei priklausomybių. Per didelis gylis gali lemti aukštą tikslumą mokymo duomenų imtyje, bet prastą – testavimo imtyje (persimokymo problema). Nors medžių gylių galima ir neriboti, tačiau standartiškai medžių gylį rekomenduojama rinktis intervale nuo 2 iki 20 kartų.
- Minimalus stebinių skaičius mazgo padalinimui (angl. *minimum number of samples required to split*). Šis parametras nurodo minimalų stebinių skaičių mazge, kad tas mazgas dar galėtų būti padalintas į dvi (ar daugiau) šakas. Jei mazge esančių stebinių skaičius yra mažesnis už numatytą, tokiu atveju mazgas yra daugiau nebedalinamas ir laikomas lapo mazgu. Mažesnė numatyta reikšmė didina persimokymo riziką, nes mazgai yra sudarinėjami net su labai mažai stebinių. Didesnis pasirinktas skaičius riboja medžio gylį ir gali lemti per daug paprastų

¹⁸ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

modelių sudarymą. Paprastai yra rekomenduojama rinktis intervale nuo 2 iki 20 stebinių, tačiau tyrėjo pasirinkimas labai priklauso nuo turimo duomenų dydžio.

- Didžiausias skaičius požymių, svarstomų kiekviename mazge padalijimui (angl. *maximum number of features considered for splitting at each node*). Parametru reguliuojama kiek požymių gali būti atsitiktinai parenkama ir vertinama kiekviename medžio mazge, kai yra sprendžiama, pagal kurį požymį reikėtų padalinti duomenis. Mažesnė reikšmė didina įvairovę tarp medžių, mažina persimokymo riziką. Visų požymių įtraukimas kelia persimokymo riziką dėl per didelio medžių panašumo, t. y., didelės koreliacijos tarp skirtingų medžių. Rekomenduojama naudoti proporcijas intervale nuo 10 % visų požymių iki 99,9 % visų požymių.

Gupta [98] išskiria pagrindinius (žr. 7 lent.) atsitiktinių miškų algoritmo privalumus ir trūkumus.

7 lentelė. Atsitiktinių miškų regresijos taikymo privalumai ir trūkumai (sudaryta pagal Gupta [102])

Privalumai	Trūkumai
Vertina netiesinius ryšius	Dėl medžių gausos, sunkiai interpretuojamas
Maža persimokymo rizika	Netinka realaus laiko prognozėms
Atsparus išskirtims	Dideli skaičiavimo resursų poreikiai
Nereikalauja požymių matavimo skalės keitimo	

Nepaisant trūkumų, atsitiktinių miškų algoritmo taikymas regresijos uždaviniuose yra laikomas galingu įrankiu, kuris padeda patikimai vertinti priklausomojo kintamojo pasikeitimus.

2.4.2. Ekstremalaus gradientinio stiprinimo algoritmas regresijai

Ekstremalaus gradientinio stiprinimo (angl. *Extreme Gradient Boosting*, XGBoost) algoritmą sukūrė Chen'as ir Guestrin'as [99]. Sprendimų medžiais grįstas ansamblinio mokymosi metodas duomenų mokslo tyrėjų dažnai minimas kaip vienas pažangiausių prižiūrimo mokymosi algoritmų, kurį sudaro objektyvinės funkcijos ir baziniai mokymosi modeliai, kurie paprastai yra sprendimų medžiai [100]. Kaip integruoto mokymosi metodas, jis yra skaluojamas gradientinio stiprinimo algoritmas, kuriame sprendimų medžiai (baziniai modeliai) generuojami nuosekliai, kai kiekvienas naujas medis mokosi iš ankstesnių modelių paklaidų. Galiausiai visų medžių prognozės yra sujungiamos, kad būtų gautas galutinis rezultatas. Vienas iš pagrindinių ekstremalaus gradientinio stiprinimo metodo privalumų yra jo gebėjimas riboti persimokymą, dėl naudojamo reguliavimo (angl. *regularization*) komponento objektyvinėje funkcijoje, kuris veikia kaip papildoma bauda už medžių sudėtingumą. Objektyvinę funkciją sudaro nuotolio funkcija, kuri įvertina modelio prognozių skirtumus nuo tikrųjų verčių, ir reguliavimo funkcija, kurioje yra taikomos baudos medžių lapų skaičiui ir svoriams:

$$\Omega(f) = \gamma Z + \frac{1}{2} \lambda \sum_j w_j^2 + \alpha \sum_i |w_j|; \quad (12)$$

čia γZ – bauda už kiekvieną medžio lapą, λ – L2 koeficientas, α – L1 koeficientas, w_j – kiekvieno medžio lapo svoris.

Shahani'is ir kt. [101] pažymi, kad taikant ekstremalaus gradientinio stiprinimo metodą regresijos uždaviniams, galutinės prognozės skaičiavimo formulė gali būti apibūdžiama:

$$\bar{y}_i = y_i^0 + \eta \sum_{k=1}^n f_k(U_i); \quad (13)$$

čia $\bar{y}_i$ – prognozuota i-toji išvestis su parametru vektoriumi, y_i^0 – originalių treniravimo duomenų parametru vidurkis, η – mokymosi greitis, U_i , n – modelyje naudojamų medžių skaičius kiekvienam f_k .

Mokymosi greitis nurodo kiek nauji medžiai prisideda prie bendro modelio atnaujinimo. Medžiai, naudojant ekstremalaus gradientinio stiprinimo algoritmu kuriamą modelį, pridedami iteratyviu būdu:

$$y_i^{-k} = y_i^{-(k-1)} + \eta f_k; \quad (14)$$

čia f_k – medžių lapų svoriai, kurie randami minimizuojant k -tojo medžio tikslinę funkciją.

kiekviename etape prie modelio yra prijungiamas medis, o k -tojo etapo prognozė yra apskaičiuojama naudojant ankstesnio etapo prognozę. Norint suformuoti medį, lapai yra dalijami į kelis naujus lapus. Šis dalijimosi mechanizmas yra aprašomas naudojant padidėjimo parametrus (angl. *gain parameters*).

$$G = \frac{1}{2} \left[\frac{O_L^2}{P_L + \lambda} + \frac{O_R^2}{P_R + \lambda} + \frac{(O_L + O_R)^2}{P_L + P_R + \lambda} \right]; \quad (15)$$

čia G – padidėjimo parametras, O_R ir P_R – dešiniojo lapo suma ir pasiskirstymas, O_L ir P_L – kairiojo lapo suma ir pasiskirstymas, λ – reguliavimo koeficientas.

Jei padidėjimo parametras artėja prie nulio, tuomet yra laikoma, jog dalijimo kriterijus nėra pagrįstas. Verta paminėti, kad didesnės reguliavimo parametrų reikšmės mažina padidėjimo reikšmę, taip užkertant kelią lapų konvergencijai (angl. *leaf convolution*).

Ekstremalaus gradiento stiprinimo algoritmas, kiek yra žinoma, iki šiol nebuvo taikytas bendro vartojimo išlaidų ir jo struktūrinių dalių vertinimo kontekste, todėl jo įtraukimas gali būti vertinamas kaip metodologijos tokioje srityje praplėtimas. Tyrime ekstremalaus gradientinio stiprinimo regresijos modelis buvo realizuotas naudojant *XGBRegressor*¹⁹ funkciją iš *XGBoost* bibliotekos. Yan'as ir kt. [96] bei Shahani'is ir kt. [101] teigia, kad nors ekstremalaus gradientinio stiprinimo regresijai pritaikytas algoritmas turi nemažai hiperparametrų pasirinkimų, tačiau jie išskiria aktualiausius:

- Medžių skaičius (angl. *number of estimators/trees*). Nurodomas medžių skaičius ansamblyje. Didesnis medžių skaičius gali padidinti modelio tikslumą, tačiau su didesne persimokymo rizika. Standartiškai pasirenkamas medžių skaičius nuo 10 iki 500.
- Didžiausias medžio gylis (angl. *maximum depth*). Skirtingai nei taikant atsitiktinių miškų metodą, naudojant ekstremalaus gradientinio stiprinimo metodą, papildomi medžiai nuosekliai pataiso ankstesniųjų klaidas, dėl to dažnai užtenka mažesnio medžių gylio. Standartiškai medžių gylį rekomenduojama rinktis intervale nuo 2 iki 10.
- Mokymosi greitis (angl. *learning rate*). Nurodo kiek stipriai modelis keičia parametrus po kiekvienos klaidos apskaičiavimo. Per didelis mokymosi greitis gali lemti per didelį prisitaikymą prie duomenyse esančio triukšmo ar peršokti optimalų sprendinį. Lėtas mokymosi greitis neleidžia modeliui persimokyti ar praleisti geriausią sprendinį, tačiau tam reikia daugiau medžių ar iteracijų. Standartiškai mokymosi greitį rekomenduojama rinktis intervale nuo 0 iki 0,5.
- Dalinė duomenų imtis (angl. *subsample*). Atsitiktinė mokymo duomenų dalis, kuri yra panaudojama kiekvieno medžio mokymui. Mažesnė duomenų imtis padidina atsitiktinumą modelio mokymosi procese, taip sumažinant medžių tarpusavio panašumą. Per maža duomenų imtis (mažesnė nei 50 %) gali lemti modelio tikslumo pablogėjimą, nes kiekvieno

¹⁹ https://xgboost.readthedocs.io/en/stable/python/python_api.html

medžio mokymuisi yra pateikiama per mažai informacijos. Standartiškai duomenų imties dalį rekomenduojama rinktis intervale nuo 0,5 iki 1.

- Kintamųjų dalis, naudojama kiekvieno medžio kūrimui (angl. *column sample by tree*). Ne visų kintamųjų naudojimas mokymosi procese lemia padidėjusį atsitiktinumą, dėl to mažėja medžių tarpusavio koreliacija ir didėja modelio generalizacija. Standartiškai kintamųjų dalį rekomenduojama rinktis intervale nuo 0,5 iki 1.
- L1 reguliavimo koeficientas svoriams (angl. *alpha coefficient for L1 regularization*). L1 bauda yra siekiama sumažinti svorių reikšmes iki nulio, todėl yra natūraliai atliekama nereikšmingų kintamųjų atranka. Koeficientą didinant yra stipriau baudžiami didesni svoriai, dėl to modelis tampa paprastesnis ir mažiau linkęs į persimokymą. Per didelės koeficiento reikšmės (pvz., > 1) gali sukelti pernelyg stiprų reguliavimą, t. y., pašalinti per daug kintamųjų ir taip pabloginti modelio gebėjimą apibendrinti. Standartiškai α koeficientą rekomenduojama rinktis intervale nuo 0 iki 1.
- L2 reguliavimo koeficientas svoriams (angl. *lambda coefficient for L2 regularization*). L2 bauda priverčia didelius svorius artėti link nulio, tačiau jų visiškai nepanaikina. Todėl ir visi požymiai lieka modelyje, tik su mažesnėmis reikšmėmis. Pernelyg didelė λ reikšmė lemia per griežtą reguliavimą, t. y., per daug sumažina svorius, dėl to modelis gali tapti per daug paprastas. Standartiškai λ koeficientą rekomenduojama rinktis intervale nuo 0 iki 1.
- Minimali instancijų svorių suma, reikalinga mazge (angl. *minimum sum of instance weight required in a child node*). Minimalus svorių kiekis, kuris turi susikaupti mazge, kad būtų galima jį padalinti. Jeigu mazgo svorių suma yra mažesnė už pasirinktą, tas mazgas yra toliau nebedalinamas. Pasirinkta maža reikšmė lemia didelį mažų mazgų kūrimąsi, kuris sukelia persimokymą. Standartiškai rekomenduojama sumą rinktis intervale nuo 1 iki 8.
- Minimalus nuostolio funkcijos sumažėjimas, reikalingas, kad įvyktų mazgo padalijimas (angl. *loss reduction needed for a further leaf split*). Didesnė γ reikšmė reiškia, kad padalijimai leidžiami tik tada, kai jie pakankamai daug sumažina nuostolį. Standartiškai rekomenduojama γ rinktis intervale nuo 0 iki 1.

Gupta [98] pabrėžia, kad ekstremalaus gradientinio stiprinimo algoritmas pasižymi aukštu prognozavimo tikslumu, gebėjimu modeliuoti sudėtingus ryšius ir įvertinti kintamųjų sąveikas, tačiau kartu yra imlus skaičiavimo resursams ir, jei netinkamai reguliuojamas, gali persimokyti.

2.5. Modelių kokybės vertinimo ir interpretavimo metodai

Modelio veikimo kokybė paprastai vertinama siekiant užtikrinti patikimus prognozavimo rezultatus. Naudojant įvairias vertinimo metrikas, tokias kaip determinacijos koeficientas, vidutinė kvadratinė paklaida ir kt., yra parenkamas optimaliausias modelis. Pagal tinkamiausią modelį yra atliekama rezultatų interpretacija, kurios įgyvendinimui, taikant mašininio mokymosi modelius, buvo papildomai panaudota SHAP analizė.

Modelių kūrimo etape, siekiant sumažinti šališkumą, buvo taikyta k-dalių kryžminė validacija. Pasirinktos modelių kokybės vertinimo metrikos buvo determinacijos koeficientas (R^2), kai tiesinės regresijos atveju papildomai – pataisytas R^2 , kvadratinės paklaidos šaknis (angl. Root Mean Squared Error, RMSE), vidutinė absoliutinė paklaida (angl. Mean Absolute Error, MAE) bei medianinė absoliutinė paklaida (angl. Median Absolute Error, MedAE). Šių metrikų metodologinis aprašas pateiktas 7 priede.

Tuo tarpu rezultatų interpretavimui buvo taikomas SHAP analizės metodas. Šaplio vertė (angl. *Shapley value*) buvo pristatyta *Lloyd Shapley*, kuris, remdamasis žaidimų teorija, tyrė kaip teisingai reikėtų paskirstyti bendrą naudą žaidimo žaidėjams pagal jų individualų indėlį [103]. SHAP (angl. *SHapley Additive exPlanations*) yra analizės metodas kiekvienam požymiui priskiriantis svarbos reikšmę konkrečiai prognozei ir šiai dienai daugeliui tyrėjų padedantis atskleisti vadinamosios „juodosios dėžės“ savybes taikant mašininio mokymosi modelius [104]. Kadangi žaidimų teorijoje kiekvienas žaidėjas gali prisijungti prie žaidimo arba likti nuošalyje, SHAP analizės kontekste yra apibrėžiama, ką reiškia, kai į modelį yra įtraukimas veiksnys ir kai nėra²⁰. Požymis yra laikomas įtrauktu į modelį, kai jo reikšmė yra žinoma, ir neįtrauktu, kai jo reikšmė nėra žinoma, tai yra laikoma dažniausiai naudojamu apibrėžimu. Norint įvertinti modelį f , kai yra žinoma tik dalis požymių S , likę požymiai yra pakeičiami jų vidutinėmis reikšmėmis pagal sąlyginę tikėtinąją reikšmę. Tai galima atlikti naudojant išraišką:

$$f_x(S) = f(x_S, E[X_{\bar{S}}]); \quad (16)$$

čia S – požymių pogrupis, kurio reikšmės yra žinomos, $[X_{\bar{S}}]$ – likę nežinomi požymiai.

Tokiu būdu yra nustatomi požymiai S , o likusių požymių reikšmės yra pakeičiamos jų tikėtinaisiais vidurkiais. Šaplio reikšmės yra taikomos mašininio mokymosi modelio sąlyginei tikėtinajai funkcijai gauti, kuri ir yra vadinama SHAP reikšme, vertinanti požymio indėlį į modeliu atliekamą prognozę. Konkretaus požymio SHAP reikšmė prognozuojant gaunama³⁰:

$$\text{SHAP reikšmė} = f(x) - E[f(x)]; \quad (17)$$

čia $f(x)$ – tai modelio prognozė konkrečiam stebėjimui, $E[f(x)]$ – vidutinė modelio išėjimo reikšmė.

SHAP reikšmė parodo, kiek modelio prognozė pasikeičia dėl konkretaus aiškinamojo kintamojo. Kitaip tariant, tai yra skirtumas tarp vidutinės modelio prognozės naudojant visus duomenis ir prognozės, atliktos žinant konkretaus kintamojo reikšmę.

SHAP analizė leidžia paaiškinti modelio prognozes interpretuojant kiekvieno kintamojo poveikį joms [30]. Vienas pagrindinių SHAP privalumų yra galimybė atlikti globalų modelio interpretavimą vizualiai. Remiantis globalia modelio interpretacija, galima lengviau suprasti taisykles ar tendencijas, pagal kurias modelis buvo mokintas naudojant visą ar dalį duomenų imties, bet ne vieną konkretų atvejį. Globali interpretacija SHAP kontekste yra pateikiama suminės diagramos (angl. *summary plot*) pavidalu, kuriame atsispindi kiekvienas duomenų taškas, reiškiantis tam tikro požymio *Shapley* reikšmę, kai taško spalva parodo to požymio reikšmę. Be to, naudojant suminę diagramą galima reitinguoti kintamuosius pagal jų indėlį į modelio prognozes. Kitas vizualizacijos įrankis yra priklausomybių grafikai (angl. *dependence plot*), kurie yra sklaidos diagramos, rodančios kaip kintamieji kartu veikia modelio prognozes. Tyrimo SHAP analizė buvo atlikta naudojant *shap*²¹ bibliotekos įrankius: *TreeExplainer*, *shap_values*, *summary_plot*, *dependence_plot*, kurie padėjo interpretuoti mašininiais mokymosi modeliais gautas prognozes bei vizualizuoti nepriklausomų kintamųjų poveikį išlaidų pasikeitimui.

²⁰https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

²¹ <https://shap.readthedocs.io/en/latest/>

išlaidomis ar jų struktūrinėmis dalimis. Siekiant sukurti kuo tikslesnius modelius, buvo pasirinkta toliau tirti tų vartojimo išlaidų priklausomybes nuo namų ūkių charakteristikų, kurių ϕ_K koreliacijos koeficientas bent su vienu požymiu buvo ne mažesnis nei 0,3, t. y., atlikti **bendrų vartojimo išlaidų, maisto ir nealkoholinių gėrimų, informacijos ir ryšių paslaugų bei draudimo ir finansinių paslaugų** išlaidų priklausomybių nuo Lietuvos namų ūkių charakteristikų poveikio vertinimą (žr. 5 pav. pabraukta). Išlaidų kategorijų, tokių kaip sveikata, transportas ir kt., atvejais yra būdingi itin silpni ϕ_K koreliacijos koeficientai su tyrime naudotais nepriklausomais kintamaisiais, dėl to pastarieji būtų nepakankami patikimai aprašyti išlaidų kitimą. Verta paminėti, kad, remiantis hibridiniu p-reikšmių skaičiavimu, apjungiančiu asimptotinį χ^2 testą bei *Monte Carlo* simuliacijas, visi veiksniai bendrų vartojimo išlaidų atveju buvo statistiškai nereikšmingi (žr. 8 priedą). Nereikšmingi ryšiai buvo nustatyti ir kitų išlaidų kategorijų – maisto ir nealkoholinių gėrimų, informacijos ir ryšių paslaugų bei draudimo ir finansinių paslaugų (žr. 8 priedą) atvejais. Kadangi atliekant tyrimą buvo naudoti sprendimų medžiais grįsti metodai, kurie pasižymi atsparumu statistiškai nereikšmingiems kintamiesiems dėl savybės nereikšmingus kintamuosius retai pasirinkti skaidymui mazguose, tyrimas buvo tęsiamas su statistiškai nereikšmingai priklausančiais kintamaisiais. Taikant ϕ_K koreliacijos koeficiento skaičiavimus tik tarp veiksmų (žr. 9 priedą) buvo nustatyta, kad tarp kintamųjų – namų ūkio tipo ir gyvenimo su partneriu, centralizuoto karšto vandens tiekimo ir pagrindinio būsto šildymo būdo, motinos gimimo šalies ir tėvo gimimo šalies – egzistuoja stiprūs tarpusavio ryšiai ($\phi_K > 0,9$), kurie gali lemti multikolinearumo problemą modelius kuriant daugialypės tiesinės regresijos metodu. Kita vertus, medžiais grįsti mašininio mokymosi metodai yra nejautrūs multikolinearumui, nes skaidymo metu pasirenka tik vieną patį svarbiausią iš tarpusavyje stipriai koreliuotų kintamųjų.

Bendras vartojimo išlaidas veikiančių požymių identifikavimas. Vienas populiariausių filtravimo metodų, skirtų patikrinti, ar tam tikri kintamieji reikšmingai veikia priklausomąjį kintamąjį, yra dispersinės analizės Fišerio statistikos testas. Siekiant korektiškai jį taikyti reikia patikrinti analizuojamiems duomenims taikomas prielaidas (žr. 5 priedą). Šapiro ir Vilko testu buvo tikrinama metodo naudojimui reikalinga normalumo prielaida, t. y., priklausomojo kintamojo reikšmių kiekviename kategorinio kintamojo lygmenyje sutapimą su teoriniu normaliuoju skirstiniu.

Naudojant programinės įrangos „Python“ Šapiro ir Vilko testui atlikti skirtą funkciją buvo gauti rezultatai (žr. 10 priedą). Bendrų vartojimo išlaidų pasiskirstymas pagal kategorinių veiksmų lygius, išskyrus darbo sutarties tipo terminuotos sutarties grupę ir motinos gimimo šalies „nenurodė“ grupę, neatitiko normalumo prielaidos, nes Šapiro ir Vilko testo p-reikšmė buvo lygi 0,00, tai yra mažiau už 5% reikšmingumo lygmenį. Dėl šios priežasties nulinė hipotezė buvo atmesta ir priimta bendra išvada, kad duomenys nėra pasiskirstę pagal normalųjį skirstinį. Tikrinant dispersijų homogeniškumo prielaidą buvo panaudotas statistinis Leveno testas (žr. 8 lent.), kuriuo buvo tikrinama hipotezė, kad skirtingų to paties kategorinio kintamojo grupių dispersijos yra panašios (homogeniškos).

Rezultatai (žr. 8 lent.) rodo, kad daugumos požymių atveju nulinę hipotezę reikia atmesti, nes Leveno statistikos p-reikšmė yra mažesnė už 0,05. Pabrėžtina, kad siekiant užtikrinti nuoseklią ir vieningą duomenų analizę, buvo laikomasi principo, kad dauguma kintamųjų turi atitikti prielaidų reikalavimus. Kadangi daugelis veiksmų neatitiko normalumo ir dispersijų homogeniškumo prielaidų, buvo nuspręsta, jog tolesnė požymių atranka naudojant ANOVA metodą būtų nekorektiška.

8 lentelė. Leveno testo rezultatai, kai priklausomas kintamasis yra bendros išlaidos

Kategorinis kintamasis	Leveno p-reikšmė	Prielaida tenkinama	Kategorinis kintamasis	Leveno p-reikšmė	Prielaida tenkinama
Gyvenamoji vieta	0,0433	Ne	Pagrindinis būsto šildymo būdas	0,0000	Ne
Regionas	0,0002	Ne	Asmenų skaičius namų ūkyje	0,0000	Ne
Namų ūkio tipas	0,0000	Ne	Namų ūkio užimamas naudingasis plotas	0,0000	Ne
Išlaikomų vaikų skaičius	0,0000	Ne	Motinos gimimo šalis	0,5765	Taip
Socialinė ekonominė grupė	0,0000	Ne	Gimimo šalis	0,5342	Taip
Pajamų šaltinis	0,0000	Ne	Pagrindinės pilietybės šalis	0,4519	Taip
Namų ūkio galvos amžius	0,0000	Ne	Tėvo gimimo šalis	0,5503	Taip
Namų ūkio galvos išsilavinimas	0,0000	Ne	Lytis	0,0000	Ne
Pajamų kvintilinė grupė	0,0000	Ne	Teisinė santuokinė padėtis	0,0000	Ne
Būsto tipas	0,0000	Ne	Gyvena su partneriu	0,0000	Ne
Būsto valdos statusas	0,0000	Ne	Pagrindinis darbas: profesija	0,0000	Ne
Centralizuotas šalto vandens tiekimas	0,1808	Taip	Šiuo metu mokosi	0,3574	Taip
Centralizuotas karšto vandens tiekimas	0,0011	Ne	Darbo sutarties tipas	0,0000	Ne

Kai dispersinės analizės prielaidos nėra tenkinamos, dažniausiai yra naudojami neparametriniai metodai, iš kurių vienas žinomiausių – Kruskalio ir Voliso testas, kurio taikymas turi vieną labai svarbią prielaidą – grupėse esančių duomenų pasiskirstymas turi būti panašios formos, t. y., pasižymėti panašiu asimetriškumu, ekscesu, modalumu ir neturėti per daug didelės dispersijos. Tyrime grupių pasiskirstymų formų panašumą buvo nuspręsta patikrinti naudojant k-imčių Andersono ir Darlingo testą, kuriuo yra formuojama nulinė hipotezė, teigianti, kad visų imčių skirstiniai yra panašūs.

9 lentelė. K-imčių Andersono ir Darlingo testo rezultatai, siekiant įvertinti, ar bendrų vartojimo išlaidų skirstiniai atskirose grupėse yra panašūs

Kategorinis kintamasis	Andersono-Darlingo statistika	p-reikšmė	Bent vienos grupės pasiskirstymo forma skiriasi
Gyvenamoji vieta	7,961721	0,001	Taip
Regionas	3,16247	0,017	Taip
Namų ūkio tipas	326,4714	0,001	Taip
Išlaikomų vaikų skaičius	234,7471	0,001	Taip
Socialinė ekonominė grupė	372,9406	0,001	Taip
Pajamų šaltinis	363,3351	0,001	Taip
Namų ūkio galvos amžius	231,2157	0,001	Taip
Namų ūkio galvos išsilavinimas	137,5677	0,001	Taip
Pajamų kvintilinė grupė	190,9641	0,001	Taip
Būsto tipas	17,19218	0,001	Taip
Būsto valdos statusas	64,90088	0,001	Taip
Centralizuotas šalto vandens tiekimas	2,581439	0,028	Taip
Centralizuotas karšto vandens tiekimas	15,92501	0,001	Taip
Pagrindinis būsto šildymo būdas	67,52947	0,001	Taip
Asmenų skaičius namų ūkyje	331,1664	0,001	Taip
Namų ūkio užimamas naudingasis plotas	127,4736	0,001	Taip
Motinos gimimo šalis	0,562437	0,222	Ne

Gimimo šalis	2,603337	0,028	Taip
Pagrindinės pilietybės šalis	5,418815	0,002	Taip
Tėvo gimimo šalis	-0,67272	0,250	Ne
Lytis	146,6227	0,001	Taip
Teisinė santuokinė padėtis	263,7969	0,001	Taip
Gyvena su partneriu	325,4734	0,001	Taip
Pagrindinis darbas: profesija	241,2492	0,001	Taip
Šiuo metu mokosi	1,130437	0,111	Ne
Darbo sutarties tipas	299,1182	0,001	Taip

Iš lentelėje pateiktos informacijos matyti, kad daugumos kategorinių kintamųjų bent vienos grupės duomenų skirstinys skiriasi nuo kitų grupių skirstinių, nes gautos Andersono ir Darlingo statistikos p-reikšmės (žr. 9 lent.), išskyrus „šiuo metu mokosi“, „tėvo gimimo šalis“ ir „motinos gimimo šalis“ kintamuosius, yra mažesnės už pasirinktą 5 % reikšmingumo lygmenį. Tai reiškia, kad daugumos kategorinių kintamųjų grupių pasiskirstymo formos skiriasi statistiškai reikšmingai, todėl taikyti Kruskalio ir Valiso testą kaip vieningą metodą visų kategorinių požymių atrankai būtų nepagrįsta.

Vartojimo išlaidoms atrinkti reikšmingus požymius buvo panaudotas kitas neparametrinis metodas, kuris ignoruoja duomenų pasiskirstymo formą, veikia net esant nevienodoms grupių imtims ir yra atsparus išskirtims, t. y., Moodo medianos testas. Taikant testą buvo formuojama nulinė hipotezė, teigianti, jog skirtingų kategorijų grupių medianos pagal vartojimo išlaidų dydį statistiškai reikšmingai nesiskiria.

10 lentelė. Medianos testo rezultatai, skirti vartojimo išlaidų skirtumams tarp kategorinių grupių vertinti

Kategorinis kintamasis	χ^2 statistikos reikšmė	p-reikšmė	Požymis yra reikšmingas
Gyvenamoji vieta	14,258	0,0008	Taip
Regionas	1,1675	0,2799	Ne
Namų ūkio tipas	823,67	0,0000	Taip
Išlaikomų vaikų skaičius	416,28	0,0000	Taip
Socialinė ekonominė grupė	791,46	0,0000	Taip
Pajamų šaltinis	767,04	0,0000	Taip
Namų ūkio galvos amžius	522,15	0,0000	Taip
Namų ūkio galvos išsilavinimas	287,46	0,0000	Taip
Pajamų kvintilinė grupė	452,26	0,0000	Taip
Būsto tipas	37,977	0,0000	Taip
Būsto valdos statusas	121,04	0,0000	Taip
Centralizuotas šalto vandens tiekimas	0,366	0,5452	Ne
Centralizuotas karšto vandens tiekimas	22,807	0,0000	Taip
Pagrindinis būsto šildymo būdas	134,16	0,0000	Taip
Asmenų skaičius namų ūkyje	516,02	0,0000	Taip
Namų ūkio užimamas naudingasis plotas	230,93	0,0000	Taip
Motinos gimimo šalis	4,8494	0,0885	Ne
Gimimo šalis	4,9072	0,0267	Taip
Pagrindinės pilietybės šalis	8,095	0,0044	Taip
Tėvo gimimo šalis	2,944	0,2294	Ne
Lytis	175,07	0,0000	Taip
Teisinė santuokinė padėtis	566,16	0,0000	Taip
Gyvena su partneriu	557,20	0,0000	Taip
Pagrindinis darbas: profesija	861,51	0,0000	Taip
Šiuo metu mokosi	4,3971	0,0360	Taip
Darbo sutarties tipas	513,45	0,0000	Taip

Rezultatai rodo, kad visų kategorinių kintamųjų, išskyrus „tėvo gimimo šalis“, „centralizuotas šalto vandens tiekimas“, „regionas“ ir „motinos gimimo šalis“, bent viena mediana tarp grupių statistiškai reikšmingai skyrėsi, nes gautos χ^2 statistikos p-reikšmės buvo mažesnės už 5 % reikšmingumo lygį. Tai reiškia, kad vartojimo išlaidų pasiskirstymas skiriasi priklausomai nuo kategorijos grupės, kas ir rodo paties kintamojo svarbumą priklausomam kintamajam. Verta paminėti, kad kuo didesnė yra χ^2 statistikos reikšmė, tuo yra labiau tikėtina, kad kategorinio kintamojo grupės tikrai skiriasi pagal išlaidų medianą, o ne tik atsitiktinai, tai kartu rodo ir to požymio didesnę poveikį bendroms vartojimo išlaidoms. Remiantis Moodo medianos testu rastomis aukščiausiomis χ^2 statistikos vertėmis (žr. 10 lent.) galima teigti, kad bendroms vartojimo išlaidoms didžiausią poveikį turėtų daryti namų ūkio galvos pasirinkta profesija, namų ūkio tipas (šeimyninė padėtis), socialinė-ekonominė grupė bei pajamų šaltinis. Remiantis testo rezultatais taip pat nustatyta, kad tokios namų ūkių charakteristikos kaip būsto tipas, karšto vandens tiekimo šaltinis, gyvenamoji vieta ar šalis, neturėtų stipriai veikti namų ūkių vartojimo išlaidų pasikeitimų. Tuo tarpu regiono, tėvo gimimo šalies, centralizuoto šalto vandens tiekimo ir motinos gimimo šalies kintamieji tolesnėje analizėje turėtų būti nebenaudojami.

Daugybiniai poriniai palyginimai buvo atlikti naudojant porinius Moodo medianos testus. Verta paminėti, kad atliekant daugybę testų vienu metu p-reikšmėms buvo taikoma Benjamino ir Hochbergo korekcija. Toliau pateikiami rezultatai (žr. 11 lent.) rodo tik labiausiai statistiškai reikšmingai tarpusavyje nesiskyrusias grupes, kai visa informacija buvo pateikta 11 priede.

11 lentelė. Porinių medianų testų rezultatai, kai pagal bendras vartojimo išlaidas nebuvo rasta skirtumų tarp kategorinių požymių grupių

Kategorinis kintamasis	Lyginamosios grupės	χ^2 statistikos reikšmė	p - reikšmė	Pakoreguota p-reikšmė
Gyvenamoji vieta	„5 didieji miestai“ ir „Kaimas“	0,3509	0,5536	0,6004
Namų ūkio tipas	„Pora be išlaikomų vaikų“ ir „Kiti namų ūkiai be išlaikomų vaikų“	0,0028	0,9578	0,9753
Pajamų šaltinis	„Pensija“ ir „Pašalpa, stipendija ir kitas pragyvenimo šaltinis“	1,9882	0,1585	0,2010
	„Darbo užmokestis“ ir „Pajamos iš savo ar šeimos verslo“	2,7225	0,0989	0,1271
Namų ūkio galvos išsilavinimas	„Vidurinis“ ir „Aukštesniojo ir aukštojo nesuteikiantis povidurinis“	0,0072	0,9323	0,9552
Pajamų kvintilinė grupė	„2-a pajamų kvintilinė grupė“ ir „1-a pajamų kvintilinė grupė“	0,0438	0,8342	0,8762
Būsto tipas	„Vieno buto namas“ ir „Butas dviejų butų ar sublokuotame name“	0,0000	1,0000	1,0000
Būsto valdos statusas	„Būsto savininkas, neturintis išmokėti paskolos“ ir „Gyvena nemokėdamas nuomos“	0,7074	0,4003	0,4579
	„Būsto savininkas, turintis neišmokėtą paskolą“ ir „Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina“	0,1219	0,7270	0,7733
Pagrindinis būsto šildymo būdas	„Centrinis šildymas iš šilumos tinklų“ ir „Nėra stacionaraus šilumos šaltinio“	1,5787	0,2089	0,2566
Gyvena su partneriu	„Taip, teisiniu pagrindu“ ir „Taip, be teisinio pagrindo“	0,0000	1,0000	1,0000
Pagrindinis darbas: profesija	„Specialistai“ ir „Įstaigų tarnautojai“	0,0000	1,0000	1,0000

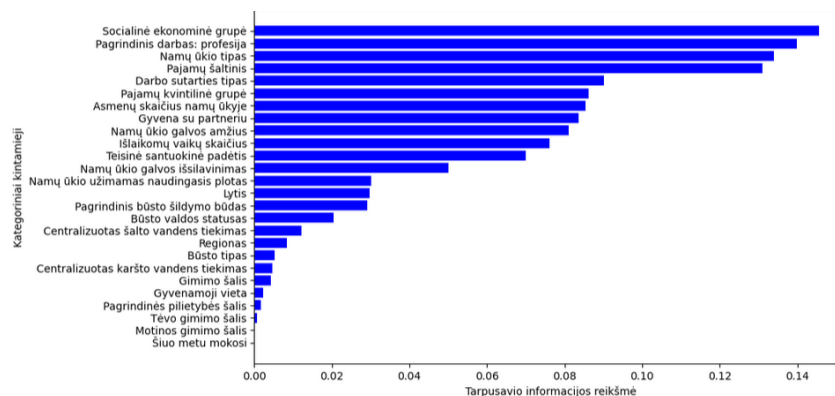
Iš lentelės matyti, kad nemažai grupių turi panašias medianas. Verta atkreipti dėmesį į tas grupes, kurių pakoreguota p-reikšmė yra artima vienetui, nes tai rodo, kad tarp tose grupėse vartojimui skirtų

išlaidų dydžių nėra jokio reikšmingo skirtumo. Grupių poros, tarp kurių medianų testas nenustatė jokio skirtumo:

- poros be vaikų ir kiti namų ūkiai be išlaikomų vaikų;
- asmenys, turintys vidurinę ir aukštesniojo ar aukštojo nesuteikiantį povidurinį išsilavinimą;
- asmenys, gyvenantys vieno buto name ir bute dviejų butų ar sublokuotame name;
- teisiniu ir be teisinio pagrindo gyvenantys su partneriais;
- yra specialistai ir įstaigų tarnautojai.

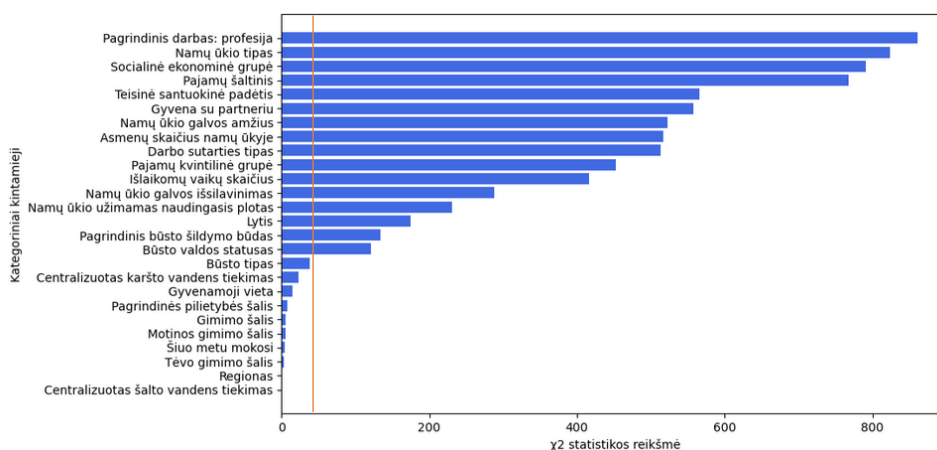
Remiantis tuo, būtų galima visas šias grupes apjungti į vieną grupę, pavyzdžiui, „poros ir kiti namų ūkiai be išlaikomų vaikų“ arba „namų ūkiai gyvenantys su partneriu“, tačiau, siekiant neprarasti svarbios informacijos, šio tyrimo kontekste grupės nebus apjunginėjamos. Įdomu tai, kad reikšmingų skirtumų pagal bendrojo vartojimo medianų reikšmes nebuvo rasta tarp namų ūkių, kurie gyvena penkiuose didžiausiuose Lietuvos miestuose bei tų, kurie gyvena kaime.

Vienas iš naujesnių įrankių požymių atrankai atlikti yra TI metodas, kuris ir buvo pritaikytas šio tyrimo metu (žr. 6 pav.). Svarbu paminėti, kad TI nėra statistinis testas, todėl jis neturi statistikos reikšmių ar aiškių procedūrų, skirtų *post – hoc* analizei atlikti.



6 pav. Tarpusavio informacijos metodo panaudojimas požymių atrankai, kai vertinamos bendros išlaidos

Remiantis TI mato rezultatais, galima teigti, kad namų ūkių socialinė-ekonominė grupė, profesija, namų ūkio tipas ir pajamų šaltinis yra svarbiausi požymiai, veikiantys vartojimo išlaidų dydį. Palyginimui, buvo vizualizuotos ir χ^2 statistikos reikšmės (žr. 7 pav.), pagal kurių dydį buvo identifikuotos svarbiausios namų ūkių charakteristikos, veikiančios bendrojo vartojimo išlaidas.



7 pav. Moodo medianos testo panaudojimas požymių atrankai, kai vertinamos bendros vartojimo išlaidos

Abu metodai rodo, kad pagal reikšmių dydžius svarbiausiai vartojimo išlaidas veikiantys kintamieji yra socialinė-ekonominė grupė, profesija, namų ūkio tipas ir pajamų šaltinis, nors ir TI metodas kaip pačią svarbiausią pabrėžia socialinę-ekonominę grupę, o medianos testas – profesiją. Abu metodai taip pat sutaria, kad gyvenamoji vieta ar būsto tipas vartojimo išlaidų pasikeitimus gali veikti silpnai. Nors abu metodai su namų ūkių kilme susijusiems kintamiesiems priskiria labai mažas reikšmes, jie nesutaria dėl dabartinio mokymosi statuso ir centralizuoto šalto vandens tiekimo veiksmų. TI metodas dabartinį namų ūkių mokymosi statusą laiko visiškai nereikšmingu, kai medianos testas jo reikšmingumo neatmeta. Tuo tarpu medianos testas centralizuoto šalto vandens tiekimą laiko mažiausiai reikšmingu veiksmu, kai TI metodas rodo, jog šis kintamasis yra svarbesnis nei centralizuotas karšto vandens tiekimas. Atsižvelgiant į tai, kad duomenyse kai kurių kategorinių kintamųjų grupės yra pasiskirsčiusios skirtingai, o priklausomojo kintamojo reikšmės pasižymi stipria teigiama asimetrija, buvo pastebėta, kad taikant TI metodą galėjo kilti priklausomybės perversinimo rizika. Dėl šios priežasties galutinė požymių atranka buvo atlikta taikant saugesnio Moodo medianos testo rezultatus.

Remiantis Moodo medianos testo rezultatais, buvo nustatyta, kad mažiausiai informacijos apie bendrąsias namų ūkių išlaidas suteiktų kintamieji, kurių χ^2 statistikos reikšmė yra mažesnė už 40 (žr. 10 lent. ir 7 pav.), t. y., „centralizuotas šalto vandens tiekimas“, „regionas“, „gimimo šalis“, „tėvo gimimo šalis“, „šiuo metu mokosi“, „motinos gimimo šalis“, „pagrindinės pilietybės šalis“, „gyvenamoji vieta“, „centralizuotas karšto vandens tiekimas“ ir „būsto tipas“. Verta paminėti, kad nors ir tarp kai kurių grupių pagal išlaidų medianas buvo nustatyti statistiškai reikšmingi skirtumai, tačiau pačios kategorijos tyrimo kontekste yra nepakankamai informatyvūs požymiai. Dėl šios priežasties išvardyti kintamieji, kuriant modelius, turėtų būti vertinami kaip pertekliniai ir šalinami iš tolimesnės analizės.

Analogiška tvarka buvo nustatyti ir **kitoms vartojimo išlaidų kategorijoms** (maisto ir nealkoholinių gėrimų, informacijos ir ryšių, draudimo ir finansinių paslaugų) svarbūs požymiai, kurių analizė pateikta 31 priede. Remiantis Moodo medianos testo rezultatais, nustatyta, kad mažiausiai informacijos apie namų ūkių išlaidas, skirtas maistui ir nealkoholiniams gėrimams, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 80 (žr. 31 priedą) t. y., „tėvo gimimo šalis“, „motinos gimimo šalis“, „gimimo šalis“, „regionas“, „šiuo metu mokosi“, „pagrindinės pilietybės šalis“, „centralizuotas šalto vandens tiekimas“, „centralizuotas karšto vandens tiekimas“, „būsto tipas“, „gyvenamoji vieta“, „būsto valdos statusas“. Tuo tarpu mažiausiai informacijos apie namų ūkių išlaidas, skirtas informacijai ir ryšiams, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 70 (žr. 31 priedą) t. y., „regionas“, „pagrindinės pilietybės šalis“, „centralizuotas karšto vandens tiekimas“, „motinos gimimo šalis“, „tėvo gimimo šalis“, „gimimo šalis“, „būsto tipas“, „centralizuotas šalto vandens tiekimas“, „gyvenamoji vieta“, „šiuo metu mokosi“, „būsto tipas“, „pagrindinis būsto šildymo būdas“. Kai mažiausiai informacijos apie namų ūkių išlaidas, skirtas draudimui ir finansinėms paslaugoms, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 9 (žr. 31 priedą) t. y., „pagrindinės pilietybės šalis“, „šiuo metu mokosi“, „regionas“, „motinos gimimo šalis“, „centralizuotas šalto vandens tiekimas“, „gimimo šalis“ ir „tėvo gimimo šalis“. Išvardyti kintamieji, kuriant išlaidų pokyčius vertinančius modelius, turėtų būti laikomi kaip pertekliniai ir šalinami iš tolimesnės analizės.

3.2. Vartojimo išlaidų priklausomybės nuo mikrolygmens veiksmų vertinimas

Vartojimo išlaidų ir jas sudarančių pasirinktų struktūrinių dalių priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimas buvo atliktas taikant geriausią, remiantis modelių kokybės vertinimo metrikomis (žr. 2.5 skyrelį ir 7 priedą), iš trijų regresijos modelių (žr. 2.4 skyrelį ir 6 priedą): klasikinės daugialypės tiesinės regresijos, atsitiktinių miškų regresijos ir ekstremalaus gradientinio stiprinimo algoritmo, pritaikyto regresijos uždavinių sprendimui.

Prieš modelių kūrimą buvo atliktas duomenų paruošimas: šalinimui išskirti mažai informatyvūs požymiai (žr. 3.1 skyrelį), priklausomojo kintamojo duomenims taikytos vinnorizacijos ribos – 1 %, 2 %, 3 % ir 5 % – pasirinktos remiantis teorija (žr. 2.3 skyrelį), tačiau atsižvelgus į tai, jog duomenyse dažnai pasitaikė artimos nuliui arba nulinės išlaidų sumos, o išskirtinai didelės išlaidos buvo reti atvejai, priklausomojo kintamojo duomenų imtis buvo apribota tik ties viršutine riba. Priklausomojo kintamojo reikšmėms taip pat buvo taikoma logaritminė transformacija su poslinkiu, kai taikant retransformaciją su korekcijos koeficientu buvo siekiama grąžinti prognozes į pradinę eurų skalę (žr. 2.3 skyrelį). Duomenims taikyta logaritminė transformacija, priartino jų pasiskirstymą prie normaliojo skirstinio varpo formos (žr. 3 priedą). Remiantis atliktais daugkartiniais porinių palyginimų rezultatais, mažai stebinių turinčios grupės, tokios kaip „kita gyvenamoji patalpa“, turinti 3 namų ūkius, ir „nėra stacionaraus šilumos šaltinio“, turinti 16 namų ūkių (žr. 29 priedą), pagal medianų reikšmes buvo pripažintos statistiškai reikšmingai nesiskiriančios nuo kitų to paties kategorinio kintamojo grupių. Mažas grupės reprezentuojančių stebinių skaičius ir jų panašumas į kitas to paties kategorinio kintamojo grupes, gali veikti modelių parametrų stabilumą, todėl šios grupės modeliavimo metu buvo nenaudojamos. Po pašalinimo liko 4315 stebiniai.

Kategoriniai kintamieji buvo konvertuoti į fiktyviuosius kintamuosius, o visa imtis padalinta į mokymui ($n = 3452$) ir testavimui ($n = 863$) skirtas dalis. Mašininio mokymosi modelių apmokymo metu hiperparametrų parinkimui buvo naudota 5 dalių kryžminė validacija, vykdyta tik treniravimo imtyje, kai testavimo duomenų rinkinys buvo naudotas tik galutiniam modelių įvertinimui.

Bajeso optimizacijos algoritmas (žr. 2.4 skyrelį) buvo taikytas atsitiktinių miškų ir ekstremalaus gradientinio stiprinimo modeliams, naudojant mokslinėje literatūroje pasiūlytus hiperparametrų paieškos intervalus (žr. 2.4.1 ir 2.4.2 skyrelius), kurie bandymų metu buvo koreguojami priklausomai nuo to, ar paieškos laukas buvo pakankamas geriausiam hiperparametrų deriniui. Tiek atsitiktinių miškų, tiek gradientinio stiprinimo modelių atveju buvo išbandyti keli Bajeso optimizavimo algoritmo metu kiekvieno hiperparametrų rinkinio vertinimui naudojami būdai, t. y., MAE, R^2 , MSE, RMSE, MedAE. Be to, modelius apmokant su ekstremalaus gradientinio stiprinimo algoritmu, taip pat buvo išbandyti skirtingi nuostolio funkcijos (angl. *objective*) variantai: kvadratinės paklaidos funkcija (angl. *squarederror*) ir absoliutinės paklaidos funkcija (angl. *absoluteerror*). Pastaroji funkcija taikoma, kai duomenyse yra išskirčių.

Bendrų vartojimo išlaidų priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimas. Anksčiau buvo nustatyta (žr. 4 priedą), kad maksimali namų ūkių bendro vartojimo išlaidų suma tarp Lietuvos gyventojų siekė 7 tūkst. Eur. per mėnesį. Taikant skirtingas viršutines vinnorizacijos metodo ribas, buvo apribotos ir skirtingos vartojimo išlaidų sumos: taikant 1 % ribą, maksimali išlaidų suma buvo apribota iki 2970 Eur, 2 % – iki 2505 Eur, 3 % – iki 2230 Eur, 4 % – iki 2093 Eur, ir 5 % – iki 1945 Eur. Bendrų vartojimo išlaidų priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimui atlikti buvo palyginti trys sukurti modeliai (žr. 12, 13 ir 14 lent.). Kiekvieno metodo atveju

buvo pateiktas pradinis modelio variantas, t. y., su neapdorotu priklausomu kintamuoju, neišmestais nereikšmingais kintamaisiais (žr. 0 skyrelį), tai leido įvertinti pradinę modelio kokybę ir ją vėliau palyginti su optimizuotu modeliu. Geriausias konkretaus metodo modelio variantas buvo tas, kuris po optimizavimo (išskirčių tvarkymo, logaritdavimo, nereikšmingų požymių pašalinimo ir hiperparametrų derinimo) pasiekė aukščiausias kokybės metrikų reikšmes tarp visų tuo metodu bandytų variantų (žr. 32 priedą).

12 lentelė. Daugialypės tiesinės regresijos modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos

	Pradinis modelio variantas	Galutinis modelio variantas
Nereikšmingi požymiai išmesti	Ne	Taip
Taikyta vinzorizacija/proc.	Ne	Taip/5%
Logarimtuotos išlaidos	Ne	Taip
F-statistikos p-reikšmė	0,00	0,00
Pašalinti požymiai, kurių VIF > 10	Ne	Taip
R²_{adj}	0,47	0,46
MAE, Eur.	316,79	328,47
RMSE, Eur.	482,33	523,68
R²	0,35	0,34
MedAE, Eur.	211,15	215,28
Jarque-Bera testo p-reikšmė	0,001	0,02
Vaito testo LM statistikos p-reikšmė	0,000	0,08
Multikolinearumo egzistavimas	13 požymių virš VIF=10	Nėra

Matyti, kad abu modeliai yra statistiškai reikšmingi (F-statistikos p-reikšmės mažesnės už 5 % reikšmingumo lygmenį), tačiau jų kokybė skiriasi. Nors pagal bendrąsias modelio kokybės metrikas (MAE, RMSE, MedAE) galutinis modelis rodo didesnes klaidas, tačiau jo kūrimo metu buvo pašalinta multikolinearumo problema, t. y., išmesti 13 kintamųjų, kurių VIF > 10 (žr. 6 priedą). Galimai svarbių, bet tarpusavyje koreliuotų kintamųjų, pašalinimas lėmė tai, kad modelio paaiškinamoji galia naudojant apmokymui skirtus duomenis šiek tiek sumažėjo ($R^2_{adj} = 0,46 < R^2_{adj} = 0,47$), kaip ir vertinimo su testavimo duomenimis atveju ($R^2 = 0,35 > R^2 = 0,34$). Pabrėžtina, kad galutinio modelio paklaidos tenkino dispersijų homoskedastiškumo prielaidą, nes Vaito testo p-reikšmė buvo didesnė nei reikšmingumo lygmuo ($0,08 > 0,05$). Verta paminėti, kad priklausomojo kintamojo logaritminė transformacija priartino duomenų pasiskirstymą prie normaliojo atvejo, tai atspindi paklaidų normalumo prielaidos tikrinimo metu, kai *Jarque-Bera* testo p-reikšmė padidėjo iki 0,02 (pradiname modelyje buvo 0,001), o tai rodo švelnesnį prielaidos pažeidimą. Dėl šių priežasčių daugialypės tiesinės regresijos galutinio modelio varianto rezultatus galima laikyti geresniais ir patikimesniais.

Visi atlikti bandymai, taikant mašininio mokymosi algoritmus, pateikti 33 priede. Atliekant bandymus buvo pastebėta, kad didinant vinzorizacijos viršutinę ribą, kartu gerėjo ir modelio kokybės metrikos (MAE, RMSE), tačiau MedAE reikšmės nemažėjo. Dėl šios priežasties buvo pasirinkta kompromisinė 3 % viršutinė riba, po kurios pritaikymo didžiausia analizuojama išlaidų suma siekė 2 224 Eur. Vertinat atsitiktinių miškų modelius (žr. 13 lent.) matyti, kad pradinis modelio variantas pagal kokybės metrikas buvo prastesnis už daugialypės tiesinės regresijos modelį. Atlikus duomenų paruošimą ir hiperparametrų optimizavimą, kurio metu buvo minimizuojama vidutinė absoliuti paklaida, atsitiktinių miškų modelio rezultatai pagerėjo, t. y., modelio aiškinamoji galia (R^2) išaugo apie 8 proc. punktus, o klaidų metrikų (MAE, RMSE, MedAE) reikšmės buvo artimesnės nuliui.

13 lentelė. Atsitiktinių miškų regresijos modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos

	Pradinis modelio variantas	Galutinis modelio variantas
Nereikšmingi požymiai išmesti	Ne	Taip
Taikyta vinzorizacija/proc.	Ne	Taip/3%
Logarimtuotos išlaidos	Ne	Taip
Hiperparametrai Bajeso optimizavimui	Netaikoma	Medžių skaičius – int* [10 ; 500]
		Minimalus stebinių skaičius lapo mazge – int [2 ; 30]
		Didžiausias medžio gylis – int [2 ; 30]
		Minimalus stebinių skaičius mazgo padalinimui – int [2 ; 20]
		Didžiausias skaičius požymių, svarstomų kiekviename mazge padalijimui – real** [0,1 ; 0,999]
Paieškos metu optimizuojama metrika	Numatytoji (R^2)	MAE
Iteracijų skaičius	Nenurodytas	150
Geriausias hiperparametrų derinys	Numatytieji, t. y., medžių skaičius: 100, naudojami požymiai: 1,0; . stebinių skaičius mazge: 1; stebiniai dalijimui: 2, kt. – nenurodyti.	Medžių skaičius: 188; min. stebinių skaičius mazge: 6; medžio gylis: 27; min. stebiniai dalijimui: 16; naudojami požymiai: 0,42.
MAE, Eur.	328,73	285,30
RMSE, Eur.	498,52	395,33
R^2	0,31	0,39
MedAE, Eur.	215,96	202,75

* sveikųjų skaičių intervalas

** tolydžių skaičių intervalas

Tuo tarpu ekstremalaus gradientinio stiprinimo algoritmo metodu gautas pradinis modelio variantas (žr. 14 lent.) buvo blogiausiai duomenis aprašantis modelis iš visų trijų metodų. Atlikus papildomą duomenų paruošimą ir modelių optimizavimą, buvo sukurtas geriausias regresijos modelis bendrųjų vartojimo išlaidų poveikio nuo mikrolygmens veiksnių vertinimui. Ekstremalaus gradientinio stiprinimo algoritmo galutinis regresijos modelio variantas pasižymėjo aukščiausiu determinacijos koeficientu tarp visų taikytų metodų, t. y., $R^2 = 0,40$, kai atsitiktinių miškų buvo 0,39, o daugialypės tiesinės regresijos – 0,34, tai reiškia, kad modelis gali paaiškinti iki 40 % Lietuvos namų ūkių bendrojo vartojimo išlaidų variacijos. Prognozavimo klaidos taip pat mažesnės, pavyzdžiui, vidutinė absoliuti paklaida ekstremalaus gradientinio stiprinimo algoritmo atveju siekė 284,39 Eur, kai atsitiktinių miškų – 285,30 Eur ir 328,30 Eur daugialypės tiesinės regresijos atveju. Galima teigti, kad išlaidų prognozės nuo tikrųjų išlaidų sumų vidutiniškai skiriasi per 284,39 eurus. Remiantis MedAE metrika galima teigti, kad pusės visų stebėtų namų ūkių prognozių paklaidos neviršijo 201,77 Eur.

Nors modelių rezultatai po atliktų papildomų veiksmų ir pagerėjo, tačiau skirtumai tarp taikytų metodų, ypač mašininio mokymosi metodų, buvo nedideli. Pabrėžtina, kad net ir pažangūs metodai, nereikalaujantys prielaidų laikymosi, pagal turimus požymius nesugebėjo paaiškinti daugiau nei 40 % bendrų namų ūkių vartojimo išlaidų variacijos Lietuvoje. Verta prisiminti, kad silpnas ryšys tarp aiškinamųjų kintamųjų ir priklausomojo kintamojo buvo pastebėtas jau ankstesnėje analizėje, t. y., vertinant ϕ_K koreliacijos koeficientų matricą (žr. 5 pav.).

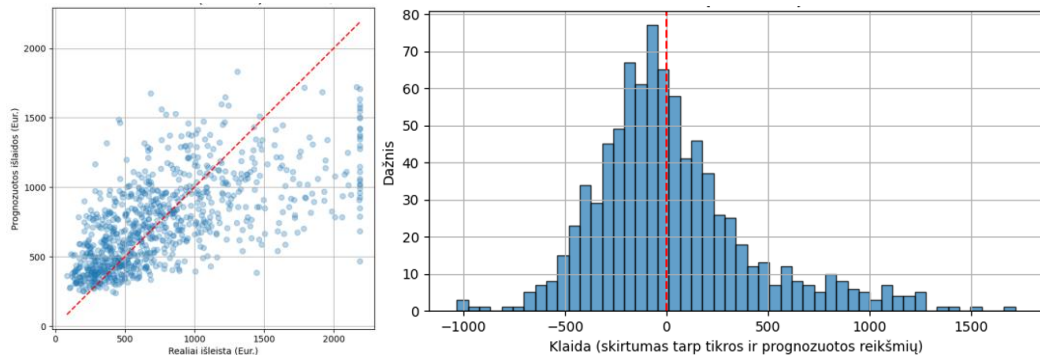
14 lentelė. Ekstremalaus gradientinio stiprinimo algoritmo metodu sukurtų regresinių modelių palyginimas, kai priklausomas kintamasis yra bendros vartojimo išlaidos

	Pradinis modelio variantas	Galutinis modelio variantas
Nereikšmingi požymiai išmesti	Ne	Taip
Taikyta vinzorizacija/proc.	Ne	Taip/3%
Logarimtuotos išlaidos	Ne	Taip
Hiperparametrai Bajeso optimizavimui	Netaikoma	Medžių skaičius – int [10 ; 1000]
		Minimali instancijų svorių suma, reikalinga mazge – real [1;8]
		Didžiausias medžio gylis – int [2 ; 20]
		Minimalus nuostolio funkcijos sumažėjimas, reikalingas, kad įvyktų mazgo padalijimas – real [1e-6 ; 2,0]
		Dalinė duomenų imtis – real [0,3 ; 1,0]
		Mokymosi greitis – int [0,0001 ; 0,1]
		Kintamųjų dalis, naudojama kiekvieno medžio kūrimui – real [0,3 ; 1,0]
		L1 reguliavimo koeficientas svoriams – real [1e-6 ; 1,0] L2 reguliavimo koeficientas svoriams – real [1e-6 ; 1,0]
Paieškos metu optimizuojama metrika	Numatytoji (R^2)	MAE
Nuostolio funkcija	Numatytoji (kvadratinė paklaida)	Absoliutinė paklaida
Iteracijų skaičius	Nenurodytas	150
Geriausias hiperparametrų derinys	Numatytieji, t. y. dauguma hiperparametrų nebuvo nurodyti.	Medžių skaičius: 811, svorių suma: 8, medžio gylis: 13; min. nuostolio funkcijos sumažėjimas: 2,0; duomenų imtis: 0,3; mokymosi greitis: 0,01, kintamųjų dalis: 1,0; L1: 0,01, L2: 0,99.
MAE, Eur.	334,01	284,39
RMSE, Eur.	508,63	391,91
R^2	0,28	0,40
MedAE, Eur.	216,58	201,77

Kadangi regresijos modelis, sukurtas naudojant ekstremalaus gradientinio stiprinimo algoritmą, pasižymėjo aukščiausia prognozavimo kokybe, t. y., aukščiausiu R^2 ir žemiausiomis MAE, RMSE bei MedAE reikšmėmis, jis buvo pasirinktas kaip pagrindinis tyrimo įrankis vertinant mikrolygmens veiksnių poveikį gyventojų bendrosioms vartojimo išlaidoms. Vertinant modelio kokybę pagal metrikas, apskaičiuotas naudojant mokymo imties duomenis, buvo gauta, kad MAE = 245,26 Eur, RMSE = 348,62 Eur, MedAE = 161,94 Eur, R^2 = 0,54, kai naudojant testavimo imtį – MAE = 284,39 Eur, RMSE = 391,91 Eur, MedAE = 201,77 Eur, R^2 = 0,40. Pakankamai nedideli kokybės metrikų skirtumai tarp mokymo ir testavimo duomenų rodo, jog modelis nebuvo pernelyg pritaikytas prie mokymo duomenyse esančio triukšmo ir išlaikė gebėjimą apibendrinti, t. y., sugebėjo pateikti panašius rezultatus naujiems duomenims.

Modelio kokybei įvertinti taip pat buvo pateikti (žr. 8 pav.) diagnostikos grafikai, kuriuose raudona punktyrinė linija žymi idealų atvejį, t. y., jei modelis prognozuotų be klaidų ir 100 % tikslumu, tai visi taškai gulėtų ant tos linijos. Sklaidos diagrama rodo, kad stebiniai nutolsta nuo šios linijos, o tai rodo didelę paklaidų dispersiją, turinčią tendenciją augti priklausomai nuo aukštesnių išlaidų reikšmių. Mažų išlaidų (iki 500 Eur) atveju taškai dažniausiai būna aukščiau linijos ir rodo, kad modeliu yra pervertinamos mažos išlaidų sumos. Tuo tarpu, ypač didelės vartojimo sumos (didesnės nei 1 500 Eur), yra dažniausiai nuvertinamos, nes dauguma taškų yra susitelkę žemiau idealios linijos. Dažniausiai skirtumai tarp tikrųjų ir prognozuotų reikšmių atsiranda, kai modeliu yra

prognozuojamos mažos išlaidų sumos (iki 500 Eur), nes paklaidų histograma (žr. 8 pav., dešinėje) rodo, jog didžioji dalis paklaidų yra susitelkusi intervale nuo – 300 Eur, iki + 300 Eur, tai reiškia, kad modelis yra linkęs pervertinti arba nuvertinti tikrąsias išlaidas per 300 Eur. Nors didelės paklaidos pasitaiko rečiau, tačiau jų dydis gali siekti ir – 1 000 Eur, ar + 1 500 Eur, tai rodo, kad modeliu prognozuotos išlaidos nuo tikrųjų gali skirtis ir per daugiau nei 1 000 Eur.



8 pav. Ekstremalaus gradientinio stiprinimo regresijos modelio diagnostika prognozuojant bendras išlaidas

Siekiant įvertinti mikrolygmens veiksnių bei jų sąveikų poveikį modelio prognozuojamoms išlaidoms, buvo naudotas SHAP metodas, paaiškinantis ekstremalaus gradientinio stiprinimo algoritmu (geriausiu modelio variantu) gautas prognozes. SHAP analizės suminė diagrama (žr. 9 pav.) padėjo identifikuoti svarbiausius veiksnius, remiantis jų vidutiniu indėliu į modelio prognozę, jį lyginant su vidutine prognoze, ir leido nustatyti kiekvieno veiksnio poveikio kryptį.



9 pav. Požymių poveikis bendroms vartojimo išlaidoms, nustatytas naudojant SHAP metodą, pritaikytą pagal ekstremalaus gradientinio stiprinimo algoritmu sukurtą regresijos modelį

SHAP reikšmė rodo vieneto poveikį modelio prognozei, kadangi analizuojamos išlaidos buvo išreikštos eurai, kiekvieno kintamojo indėlis taip pat buvo matuojamas eurai. Grafiko (žr. 9 pav.) Y ašyje nuo svarbiausios iki mažiausiai aktualios buvo išrikiuotos modelyje naudotos namų ūkių

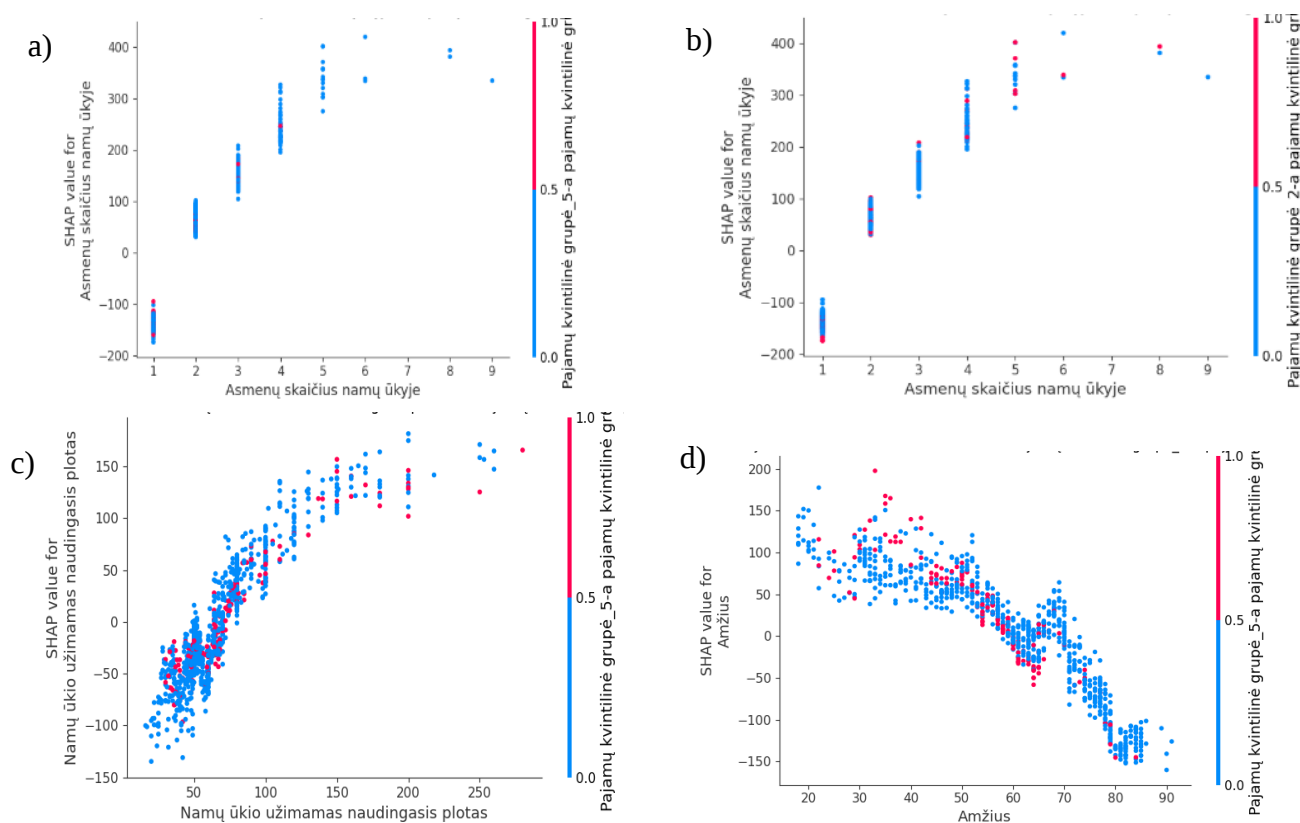
charakteristikos. *X* ašyje yra matuojamas kiekvieno požymio poveikis išlaidų prognozavimui, lyginant su vidutine modelio prognoze t. y., teigiama SHAP reikšmė rodo, kad požymis padidina prognozuojamas išlaidas virš bazinės vertės, o neigiama – kad mažina. Taškai yra stebiniai, t. y., konkretus apklausoje dalyvavęs Lietuvos namų ūkis. Grafiko dešinėje esanti legenda rodo konkretaus požymio reikšmės dydį pagal spalvą. Raudona spalva atitinka analizuojamo aiškinamojo kintamojo didelę (angl. *high*) reikšmę, pavyzdžiui, daug asmenų namų ūkyje, kai mėlyna – mažą (angl. *low*) reikšmę, pavyzdžiui, vienas asmuo visame namų ūkyje. Fiktyviųjų kintamųjų atveju raudona spalva žymi požymio buvimą (reikšmę 1), o mėlyna – jo nebuvimą (reikšmę 0).

Vidutinė prognozuota mėnesinė vartojimo išlaidų suma siekė 691,15 Eur. Suminė diagrama (žr. 9 pav.) rodo, kad svarbiausias vartojimo išlaidas veikiantis požymis yra asmenų skaičius namų ūkyje, kai didesnis narių skaičius (raudoni taškai) mėnesines vartojimo išlaidas gali padidinti daugiau nei per 250 Eur, o mažesnis skaičius (mėlyni taškai) – sumažinti apie 50 Eur. Pajamų kvintilinės grupės (ypač 5-a ir 4-a) taip pat stipriai veikia išlaidų lygį, modelio prognozės rodo tai, kad 5-ai pajamų kvintilinei grupei priklausantys namų ūkiai gali turėti daugiau nei 250 eurų didesnes prognozuojamas bendrojo vartojimo išlaidas, palyginti su 1-ai pajamų kvintilinei grupei priklausančiais namų ūkiais. Namų ūkio galvos amžius taip pat yra svarbus veiksnys, tačiau šiuo atveju yra matyti atvirkštinė tendencija, nes vyresnis amžius siejasi su mažesnėmis prognozuojamomis vartojimo išlaidomis (apie 50 Eur. mažesnėmis nei vidutinė prognozuota išlaidų suma). Tai gali atspindėti mažesnes vyresnio amžiaus asmenų vartojimo galimybes, ką patvirtina ir aštunta pagal svarbumą namų ūkio charakteristika – pajamų šaltinis iš pensijos. Matyti, kad namų ūkiai, kurių pagrindinis pajamų šaltinis yra pensija (raudona spalva), dažniausiai turi neigiamas vartojimo išlaidų SHAP reikšmes, t. y., pensija, kaip pagrindinis pajamų šaltinis, mažina prognozuojamas vartojimo išlaidas apie 20 Eur, lyginant su gyventojais, kurių pagrindinis pajamų šaltinis yra darbo užmokestis. Verta pažymėti, kad senatvės pensininkai buvo antra pagal dydį grupė tarp visų apklausos dalyvių (žr. 29 priedą). Namų ūkių užimamas naudingasis plotas yra ketvirtas pagal svarbumą požymis, rodantis, kad namų ūkiai, kurių būstai užima didesnę naudingąją plotą, prie prognozuojamų vartojimo išlaidų prisideda teigiamai (raudoni taškai dešinėje pusėje), tai gali būti susiję su didesnėmis būsto išlaikymo sąnaudomis ar pasirinktu gyvenimo būdu.

Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas (žr. 10 pav.). Šios diagramos padeda suprasti veiksmų įtaką modeliui atliktoms išlaidų prognozėms, kartu įvertinant ir sąveiką su kitu veiksmu. *X* ašyje yra pasirinkto požymio reikšmės, *Y* ašyje – SHAP reikšmės (Eur), parodančios kiek pasirinktas požymis prisideda prie bendrų vartojimo išlaidų prognozavimo. Spalvų legenda šiuo atveju parodo kito pasirinkto požymio reikšmes, t. y., raudona – jei kito kintamojo reikšmės yra didelės (arba egzistuoja), ir mėlyna – jei mažos (arba neegzistuoja).

Nors abiejuose grafikuose (žr. 10 pav. a ir b) galima matyti aiškią teigiamą priklausomybę, t. y., kuo didesnis namų ūkis, tuo didesnės ir prognozuojamos vartojimo išlaidos, nepriklausomai nuo pajamų lygio, tačiau šis ryšys nėra visiškai tiesinis. Matyti, kad nuo 1 iki 5 namų ūkyje esančių asmenų skaičiaus, SHAP reikšmės (išlaidos) auga staigiai, tačiau kai asmenų skaičius namų ūkyje pasiekia 6, tada vartojimo išlaidų suma nustoja didėti, tai rodo, kad ypač dideli namų ūkiai nebūtinai vartojimui skirs didesnes išlaidas. Remiantis grafikais (žr. 10 pav. a ir b) galima sakyti, jog mažesni namų ūkiai (iki 3 asmenų) dažniau priklauso 5-ai pajamų kvintilinei grupei, kai didesni namų ūkiai (iki 8 asmenų) dažniau yra priskiriami 2-ai pajamų kvintilinei grupei. Be to, pagal prognozuotas SHAP reikšmes išlaidoms (žr. 10 pav. b) matyti, kad 5 asmenų namų ūkis, priklausantis 2-ai pajamų kvintilinei grupei,

per mėnesį vartojimo išlaidoms per mėnesį vidutiniškai skirtų apie 400 eurų daugiau nei namų ūkis, priklausantis 1-ai kvintilinei grupei.



10 pav. SHAP priklausomybės diagramos, vertinančios kintamųjų tarpusavio sąveikas prognozuojamų bendrų vartojimo išlaidų atžvilgiu, kai vienas iš veiksnių yra 5-a pajamų kvintilinė grupė

SHAP priklausomybės diagrama (žr. 10 pav., c), vizualizuojanti, kaip namų ūkio užimamas naudingasis būsto plotas veikia prognozuojamas bendrojo vartojimo išlaidas, atsižvelgiant į tai, ar namų ūkis priklauso 5-ai pajamų kvintilinei grupei (raudona spalva) ar nepriklauso (mėlyna) rodo, kad didesnis namų ūkių naudingasis plotas yra susijęs su bendrojo vartojimo išlaidų augimu. Verta paminėti, kad ypač didelio naudingojo ploto būstai (daugiau nei 200 m²) vartojimo išlaidų nedidina taip sparčiai kaip būstai, užimantys nuo 0 m² iki 150 m² naudingojo ploto, tai gali rodyti, kad Lietuvos namų ūkių vartojimo augimas yra ribinis ir, pasiekus tam tikrą užimamo naudingojo ploto ribą, šio požymio poveikis bendroms išlaidoms pradeda mažėti. Pastebėtina, kad namų ūkiai, priklausantys 5-ai kvintilinei grupei (raudoni taškai) dažniau pasitaiko, kai naudingasis užimamas plotas yra didesnis nei 150 m², tai rodo, kad turtingesni namų ūkiai ne tik yra labiau linkę gyventi didesnio naudingojo ploto būstuose, bet ir pasižymi aukštesniu vartojimo lygiu.

SHAP priklausomybės diagrama (žr. 10 pav., d), vaizduojanti, kaip namų ūkio galvos amžius veikia bendro vartojimo išlaidų prognozę, atsižvelgiant į tai, ar namų ūkis priklauso 5-ai pajamų kvintilinei grupei (raudona spalva) ar nepriklauso (mėlyna) rodo, jog ryšys tarp bendrų vartojimo išlaidų ir amžiaus yra netiesinis. Nors bendrai buvo teigiama, kad vyresnis amžius siejasi su mažesnėmis prognozuojamomis vartojimo išlaidomis, tačiau žiūrint detaliau matyti, kad jos yra teigiamos, kai namų ūkio galvos amžius nesiekia 50 metų. Tai reiškia, kad jaunesni asmenys linkę išleisti daugiau nei prognozuotas vidurkis. Taip pat pastebimas ir padidėjęs vartojimo išlaidų lygis, kai namų ūkio galvos amžius patenka į 60 – 68 metų intervalą, tik po kurio seka staigus prognozuotų išlaidų sumažėjimas (neigiamos SHAP reikšmės), o tai ir parodo, kad ryšys nėra visiškai tiesinis. Pabrėžtina,

kad asmenys, priklausantys 5-ai kvintilinei grupei (raudoni taškai), dažniau pasitaiko iki 55 metų, o tai rodo, jog aukštesnes pajamas uždirbantys gyventojai yra jaunesnio arba vidutinio amžiaus. Galima teigti, kad asmenų amžius bendrąsias vartojimo išlaidas veikia ne tik tiesiogiai, bet ir netiesiogiai per priklausymą konkrečiai pajamų kvintilinei grupei.

Vadovaujantis analogišku principu, t. y., taikant daugialypės tiesinės regresijos, atsitiktinių miškų ir gradientinio stiprinimo algoritmo metodus, buvo kuriami regresijos modeliai ir kitų vartojimo išlaidų kategorijų – **maisto ir nealkoholinių gėrimų, informacijos ir ryšių, draudimo bei finansavimo paslaugų** – pokyčiams aprašyti. Sukurtų modelių tarpusavio palyginimas bei interpretavimas taikant SHAP metodą yra pateikti 37 priede.

Nustatyta, kad regresijos modelis, paremtas ekstremalaus gradientinio stiprinimo algoritmu ir skirtas maisto ir nealkoholinių gėrimų vartojimo išlaidų prognozavimui, pasižymėjo aukščiausiu determinacijos koeficientu tarp visų taikytų metodų, t. y., $R^2 = 0,31$, kai atsitiktinių miškų atveju $R^2 = 0,30$, o tiesinės regresijos – $R^2 = 0,25$, tai reiškia, kad geriausias modelis galėjo paaiškinti iki 31 % Lietuvos namų ūkių maisto ir nealkoholinių gėrimų vartojimo išlaidų variacijos. RMSE, MAE bei MedAE paklaidos taip pat buvo mažesnės, pavyzdžiui, vidutinė absoliuti paklaida ekstremalaus gradientinio stiprinimo algoritmo atveju siekė 97,78 Eur, kai atsitiktinių miškų – 98,1 Eur, ir 107,14 Eur – daugialypės tiesinės regresijos atveju. Verta papildomai paminėti, kad sukurto daugialypės tiesinės regresijos modelio liekanos taip pat neatitiko esminių klasikinei regresijai taikomų prielaidų (Vaito testo p-reikšmė = 0,00; *Jarque-Bera* testo p-reikšmė = 0,00). Remiantis SHAP analizės sumine diagrama, nustatyta, kad vidutinė prognozuota mėnesinė maisto ir nealkoholinių gėrimų vartojimo išlaidų suma siekė 233,58 Eur, o prognozuojant šią išlaidų kategoriją svarbiausi veiksniai buvo asmenų skaičius namų ūkyje, užimamas naudingasis plotas, priklausymas konkrečiai pajamų kvintilinei grupei bei namų ūkio galvos amžius. Visų išvardintų veiksnių poveikio kryptis yra tokia pati, kaip ir bendrųjų vartojimo išlaidų atveju. Vienas iš svarbesnių, ir nuo bendrųjų vartojimo išlaidų prognozės modelio besiskiriančių veiksnių, yra namų ūkio teisinė santuokinė padėtis, kuri rodo, jog santuokoje gyvenantys namų ūkiai teigiamai veikia maisto ir gėrimų vartojimo išlaidas (teigiamos SHAP reikšmės). Vedę (ištekėjusios) gyventojai prie tiriamos išlaidų kategorijos prognozuojamos sumos vidutiniškai prisideda apie 40 Eur, lyginant juos su išsituokusiais namų ūkiais, kurie buvo referencinė grupė. Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas, kurios parodė nevisiškai tiesinį ryšį tarp asmenų skaičiaus namų ūkyje ir vartojimo išlaidų. Buvo matyti, kad SHAP reikšmės viršijo 200 Eur nuo vidurkio ribą, kai namų ūkį sudarė 6 asmenys, tačiau didesniuose nei 6 asmenų namų ūkiuose didesnis išlaidų kiekis nebuvo pastebėtas. Taip pat pasitvirtino ir anksčiau nustatytas netiesinis ryšys tarp vartojimo išlaidų ir namų ūkio galvos amžiaus.

Vertinant informacijos ir ryšių paslaugų vartojimo išlaidas, geriausiu modeliu buvo pripažintas ekstremalaus gradientinio stiprinimo galutinis variantas (be išmestų nesvarbių kintamųjų, bet taikyta vinzorizacija ir logaritmavimu), kurio $R^2 = 0,38$ ir kuris aplenkė tiek atsitiktinių miškų ($R^2 = 0,36$), tiek daugialypės tiesinės regresijos ($R^2 = 0,25$) modelius. Tai reiškia, kad geriausias modelis galėjo paaiškinti iki 38 % informacijos ir ryšių paslaugų vartojimo išlaidų dispersijos. Vertinant modelio kokybę pagal klaidų metrikas, ekstremalaus gradientinio stiprinimo modelis taip pat buvo panašesnis – jo MAE, RMSE ir MedAE reikšmės, lyginant su kitais sukurtais modeliais, buvo arčiausiai nulio, kai vidutinė absoliuti paklaida siekė 17,69 Eur, o atsitiktinių miškų modelio atveju – 17,93 Eur ir 23,04 Eur – daugialypės tiesinės regresijos atveju. Vidutinė prognozuota mėnesinė informacijos ir ryšių paslaugų vartojimo išlaidų suma siekė 34,74 Eur. Remiantis SHAP suminės diagramos informacija, teigiama, kad informacijos ir ryšių vartojimo išlaidas stipriausiai veikia namų ūkio

galvos amžius, kurio poveikį galima vertinti kaip atvirkštinį. Trečias pagal svarbą kintamasis yra vieno gyvenančio asmens fiktyvusis požymis, kurio buvimas yra susijęs su mažesnėmis vartojimo išlaidomis, skirtomis informacijai ir ryšiams. Pensijos, kaip pagrindinio pajamų šaltinio, poveikis vartojimo išlaidoms yra panašus į poveikį, kuris buvo nustatytas vieno gyvenančio asmens atveju, tai gali būti paaiškinama tuo, kad pensinio amžiaus asmenys dažnai gyvena vieni. Aukštesnįjį ar aukštąjį išsilavinimą turinčių asmenų poveikis informacijai ir ryšiams skirtoms išlaidoms yra teigiamas, ir rodo, kad tokį išsilavinimo lygį turintys asmenys, prognozuojant išlaidas, nuo vidurkio prisidėtų per 5 Eur daugiau, lyginant su asmenimis, kurių išsilavinimas yra aukštesniojo ir aukštojo nesuteikiantis pavidurinis. Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas, kurios parodė, jog didžiausios teigiamos SHAP reikšmės, t. y., prognozuojamos informacijai ir ryšiams skirtos išlaidos, viršijančios vidurkį daugiau nei 10 Eur pasitaiko tarp jaunų (apie 20 – 40 metų) ir vienu gyvenančių asmenų. Verta pabrėžti, aukštesnio išsilavinimo teigiamas poveikis informacijos ir ryšių paslaugų vartojimui atsispindėjo net ir tarp vyresnio amžiaus gyventojų.

Vertinant draudimo ir finansinių paslaugų vartojimo išlaidas, geriausiu modeliu buvo pripažintas atsitiktinių miškų regresijos galutinis variantas (pašalinti nesvarbūs kintamieji, taikyta 3 % vinzorizacija, bet nenaudota logaritminė transformacija), kurio $R^2 = 0,31$ ir kuris aplenkė tiek ekstremalaus gradientinio stiprinimo ($R^2 = 0,30$), tiek daugialypės tiesinės regresijos ($R^2 = 0,26$) modelius. Tai reiškia, kad geriausias modelis galėjo paaiškinti iki 31 % draudimo ir finansinių paslaugų išlaidų dispersijos. Vertinant modelio kokybę pagal klaidų metrikas, atsitiktinių miškų modelis taip pat yra panašesnis – jo MAE, RMSE ir MedAE reikšmės tarp visų modelių buvo arčiausiai nulio, kai vidutinė absoliuti paklaida siekė 8,59 Eur, o ekstremalaus gradientinio algoritmo atveju – 8,88 Eur ir 10,13 Eur – daugialypės tiesinės regresijos atveju. Vidutinė prognozuota mėnesinė draudimo ir finansinių paslaugų vartojimo išlaidų suma siekė 11,11 Eur. Remiantis SHAP suminės diagramos informacija, nustatyta, kad didžiausią poveikį prognozuojamoms draudimo ir finansinių paslaugų išlaidoms turi namų ūkio užimamas naudingasis plotas, pensija kaip pagrindinis pajamų šaltinis bei priklausymas 5-ai pajamų kvintilinei grupei, kai didesnis užimamas naudingasis plotas ir priklausymas 5-ai kvantilinei grupei turi teigiamą poveikį, nes gali padidinti išlaidas daugiau nei per 10 Eur nuo vidurkio, o gaunamos pajamos iš pensijos – atvirkštinį. Svarbų poveikį finansinių paslaugų vartojimui taip pat turi ir didesnis asmenų skaičius namų ūkyje bei aukštesnis išsilavinimas, pagal kuriuos prognozuojant vidutinė išlaidų suma didėtų apie 5 eurus. Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas, kurios atskleidė tarp draudimo ir finansinių paslaugų vartojimo išlaidų bei namų ūkių užimamo naudingojo ploto egzistuojantį netiesinį ryšį.

3.3. Mikrolygmens veiksmų poveikio vartojimo išlaidoms tyrimo apibendrinimas

Remiantis ϕ_K koreliacijos koeficientais buvo nustatyta, kad Lietuvos namų ūkių kontekste tiriami nepriklausomi kintamieji yra silpnai arba vidutiniškai susiję su skirtingomis vartojimo išlaidų kategorijomis. Stipriausi ryšiai ($\phi_K > 0,3$) buvo fiksuoti tarp analizės veiksmų ir bendrų vartojimo išlaidų, maisto ir nealkoholinių gėrimų, informacijos ir ryšių paslaugų bei draudimo ir finansinių paslaugų išlaidų, todėl su jomis ir buvo sudaryti modeliai, vertinantys išlaidų priklausomybę nuo Lietuvos namų ūkių charakteristikų.

Remiantis atlikta požymių atranka taikant filtravimo metodus, t. y., ϕ_K koreliacijos koeficientą, Moodo medianos testą ir tarpusavio informacijos matą, ir atsižvelgus į turimų duomenų pasiskirstymo

pobūdį, pagrindiniu metodu buvo pasirinktas Moodo medianos testas, kuriuo buvo nustatyta, jog kintamieji, tokie kaip „regionas“, „gimimo šalis“, „tėvų gimimo šalis“, „šiuo metu mokosi“, „gyvenamoji vieta“, bei centralizuotas šalto ar karšto vandens tiekimas buvo pripažinti nepakankamai informatyviais požymiais norint vertinti vartojimo pasikeitimus tiek bendrose išlaidose, tiek atskirose maisto ir nealkoholinių gėrimų, informacijos ir ryšių, bei finansinių paslaugų kategorijose. Tuo tarpu informatyvios namų ūkių charakteristikos buvo namų ūkio tipas, pajamų kvintilinė grupė, pajamų šaltinis, socialinė-ekonominė grupė, amžius, asmenų skaičius namų ūkyje.

Kiekvienai iš pasirinktų vartojimo išlaidų kategorijų buvo sukurtas modelis, vertinantis jų priklausomybę nuo analizės požymių. Modelių kūrimui buvo taikyti trys regresinių uždavinių sprendimui naudojami metodai: klasikinė daugialypė tiesinė regresija bei du pažangūs mašininio mokymosi algoritmai – atsitiktinių miškų regresija ir ekstremalaus gradientinio stiprinimo algoritmas. Atsižvelgiant į išlaidų pasiskirstymo asimetriškumą, priklausomajam kintamajam buvo taikytos logaritminė transformacija su poslinkiu ir viršutinės ribos vinzorizacija. Modelių kokybė buvo vertinama naudojant metrikas: determinacijos koeficientą (R^2), vidutinę absoliučią paklaidą (MAE), šaknies vidutinę kvadratinę paklaidą (RMSE) ir medianinę absoliučią paklaidą (MedAE). Nustatyta, kad daugialypė tiesinė regresija visų tirtų išlaidų variantų atveju pasirodė prasčiausiai, ypač dėl jai taikomų prielaidų neatitikimo t. y., liekanų homoskedastiškumo ir normalumo. Poveikio vertinimas buvo atliekamas taikant mašininio mokymosi modelius, nes jie pasiekė aukščiausius tikslumo rezultatus (aukščiausia pasiekta R^2 reikšmė siekė 0,40). Mašininio mokymosi modelių hiperparametrų optimizavimui buvo taikytas Bajeso optimizavimo algoritmas. Vertinant bendras vartojimo išlaidas, taip pat išlaidas maistui ir nealkoholiniams gėrimams ir informacijos ir ryšių paslaugoms, geriausios modelio kokybės vertinimo metrikų reikšmės buvo pasiektos taikant ekstremalaus gradientinio stiprinimo algoritmą (atitinkamai $R^2 = 0,40$; $R^2 = 0,31$; $R^2 = 0,38$), o tuo tarpu draudimo ir finansinių paslaugų išlaidų atveju optimaliausiu metodu buvo pripažintas atsitiktinių miškų regresijos modelis ($R^2 = 0,30$).

Mašininio modelių rezultatų interpretavimas buvo vykdomas naudojant SHAP metodą. Vertinant bendrųjų vartojimo išlaidų priklausomybę nuo mikrolygmens veiksnių buvo atrasta, kad svarbiausias veiksnys yra asmenų skaičius namų ūkyje, ir kuo jis didesnis – tuo didesnės yra ir vartojimui skiriamos išlaidos (teigiamas poveikis), visgi detalesnė priklausomybių diagrama parodė, kad kai namų ūkyje yra nuo 1 iki 5 asmenų, išlaidos auga staigiai, tačiau kai asmenų skaičius namų ūkyje pasiekia 6, tada vartojimo išlaidų suma nustoja didėti, o tai rodo, kad ryšys nėra visiškai tiesinis. Gyventojai, priklausantys aukštesnei pajamų kvintilei (5-ai ar 4-ai) yra linkę išleisti apie 250 Eur daugiau nei bendras išlaidų vidurkis. Tuo tarpu namų ūkio galvos amžius turi nevienareikšmį poveikį: jaunesni asmenys (iki 55 metų) yra linkę išleisti daugiau, kai 60 – 70 metų asmenys yra linkę išleisti panašiai kaip 40 – 50 metų asmenys, nors vyresni kaip 70 metų ypač mažina vartojimui išleidžiamas sumas. Buvo pastebėta, kad namų ūkių užimamas didesnis naudingasis plotas yra susijęs su didesnėmis išlaidomis, tačiau teigiamas poveikis nusistovi, kai plotas tampa didesnis nei 200 m². Verta paminėti, kad teigiamas poveikis bendrųjų vartojimo išlaidų atveju buvo pastebėtas ir dėl tokių veiksnių kaip: būsto šildymas ne krosniniu būdu, jei pora turi išlaikomų vaikų, jei namų ūkio galva gyvena santuokoje, asmuo turi aukštąjį ar aukštesnįjį išsilavinimo lygį, bei dirba specialisto darbą.

Vertinant maisto ir nealkoholinių gėrimų vartojimo išlaidų priklausomybę nuo mikrolygmens veiksnių taikant suminę SHAP diagramą buvo nustatyta, kad svarbiausią įtaką prognozuojant išlaidas turėjo asmenų skaičius namų ūkyje, užimamas naudingasis būsto plotas, 5-a pajamų kvintilinė grupė, namų ūkio galvos amžius ir teisinė santuokinė padėtis. Kaip ir bendrųjų vartojimo išlaidų atveju,

didesnis namų ūkio narių skaičius, didesnis būsto plotas ir priklausymas aukštesnei pajamų kvintilinei grupei išlaidas didino, o vyresnis amžius buvo susijęs su mažesnėmis išlaidomis. Verta paminėti, kad priklausomybės diagramos papildomai parodė anksčiau paminėtus netiesinius ryšius. Tiriant maisto ir nealkoholinių gėrimų vartojimo išlaidas nustatyta, kad santuokoje gyvenantys asmenys išlaidoms gali skirti apie 40 Eur daugiau nei išsituokę asmenys, kurie buvo referencinė grupė. Tokie veiksniai kaip naudojimas centriniu šildymu iš šilumos tinklų, vidurinis išsilavinimo lygis, įrenginių ir mašinų operatoriaus profesija, taip pat yra susiję su mažesnėmis išlaidomis maisto ir nealkoholinių gėrimų vartojimui.

Vertinant informacijos ir ryšių paslaugų vartojimo išlaidų priklausomybę nuo mikrolygmens veiksnių, taikant suminę SHAP diagramą buvo nustatyta, kad svarbiausią įtaką prognozuojamoms išlaidoms turėjo namų ūkio galvos amžius, vieno gyvenančio asmens fiktyvusis požymis, pensija kaip pagrindinis pajamų šaltinis, kurie turėjo atvirkštinį poveikį išlaidoms, kai asmenų skaičius namų ūkyje, namų ūkių užimamas naudingasis plotas, išsilavinimo lygis bei priklausymas aukštesnei pajamų kvintilei – teigiamą. Verta paminėti, kad priklausomybės diagramos ir šiuo atveju rodė netiesinius ryšius tarp išlaidų ir kiekybinių aiškinamųjų kintamųjų. Papildoma įžvalga buvo, kad aukštesnio išsilavinimo teigiamas poveikis informacijos ir ryšių paslaugų vartojimui atsispindėjo net ir tarp vyresnio amžiaus gyventojų.

Vertinant draudimo ir finansinių paslaugų vartojimo išlaidų priklausomybę nuo mikrolygmens veiksnių, taikant suminę SHAP diagramą buvo nustatyta, kad svarbiausi požymiai buvo namų ūkio užimamas naudingasis plotas, pensija kaip pagrindinis pajamų šaltinis, priklausymas 5-ai pajamų kvintilinei grupei, asmenų skaičius namų ūkyje ir išsilavinimo lygis. Didesnis naudingojo ploto dydis, aukštesnis išsilavinimo lygis ir priklausymas turtingesnei pajamų grupei turi teigiamą poveikį išlaidoms, o pensija kaip pagrindinis pajamų šaltinis – atvirkštinį.

Mašininio mokymosi modeliai pasirodė esantys pranašesni už daugialypės tiesinės regresijos modelius, o tai sutapo su ankstesniais tyrimais [30, 60]. Tai, kad pajamos yra vienas iš svarbiausių veiksnių, o amžius ir vartojimo išlaidos yra susieti netiesiniais ryšiais buvo atrasta ir kitų tyrimų metu, t. y., naudojant Malaizijos [30] ir Rusijos [23] namų ūkių duomenis. Atliktas tyrimas Lietuvos namų ūkių atveju parodė, kad jokiame vertinimo modeliui nepavyko paaiškinti daugiau kaip 40 % vartojimo išlaidų variacijos. Žemi modelių tikslumai dėl išlaidų modeliavimo sudėtingumo buvo pastebėti ir kitų tyrėjų darbuose. Mignouna ir kt. [57], taikydami kvantilinės regresijos metodą vartojimo išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimui pasiekė tik pseudo $R^2 = 0,07$ reikšmę. Puttanapong'as ir Lim'as [60] tyrime maksimalus determinacijos koeficientas siekė 0,41 reikšmę, nors ir buvo taikyti pažangūs mašininio mokymosi metodai, tokie kaip atsitiktiniai miškai, neuroniniai tinklai ir atraminių vektorių metodas. Tuo tarpu Lee'ė, Ong'as ir Lee'ė [30] savo tyrime nustatė, jog namų ūkių gaunamos pajamos, išreikštos pinigine forma, paaiškina apie 54,39 % visos bendrų išlaidų variacijos. Atsižvelgiant į tai, jog Lietuvos namų ūkių tyrimo kontekste pajamos buvo ne skaitinis, o kategorinis kintamasis, t. y., pajamų kvintilinės grupės, galėtų paaiškinti mažesnių modelių tikslumą dėl reikšmingos informacijos dalies praradimo.

Išvados

1. Namų ūkių vartojimą lemia mikrolygmens veiksniai, makroekonominė šalies padėtis ir išorinė aplinka. Mikrolygmeniu svarbiausias veiksnys yra disponuojamosios pajamos, kiti veiksniai – užimtumas, turimas turtas, taupymo elgsena, polinkis skolintis, namų ūkių amžius, sudėtis, lytis, išsilavinimo lygis, gyvenimo būdas, profesijos pasirinkimas ar psichologiniai požymiai. Makrolygmeniu vartojimą veikia palūkanų normos, infliacija ir jos lūkesčiai, būsto kainos ar valdžios politika.
2. Tyrime buvo naudoti 2021 m., Lietuvos namų ūkių biudžetų tyrimo duomenys, apimantys 4334 stebinius ir 40 kintamųjų, iš kurių 14 buvo priklausomieji – vartojimo išlaidas atspindintys kintamieji pagal skirtingas kategorijas, kurių pasiskirstymas pasižymėjo didele dispersija ir dešiniąja asimetrija. Požymių svarba vartojimo išlaidoms buvo vertinta taikant *Phik* koreliacijos koeficientus, Moodo medianos testą, TI matą bei kitus statistinius metodus, priklausomai nuo prielaidų. Išskirčių ir nulinių reikšmių įtaka mažinta naudojant vinzorizacijos ir logaritminio transformavimo metodus. Modelių kūrimui buvo naudoti daugialypės tiesinės regresijos, atsitiktinių miškų bei ekstremalaus gradientinio stiprinimo metodai. Tiesinės regresijos tinkamumas buvo vertintas pagal prielaidas, o mašininio mokymosi modelių hiperparametrai buvo optimizuojami Bajeso metodu. Modelių palyginimui naudotos determinacijos koeficiento, kvadratinės, vidutinės absoliutinės ir medianinės absoliutinės paklaidų metrikos.
3. Remiantis *Phik* koreliacijos koeficientais, tik keturios vartojimo išlaidų kategorijos – bendros, maisto, ryšių ir finansinių paslaugų – turėjo bent vieną veiksnį, kurio ryšio su išlaidomis stiprumas viršijo 0,3, todėl jos ir buvo pasirinktos tolimesnei analizei. Moodo medianos testu buvo nustatyta, kad bendras ir maisto vartojimo išlaidas mažiausiai paaiškinantys tokie veiksniai kaip centralizuoto vandens tiekimo būdas, namų ūkio kilmės rodikliai, su gyvenamąja vieta susiję rodikliai, dabartinio mokymosi statusas ir būsto tipas, kai ryšių išlaidų atveju papildomai buvo pagrindinis būsto šildymo būdas, o finansinių paslaugų atveju – tie patys, kaip ir bendroms išlaidoms, išskyrus su gyvenamąja vieta bei būsto tipu susijusius veiksnius. Visoms tirtoms išlaidų kategorijoms svarbūs požymiai buvo namų ūkio tipas, pajamų kvintilinė grupė ir šaltinis, socialinė-ekonominė grupė, amžius bei asmenų skaičius namų ūkyje.
4. Bendrų, maisto ir ryšių paslaugų vartojimo išlaidų variaciją geriausiai paaiškino ekstremalaus gradientinio stiprinimo regresijos modelis (determinacijos koeficientų reikšmės atitinkamai 0,40; 0,31; 0,38), o draudimo – atsitiktinių miškų modelis (determinacijos koeficiento reikšmė = 0,30), taikant 3 % vinzorizaciją, ir, kai kuriais atvejais, požymių atranką ar logaritminę transformaciją. Remiantis SHAP analize buvo nustatyta, kad nepriklausomai nuo tirtos išlaidų kategorijos, svarbiausi požymiai jų augimui buvo didesnis namų ūkio narių skaičius bei aukštesnė pajamų kvintilinė grupė, o mažėjimui – didesnis namų ūkio galvos amžius. Be to, atrasta, kad tiek užimamas naudingasis būsto plotas, tiek amžius, tiek gyventojų skaičius namų ūkyje yra susiję su vartojimo išlaidomis nevysiškai tiesiniais ryšiais. Išsiaiškinta, kad teigiamo poveikio aukštojo išsilavinimo požymis buvo svarbesnis vertinant ryšių ir finansinių paslaugų išlaidų kitimą, kai tuo tarpu pensijos, kaip pagrindinio pajamų šaltinio egzistavimas, mažino prognozuojamas išlaidas visose kategorijose, išskyrus maisto išlaidas.

Pripažįstama, kad vien mikrolygmens veiksnių nepakanka paaiškinti vartojimo elgsenos ir apie 60 % likusios vartojimo išlaidų variacijos galėtų paaiškinti papildomi požymiai, tokie kaip psichologiniai, makroekonominiai ar kiti išoriniai aspektai, todėl yra rekomenduojama tolimesniuose tyrimuose atsižvelgti į tiriamų požymių papildymą. Be to, iš literatūros analizės buvo pastebėta, kad didesni modelių tikslumai buvo pasiekti vertinant išlaidas laike, todėl tolesniuose tyrimuose būtų patartina išbandyti laiko eilučių duomenis. Taip pat yra patartina atkreipti dėmesį į duomenų iš apklausų šališkumą, kuris atsiranda dėl nenuoširdžių atsakymų arba nereprezentatyvios respondentų imties (Lietuvos atveju – 40 % visų respondentų buvo pensininkai). Atliekant tolimesnius tyrimus yra siūloma išbandyti dar pažangesnius mašininio mokymosi algoritmus arba kvantilinę regresiją.

Studijų metu taip pat buvo parengtas tiriamasis darbas, kuris buvo pristatytas 2024 m. Lietuvos mokslo tarybos organizuotoje mokslinės praktikos konferencijoje „Mokslo vasara“. Dalyvavimo įrodymas pateiktas 38 priede.

Literatūros sąrašas

1. BAKRI, Syazwani, Mohd, et al. Understanding Behavior of Consumption Expenditure of Households. *International Business Education Journal*. 2017, 10(1), 43-52. Prieiga per: doi: <http://dx.doi.org/10.37134/ibej.vol10.5.2017>
2. MAGRABI, Frances M., et al. *The Economics Of Household Consumption*. New York: Greenwood, 1991. ISBN 0-275-94113-2
3. RAMANAUSKAS, Tomas ir Audronė JAKAITIENĖ. Namų ūkių vartojimo modeliavimas. *Pinigų studijos* [interaktyvus]. Vilnius: Lituania, 2007, 1, 25-44 [žiūrėta 2025-04-02]. ISSN 1392-2637. Prieiga per: <https://www.lituanistika.lt/content/14579>
4. KŠIVICKIENĖ, Daiva. Skurdo apraiškos Lietuvos kaime. *Žemės ūkio mokslai* [interaktyvus]. Vilnius: Lituania, 2006, 1, 81-91 [žiūrėta 2025-04-03]. ISSN 1392-0200. Prieiga per: <https://www.lituanistika.lt/content/47460>
5. ISMAL-ABU, Khalid, et al: *Counting the world's poor: Back to Engel's law. Economic and Social Commission for Western Asia* [interaktyvus]. 2022 [žiūrėta 2025-04-02]. Prieiga per: <https://www.unescwa.org/sites/default/files/pubs/pdf/counting-world-poor-engel-law-english.pdf>
6. ŠČUPAKOVAITĖ, Evelina. Namų ūkio vartojimo kitimas ekonomikos augimo sąlygomis: Lietuvos atvejis. *Studentų mokslinės konferencijos straipsnių rinkinys*. 2018. ISSN 1822-9913. Prieiga per: <http://evf.asu.lt/wp-content/uploads/sites/4/2018/04/Jaunasis-mokslininkas-2018-STRAIPSNAI.pdf>
7. LI, Timothy, POTTERS, Charles: *Consumer Spending Definition, Measurement, and Importance, Economics* [interaktyvus]. 2023 [žiūrėta 2025-04-03]. Prieiga per: <https://www.investopedia.com/terms/c/consumer-spending.asp>
8. BOETTEKE, Peter J, NEWMAN, Patrick. The Consequences of Keynes. *Journal of Markets and Morality, Forthcoming*. 2016, 12. Prieiga per: doi: <https://dx.doi.org/10.2139/ssrn.2803980>
9. BIROL, Ozlen., GENCER, Aysen. The Keynesian system. *Journal on Business Review*. 2014, 3(26). Prieiga per: doi: <https://doi.org/10.7603/s40706-014-0026-6>
10. HORWITZ, S: *Say's Law of Markets: an Austrian Appreciation* [interaktyvus]. 2003 [žiūrėta 2025-04-05]. Prieiga per: <http://myslu.stlawu.edu/~shorwitz/Papers/Say's%20Law-Elgar.pdf>
11. GEDMINAITĖ, Rūta: *Lietuvos namų ūkių pajamų ir jų nelygybės analizė Baltijos šalių kontekste* [interaktyvus]. 2015 [žiūrėta 2025-03-15]. Prieiga per: <https://epublications.vu.lt/object/elaba:8925446/8925446.pdf>
12. PIEKUT, Marlena, PIEKUT, Kamil. Changes in Patterns of Consumer Spending in European Households. *Journal of Sustainability*. 2022, 14(19), 12794. Prieiga per: doi: <https://doi.org/10.3390/su141912794>
13. KOROL, Tomasz. Evaluation of the Macro- and Micro-Economic Factors Affecting the Financial Energy of Households. *Energies*. 2021, 14(12), 3512, doi: <https://doi.org/10.3390/en14123512>
14. JAJIĆ, Iavn, HERCEG, Tomislav, BACH, Mirjana. Deployment of the Microeconomic Consumer Theory in the Artificial Neural Networks Modelling: Case of Organic Food Consumption. *Mathematics*. 2022, 10(17), 3215, doi: <https://doi.org/10.3390/math10173215>
15. HAUSER, John R., URBAN, Glen, L. The Value Priority Hypotheses for Consumer Budget Plans. *Journal of Consumer Research*. 1986, 12(4), 446-462. Prieiga per: doi: <https://doi.org/10.1086/208529>

16. CIARLI, Tommaso, LORENTZ, Andre, SAVONA, Maria, VALENTE, Marco. The Effect Of Consumption And Production Structure On Growth And Distribution. A Micro To Macro Model. *Microeconomica*. 2009, 180-218, doi: <https://doi.org/10.1111/j.1467-999X.2009.04069.x>
17. LIU, Ting, SHRYANE, Nick, ELLIOT, M. Micro-macro multilevel analysis of day-to-day lifestyle and carbon emissions in UK multiple occupancy households. *Sustainable Production and Consumption*. 2023, 13-29. Prieiga per: doi: <https://doi.org/10.1016/j.spc.2023.04.023>.
18. CARBONELL, Ada, BERGH, Jeroen. A Micro-Econometric Analysis of Determinants of Unsustainable Consumption in The Netherlands. *Environmental and Resource Economics*. 2004, 27(4), 367-389. Prieiga per: doi: <http://dx.doi.org/10.1023/B:EARE.0000018514.98541.8b>
19. ADRIAN, John, DANIEL, Raymond. Impact of Socioeconomic Factors on Consumption of Food in the United States. *Agricultural Economics*. 1976, doi: <https://doi.org/10.2307/1238574>
20. FREIDERIKS, Elisha, Karen STENNER, and Elizabeth HOBMAN. The Socio-Demographic and Psychological Predictors of Residential Energy Consumption. *Energies*. 2015, 8(1), 573-609. Prieiga per: doi: <https://doi.org/10.3390/en8010573>
21. BIANCA, Vainikka: *Psychological Factors Influencing Consumer Behaviour* [interaktyvus]. 2015 [žiūrėta 2025-04-06]. Prieiga per: <https://www.theseus.fi/handle/10024/96405>
22. DING, Yongxia et al. Factors Influencing the Spatial Difference in Household Energy Consumption in China. *Sustainability*. 2016, 8(12), 1285, doi: <https://doi.org/10.3390/su8121285>
23. ROSHCHINA, Yana. To Drink or Not to Drink: The Microeconomic Analysis of Alcohol Consumption in Russia. *Higher School of Economics Research Paper*. 2013, 29. Prieiga per: doi: <https://dx.doi.org/10.2139/ssrn.2286169>
24. GANONG, Peter, NOEL, Pasca: *How Does Unemployment Affect Consumer Spending?* [interaktyvus]. 2016 [žiūrėta 2025-04-07]. Prieiga per: https://scholar.harvard.edu/files/ganong/files/ganong_jmp_unemployment_spending.pdf
25. POLOGEORGIS, Nicolas: *Employability, the Labor Force, and the Economy* [interaktyvus]. 2023 [žiūrėta 2025-04-08]. Prieiga per: <https://www.investopedia.com/articles/economics/12/employability-labor-force-economy.asp>
26. NAMAZKHAN, Maliheh, ALBERS, Casper, STEG, Linda. A decision tree method for explaining household gas consumption: The role of building characteristics, socio-demographic variables, psychological factors and household behaviour. *Renewable and Sustainable Energy Reviews*. 2020, 119, ISSN 1364-0321, doi: <https://doi.org/10.1016/j.rser.2019.109542>
27. BUGHEANU, Alexandru., STRACHINARU, Adina. Financial Spending Behavior Patterns Based on Education, Gender and Age. *Studies in Business and Economics*. 2020, 15 (2), 62-68. Prieiga per: doi: <https://doi.org/10.2478/sbe-2020-0025>
28. ARAPOVA, Ekaterina. Determinants Of Household Final Consumption Expenditures In Asian Countries: A Panel Model. *Applied Econometrics and International Development* [interaktyvus]. 2018, 18(1) 122-140. Prieiga per: https://ideas.repec.org/a/eea/aeinde/v18y2018i1_8.html
29. YORK, Richard. Energy consumption in European Union Nations. *Social Science Research*. 2007, 36(3), 855-872. Prieiga per: doi: <https://doi.org/10.1016/j.ssresearch.2006.06.007>
30. LEE, En, Thian Song ONG, and Yvonnne LEE. Evaluating Household Consumption Patterns: Comparative Analysis Using Ordinary Least Squares and Random Forest Regression Models. *HighTech and Innovation Journal*. 2024, 5(2), 489-507. Prieiga per: doi: <https://doi.org/10.28991/HIJ-2024-05-02-019>

31. LUND, P. J. and B. J. DERRY. Household Food Consumption: The Influence Of Household Characteristics. *Journal of Agricultural Economics*. 1985, 36(1), 41-58. Prieiga per: doi: <https://doi.org/10.1111/j.1477-9552.1985.tb00155.x>
32. VARLAMOVA, J. and N. LARIONAVA. Macroeconomic and Demographic Determinants of Household Expenditures in OECD Countries. *Procedia Economics and Finance*. 2015, 24, 727-733. Prieiga per: doi: [https://doi.org/10.1016/S2212-5671\(15\)00686-3](https://doi.org/10.1016/S2212-5671(15)00686-3)
33. PAVLISA, Karina et al. Structures of Consumption and Professional Identity: An Analysis of the French Household Budget Survey. *Sociology*. 2024, 58(6), 1451-1471. Prieiga per: doi: <https://doi.org/10.1177/00380385241254310>
34. EDINAK, E. A. and A.A. SHIROV. Assessment of the Relationship between the Qualification Structure of Employment and Household Consumption. *Studies on Russian Economic Development*. 2021, 32, 593-602. Prieiga per: doi: <https://doi.org/10.1134/S1075700721060046>
35. KHAMIS, Abd, Alghani, et al. Alcohol Consumption Patterns: A Systematic Review of Demographic and Sociocultural Influencing Factors. *Public Health*. 2022, 19(13), 8103. Prieiga per: doi: <https://doi.org/10.3390/ijerph19138103>
36. SERMAN, Zehra and Julian SIMS. How Social Media Influencers Affect Consumers Purchase Habit?. *UK Academy for Information Systems Conference Proceedings* [interaktyvus]. London: ELibrary, 2020, 10 [žiūrėta 2025-04-15]. Prieiga per: <https://aisel.aisnet.org/ukais2020/10/>
37. SUN, Shiqiang and Pingshang LV. The Role Of Tax Regulation On Private Consumption: An Analysis based on China's Tax System Structure. *Economics*. 2023, 66. Prieiga per: doi: <http://dx.doi.org/10.21203/rs.3.rs-3323328/v1>
38. MAY, Diego, Gabriela NODARI, and Daniel REES. Wealth and Consumption. Reserve Bank of Australia. *Australian Economic Review*. 2020, 53(1), 105-117. Prieiga per: doi: 10.1111/1467-8462.12364
39. PETTINGER, Teyvan: *Causes of Consumer Spending* [interaktyvus]. 2017 [žiūrėta 2025-04-16]. Prieiga per: <https://www.economicshelp.org/blog/396/economics/consumer-spending-its-causes-and-effects/>
40. SOMAN, Dilip and Amar CHEEMA. The Effect of Credit on Spending Decisions: The Role of the Credit Limit and Credibility. *Marketing Science*. 2002, 21(1), 32-53. Prieiga per: doi: <https://doi.org/10.1287/mksc.21.1.32.155>
41. LI, Jie et al. The effects of online consumer credit on household consumption level and structure: Evidence from China. *The Journal of Consumer Affairs*. 2021, 1614-1632. Prieiga per: doi: <https://doi.org/10.1111/joca.12390>
42. KOZLOV, Roman. The Effect of Interest Rate Changes on Consumption. *Economies*. 2015, 11(1), 23. Prieiga per: doi: <https://doi.org/10.3390/economies11010023>
43. MANASSEH, Charles O., et al. The Effects of Interest and Inflation Rates on Consumption Expenditure. *International Journal of Economics and Financial Issues* [interaktyvus]. 2018, 8(4), 32-38 [žiūrėta 2025-04-20]. ISSN 2146-4138. Prieiga per: EconJournals.
44. RADU, Ioana, KENNY, Geoff, REUTER, Andreas. Inflation Expectations, Consumption And The Lower Bound: Micro Evidence From A Large Multi-Country Survey. *Journal of Monetary Economics*. 2021, 120-134. Prieiga per: doi: <https://doi.org/10.1016/j.jmoneco.2020.03.005>
45. COIBION, Oliver, GEORGARAKOS, Dimitris, GORODNICHENKO, Yuriy, ROOIJ, Maarten How Does Consumption Respond to News about Inflation?. *American Economic Journal: Macroeconomics*. 2023, 15(3), 109-52. Prieiga per: doi: 10.1257/mac.20200445

46. MANDEYA, Shelton and Ho SIN-YU. Inflation, Inflation Uncertainty and the Economic Growth Nexus: A Review of the Literature. *Folia Oeconomica Stetinensia*. 2022, 1(19), 172-190. Prieiga per: <https://www.ceeol.com/search/article-detail?id=1051291>
47. CHARALAMPAKIS, Evangelos, FAGANDINI, Bruno, HENKEL, Lukas, OSBAT, Chiara. The Impact Of The Recent Rise In Inflation On Low-Income Households. *European Central Bank*. 2022, 7. Prieiga per: <https://ideas.repec.org/a/ecb/ecbbox/202200074.html>
48. GHOSH, Sudeshna., et al. Consumer Confidence and Consumer Spending in Brazil. *Journal of Economic Theory and Practice*. 2020, 20 (1), doi: <https://doi.org/10.1177/0976747919898906>
49. HALL, Robert. Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*. 1978, 86(6). Prieiga per: <https://www.journals.uchicago.edu/doi/abs/10.1086/260724>
50. BERGER, David, et al. House Prices and Consumer Spending. *The Review of Economic Studies*. 2017, 85(3), 1502-1542. Prieiga per: doi: <https://doi.org/10.1093/restud/rdx060>
51. ANDERSEN, Hendrik and Soren-Leth PETERSEN: *How Do House Prices Drive Spending?* [interaktyvus]. 2019 [žiūrėta 2025-04-15]. Prieiga per: <https://www.econ.ku.dk/cebi/research/how-do-house-prices-drive-spending/>
52. ALM, James, and Asmaa El GANAINY. Value-added Taxation And Consumption. *International Tax and Public Finance*. 2013, 20, 105-128, doi: <https://doi.org/10.1007/s10797-012-9217-0>
53. MAISON, Dominika. What Shapes Spending Styles?. *Rozprawy Ubezpieczeniowe. Konsument Na Rynku Usług*. 2018, 29, 3-26. Prieiga per: <https://www.ceeol.com/search/article-detail?id=745892>
54. PAIK, Songyi. The U.S. Consumption Analysis: Using a Linear Regression Model. *Proceedings of Great Day*. 2016, 21. Prieiga per: <https://knightscholar.geneseo.edu/proceedings-of-great-day/vol2015/iss1/21>
55. SINGH, Jay and Khushi KUMAR. Socio-Economic Dynamics of Energy: A Micro-Level Regression Approach to Domestic Consumption in India. *Journal of Energy and Environmental Policy Options* [interaktyvus]. Kerala: Resdojournals, 2018, 1(3), 60-64 [žiūrėta 2025-04-17]. ISSN 3006-5712. Prieiga per: <https://resdojournals.com/index.php/JEEPO/article/view/20>
56. KUMAR, Jyothi B et al. Logistic regression for polymorphic malware detection using ANOVA F-test. *International Conference on Innovations in Information*. 2017, 1-5. Prieiga per: doi: <https://doi.org/10.1109/ICIIECS.2017.8275880>
57. MIGNOUNA, D. B., et al. A Microeconomic Analysis of Household Consumption Expenditure Determinants of Nigeria and Ghana. *Tropicultura*. 2015, 33(3), 226-237. Prieiga per: <https://hdl.handle.net/10568/72997>
58. MONDIANA, Y. Q., ZAIRINA, A., SARI, R. K. Quantile regression modeling to predict extreme precipitation. *Journal of Physics*. 2021, 1918, doi: 10.1088/1742-6596/1918/4/042031
59. CAGLAYAN, Ebru, and Melek ASTAR. A Microeconomic Analysis of Household Consumption Expenditure Determinants for Both Rural and Urban Areas in Turkey. *Health Expenditures*. 2012. Prieiga per: <https://www.academia.edu/download/108863080/4.pdf>
60. PUTTANAPONG, Nattapong and Siphath LIM. Predicting Household Expenditure Using Machine Learning Techniques: A Case of Cambodia. *Nakhara Journal of Environmental Design and Planning*. 2024, 23(3), 421. Prieiga per: doi: <http://dx.doi.org/10.54028/NJ202423421>
61. MOSCA, Edoardo, et al. SHAP-Based Explanation Methods: A Review for NLP Interpretability. In *Proceedings of the 29th International Conference on Computational Linguistics* [interaktyvus].

- Gyeongju: International Committee on Computational Linguistics, 2022, pp. 4593-4603 [žiūrēta 2025-04-17]. Prieiga per: <https://aclanthology.org/2022.coling-1.406/>
62. RIVAS-PEREA, Pablo, et al. Support Vector Machines for Regression: A Succinct Review of Large-Scale and Linear Programming Formulations. *International Journal of Intelligence Science*. 2013, 3, 5-14. Prieiga per: doi: <http://dx.doi.org/10.4236/ijis.2013.31002>
 63. CIABURRO, Giuseppe, and Balaji VENKATESWARAN. *Neural Networks*. Birmingham: Packt 2017. ISBN 978-1-78839-787-2.
 64. ERGUN, Selim, Phebe OWUSU, and Maria RIVAS. Determinants of renewable energy consumption in Africa. *Environmental Science and Pollution Research*. 2019, 26, 15390–15405. Prieiga per: doi: <https://doi.org/10.1007/s11356-019-04567-7>
 65. ROSSUM, Guido: *Python Tutorial. Release 3.10.11* [interaktyvus]. 2023 [žiūrēta 2025-04-18]. Prieiga per: <https://scicomp.ethz.ch/public/manual/Python/3.10.4/tutorial.pdf>
 66. KAMPEZIDOU, Styliani I., et al. Fundamental Components and Principles of Supervised Machine Learning Workflows with Numerical and Categorical Data. *Artificial Intelligence and Data Science for Engineering Improvements*. 2024, 5(1), 384-416. Prieiga per: doi: <https://doi.org/10.3390/eng5010021>
 67. ZHENG, Alice, and Amanda CASARI. *Feature Engineering for Machine Learning. Principles and Techniques for Data Scientists*. Sebastopol: O'Reilly Media, 2018. ISBN 978-1491953242.
 68. BROWNLEE, Jason: *How to Choose a Feature Selection Method For Machine Learning* [interaktyvus]. 2020 [žiūrēta 2025-04-20]. Prieiga per: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>
 69. AMBUSAI, Mohammed, A., and Xiangjain, Nanda, Priyadarsi HE. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. *IEEE Transactions on Computers*. 2016, 65(10), 2986 – 2998, doi: <https://doi.org/10.1109/TC.2016.2519914>
 70. WANG, Shuihua, et al. Pathological Brain Detection By Artificial Intelligence In Magnetic Resonance Imaging Scanning. *Progress In Electromagnetics Research*. 2016, 156, 105-133. Prieiga per: doi: <http://dx.doi.org/10.2528/PIER16070801>
 71. SIRAJ, Muhammad, Tohari AHMAD, and Royyana IJTIHADIE. Analyzing ANOVA F-test and Sequential Feature Selection for Intrusion Detection Systems. *International Journal of Advances in Soft Computing and its Applications*. 2022, 14(2), 185-195. Prieiga per: doi: <https://doi.org/10.15849/IJASCA.220720.13>
 72. RIFFENBURGH, Robert H. Tests on the Distribution Shape of Continuous Data. *Statistics in Medicine*. 2006, 369-386. Prieiga per: doi: <https://doi.org/10.1016/B978-012088770-5/50060-5>
 73. ZHOU, Y., Yiyang ZHU, and Weng Kee WONG. Statistical tests for homogeneity of variance for clinical trials and recommendations. *Contemp Clin Trials Commun*. 2023, 33,:101119. Prieiga per: doi: <https://doi.org/10.1016/j.conctc.2023.101119>
 74. DOQUIRE, Gauthier, VERLEYSEN, Michel. Mutual information-based feature selection for multilabel classification. *Neurocomputing*. 2013, 122, 148-155. Prieiga per: doi: <https://doi.org/10.1016/j.neucom.2013.06.035>
 75. PENG, Hanchuan, LONG, Fuhui, DING, Cris. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *Eee Transactions On Pattern Analysis And Machine Intelligence*. 2005, 27(8), 1226-1236, Prieiga per: doi: <https://doi.org/10.1109/tpami.2005.159>

76. KINNEY, Justin B., ATWAL, Gurinder S. Equitability, Mutual Information, and the Maximal Information Coefficient. *Statistics*. 2014, 111(9), 3354-3359. Prieiga per: doi: <https://doi.org/10.1073/pnas.1309933111>
77. HERNANDEZ, Damian G., and Ines SAMEGO. Estimating the Mutual Information between Two Discrete, Asymmetric Variables with Limited Samples. *Entropy*. 2019, 21(6), 623. Prieiga per: doi: <https://doi.org/10.3390/e21060623>
78. MCCLENAGHAN, Eliot: *The Kruskal–Wallis Test* [interaktyvus]. 2024 [žiūrėta 2025-04-15]. Prieiga per: <https://www.technologynetworks.com/informatics/articles/the-kruskal-wallis-test-370025>
79. DURIS Viliam, TIRPAKOVA, Anna. A survey on the global optimization problem using Kruskal–Wallis test. *Annales Mathematicae et Informaticae*. 2020, 52, 281-298. Prieiga per: doi: <https://doi.org/10.33039/ami.2020.05.004>
80. GLEASON, Jamie G: *Comparative Power Of The Anova, Approximate Randomization Anova, And Kruskal-Wallis Test* [interaktyvus]. 2013 [žiūrėta 2025-04-21]. Prieiga per: https://digitalcommons.wayne.edu/oa_dissertations/658/
81. ENGMANN, Sonja, COUSINEAU, Denis. Comparing Distributions: The Two-Sample Anderson-Darling Test As An Alternative To The Kolmogorov-Smirnoff Test. *Quantitative Methods Inquires*. 2011, 6(3), 17. Prieiga per: https://www.jaqm.ro/issues/volume-6,issue-3/pdfs/1_engmann_cousineau.pdf
82. SCHEFF, Stephen W. Nonparametric Statistics. Fundamental Statistical Principles for the Neurobiologist. *A Survival Guide*. 2016, 157-182. Prieiga per: doi: <https://doi.org/10.1016/B978-0-12-804753-8.00008-7>
83. HUSSAIN, Ibrar, et al. Optimal features selection in the high dimensional data based on robust technique: Application to different health database. *Heliyon*. 2024, 10(17). Prieiga per: doi: <https://doi.org/10.1016/j.heliyon.2024.e37241>
84. SMIDA, Zaineb et al. A median test for functional data. *Journal of Nonparametric Statistics*. 2022, 34(2), 520-553. Prieiga per: doi: <https://doi.org/10.1080/10485252.2022.2064997>
85. FREIDLIN, Boris, GASTWIRTH, Joseph L. Should the Median Test be Retired from General Use?. *The American Statistician*. 2012, 54(3) 161-164. Prieiga per: doi: <https://doi.org/10.1080/00031305.2000.10474539>
86. OTI, Eric, Michael OLUSOLA, and Perewarebo ESEMOKUMO. Statistical Analysis of the Median Test and the Mann-Whitney U Test. *International Journal of Advanced Academic Research* [interaktyvus]. Ekowe: Ijaar publishing, 2021, 7(9), 44-51 [žiūrėta 2025-04-24]. ISSN 2488-9849. Prieiga per: <https://www.ijaar.org/articles/v7n9/ste/ijaares-v7n9-Sep21-p7902.pdf>
87. TASCÓN, Maria. T., and Francisco J. CASTANO. Selection of Variables in Small Business Failure Analysis: Mean Selection vs. Median Selection. *Network of Scientific Journals* [interaktyvus]. Empresa: P' aginas, 2017, 54-88 [žiūrėta 2015-08-07]. E-ISSN 1886-516X. Prieiga per: doi: <https://doi.org/10.46661/revmetodoscuanteconempresa.2880>
88. MANGIAFICO, Salvatore: *Functions to Support Extension Education Program Evaluation* [interaktyvus]. 2017 [žiūrėta 2025-04-27]. Prieiga per: <https://pbil.univ-lyon1.fr/CRAN/web/packages/rcompanion/rcompanion.pdf>
89. LESACK, Kyle, and Christopher NAUGLER. An open-source software program for performing Bonferroni and related corrections for multiple comparisons. *Journal of Pathology Informatics*. 2011, 2(1), 52. Prieiga per; doi: <https://doi.org/10.4103/2153-3539.91130>

90. BAAK, M., et al. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. *Computational Statistics & Data Analysis*. 2020, 152, 107043. Prieiga per: doi: <https://doi.org/10.1016/j.csda.2020.107043>
91. NYITRAI, Tamas, and Miklós VIRÁG. The effects of handling outliers. *Socio-Economic Planning Sciences*. 2019, 67, 34-42. Prieiga per: doi: <https://doi.org/10.1016/j.seps.2018.08.004>
92. GUSSENBAUGER, Johannes: *Robust Statistical Methods for Outlier Detection* [interaktyvus]. 2015 [žiūrėta 2025-04-27]. Prieiga per: 84. Prieiga per: <https://repositum.tuwien.at/bitstream/20.500.12708/14511/2/Gussenbauer%20Johannes%20-%202015%20-%20Robust%20statistical%20methods%20for%20outlier%20detection...pdf>
93. BELLÉGO Christophe, BENATIA, David PAPE, Louis-Daniel. Dealing with Logs and Zeros in Regression Models. *Série des Documents de Travail*. 2022, 71. Prieiga per: doi: <https://dx.doi.org/10.2139/ssrn.3444996>
94. WOOLDRIGE, Jeffery M. *Econometrics. A Modern Approach*. 5th Edition [interaktyvus]. Mason: Michigan State University, 2012 [žiūrėta 2025-04-28]. ISBN 1-111-53104-8. Prieiga per: https://cbpbu.ac.in/userfiles/file/2020/STUDY_MAT/ECO/2.pdf
95. SALAZAR, Jose J., et al. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *Journal of Petroleum Science and Engineering*. 2022, 209, 109885. Prieiga per: doi: <https://doi.org/10.1016/j.petrol.2021.109885>
96. YAN, Peiliang et al. The potential of machine learning to predict melting response time of phase change materials in triplex-tube latent thermal energy storage systems. *Applied Energy*. 2025, 390, 125863. Prieiga per: doi: <https://doi.org/10.1016/j.apenergy.2025.125863>
97. WEISBERG, Sanford. *Applied Linear Regression*. 3rd Edition. Minnesota: Wiley, 2005. ISBN 0-471-66379-4.
98. GUPTA, Shubhy. Evaluating the Economic Impact of Sales Promotions through Predictive Analytics Techniques. *SSRN*. 2023, 19. Prieiga per: doi: <https://dx.doi.org/10.2139/ssrn.4934279>
99. CHEN, Tianqi, and Carlos GUEESTRIN. XGBoost: A Scalable Tree Boosting System. *International Conference on Knowledge Discovery and Data Mining*. 2016, 785-794. Prieiga per: doi: <https://doi.org/10.1145/2939672.2939785>
100. JANGARAJ, Avanija, SUNITHA, Gurram, MADHAVi, Reddy K, KORA, Padmavathi. Prediction of House Price Using XGBoost. *Turkish Journal of Computer and Mathematics Education* [interaktyvus]. Tirupati: Research article, 2021, 12(2), 2151-2155. Prieiga per: https://www.researchgate.net/publication/350810698_Prediction_of_House_Price_Using_XGBoost_Regression_Algorithm
101. SHAHANI, Niaz Muhammad, et al. Developing an XGBoost Regression Model for Predicting Young's Modulus. *Advances in Modeling*. 2021, 9. Prieiga per: doi: <https://doi.org/10.3389/feart.2021.761990>
102. LI, Tingyang, STEIN, Joshua, NALLASAMY, Nambi. Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery. *Clinical science*. 2023, 107, 1066-1071. Prieiga per: doi: <http://dx.doi.org/10.1136/bjophthalmol-2021-320599>
103. ROTH, Alvin E. The Shapley value. *Essays in Honor of Lloyd S. Shapley*. 2009, 1-28. Prieiga per: doi: <https://doi.org/10.1017/CBO9780511528446.002>
104. LUNDBERG, Scott, and Su-In LEE. A Unified Approach to Interpreting Model Predictions. *Computer Science*. 2017. Prieiga per: doi: <https://doi.org/10.48550/arXiv.1705.07874>

Priedai

1 Priedas. Analizuojamų duomenų dydis

```
print(df.columns)
print("Eilučių skaičius:", df.shape[0])
print("Stulpelių skaičius:", df.shape[1])

Index(['Visos namų ūkio vartojimo išlaidos (mėnesinės)',
      'Maistas ir nealkoholiniai gėrimai',
      'Alkoholiniai gėrimai, tabakas ir narkotikai', 'Apranga ir avalynė',
      'Būstas, vanduo, elektra, dujos ir kitas kuras',
      'Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra',
      'Sveikata', 'Transportas', 'Informacija ir ryšiai',
      'Poilsis, sportas ir kultūra', 'Švietimo paslaugos',
      'Restoranai ir apgyvendinimo paslaugos',
      'Draudimas ir finansinės paslaugos',
      'Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos',
      'Amžius', 'Gyvenamoji vieta', 'Regionas', 'Namų ūkio tipas',
      'Išlaikomų vaikų skaičius', 'Socialinė ekonominė grupė',
      'Pajamų šaltinis', 'Namų ūkio galvos išsilavinimas',
      'Pajamų kvintilinė grupė', 'Būsto tipas', 'Būsto valdos statusas',
      'Centralizuotas šalto vandens tiekimas',
      'Centralizuotas karšto vandens tiekimas',
      'Pagrindinis būsto šildymo būdas', 'Asmenų skaičius namų ūkyje',
      'Namų ūkio užimamas naudingasis plotas', 'Motinos gimimo šalis',
      'Gimimo šalis', 'Pagrindinės pilietybės šalis', 'Tėvo gimimo šalis',
      'Lytis', 'Teisinė santuokinė padėtis', 'Gyvena su partneriu',
      'Pagrindinis darbas: profesija', 'Šiuo metu mokosi',
      'Darbo sutarties tipas'],
      dtype='object')
Eilučių skaičius: 4334
Stulpelių skaičius: 40
```

2 Priedas. Trūkstančių reikšmių patikrinimas

```
missing_values = df.isna().sum()

# Spausdiname rezultata
print(missing_values)
```

Visos namų ūkio vartojimo išlaidos (mėnesinės)	0
Maistas ir nealkoholiniai gėrimai	0
Alkoholiniai gėrimai, tabakas ir narkotikai	0
Apranga ir avalynė	0
Būstas, vanduo, elektra, dujos ir kitas kuras	0
Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra	0
Sveikata	0
Transportas	0
Informacija ir ryšiai	0
Poilsis, sportas ir kultūra	0
Švietimo paslaugos	0
Restoranai ir apgyvendinimo paslaugos	0
Draudimas ir finansinės paslaugos	0
Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos	0
Amžius	0
Gyvenamoji vieta	0
Regionas	0
Namų ūkio tipas	0
Išlaikomų vaikų skaičius	0
Socialinė ekonominė grupė	0
Pajamų šaltinis	0
Namų ūkio galvos išsilavinimas	0
Pajamų kvintilinė grupė	0
Būsto tipas	0
Būsto valdos statusas	0
Centralizuotas šalto vandens tiekimas	0
Centralizuotas karšto vandens tiekimas	0
Pagrindinis būsto šildymo būdas	0
Asmenų skaičius namų ūkyje	0
Namų ūkio užimamas naudingasis plotas	0
Motinos gimimo šalis	0
Gimimo šalis	0
Pagrindinės pilietybės šalis	0
Tėvo gimimo šalis	0
lytis	0
Teisinė santuokinė padėtis	0
Gyvena su partneriu	0
Pagrindinis darbas: profesija	0
Šiuo metu mokosi	0
Darbo sutarties tipas	0

The figure displays a 4x4 grid of histograms, comparing the frequency distribution of various categories in their original form (left column) and after a logarithmic transformation (right column). The categories are listed in the rows, and the transformations are listed in the columns.

Categories (Rows):

- Visos namų ūkio vartojimo išlaidos (mėnesinės)
- Būstas, vanduo, elektra, dujos ir kitas kuras
- Poilsis, sportas ir kultūra
- Maistas ir nealkoholiniai gėrimai
- Transportas
- Informacija ir ryšiai
- Draudimas ir finansinės paslaugos
- Asmens priežiūra, socialinė apsauga ir įvairios prekės ir paslaugos

Transformations (Columns):

- Originalių reikišmių histograma
- Logaritmuotų reikišmių histograma
- Originalių reikišmių histograma
- Logaritmuotų reikišmių histograma

The histograms show the frequency distribution of values for each category. The x-axis represents the value of the variable, and the y-axis represents the frequency (Dažnis). The log-transformed versions generally show a more normal distribution, indicating that the logarithmic transformation is effective in reducing skewness.

4 Priedas. Lietuvos namų ūkių pasiskirstymo pagal išlaikomų vaikų skaičių aprašomoji analizė ir struktūrinė analizė pagal išlaidų kategorijas

Taikant „pandas“ bibliotekos funkcijas, tokias kaip *groupby*, *agg*, *quantile* ar *describe*, buvo apskaičiuotos išlaidų skaitinės charakteristikos, sugrupuotos pagal kategorinio kintamojo lygius, arba įvertintos pagal visą kiekybinį kintamąjį. Pateiktos skaitinės charakteristikos: *N* – stebėjimų skaičius, *vid.* – aritmetinis vidurkis, *std* – standartinis nuokrypis, parodantis duomenų išsibarstymą, t. y., kiek reikšmės yra vidutiniškai nutolusios nuo vidurkio, *min* – mažiausia reikšmė, 25 % – pirmasis kvartilis, kuris reiškia, kad 25 % visų duomenų yra mažesni už tą reikšmę, kai 50 % ir 75 % yra antrasis (mediana) bei trečiasis kvartiliai, kurie atitinkamai reiškia, kad 50 % ir 75 % duomenų yra mažesni už apskaičiuotas reikšmes, *max* – didžiausia reikšmė.

Lietuvos namų ūkių vartojimo išlaidų ir skaitinių aiškinamųjų kintamųjų aprašomoji statistika

Kintamasis	Vid.	Std	Min	25 %	50 %	75 %	Max
Visos namų ūkio vartojimo išlaidos (mėnesinės), Eur.	792,74	615,55	56,23	382,34	630,5	1016,54	7984,83
Maistas ir nealkoholiniai gėrimai, Eur.	257,53	180,29	0	131,14	218,4	339,4	2053,07
Alkoholiniai gėrimai, tabakas ir narkotikai, Eur.	20,19	54,96	0	0	0	1,94	744,46
Apranga ir avalynė, Eur.	45,83	104,97	0	0	0	45	1186,25
Būstas, vanduo, elektra, dujos ir kitas kuras, Eur.	130,52	124,95	0	71,22	106,33	152,38	3339,30
Būsto apstatymo, namų ūkio įranga ir kasdienė namų priežiūra, Eur.	41,91	103,08	0	0	13,98	48,47	3691,87
Sveikata, Eur.	61,57	127,86	0	0	15	79,57	1999,98
Transportas, Eur.	87,64	242,03	0	0	0	101,76	5291,96
Informacija ir ryšiai, Eur.	42,29	41,64	0	17	32	53	558,32
Poilsis, sportas ir kultūra, Eur.	27,21	62,19	0	0	2,17	25	1121,28
Švietimo paslaugos, Eur.	4,73	36,99	0	0	0	0	1035
Restoranai ir apgyvendinimo paslaugos, Eur.	31,74	91,61	0	0	0	10,64	1568,69
Draudimas ir finansinės paslaugos, Eur.	12,17	18,64	0	0	6,4	15,83	213,45
Asmens priežiūra, socialinė apsauga ir įvairios paslaugos, Eur.	29,42	59,39	0	0	6,67	30,85	778,57
Amžius, metais	58,86	16,74	18	48	61	72	96
Asmenų skaičius namų ūkyje, asm.	1,96	1,1	1	1	2	2	9
Namų ūkio užimamas naudingasis plotas, m ²	71,41	37,77	12	49	62	81	400

Kaip rodo aprašomoji statistika, bendros mėnesinės vartojimo išlaidos tarp Lietuvos namų ūkių yra itin nevienodos, nes nors vidurkis siekia 792,74 Eur, tačiau standartinis nuokrypis yra 615,55 Eur, kai maksimali išleidžiama suma vartojimui viršija 7984 Eur, o minimaliai išleidžiama – vos 56 Eur. Tai rodo labai platų duomenų pasiskirstymo intervalą ir pakankamai didelį stebėjimų nutolimą nuo vidurkio, indikuojantį apie rezultatų kokybei įtakos turinčias išskirtis. Tokios tendencijos matyti ir su kitomis išlaidų kategorijomis, ypač alkoholiniams gėrimams, aprangai ir avalynei, švietimo paslaugoms, transportui, restoranams ir apgyvendinimui skirtomis išlaidomis, kurioms 50 % visų namų ūkių iš viso neskyrė savo pajamų, kai vidurkis siekė atitinkamai apie 20 Eur., 45 Eur., 4 Eur., 87 Eur., 31 Eur. per mėn. Prie maksimaliai išleidžiančių namų ūkių nepriartėja ir 75 % apklausoje dalyvavusių namų ūkių, kurie, pavyzdžiui, maistui ir nealkoholiniams gėrimams per mėnesį

vidutiniškai išleidžia ne daugiau 1016,54 Eur, nors maksimali suma siekia beveik 8 tūkst. eurų. Pastebima, kad apklausoje dalyvavo tik pilnamečiai asmenys (minimalus amžius – 18 metų), kai bendrai daugiausiai atsakymų pateikė vyresnio amžiaus asmenys, nes vidutinis respondentų amžius siekė beveik 59 metus. Nors apklausoje pasitaikė ir didelių namų ūkių (iki 9 asmenų), vidutinis Lietuvoje namų ūkio dydis yra apie 2 asmenis, kai tuo tarpu didžioji dalis respondentų (75 %) gyvena mažesniuose nei 81 m² naudingojo ploto turinčiuose būstuose.

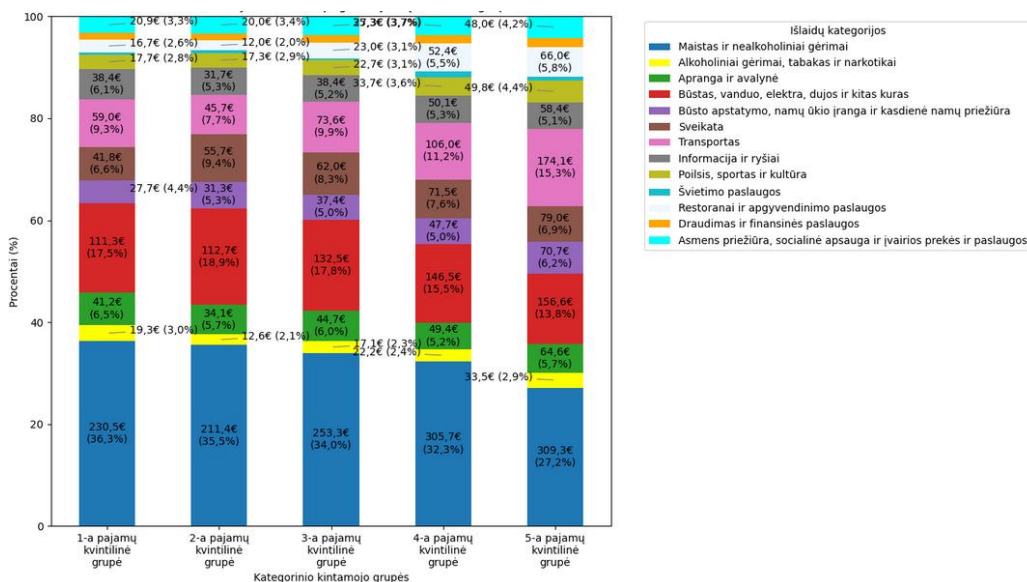
Remiantis Lee'ė, Ong'as ir Lee'ė [30] bei kitų mokslininkų išvadomis, svarbiausi namų ūkių vartojimo išlaidas veikiantys aiškinamieji kintamieji yra pajamos ir namų ūkio dydis, dėl to aprašomoji vartojimo išlaidų analizė pagal kategorinius kintamuosius buvo pateikta tik šių veiksmų atžvilgiu. Aprašomoji statistika pagal kitus kategorinius kintamuosius pateikta 29 priede.

Namų ūkių pajamų dydį nusakančio požymio grupės yra sudarytos padalijus namų ūkius į penkias grupes pagal didėjimo tvarką išrikiuotas disponuojamąsias pajamas vienam namų ūkio nariui. Pirmąją penktadalį, t. y. pirmąją kvintilinę grupę, sudaro asmenys, kurių disponuojamosios pajamos vienam namų ūkio nariui yra mažiausios (mažesnės už pirmąją 20 % kvintilį). Antrajai kvintilinei grupei yra priskiriami tie, kurių pajamos yra didesnės už pirmąją kvintilį, bet mažesnės už antrąją kvintilį (nuo 20 % iki 40 %), kai penktoji kvintilinė grupė apima 20 % gyventojų, kurių namų ūkyje esantys asmenys priklauso aukščiausiam pajamų lygiui, t. y., turtingiausi namų ūkiai. Oficialiosios statistikos portalo (šaltinis: <https://osp.stat.gov.lt/lietuvas-gyventoju-pajamos-ir-gyvenimo-salygos-2022/namu-ukiu-pajamos/pajamu-koncentracija>) duomenimis, 2021 m. ekvivalentinės piniginės disponuojamosios pajamos skurdžiausioje (pirmojoje kvintilinėje) namų ūkių grupėje siekė iki 346 Eur, antrojoje – nuo 347 iki 585 Eur, trečiojoje – nuo 586 iki 807 Eur, ketvirtojoje – nuo 808 iki 1094 Eur, o penktojoje – daugiau nei 1095 Eur. Verta paminėti, kad daugiausiai namų ūkių pateko į antrąją kvintilinę grupę ir sudarė 25 % visų tirtų namų ūkių. Vartojimo išlaidų aprašomoji statistika pagal namų ūkių gaunamų pajamų lygį rodo, kad didėjant pajamoms (pereinant į aukštesnę kvintilinę grupę), auga ir vartojimo išlaidos. Visgi pastebima ir priešinga tendencija – tarp pirmosios ir antrosios kvintilinių grupių, nes antrojoje kvintilinėje grupėje 75 % namų ūkių per mėnesį išleidžia ne daugiau kaip 734 Eur, kai pirmojoje grupėje šis rodiklis yra aukštesnis, t. y., iki 858,4 Eur, nepaisant priklausymo žemesnei pajamų kvintilinei grupei. Kaip pabrėžia Ramanauskas ir Jakaitienė [3], toks neatitikimas tarp oficialių pajamų ir vartojimo gali būti paaiškintas namų ūkių pajamų generavimu iš šešėlinės veiklos. Kitų kvintilinių grupių aprašomoji statistika pagal vartojimo išlaidas atskleidžia tipiškesnius rezultatus, pvz., medianos reikšmės rodo, kad 50 % namų ūkių, esančių pirmojoje pajamų kvintilinėje grupėje per mėnesį išleidžia ne daugiau 291,2 Eur, kai tuo tarpu penktojoje grupėje – ne daugiau 621 Eur.

Lietuvos namų ūkių pasiskirstymo pagal namų ūkio pajamų kvintilinę grupę ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vid. (Eur)	Std (Eur)	Min (Eur)	25 % (Eur)	50 % (Eur)	75 % (Eur)	Max (Eur)
1-a pajamų kvintilinė grupė	751	634,6	473,4	503	82,2	291,2	858,4	5416,4
2-a pajamų kvintilinė grupė	1084	595,2	468,6	438,6	94,9	296,3	734	3145,2
3-a pajamų kvintilinė grupė	1009	745,5	608,3	501,2	64,0	401	936,4	3744,5
4-a pajamų kvintilinė grupė	701	946	779,7	629,7	113,4	528,2	1186,4	4491,2
5-a pajamų kvintilinė grupė	789	1138,7	911,8	826,8	56,2	621	1414,1	7984,8

Vartojimo išlaidų struktūros analizė rodo, kad didėjant namų ūkio pajamoms, vartojimui skiriamų lėšų dalis, tenkanti maistui, būsto išlaikymui, aprangai ir avalynei, mažėja, o išlaidų dalis, skirta transportui, restoranams, poilsiui bei kitoms prekėms ir paslaugoms – didėja. Verta pastebėti, kad penktoje pajamų kvintilinėje grupėje esantys asmenys struktūriškai didesnę nei kitos grupės savo išlaidų dalį skiria transportui, būsto priežiūrai bei poilsiui. Pastebėtina, kad 1-ai pajamų kvintilinei grupei priskiriami asmenys maistui ir nealkoholiniams gėrimams skiria apie 36 % visų vartojimo išlaidų, o tai gali reikšti, kad Lietuvos gyventojai nėra skurstantys, nes ši dalis nesiekia teorijoje [5] nurodytos 60 % ribos, kuri yra siejama su absoliučiu skurdu.

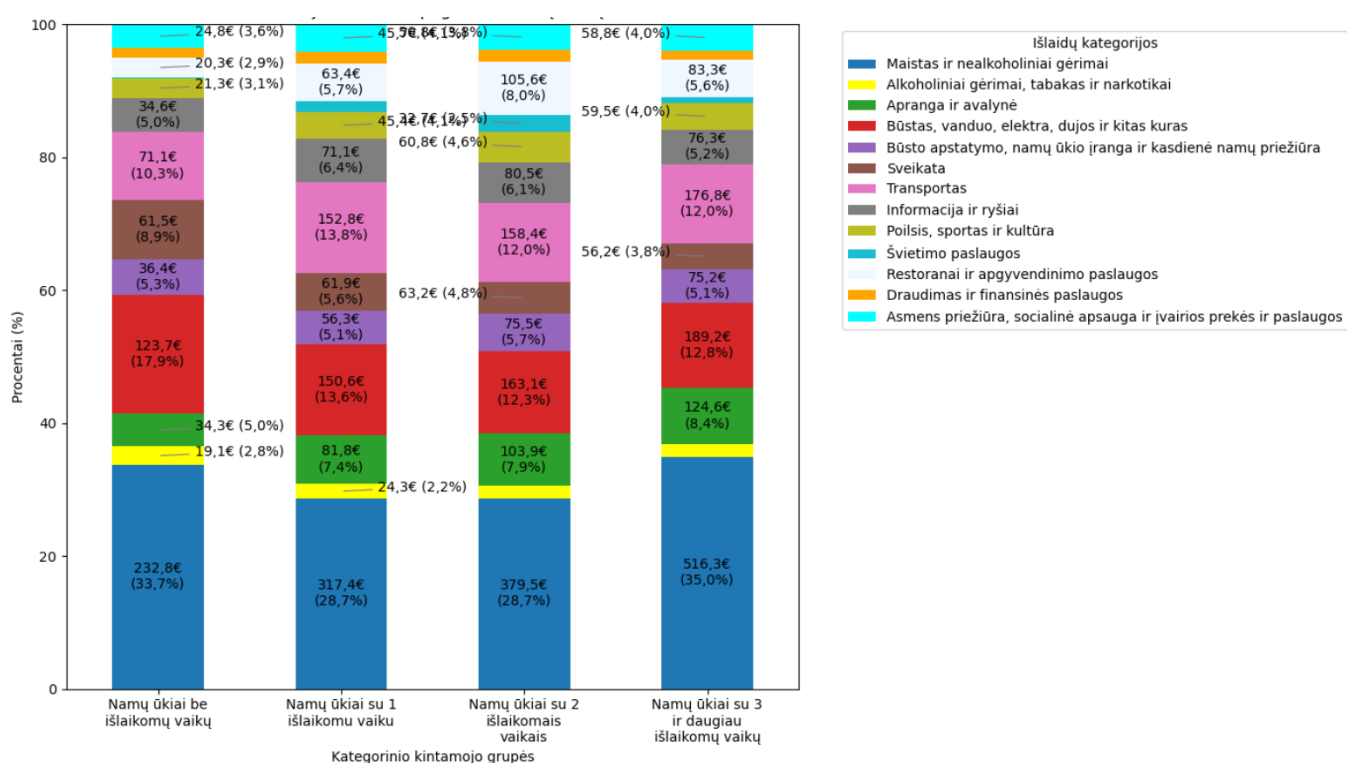


Lietuvos namų ūkių 2021 m. vidutinių vartojimo išlaidų per mėnesį struktūra pagal namų ūkių pajamų kvintilinę grupę

Namų ūkio dydį, kaip kategorinį kintamąjį, gali atspindėti išlaikomų vaikų skaičius. Apklausoje dalyvavusių namų ūkių pasiskirstymas pagal išlaikomų vaikų skaičių rodo, kad 80,9 % visų namų ūkių neturėjo išlaikomų vaikų, kai tuo tarpu namų ūkiai su vienu išlaikomu vaiku sudarė mažesnę dalį (10,3 %), o mažiausiai reprezentuojama grupė – namų ūkiai su 3 ir daugiau vaikų, t. y., 2,3 % visų apklausoje dalyvavusių namų ūkių. Verta paminėti, kad nemažą namų ūkių dalį be išlaikomų vaikų gali sudaryti senatvės pensininkų grupė, kuri duomenų imtyje sudaro 40,7 % visų namų ūkių. Pastebėtina, kad kuo daugiau namų ūkyje yra vaikų, tuo didesnės yra ir vidutinės vartojimo išlaidos. Pavyzdžiui, 50 % visų namų ūkių, kurie neturi vaikų, per mėnesį išleidžia ne daugiau kaip 553,2 Eur, kai tuo tarpu namų ūkiai su trimis ar daugiau vaikų – 1321 Eur. Taip pat nors vidutinės vartojimo išlaidos auga didėjant išlaikomų vaikų skaičiui, tačiau maksimalios išlaidų reikšmės yra didesnės tarp namų ūkių, kurie turi vieną arba visai neturi vaikų. Detalesnė vartojimo išlaidų struktūra rodo, kad ypač daug lėšų maistui ir nealkoholiniams gėrimams įsigyti skiria namų ūkiai, turintys tris ar daugiau vaikų (35 % nuo visų išlaidų) ir namų ūkiai be vaikų (33,7 % nuo visų išlaidų), o tai galėjo lemti jų sudėtyje buvus gausius skaičius pensinio amžiaus gyventojų. Namų ūkiai su vienu ar dviem išlaikomais vaikais yra labiau linkę savo vartojimo lėšų dalį skirti pramogoms, t. y., restoranams, poilsiui, sportui, nei kitos dvi grupės. Tuo tarpu namų ūkiai, auginantys vieną vaiką, struktūriškai didesnę savo biudžeto dalį skiria transportui nei kitos grupės – vidutiniškai apie 13,8 % visų mėnesinių išlaidų, kai kitų grupių dalis siekia iki 12 %.

Lietuvos namų ūkių pasiskirstymo pagal išlaikomų vaikų skaičių ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupė	N	Vid. (Eur.)	Std (Eur.)	Min (Eur.)	25 % (Eur.)	50 % (Eur.)	75 % (Eur.)	Max (Eur.)
Namų ūkiai be išlaikomų vaikų	3505	690,4	553,2	538,3	56,2	349,1	867,9	7984,8
Namų ūkiai su 1 išlaikomu vaiku	446	1107,5	957,7	687,9	174,8	645,4	1365,7	5474,2
Namų ūkiai su 2 išlaikomais vaikais	285	1323,4	1165,3	740,5	184,8	780,9	1589,4	4475,2
Namų ūkiai su 3 ir daugiau išlaikomų vaikų	98	1476,1	1321	742,1	305	1001,8	1793,8	4491,2



Lietuvos namų ūkių 2021 m. vidutinių vartojimo išlaidų per mėnesį struktūra pagal išlaikomų vaikų skaičių

5 Priedas. Dispersijos analizė bei Kruskalio ir Voliso testas kaip papildomi požymių atrankos metodai

Dispersijos analizė vertina, kiek kategorinis požymis prisideda prie priklausomo kintamojo dispersijos paaiškinimo. Vienfaktorė ANOVA (angl. *Analysis of Variance*) yra dažniausiai naudojama dispersijos analizės forma, skirta tikrinti, ar tikslo kintamojo vidurkiai skiriasi tarp daugiau nei dviejų grupių, apibrėžiamų viena kategorine savybe [56]. ANOVA Fišerio (toliau – F-statistika) yra apskaičiuojama pagal šį santykį:

$$F = \frac{MS_{\text{tarpgrupinė}}}{MS_{\text{vidinė}}};$$

čia $MS_{\text{tarpgrupinė}}$ yra tarpgrupinė dispersija, o $MS_{\text{vidinė}}$ – vidinė dispersija.

Tarpgrupinė dispersija matuoja skirtumus tarp grupių vidurkių ir bendrojo vidurkio, o vidinė dispersija matuoja variaciją kiekvienos grupės viduje. Pagal gautą F-statistikos reikšmę yra vertinamos hipotezės: H_0 , teigianti, kad kokybinis požymis neturi poveikio priklausomam kintamajam ir H_1 – požymis turi reikšmingą poveikį. Vertinant pagal F-statistiką gautą p-reikšmę, ji yra lyginama su nustatytu reikšmingumo lygmeniu (α , pvz., 0,05), jei ji mažesnė už reikšmingumo lygmenį, tai yra priimama H_1 [71]. Pažymėtina, kad didelė F-statistikos reikšmė rodo, jog požymis yra stipriai susijusi su tikslo kintamuoju ir labiau paaiškina jo skirtumus tarp grupių. Verta pabrėžti, kad ANOVA analizė nėra skirta nustatyti, kokį konkretų poveikį daro kintamasis, o tik patikrinti, ar jis statistiškai reikšmingai veikia priklausomąjį kintamąjį. ANOVA analizei atlikti naudota *scipy* bibliotekos `f_oneway()` funkcija (šaltinis: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html).

Svarbu suprasti, kad ANOVA F-statistika, kaip ir visi ANOVA variantai, yra parametrinis metodas, todėl privalo tenkinti nepriklausomumo, tikslo kintamojo duomenų pasiskirstymo pagal normalųjį skirstinį ir grupių dispersijų lygumo prielaidas:

- nepriklausomumo prielaida yra tenkinama, kai yra užtikrinama, jog naudojamų duomenų imtis yra nepriklausoma;
- normalumo prielaida yra tenkinama, kai priklausomojo kintamojo reikšmės kiekviename kategorijos lygmenyje sutampa su teoriniu normaliuoju skirstiniu, kuris primena varpo formą. Jei duomenys nėra normaliai pasiskirstę, nėra galima taikyti metodų, kuriems būtina prielaida apie Gauso skirstinio paklaidas. Duomenų apytikslį normalumą galima patikrinti keliais testais, pavyzdžiui, *Chi*-kvadrato gerumo atitikimo testu, Kolmogorovo ir Smirnov testu ar Šapiro ir Vilko testu (angl. *Shapiro-Wilk test*), kai pastarasis yra naujausias iš minėtų. Riffenburgh'o [72] teigimu, didesnio sudėtingumo testai yra geresni, o tokie dažniausiai būna naujesni metodai. Be to, Šapiro ir Vilko testas yra mažiau konservatyvus, palyginti su anksčiau minėtais testais. Naudojant Šapiro ir Vilko testą yra tikrinamos hipotezės: H_0 – populiacija X yra pasiskirsčiusi pagal normalųjį skirstinį, t. y., normalųjį Gauso pasiskirstymą su tikroju vidurkiu μ ir dispersija σ^2 , žymima – $N(\mu, \sigma^2)$, ir alternatyvioji (H_1): populiacija X nėra normaliai pasiskirsčiusi. Statistinio dydžio W formulė:

$$W = \frac{\sum_{i=1}^m a_{i,n} (x_{n+1-i} - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2};$$

čia m – didžiausias sveikasis skaičius, kuris yra mažesnis arba lygus $n/2$, $\bar{x}$ – imties vidurkis, $a_{i,n}$ – skaičiavimo parametras

Imtis $x_1, x_2, \dots, x_n$ (n – imties dydis) turėtų būti paimta iš populiacijos X ir išrikiuota didėjimo tvarka. Kai reikšmingumo lygis yra α (dažniausiai – 0,05), tai nulinė hipotezė atmetama, kai $W \leq W_\alpha$ (kai Šapiro ir Vilko testo W statistikos tikimybės reikšmė yra mažesnė arba lygi 0,05) Jei $W > W_\alpha$, tai nėra pakankamai įrodymų H_0 atmetimui ir galima daryti išvadą, kad populiacija X yra normaliai pasiskirsčiusi (kai Šapiro ir Vilko testo W statistikos tikimybės reikšmė yra didesnė už 0,05). Normalumo testas pagal Šapiro ir Vilko kriterijų atliktas taikant *scipy.stats* bibliotekos *shapiro()* funkciją (šaltinis: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>).

- Dispersija matuoja, kiek plačiai kintamojo reikšmės yra pasiskirsčiusios aplink jo vidurkį. Jei dispersijos skirtingose kategorinio kintamojo grupėse labai skiriasi, t. y., egzistuoja duomenų heteroskedastiškumas, parametriniu metodu gauti rezultatai nėra patikimi ir gali lemti klaidingas statistines išvadas [73]. Vienas populiariausių testų yra Leveno (angl. *Levene*) testas, kuris nėra labai jautrus išskirtinėms reikšmėms, nes nepriklauso nuo imčių dispersijų. Tyrėjai pabrėžia, kad Leveno testu yra daroma prielaida, jog duomenys Y_{ij} (i – grupės numeris, j – konkretaus stebėjimo numeris toje grupėje) yra nepriklausomi ir normaliai pasiskirstę. Apibrėžus, kad $Z_{ij} = |Y_{ij} - \bar{Y}_i|$, $\bar{Z}_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij} / N$, $\bar{Z}_{i.} = \sum_{j=1}^{n_i} Z_{ij} / n_i$, Leveno testo statistika homogeniškumui tarp grupių vertinti yra gaunama pagal formulę:

$$W = \frac{(N-k)}{k-1} \frac{\sum_{i=1}^k n_i (Z_{i.} - \bar{Z}_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - Z_{i.})^2};$$

čia N – imties dydis, k – grupių skaičius, n_i – i -osios grupės imties dydis

Jei esant pasirinktam reikšmingumo lygiui α , W statistikos reikšmė yra didesnė nei kritinė reikšmė ($W > W_{\alpha, k-1, N-k}$), tai nulinė hipotezė, teigianti, kad dispersijos tarp grupių yra homogeniškos, turi būti atmesta (Leveno statistikos tikimybės reikšmė yra mažesnė už 0,05). Priešingu atveju, t. y., kai Leveno statistikos p -reikšmė (tikimybė) yra didesnė arba lygi 0,05, yra laikoma, kad turimi duomenys nesuteikia pagrindo manyti, jog grupių dispersijos reikšmingai skiriasi. Leveno testas atliktas taikant *scipy.stats* bibliotekos *levене()* funkciją (šaltinis: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levене.html>)

Prielaidų tenkinimas yra svarbus aspektas, nulemiantis, ar tyrime yra tikslinga naudoti įprastus parametrinius testus, tokius kaip ANOVA.

Kruskalio ir Voliso testas yra neparametrinis statistinis metodas, skirtas įvertinti, ar duomenų skirstinių padėtis skiriasi, t. y., ar bent viena grupė turi tendenciją būti didesnė ar mažesnė už kitas. Tai yra ANOVA alternatyva, kai duomenys neatitinka normalumo ar dispersijų homogeniškumo prielaidų [78]. Testas naudotinas, kai yra vertinamos daugiau nei dvi grupės. Jei grupių skirstinių forma yra vienoda ir jos skiriasi tik padėties parametru, galima daryti išvadas apie grupių medianas. Dėl šios priežasties Kruskalio ir Voliso testas dažnai vadinamas „medianų lyginimo metodu“. Kruskalio ir Voliso testo atveju yra formuluojamos dvi hipotezės: nulinė hipotezė (H_0), teigianti, kad visų grupių medianos yra vienodos, ir alternatyvioji hipotezė (H_1), kuri nurodo, jog bent vienos grupės mediana skiriasi nuo kitų. Kruskalio ir Voliso kriterijaus statistika apskaičiuojama:

$$H = \left(\frac{12}{N(N+1)} \right) \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1);$$

čia N – bendras visų stebėjimų skaičius, k – grupių (požymių) skaičius, n_i – stebėjimų skaičius i -toje grupėje ir R_i – rangų suma i -toje grupėje.

Apskaičiuota H statistika yra rangų skirtumų tarp grupių matas. Ši reikšmė lyginama su χ^2 pasiskirstymu, kur laisvės laipsnių skaičius yra $k-1$ [79]. Gautos H reikšmės dydis rodo, kiek reikšmingi yra rangų skirtumai tarp grupių. Pagal šią statistiką apskaičiuota p -reikšmė leidžia įvertinti nulinės hipotezės priėmimą ar atmetimą. Jei p -reikšmė yra mažesnė už 0,05 (pasirinktą reikšmingumo lygmenį), tai rodo, kad bent viena grupė turi medianos reikšmę, kuri reikšmingai išsiskiria iš kitų, ir požymis yra laikomas reikšmingai susijęs su priklausomu kintamuoju. Testo taikymui naudota $kruskal()$ ²² funkcija iš *scipy.stats* bibliotekos.

Kruskalio ir Voliso testui nėra reikalinga normalumo prielaida, tačiau taikant jį turi galioti kitos dvi prielaidos: stebėjimai turi būti nepriklausomi, o grupių skirstinių formos – panašios. Gleason'as [80] teigia, kad grupėse esančių duomenų pasiskirstymas yra laikomas panašios formos kai duomenų dispersija tarp skirtingų grupių nėra labai skirtinga, skirstinių vidurkio, medianos ir modos padėtys yra artimos (panašus asimetriškumas), turi panašų ekscesą (sutampantį pasiskirstymo viršūnės tipą) ir turi sutampantį modų, t. y., viršūnių, skaičių. Pažeista duomenų pasiskirstymo grupėse formos panašumo prielaida gali reikšti, kad skirtumas tarp grupių gali būti klaidingai suprastas kaip medianų skirtumas, kai iš tikrųjų jį galėjo lemti skirtingos pasiskirstymo formos. Esant labai skirtingiems grupių pasiskirstymams, testas gali prarasti galią tiksliai nustatyti tikrąjį skirtumą tarp grupių ir lemti klaidinančius rezultatus. Skirstinių palyginimui dažnai yra naudojamas Andersono ir Darlingo, t. y., AD testas [81]. AD testas suteikia didesnę svorį duomenų uodegoms, todėl geriau nustato skirstinių skirtumus, kai pastarieji turi ekstremalių reikšmių. Be to, AD testas gali patikimiau nustatyti skirstinių skirtumus net ir su mažesne imtimi. KS testas taip pat pasižymi didesniu konservatyvumu, kas reiškia, jog jis dažnai neaptinka reikšmingų skirtumų, nors ir jie iš tikrųjų egzistuoja. K -imčių Andersono ir Darlingo testu galima palyginti daugiau nei dvi imtis vienu metu. Šis testas tikrina, ar duomenys iš skirtingų grupių atitinka tą patį bendrą skirstinį, t. y., iš anksto nėra apibrėžta teorinė skirstinio forma. K -imčių AD tinkamumo kriterijaus statistikos reikšmės formulė:

$$A_{kN}^2 = \sum_{i=1}^k n_i \int_{B_N} \frac{\{F_{n_i}(x) - H_N(x)\}^2}{H_N(x)\{1 - H_N(x)\}} dH_N(x);$$

čia k – imčių skaičius, n_i – i -osios imties dydis, $F_{n_i}(x)$ – i -osios imties empirinė kaupiamoji skirstinio funkcija, $H_N(x)$ – bendra sujungtų imčių kaupiamoji skirstinio funkcija, B_N – reikšmių sritis, kai $H_N(x) < 1$.

Naudojant šią formulę k skirtingų imčių empirinės kaupiamosios skirstinio funkcijos yra lyginamos su bendra sujungta imtimi $H_N(x)$. Jei visos tiriamos imtys yra pasiskirsčiusios pagal tą patį skirstinį, tai jų empirinės kaupiamosios funkcijos bus panašios į bendrą sujungtų imčių kaupiamąją skirstinio funkciją ir todėl k -imčių Andersono ir Darlingo tinkamumo kriterijaus statistikos reikšmė (A_{kN}^2) bus labai maža, o pagal ją paskaičiuota p -reikšmė didesnė nei pasirinktas reikšmingumo lygmuo [81]. Tokiu atveju būtų priimta nulinė hipotezė, teigianti, kad visų imčių skirstiniai panašūs. Priešingu atveju, t. y. alternatyvios hipotezės atveju, jei gauta A_{kN}^2 reikšmė yra didelė, o pagal ją paskaičiuota p -reikšmė mažesnė už pasirinktą reikšmingumo lygmenį, tai reikštų, kad bent vienos imties skirstinys

²² <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html>

skiriasi nuo kitų grupių imčių ir skirtinių formos grupėse nėra panašios. Taikant programinę įrangą buvo naudota *anderson_ksamp()*²³ funkcija iš *scipy.stats* bibliotekos.

²³ https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.anderson_ksamp.html

6 Priedas. Daugialypės tiesinės regresijos metodologija

Paprasta ir klasikinė jos lygtis buvo aptarta mokslinės literatūros metu (žr. 1.4. skyrelį ir (3) formulę). Daugialypės tiesinės regresijos modelis buvo sukurtas taikant *OLS*²⁴ funkciją iš *statsmodels* bibliotekos. Metodo taikymo prielaidos (Weisberg'as [97], Paik [54]):

- Nėra multikolinearumo problemos. Multikolinerumo egzistavimas reiškia, kad nepriklausomi kintamieji yra vienas su kitu stipriai tarpusavyje susiję, dėl to yra iškraipomas nepriklausomų kintamųjų statistinis reikšmingumas. Tokiu atveju nėra įmanoma tiksliai nustatyti kuris kintamasis yra atsakingas už priklausomojo kintamojo pokytį, o t-testu gauti rezultatai tampa nepatikimais. Be to, dėl multikolinearumo regresijos koeficientai taip pat tampa nestabilūs. Jei vienas iš kintamųjų gali visiškai paaiškinti kito dispersiją, tai nėra tikslinga abu kintamuosius įtraukti į regresijos modelį. Literatūroje plačiai taikomas rodiklis, skirtas multikolinearumo tikrinimui:

$$VIF_i = \frac{1}{1 - R_i^2};$$

čia R_i^2 – determinacijos koeficientas, rodantis, kaip konkretus aiškinamasis kintamasis yra paaiškinamas kitais nepriklausomais kintamaisiais po atskirų regresinių lygčių taikymo.

Dispersijos didinimo koeficientas (angl. *Variance Inflation Factor*, VIF) interpretuojamas taip, kad kuo yra didesnė VIF reikšmė, tuo labiau tikėtina, jog egzistuoja multikolinearumo problema. Weisberg'as [97] teigia, kad dažniausiai VIF reikšmės, viršijančios 10, rodo stiprų ir problematišką multikolinearumą, dėl to yra rekomenduojama pašalinti vieną iš stipriai tarpusavyje susijusių aiškinamųjų kintamųjų. Tyrime buvo naudojama *variance_inflation_factor*²⁵ funkcija iš *statsmodels* bibliotekos.

- Homoskedastiškumas. Jis egzistuoja tada, kai modelio paklaidos turi pastovią dispersiją visame prognozuotame intervale, o heteroskedastiškumas – kai paklaidų dispersija yra nelygi ir keičiasi priklausomai nuo tikslo kintamojo reikšmių. Homoskedastiškumo prielaidos pažeidimas reiškia, kad reikšmingumo testai (t-testas, F-testas) gali klaidingai įvertinti ryšių statistinį reikšmingumą. Vienas dažniausiai naudojamų testų – Vaito testas, pagrįstas originalios regresijos liekanų kvadratų regresija su tais pačiais pradinės regresijos nepriklausomais kintamaisiais, įtraukiant jų kvadratus ir kryžmines sandaugas. Nulinė Vaito testo hipotezė teigia, kad modelio liekanų dispersija yra pastovi, o alternatyvioji – kad ji kinta.

$$LM = n \times R_f^2;$$

čia n – stebinių skaičius, R_f^2 – determinacijos koef. iš liekanų regresijos.

Vaito testas yra priskiriamas Lagranžo daugiklio (LM) testų grupei, kurie remiasi χ^2 statistika. Pagal Vaito testą yra apskaičiuojama p-reikšmė, t. y. tikimybė, kad χ^2 statistika su atitinkamais laisvės laipsniais bus didesnė už gautą LM reikšmę. Jei p-reikšmė yra mažesnė už pasirinktą reikšmingumo lygį (p-reikšmė < 0,05), nulinė hipotezė turi būti atmesta ir priimta alternatyvioji hipotezė, t. y., modelio liekanos pasižymi heteroskedastiškumu. Statistikai įvertinti tyrime buvo naudota *het_white*²⁶ funkcija iš *statsmodels* bibliotekos.

²⁴ https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html

²⁵ https://www.statsmodels.org/dev/generated/statsmodels.stats.outliers_influence.variance_inflation_factor.html

²⁶ https://www.statsmodels.org/stable/generated/statsmodels.stats.diagnostic.het_white.html

- Normalumas. Paklaidų skirstinys turi atitikti normalųjį pasiskirstymą. Normalumo prielaidos pažeidimas taip pat gali iškraipyti statistinių reikšmingumo testų rezultatus ir sumažinti modelio išvadų patikimumą. Ekonometrijos moksle duomenų normalumo prielaidai patikrinti dažniausiai naudoja *Jarque-Bera* (JB) testą, kurio formulė:

$$JB = \frac{n}{6} \cdot \left(S^2 + \frac{(K-3)^2}{4} \right);$$

čia n – imties dydis, S – imties asimetrijos koef., K – imties eksceso koef.

Jei duomenyse egzistuoja normalumas, tai JB testo reikšmė turi būti maža, o pagal ją gauta p-reikšmė, turėtų būti didesnė nei reikšmingumo lygmuo (p-reikšmė $> 0,05$), kas reikštų, jog nėra pagrindo atmesti nulinę hipotezę ir manyti, kad paklaidos nėra normaliai pasiskirsčiusios. Statistikai įvertinti tyrime buvo naudota *jarque_bera*²⁷ funkcija iš *scipy* bibliotekos.

- Autokoreliacijos nebuvimas. Modelio paklaidos neturėtų priklausyti viena nuo kitos laiko ar eilučių tvarka. Kadangi tyrime yra naudojami ne laiko eilučių duomenys ir apklausos rezultatai buvo pateikti visiškai atsitiktine tvarka, ši prielaida buvo neaktuali.

Modelio tinkamumo testai:

- F-testas. Vertinamas sukurto modelio reikšmingumas, kuris nurodo ar bent vienas iš modelyje esančių nepriklausomų kintamųjų gali paaiškinti priklausomojo kintamojo kitimą. F-testo reikšmė yra apskaičiuojama:

$$F = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k};$$

čia k – nepriklausomų kintamųjų skaičius, n – visi stebiniai, R^2 – determinacijos koef.

Reikšmingumas yra vertinamas pagal apskaičiuotą F-testo reikšmę ir jai atitinkamą p-reikšmę, kuri yra lyginama su 5 % reikšmingumo lygmeniu ($\alpha = 0,05$). Nulinė hipotezė – determinacijos koeficientas populiacijoje yra lygus nuliui, t. y., modelis visiškai nepajėgia paaiškinti priklausomojo kintamojo dispersijos. Atmesta nulinė hipotezė (p-reikšmė $< 0,05$) reiškia, kad modelis yra reikšmingas, t. y., R^2 populiacijoje statistiškai reikšmingai skiriasi nuo nulio.

- t-testas. Taikomas nustatant statistiškai reikšmingai prie priklausomojo kintamojo prognozės prisidėjusius nepriklausomus kintamuosius. Nulinė hipotezė teigia, kad gauti nepriklausomų kintamųjų koeficientai statistiškai reikšmingai nesiskiria nuo nulio, t. y., veiksnys neturi statistiškai reikšmingos įtakos priklausomam kintamajam. T-testo statistikos formulė:

$$t = \frac{b_j}{s_{b_j}};$$

čia b_j – j -ojo kintamojo regresijos koeficientas, s_{b_j} – to koeficiento standartinė paklaida.

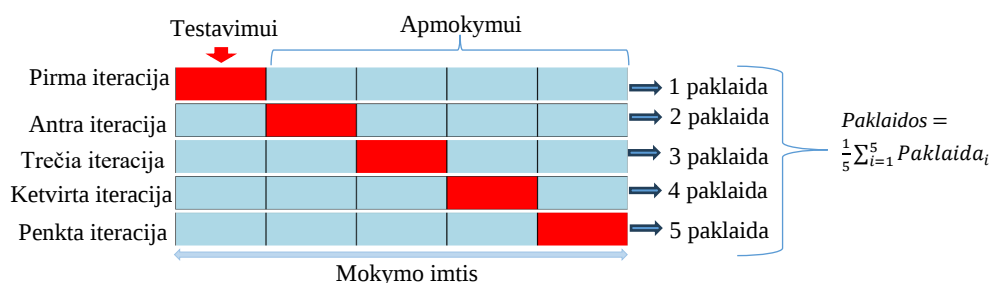
Jei t-testo atitinkama p-reikšmė mažesnė už reikšmingumo lygmenį (p-reikšmė $< 0,05$), reiškia analizuotas koeficientas statistiškai reikšmingai skiriasi nuo nulio.

Jei modelis atitinka visas prielaidas ir yra reikšmingas, laikoma, kad jis yra tinkamas ir patikimas interpretuojant veiksmų poveikį priklausomam kintamajam.

²⁷ https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.jarque_bera.html

7 Priedas. Modelių kokybės vertinimo metodai (validacija ir paklaidos)

Kad sukurti modeliai būtų kuo mažiau šališki buvo naudojama vidinė (angl. *k-fold cross-validation*). Toks vertinimas buvo taikytas kartu su Bajeso optimizavimo algoritmu naudojant mokymo imties duomenis. Yan'as ir kt. [96] pastebi, kad kai kryžminė validacija yra derinama kartu su hiperparametrų paieškos metodais, ji tampa vidiniu validacijos metodu, t. y., kiekvienam hiperparametrų rinkiniui yra taikoma tokia validacija, o optimalus derinys yra parenkamas pagal vidutinę kryžminės validacijos paklaidos įvertį. Shahani'is ir kt. [101] teigia, kad atliekant vidinę kryžminę validaciją, visas mokymosi rinkinys yra padalijamas į k vienodo dydžio blokus (žr. 6 pav.). Dalyje i -ojoje iteracijoje vienas iš k blokų yra parenkamas kaip testavimui skirta dalis, kai likę $k-1$ blokai – apmokymui. Po kiekvienos iteracijos yra įvertinama pasirinkta metrika (pvz., RMSE, MAE, R^2), ir pagal duomenis, esančius testavimo bloke, yra gaunama vienos iteracijos paklaida. Visos gautos paklaidos yra sumuojamos ir padalijamos iš k , kad būtų apskaičiuota bendra kryžminės validacijos paklaida. Tokiu būdu yra vertinamas modelio gebėjimas prognozuoti naujus, nematytus duomenis kiekvienoje kartoje atskirai, ir kartu duomenų panaudojimas yra visapusiškiau išnaudojamas.



Kryžminės validacijos veikimo principas. (Sukurta pagal Sahani'į ir kt. [105])

Apmokytų modelių veikimui įvertinti buvo naudota iš anksto atskirta testavimo imtis, su kurios duomenimis buvo apskaičiuotos įvairios metrikos, skirtos įvertinti modelio gebėjimą apdoroti nematytus duomenis. Yan'o ir kt. [96] teigimu, tarp mokslininkų plačiai taikomas modelio įvertinimo metodas yra determinacijos koeficientas, žymimas R^2 . Jis matuoja bendrą nepriklausomų kintamųjų paaiškinimą prognozuojamo priklausomo kintamojo dispersiją. R^2 gali įgyti reikšmes tarp 0 ir 1, o jo formulė:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - h_i(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2};$$

čia y_i – tikroji reikšmė, $h_i(x_i)$ – modelio prognozė, $\bar{y}$ – visų tikrųjų reikšmių vidurkis, n – stebinių skaičius.

- Klasikinėje tiesinėje regresijoje, pridėdant vis daugiau nepriklausomų kintamųjų į modelį, determinacijos koeficiento reikšmė natūraliai didėja, tačiau tai nebūtinai reiškia, kad tai iš tikrųjų pagerina prognozės tikslumą. Dėl šios priežasties buvo sukurtas pataisytas R^2 (angl. *adjusted*), kuris tinkamiau vertina modelio tikslumą, nes yra atsižvelgiama į naudojamų nepriklausomų kintamųjų skaičių:

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2);$$

čia p – nepriklausomų kintamųjų skaičius

Vidutinė kvadratinė paklaida (angl. *Mean Squared Error*, MSE) taip pat naudojama modelio veikimo vertinimui. Ja yra apskaičiuojamas vidutinis kvadratinis skirtumas tarp prognozuotos reikšmės $\hat{y}_i$ ir

tikrosios reikšmės y_i . Gauta šios metrikos reikšmė gali būti tik teigiama bei sunkiai interpretuojama. Modelio veikimas yra pripažįstamas tuo geresniu, kuo MSE reikšmė yra artimesnė nuliui.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 ,$$

Sahani'is ir kt. [101] teigia, kad dar vienas dažnai taikomas modelių kokybės vertinimo metodas yra kvadratinės paklaidos šaknis (angl. *Root Mean Squared Error*, RMSE). Gauta šios metrikos reikšmė lengvai interpretuojama, nes yra pateikiama originaliais kintamojo vienetais. Pabrėžtina, kad ši metrika yra ypač jautri duomenyse esančioms išskirtims, nes net vienas stipriai nukrypęs taškas gali stipriai išauginti RMSE reikšmę, taip iškreipiant tikrąjį modelio tikslumą. Kuo kvadratinės paklaidos šaknies reikšmė yra artimesnė nuliui, tuo geriau yra vertinamos modelio prognozavimo galimybės. Formulė:

$$RMSE = \sqrt{MSE} ,$$

Li'is, Stein'as ir Nallasamy'is [102] kaip vertinimo metodiką išskiria ir rečiau naudojamą vidutinę absoliutinę paklaidą (angl. *Mean Absolute Error*, MAE), kuri pasižymi didesniu atsparumu išskirtims nei RMSE metrika. MAE parodo kiek vidutiniškai prognozės nukrypsta nuo tikrųjų reikšmių, jų originalioje matavimo skalėje. Metrika yra tinkama naudoti, kai duomenyse egzistuoja nulinės reikšmės. Pripažinta, kad kuo šios metrikos reikšmė yra artimesnė nuliui, tuo geresnis yra ir vertinamo modelio veikimas. Formulė:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| ,$$

Medianinė absoliutinė paklaida (angl. *Median Absolute Error*, MedAE) yra visų absoliučių paklaidų mediana. Ji parodo, kokia yra tipinė paklaida tarp prognozuotų ir tikrųjų reikšmių, t. y., reikšmė, kuri neviršija 50 % visų absoliučių paklaidų. Tai viena iš labiausiai išskirtims atsparių vertinimo metrikų, kurios formulė:

$$MedAE = median(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, \dots, |y_n - \hat{y}_n|);$$

čia *median()* – funkcija, grąžinanti paklaidų medianos reikšmę.

Vertinimo metrikos tyrime buvo apskaičiuotos naudojant *r2_score()*, *mean_squared_error()*, *mean_absolute_error()*, *median_absolute_error()* funkcijas iš *scikit-learn* bibliotekos. Modelių veikimo įvertinimui yra svarbu naudoti kelias skirtingas metrikas, nes kiekviena iš jų atskleidžia skirtingus modelių kokybės aspektus. Remiantis metrikų įverčių lyginimu, yra pasirenkamas geriausiai duomenis aprašantis modelis, kuris vėliau naudojamas interpretacijoms atlikti.

8 Priedas. Veiksnių priklausomybės nuo išlaidų ir jų struktūrinių dalių statistinio reikšmingumo vertinimas

--- Visos namų ūkio vartojimo išlaidos (mėnesinė) ---		--- Maistas ir nealkoholiniai gėrimai ---	
variable	p_value	variable	p_value
Pagrindinis darbas: profesija	0.393	Asmenų skaičius namų ūkyje	2.4e-09
Gyvena su partneriu	0.419	Gyvena su partneriu	0.0399
Socialinė ekonominė grupė	0.407	Namų ūkio tipas	0.00907
Pajamų šaltinis	0.432	Pagrindinis darbas: profesija	0.025
Namų ūkio tipas	0.427	Teisinė santuokinė padėtis	0.855
Darbo sutarties tipas	0.744	Išlaikomų vaikų skaičius	0.0044
Amžius	0.832	Namų ūkio užimamas naudingasis plotas	0.505
Asmenų skaičius namų ūkyje	0.913	Pajamų šaltinis	0.398
Išlaikomų vaikų skaičius	0.375	Socialinė ekonominė grupė	0.233
Pajamų kvintilinė grupė	0.479	Pajamų kvintilinė grupė	0.373
Teisinė santuokinė padėtis	0.518	Darbo sutarties tipas	0.798
Namų ūkio galvos išsilavinimas	0.519	Lytis	0.392
Namų ūkio užimamas naudingasis plotas	0.495	Amžius	1
Būsto valdos statusas	0.371	Namų ūkio galvos išsilavinimas	0.9
Lytis	0.467	Centralizuotas karšto vandens tiekimas	0.387
Šiuo metu mokosi	0.937	Pagrindinis būsto šildymo būdas	0.646
Pagrindinis būsto šildymo būdas	0.671	Būsto tipas	0.000914
Regionas	0.469	Gyvenamoji vieta	0.577
Būsto tipas	0.343	Būsto valdos statusas	0.997
Gyvenamoji vieta	0.493	Centralizuotas šalto vandens tiekimas	0.207
Centralizuotas karšto vandens tiekimas	0.492	Pagrindinės pilietybės šalis	0.136
Pagrindinės pilietybės šalis	0.264	Gimimo šalis	0.397
Centralizuotas šalto vandens tiekimas	0.465	Regionas	0.576
Motinos gimimo šalis	0.481	Motinos gimimo šalis	0.755
Gimimo šalis	0.37	Tėvo gimimo šalis	0.763
Tėvo gimimo šalis	0.418	Šiuo metu mokosi	0.488

--- Alkoholiniai gėrimai, tabakas ir narkotikai ---		--- Apranga ir avalynė ---	
variable	p_value	variable	p_value
Lytis	0.00466	Amžius	1
Pagrindinis darbas: profesija	5.63e-73	Pagrindinis darbas: profesija	2.1e-142
Socialinė ekonominė grupė	6.49e-05	Išlaikomų vaikų skaičius	1.18e-160
Darbo sutarties tipas	1.03e-11	Asmenų skaičius namų ūkyje	0
Pajamų šaltinis	0.00015	Namų ūkio tipas	5.53e-71
Pajamų kvintilinė grupė	0.00845	Pajamų šaltinis	6.38e-18
Asmenų skaičius namų ūkyje	6.48e-229	Darbo sutarties tipas	7.68e-39
Gyvena su partneriu	3.05e-17	Socialinė ekonominė grupė	5.05e-18
Teisinė santuokinė padėtis	1	Namų ūkio galvos išsilavinimas	1
Namų ūkio tipas	1.26e-05	Motinos gimimo šalis	7.93e-61
Pagrindinis būsto šildymo būdas	9.76e-14	Tėvo gimimo šalis	8.98e-48
Pagrindinės pilietybės šalis	1.38e-11	Būsto valdos statusas	8.73e-36
Gyvenamoji vieta	0.55	Gimimo šalis	0.0614
Būsto valdos statusas	3.43e-40	Gyvena su partneriu	1.03e-07
Amžius	1	Pajamų kvintilinė grupė	0.00037
Namų ūkio galvos išsilavinimas	1	Gyvenamoji vieta	0.102
Regionas	0.963	Teisinė santuokinė padėtis	0.991
Išlaikomų vaikų skaičius	3.39e-20	Pagrindinės pilietybės šalis	1.16e-45
Būsto tipas	6.23e-10	Namų ūkio užimamas naudingasis plotas	0.0587
Gimimo šalis	0.023	Centralizuotas karšto vandens tiekimas	0.324
Namų ūkio užimamas naudingasis plotas	0.215	Lytis	0.191
Centralizuotas karšto vandens tiekimas	0.625	Centralizuotas šalto vandens tiekimas	0.993
Centralizuotas šalto vandens tiekimas	0.314	Regionas	0.888
Motinos gimimo šalis	0.0092	Būsto tipas	1
Tėvo gimimo šalis	0.0133	Pagrindinis būsto šildymo būdas	1
Šiuo metu mokosi	1	Šiuo metu mokosi	0.000241

--- Bustas, vanduo, elektra, dujos ir kitas kuras	
variable	p_value
Būsto valdos statusas	3.1e-14
Namų ūkio užimamas naudingasis plotas	0.244
Pajamų šaltinis	0.729
Asmenų skaičius namų ūkyje	1
Pagrindinis darbas: profesija	0.835
Socialinė ekonominė grupė	0.87
Pajamų kvintilinė grupė	0.394
Gyvena su partneriu	0.0903
Išlaikomų vaikų skaičius	0.561
Darbo sutarties tipas	0.836
Šiuo metu mokosi	0.659
Namų ūkio tipas	0.599
Būsto tipas	1
Pagrindinis būsto šildymo būdas	4.02e-06
Teisinė santuokinė padėtis	0.316
Centralizuotas karšto vandens tiekimas	0.000178
Namų ūkio galvos išsilavinimas	0.418
Lytis	0.29
Pagrindinės pilietybės šalis	0.00759
Regionas	0.00186
Amžius	0.999
Gyvenamoji vieta	0.0197
Gimimo šalis	2.42e-05
Motinos gimimo šalis	0.000147
Centralizuotas šalto vandens tiekimas	0.928
Tėvo gimimo šalis	2.45e-08

--- Būsto apstatymo, namų ūkio įranga ir kasdienė	
variable	p_value
Šiuo metu mokosi	9.64e-19
Amžius	1
Namų ūkio tipas	2.43e-29
Darbo sutarties tipas	2.75e-38
Pagrindinis darbas: profesija	9.89e-66
Išlaikomų vaikų skaičius	6.7e-67
Regionas	0.745
Asmenų skaičius namų ūkyje	2.24e-218
Būsto valdos statusas	1.06e-05
Pajamų šaltinis	0.276
Socialinė ekonominė grupė	0.0554
Gyvena su partneriu	2e-05
Pajamų kvintilinė grupė	0.0362
Namų ūkio užimamas naudingasis plotas	0.0485
Teisinė santuokinė padėtis	1
Gyvenamoji vieta	0.196
Namų ūkio galvos išsilavinimas	1
Centralizuotas šalto vandens tiekimas	0.972
Lytis	0.398
Pagrindinis būsto šildymo būdas	0.945
Būsto tipas	1
Centralizuotas karšto vandens tiekimas	0.218
Gimimo šalis	0.941
Motinos gimimo šalis	9.01e-06
Pagrindinės pilietybės šalis	0.223
Tėvo gimimo šalis	0.0519

--- Sveikata ---	
variable	p_value
Gyvena su partneriu	0.463
Pajamų kvintilinė grupė	0.237
Gyvenamoji vieta	0.0227
Namų ūkio tipas	0.291
Pagrindinis darbas: profesija	1
Teisinė santuokinė padėtis	0.729
Centralizuotas šalto vandens tiekimas	1
Namų ūkio galvos išsilavinimas	0.000311
Asmenų skaičius namų ūkyje	1
Pagrindinis būsto šildymo būdas	1
Būsto tipas	0.999
Pajamų šaltinis	1
Išlaikomų vaikų skaičius	0.996
Socialinė ekonominė grupė	0.993
Regionas	0.981
Amžius	1
Centralizuotas karšto vandens tiekimas	0.212
Būsto valdos statusas	1
Gimimo šalis	2.73e-06
Motinos gimimo šalis	2.87e-38
Namų ūkio užimamas naudingasis plotas	0.508
Pagrindinės pilietybės šalis	0.888
Lytis	0.238
Tėvo gimimo šalis	6.05e-39
Šiuo metu mokosi	1
Darbo sutarties tipas	0.000618

--- Transportas ---	
variable	p_value
Pagrindinis darbas: profesija	5.42e-150
Darbo sutarties tipas	4.33e-77
Gyvena su partneriu	6.59e-34
Asmenų skaičius namų ūkyje	6.02e-129
Pajamų šaltinis	2.89e-21
Socialinė ekonominė grupė	3.34e-18
Namų ūkio tipas	1.36e-40
Pajamų kvintilinė grupė	0.00191
Būsto valdos statusas	2.75e-40
Išlaikomų vaikų skaičius	2.57e-107
Lytis	2.13e-06
Namų ūkio užimamas naudingasis plotas	0.0431
Teisinė santuokinė padėtis	0.999
Būsto tipas	1
Namų ūkio galvos išsilavinimas	1
Gyvenamoji vieta	0.185
Centralizuotas šalto vandens tiekimas	0.0582
Gimimo šalis	1
Regionas	0.962
Amžius	6.61e-05
Centralizuotas karšto vandens tiekimas	0.141
Pagrindinis būsto šildymo būdas	9.08e-05
Motinos gimimo šalis	0.327
Tėvo gimimo šalis	0.491
Pagrindinės pilietybės šalis	0.00179
Šiuo metu mokosi	7.45e-48

--- Informacija ir ryšiai ---

variable	p_value
Amžius	0.976
Pagrindinis darbas: profesija	6.01e-36
Išlaikomų vaikų skaičius	4.12e-63
Namų ūkio tipas	2.69e-54
Pajamų šaltinis	2.35e-09
Socialinė ekonominė grupė	1.8e-14
Asmenų skaičius namų ūkyje	4.39e-142
Darbo sutarties tipas	4.94e-39
Gyvena su partneriu	1.95e-13
Būsto valdos statusas	6.76e-29
Pajamų kvintilinė grupė	2.39e-05
Namų ūkio galvos išsilavinimas	0.497
Teisinė santuokinė padėtis	9.66e-08
Namų ūkio užimamas naudingasis plotas	0.0155
Lytis	0.0059
Šiuo metu mokosi	1.17e-27
Pagrindinis būsto šildymo būdas	0.928
Gyvenamoji vieta	0.00791
Gimimo šalis	0.000355
Centralizuotas karšto vandens tiekimas	0.000879
Būsto tipas	1
Regionas	0.231
Motinos gimimo šalis	2.73e-05
Centralizuotas šalto vandens tiekimas	1
Pagrindinės pilietybės šalis	1.02e-16
Tėvo gimimo šalis	1.05e-20

--- Švietimo paslaugos ---

variable	p_value
Amžius	3.03e-38
Išlaikomų vaikų skaičius	3.63e-235
Asmenų skaičius namų ūkyje	5.21e-229
Namų ūkio tipas	1.09e-89
Pagrindinis darbas: profesija	3.29e-51
Pajamų šaltinis	1.57e-08
Šiuo metu mokosi	5.28e-104
Darbo sutarties tipas	1.93e-17
Socialinė ekonominė grupė	2.05e-08
Būsto valdos statusas	5.03e-83
Gyvena su partneriu	4.71e-10
Būsto tipas	0.245
Pajamų kvintilinė grupė	0.0386
Lytis	0.222
Namų ūkio užimamas naudingasis plotas	0.0207
Gyvenamoji vieta	0.542
Centralizuotas šalto vandens tiekimas	1
Regionas	0.0297
Namų ūkio galvos išsilavinimas	0.999
Motinos gimimo šalis	0.0641
Centralizuotas karšto vandens tiekimas	0.211
Pagrindinis būsto šildymo būdas	1
Gimimo šalis	0.994
Teisinė santuokinė padėtis	0.999
Pagrindinės pilietybės šalis	4.25e-09
Tėvo gimimo šalis	0.996

--- Restoranai ir apgyvendinimo paslaugos ---

variable	p_value
Amžius	0
Darbo sutarties tipas	5.03e-77
Asmenų skaičius namų ūkyje	0
Pagrindinis darbas: profesija	5.18e-178
Namų ūkio tipas	1.88e-126
Išlaikomų vaikų skaičius	8.57e-260
Socialinė ekonominė grupė	1.83e-28
Pajamų šaltinis	1.03e-31
Pajamų kvintilinė grupė	3.34e-07
Gyvena su partneriu	1.66e-09
Gyvenamoji vieta	0.182
Namų ūkio galvos išsilavinimas	1
Būsto valdos statusas	5.8e-171
Teisinė santuokinė padėtis	0.77
Lytis	0.035
Regionas	0.0117
Namų ūkio užimamas naudingasis plotas	0.184
Būsto tipas	1
Pagrindinės pilietybės šalis	2.48e-27
Pagrindinis būsto šildymo būdas	4.2e-14
Gimimo šalis	1
Centralizuotas karšto vandens tiekimas	0.234
Motinos gimimo šalis	2.24e-26
Centralizuotas šalto vandens tiekimas	0.916
Tėvo gimimo šalis	2.81e-10
Šiuo metu mokosi	2e-93

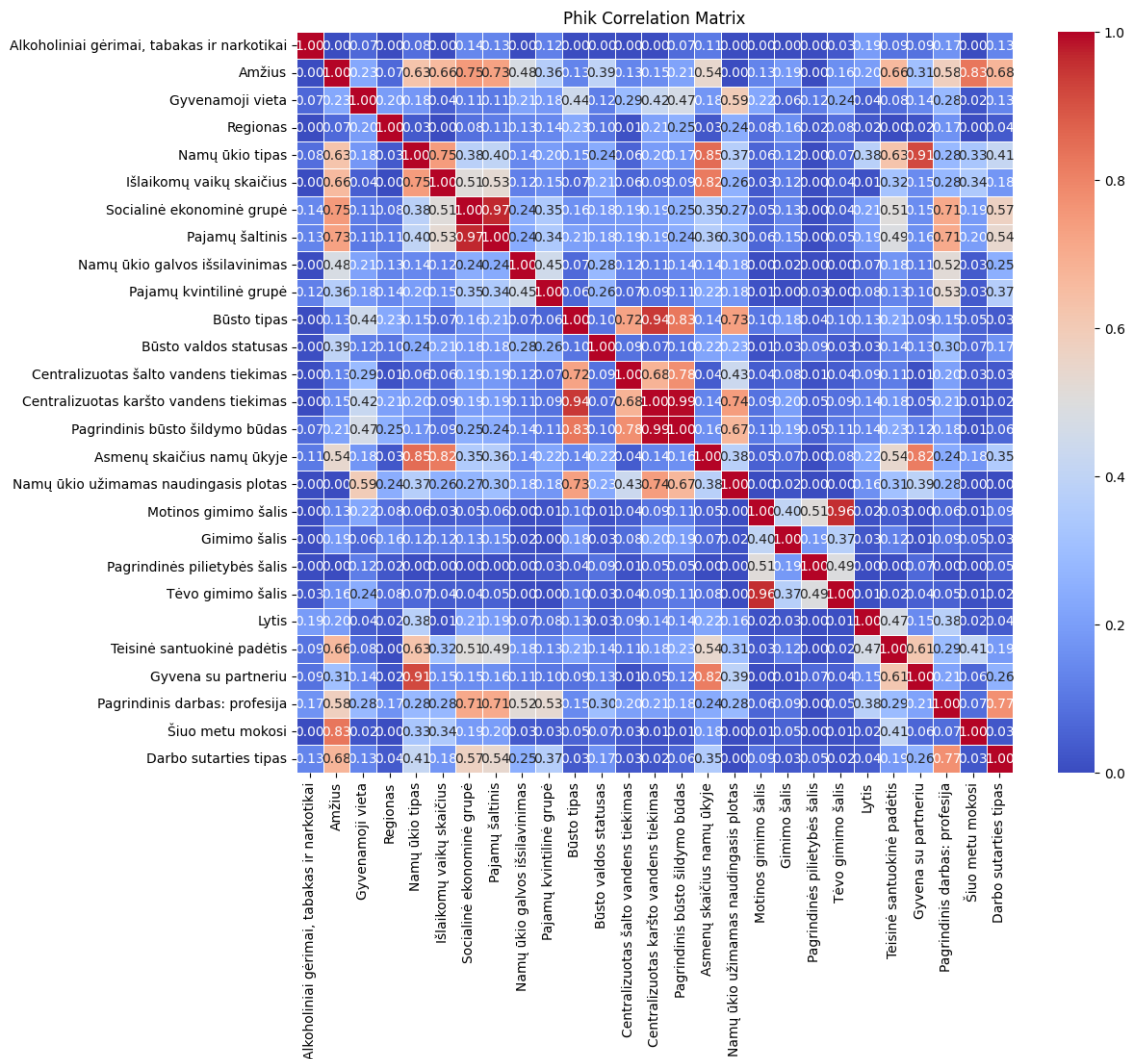
--- Draudimas ir finansinės paslaugos ---

variable	p_value
Pagrindinis darbas: profesija	6.77e-111
Namų ūkio užimamas naudingasis plotas	7.2e-08
Darbo sutarties tipas	2.04e-106
Socialinė ekonominė grupė	3.41e-13
Pajamų šaltinis	1.55e-19
Pajamų kvintilinė grupė	5.1e-06
Gyvena su partneriu	1.04e-27
Amžius	1
Namų ūkio tipas	2.41e-40
Asmenų skaičius namų ūkyje	8.08e-256
Namų ūkio galvos išsilavinimas	1
Išlaikomų vaikų skaičius	1.43e-50
Būsto valdos statusas	7.87e-07
Teisinė santuokinė padėtis	0.305
Lytis	5.41e-05
Būsto tipas	1
Pagrindinis būsto šildymo būdas	0.0113
Centralizuotas karšto vandens tiekimas	0.0116
Gyvenamoji vieta	0.0384
Šiuo metu mokosi	1.29e-31
Regionas	4.51e-08
Pagrindinės pilietybės šalis	0.983
Centralizuotas šalto vandens tiekimas	0.721
Motinos gimimo šalis	1
Gimimo šalis	0.999
Tėvo gimimo šalis	0.812

--- Asmens priežiūra, socialinė apsauga ir įvairio	
variable	p_value
Darbo sutarties tipas	3.81e-52
Amžius	0.708
Pagrindinis darbas: profesija	3.75e-107
Asmenų skaičius namų ūkyje	2.5e-208
Pajamų šaltinis	0.000343
Socialinė ekonominė grupė	0.00432
Išlaikomų vaikų skaičius	1.3e-96
Namų ūkio tipas	6.99e-42
Pajamų kvintilinė grupė	0.00344
Gyvena su partneriu	2.14e-07
Būsto valdos statusas	6.48e-52
Teisinė santuokinė padėtis	1
Namų ūkio galvos išsilavinimas	1
Gyvenamoji vieta	0.225
Pagrindinis būsto šildymo būdas	1
Centralizuotas karšto vandens tiekimas	0.272
Pagrindinės pilietybės šalis	6.24e-23
Būsto tipas	1
Namų ūkio užimamas naudingasis plotas	0.207
Centralizuotas šalto vandens tiekimas	1
Regionas	0.9
Lytis	0.275
Gimimo šalis	0.585
Motinos gimimo šalis	3.14e-17
Tėvo gimimo šalis	0.00254
Šiuo metu mokosi	0.0327

--- Poilsis, sportas ir kultūra ---	
variable	p_value
Išlaikomų vaikų skaičius	6.9e-92
Darbo sutarties tipas	1.38e-58
Socialinė ekonominė grupė	0.0889
Pajamų šaltinis	0.00622
Pagrindinis darbas: profesija	2.38e-70
Amžius	1
Namų ūkio tipas	9.59e-38
Asmenų skaičius namų ūkyje	0
Gyvena su partneriu	4.97e-07
Pajamų kvintilinė grupė	0.00676
Būsto valdos statusas	1.35e-32
Teisinė santuokinė padėtis	1
Namų ūkio užimamas naudingasis plotas	0.00604
Namų ūkio galvos išsilavinimas	1
Gyvenamoji vieta	0.376
Lytis	0.224
Pagrindinis būsto šildymo būdas	6.52e-14
Pagrindinės pilietybės šalis	6.18e-26
Regionas	0.86
Centralizuotas karšto vandens tiekimas	0.196
Būsto tipas	1
Centralizuotas šalto vandens tiekimas	0.477
Gimimo šalis	0.993
Motinos gimimo šalis	1
Tėvo gimimo šalis	1
Šiuo metu mokosi	3.52e-08

9 Priedas. Koreliacijos koeficientų (ϕ_k) matrica tarp nepriklausomų kintamųjų



10 Priedas. Normalumo tikrinimo testų rezultatai, kai priklausomas kintamasis yra bendras vartojimo išlaidas

Gyvenamoji vieta = 5 didieji miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gyvenamoji vieta = Kiti miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gyvenamoji vieta = Kaimas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Regionas = Sostinės regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Regionas = Vidurio ir vakarų Lietuvos regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Vienas gyvenantis asmuo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Kiti namų ūkiai su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Pora be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Pora su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Vienas suaugęs asmuo su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio tipas = Kiti namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Išlaikomų vaikų skaičius = Namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Išlaikomų vaikų skaičius = Namų ūkiai su 1 išlaikomu vaiku: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Išlaikomų vaikų skaičius = Namų ūkiai su 2 išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Išlaikomų vaikų skaičius = Namų ūkiai su 3 ir daugiau išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Socialinė ekonominė grupė = Senatvės pensininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Socialinė ekonominė grupė = Samdomieji darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Socialinė ekonominė grupė = Bedarbis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Socialinė ekonominė grupė = Savarankiškai dirbantys asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų šaltinis = Pensija: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų šaltinis = Darbo užmokestis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų šaltinis = Pašalpa, stipendija ir kitas pragyvenimo šaltinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų šaltinis = Pajamos iš savo ar šeimos verslo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos amžius = 60 metų ir vyresni: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos amžius = 50-59 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos amžius = 40-49 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos amžius = iki 30 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos amžius = 30-39 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos išsilavinimas = Vidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos išsilavinimas = Aukštesniojo ir aukštojo nesuteikiantis pavidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos išsilavinimas = Aukštesnysis ir aukštasis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos išsilavinimas = Pradinis, nebaigę pradinės mokyklos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio galvos išsilavinimas = Pagrindinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų kvintilinė grupė = 2-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų kvintilinė grupė = 1-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų kvintilinė grupė = 5-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų kvintilinė grupė = 4-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pajamų kvintilinė grupė = 3-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu

Būsto tipas = Butas daugiabučiame name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto tipas = Vieno buto namas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto tipas = Butas dviejų butų ar sublokuotame name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto valdos statusas = Būsto savininkas, neturintis išmokėti paskolos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto valdos statusas = Būsto savininkas, turintis neišmokėtą paskolą: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto valdos statusas = Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: Shapiro-Wilk p = 0.0089 ✗ Nenormalu
 Būsto valdos statusas = Gyvena nemokėdamas nuomos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Būsto valdos statusas = Nuomojasi šį būstą mažesne nei rinkos kaina: Shapiro-Wilk p = 0.0001 ✗ Nenormalu
 Centralizuotas šalto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Centralizuotas šalto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Centralizuotas karšto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Centralizuotas karšto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis būsto šildymo būdas = Centrinis šildymas iš šilumos tinklu: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis būsto šildymo būdas = Centrinis šildymas iš vietinio šilumos šaltinio: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis būsto šildymo būdas = Nėra stacionaraus šilumos šaltinio: Shapiro-Wilk p = 0.6184 ✓ Normalu
 Pagrindinis būsto šildymo būdas = Krosnis, židynys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Asmenų skaičius namų ūkyje = 1-2 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Asmenų skaičius namų ūkyje = 3-4 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Asmenų skaičius namų ūkyje = 5 ir daugiau: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio populiacijos svoris = 201-400: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio populiacijos svoris = >600: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio populiacijos svoris = 401-600: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio populiacijos svoris = 101-200: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio populiacijos svoris = <100: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio užimamas naudingasis plotas = <50 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio užimamas naudingasis plotas = 50-99 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio užimamas naudingasis plotas = 100-199 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Namų ūkio užimamas naudingasis plotas = >200 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Motinos gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Motinos gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Motinos gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.3563 ✓ Normalu
 Gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinės pilietybės šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinės pilietybės šalis = Kita šalis: Shapiro-Wilk p = 0.0226 ✗ Nenormalu
 Tėvo gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Tėvo gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Tėvo gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0160 ✗ Nenormalu
 Lytis = Moteris: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Lytis = Vyras: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Teisinė santuokinė padėtis = Išsituokęs (-usi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Teisinė santuokinė padėtis = Našlys (-ė): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Teisinė santuokinė padėtis = Nevedęs (netekėjusi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Teisinė santuokinė padėtis = Vedęs (išteklėjusi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gyvena su partneriu = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gyvena su partneriu = Taip, teisiniu pagrindu: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Gyvena su partneriu = Taip, be teisinio pagrindo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Netaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Specialistai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Įstaigų tarnautojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Technikai ir jaunesnieji specialistai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Nenurodė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Vadovai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Kvalifikuoti darbininkai ir amatininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Paslaugų sektoriaus darbuotojai ir pardavėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Įrenginių ir mašinų operatoriai ir surinkėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Pagrindinis darbas: profesija = Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Šiuo metu mokosi = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Šiuo metu mokosi = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Darbo sutarties tipas = Netaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Darbo sutarties tipas = Neterminuota sutartis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
 Darbo sutarties tipas = Terminuota sutartis: Shapiro-Wilk p = 0.1452 ✓ Normalu

11 Priedas. Daugkartiniai palyginimai naudojant medianos testą, kai priklausomas kintamasis yra bendras vartojimo išlaidas

Tikrinamas kintamasis: Gyvenamoji vieta
5 didieji miestai vs Kiti miestai: chi tatistika = 10.9061, p-reikšmė = 0.0010, pakoreguota p-reikšmė = 0.0016

5 didieji miestai vs Kaimas: chi tatistika = 0.3509, p-reikšmė = 0.5536, pakoreguota p-reikšmė = 0.6004

Kiti miestai vs Kaimas: chi tatistika = 9.1055, p-reikšmė = 0.0025, pakoreguota p-reikšmė = 0.0042

Tikrinamas kintamasis: Regionas
Sostinės regionas vs Vidurio ir vakarų Lietuvos regionas: chi tatistika = 1.1675, p-reikšmė = 0.2799, pakoreguota p-reikšmė = 0.3269

Tikrinamas kintamasis: Namų ūkio tipas
Vienas gyvenantis asmuo vs Kiti namų ūkiai su išlaikomais vaikais: chi tatistika = 120.0434, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo vs Pora be išlaikomų vaikų: chi tatistika = 361.7469, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo vs Pora su išlaikomais vaikais: chi tatistika = 456.9172, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 84.8019, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 144.4669, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Kiti namų ūkiai su išlaikomais vaikais vs Pora be išlaikomų vaikų: chi tatistika = 30.1723, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Kiti namų ūkiai su išlaikomais vaikais vs Pora su išlaikomais vaikais: chi tatistika = 6.5065, p-reikšmė = 0.0103, pakoreguota p-reikšmė = 0.0157

Kiti namų ūkiai su išlaikomais vaikais vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 7.2730, p-reikšmė = 0.0070, pakoreguota p-reikšmė = 0.0110

Kiti namų ūkiai su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 18.2466, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora be išlaikomų vaikų vs Pora su išlaikomais vaikais: chi tatistika = 163.3926, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora be išlaikomų vaikų vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 5.4019, p-reikšmė = 0.0201, pakoreguota p-reikšmė = 0.0297

Pora be išlaikomų vaikų vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 0.0028, p-reikšmė = 0.9578, pakoreguota p-reikšmė = 0.9753

Pora su išlaikomais vaikais vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 36.3074, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 143.1709, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas suaugęs asmuo su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 4.0299, p-reikšmė = 0.0447, pakoreguota p-reikšmė = 0.0592

Tikrinamas kintamasis: Išlaikomų vaikų skaičius
Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 1 išlaikomu vaiku: chi tatistika = 151.8506, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 204.2125, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 79.4438, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai su 1 išlaikomu vaiku vs Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 18.2869, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai su 1 išlaikomu vaiku vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 23.0131, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai su 2 išlaikomais vaikais vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 5.0846, p-reikšmė = 0.0241, pakoreguota p-reikšmė = 0.0348

Tikrinamas kintamasis: Socialinė ekonominė grupė
Senatvės pensininkai vs Samdomieji darbuotojai: chi tatistika = 644.6525, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Senatvės pensininkai vs Bedarbiai: chi tatistika = 5.2260, p-reikšmė = 0.0223, pakoreguota p-reikšmė = 0.0325

Senatvės pensininkai vs Savarankiškai dirbantys asmenys: chi tatistika = 197.5072, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Samdomieji darbuotojai vs Bedarbiai: chi tatistika = 97.9752, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Samdomieji darbuotojai vs Savarankiškai dirbantys asmenys: chi tatistika = 9.9882, p-reikšmė = 0.0016, pakoreguota p-reikšmė = 0.0027

Bedarbiai vs Savarankiškai dirbantys asmenys: chi tatistika = 128.5127, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Pajamų šaltinis
Pensija vs Darbo užmokestis: chi tatistika = 681.0441, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pensija vs Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 1.9882, p-reikšmė = 0.1585, pakoreguota p-reikšmė = 0.2010

Pensija vs Pajamos iš savo ar šeimos verslo: chi tatistika = 134.0961, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Darbo užmokestis vs Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 57.6343, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Darbo užmokestis vs Pajamos iš savo ar šeimos verslo: chi tatistika = 2.7225, p-reikšmė = 0.0989, pakoreguota p-reikšmė = 0.1271

Pašalpa, stipendija ir kitas pragyvenimo šaltinis vs Pajamos iš savo ar šeimos verslo: chi tatistika = 74.1306, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Namų ūkio galvos amžius
60 metų ir vyresni vs 50-59 metų: chi statistika = 189.2694, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni vs 40-49 metų: chi statistika = 261.9445, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni vs iki 30 metų: chi statistika = 44.0553, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni vs 30-39 metų: chi statistika = 195.1724, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-59 metų vs 40-49 metų: chi statistika = 16.2284, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0001

50-59 metų vs iki 30 metų: chi statistika = 0.5167, p-reikšmė = 0.4723, pakoreguota p-reikšmė = 0.5223

50-59 metų vs 30-39 metų: chi statistika = 27.6706, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

40-49 metų vs iki 30 metų: chi statistika = 14.3327, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0003

40-49 metų vs 30-39 metų: chi statistika = 0.2337, p-reikšmė = 0.6288, pakoreguota p-reikšmė = 0.6774

iki 30 metų vs 30-39 metų: chi statistika = 15.2926, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002

Tikrinamas kintamasis: Namų ūkio galvos išsilavinimas
Vidurinis vs Aukštesniojo ir aukštojo nesuteikiantis povidurinis: chi statistika = 0.0072, p-reikšmė = 0.9323, pakoreguota p-reikšmė = 0.9552

Vidurinis vs Aukštesnysis ir aukštasis: chi statistika = 70.3146, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis vs Pradinis, nebaigę pradinės mokyklos: chi statistika = 44.9046, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis vs Pagrindinis: chi statistika = 38.5103, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Aukštesnysis ir aukštasis: chi statistika = 98.6359, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Pradinis, nebaigę pradinės mokyklos: chi statistika = 46.1371, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Pagrindinis: chi statistika = 41.8356, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis vs Pradinis, nebaigę pradinės mokyklos: chi statistika = 83.9984, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis vs Pagrindinis: chi statistika = 143.1453, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pradinis, nebaigę pradinės mokyklos vs Pagrindinis: chi statistika = 8.5024, p-reikšmė = 0.0034, pakoreguota p-reikšmė = 0.0055

Tikrinamas kintamasis: Būsto valdos statusas
Būsto savininkas, neturintis išmokėti paskolos vs Būsto savininkas, turintis neišmokėtą paskolą: chi statistika = 107.5555, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, neturintis išmokėti paskolos vs Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi statistika = 4.4075, p-reikšmė = 0.0358, pakoreguota p-reikšmė = 0.0485

Būsto savininkas, neturintis išmokėti paskolos vs Gyvena nemokėdamas nuomos: chi statistika = 0.7074, p-reikšmė = 0.4003, pakoreguota p-reikšmė = 0.4579

Būsto savininkas, neturintis išmokėti paskolos vs Nuomojasi šį būstą mažesne nei rinkos kaina: chi statistika = 4.2254, p-reikšmė = 0.0398, pakoreguota p-reikšmė = 0.0532

Būsto savininkas, turintis neišmokėtą paskolą vs Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi statistika = 0.1219, p-reikšmė = 0.7270, pakoreguota p-reikšmė = 0.7733

Būsto savininkas, turintis neišmokėtą paskolą vs Gyvena nemokėdamas nuomos: chi statistika = 76.4777, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, turintis neišmokėtą paskolą vs Nuomojasi šį būstą mažesne nei rinkos kaina: chi statistika = 31.9975, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina vs Gyvena nemokėdamas nuomos: chi statistika = 5.0397, p-reikšmė = 0.0248, pakoreguota p-reikšmė = 0.0354

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina vs Nuomojasi šį būstą mažesne nei rinkos kaina: chi statistika = 7.4828, p-reikšmė = 0.0062, pakoreguota p-reikšmė = 0.0099

Gyvena nemokėdamas nuomos vs Nuomojasi šį būstą mažesne nei rinkos kaina: chi statistika = 0.5343, p-reikšmė = 0.4648, pakoreguota p-reikšmė = 0.5175

Tikrinamas kintamasis: Centralizuotas šalto vandens tiekimas
Taip vs Ne: chi statistika = 0.3660, p-reikšmė = 0.5452, pakoreguota p-reikšmė = 0.5953

Tikrinamas kintamasis: Centralizuotas karšto vandens tiekimas
Taip vs Ne: chi statistika = 22.8071, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Pagrindinis būsto šildymo būdas
Centrinis šildymas iš šilumos tinklu vs Centrinis šildymas iš vietinio šilumos šaltinio: chi statistika = 63.0683, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Centrinis šildymas iš šilumos tinklu vs Nėra stacionaraus šilumos šaltinio: chi statistika = 1.5787, p-reikšmė = 0.2089, pakoreguota p-reikšmė = 0.2566

Centrinis šildymas iš šilumos tinklu vs Krosnis, židiny: chi statistika = 37.7024, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Centrinis šildymas iš vietinio šilumos šaltinio vs Nėra stacionaraus šilumos šaltinio: chi statistika = 0.0632, p-reikšmė = 0.8016, pakoreguota p-reikšmė = 0.8472

Centrinis šildymas iš vietinio šilumos šaltinio vs Krosnis, židiny: chi statistika = 96.2111, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nėra stacionaraus šilumos šaltinio vs Krosnis, židiny: chi statistika = 5.2185, p-reikšmė = 0.0223, pakoreguota p-reikšmė = 0.0325

Tikrinamas kintamasis: Asmenų skaičius namų ūkyje
 1-2 asmenys vs 3-4 asmenys: chi tatistika = 407.2261, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 1-2 asmenys vs 5 ir daugiau: chi tatistika = 129.5267, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 3-4 asmenys vs 5 ir daugiau: chi tatistika = 9.2881, p-reikšmė = 0.0023, pakoreguota p-reikšmė = 0.0038

Tikrinamas kintamasis: Namų ūkio užimamas naudingasis plotas
 <50 m² vs 50-99 m²: chi tatistika = 91.7336, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 <50 m² vs 100-199 m²: chi tatistika = 188.6422, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 <50 m² vs >200 m²: chi tatistika = 18.7707, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 50-99 m² vs 100-199 m²: chi tatistika = 67.3197, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 50-99 m² vs >200 m²: chi tatistika = 11.2301, p-reikšmė = 0.0008, pakoreguota p-reikšmė = 0.0014

 100-199 m² vs >200 m²: chi tatistika = 0.6746, p-reikšmė = 0.4115, pakoreguota p-reikšmė = 0.4674

Tikrinamas kintamasis: Motinos gimimo šalis
 Lietuva vs Kita šalis: chi tatistika = 1.2095, p-reikšmė = 0.2714, pakoreguota p-reikšmė = 0.3192

 Lietuva vs Nenurodė: chi tatistika = 2.9175, p-reikšmė = 0.0876, pakoreguota p-reikšmė = 0.1134

 Kita šalis vs Nenurodė: chi tatistika = 4.8200, p-reikšmė = 0.0281, pakoreguota p-reikšmė = 0.0391

Tikrinamas kintamasis: Gimimo šalis
 Lietuva vs Kita šalis: chi tatistika = 4.9072, p-reikšmė = 0.0267, pakoreguota p-reikšmė = 0.0375

Tikrinamas kintamasis: Pagrindinės pilietybės šalis
 Lietuva vs Kita šalis: chi tatistika = 8.0950, p-reikšmė = 0.0044, pakoreguota p-reikšmė = 0.0071

Tikrinamas kintamasis: Tėvo gimimo šalis
 Lietuva vs Kita šalis: chi tatistika = 1.9850, p-reikšmė = 0.1589, pakoreguota p-reikšmė = 0.2010

 Lietuva vs Nenurodė: chi tatistika = 0.6040, p-reikšmė = 0.4370, pakoreguota p-reikšmė = 0.4918

 Kita šalis vs Nenurodė: chi tatistika = 1.3683, p-reikšmė = 0.2421, pakoreguota p-reikšmė = 0.2930

Tikrinamas kintamasis: Lytis
 Moteris vs Vyras: chi tatistika = 175.0753, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Teisinė santuokinė padėtis
 Išsituokęs (-usi) vs Našlys (-ė): chi tatistika = 37.1902, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Išsituokęs (-usi) vs Nevedęs (netekėjusi): chi tatistika = 16.9990, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0001

 Išsituokęs (-usi) vs Vedęs (ištekėjusi): chi tatistika = 191.9892, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Našlys (-ė) vs Nevedęs (netekėjusi): chi tatistika = 65.3327, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Našlys (-ė) vs Vedęs (ištekėjusi): chi tatistika = 454.2086, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Nevedęs (netekėjusi) vs Vedęs (ištekėjusi): chi tatistika = 69.0989, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Gyvena su partneriu
 Ne vs Taip, teisiniu pagrindu: chi tatistika = 538.1910, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Ne vs Taip, be teisinio pagrindo: chi tatistika = 65.3247, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Taip, teisiniu pagrindu vs Taip, be teisinio pagrindo: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Tikrinamas kintamasis: Pagrindinis darbas: profesija
Netaikoma vs Specialistai: chi tatistika = 322.3998, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Įstaigų tarnautojai: chi tatistika = 46.1112, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Technikai ir jaunesnieji specialistai: chi tatistika = 89.7837, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Nenurodė: chi tatistika = 47.2887, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Vadovai: chi tatistika = 151.0726, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 201.1669, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 62.2170, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 157.9267, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 37.9589, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai vs Įstaigų tarnautojai: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Specialistai vs Technikai ir jaunesnieji specialistai: chi tatistika = 0.1643, p-reikšmė = 0.6852, pakoreguota p-reikšmė = 0.7336

Specialistai vs Nenurodė: chi tatistika = 43.3106, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai vs Vadovai: chi tatistika = 15.5216, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002

Specialistai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 1.0205, p-reikšmė = 0.3124, pakoreguota p-reikšmė = 0.3598

Specialistai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 26.4098, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 4.6210, p-reikšmė = 0.0316, pakoreguota p-reikšmė = 0.0432

Specialistai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.5995, p-reikšmė = 0.4388, pakoreguota p-reikšmė = 0.4918

Įstaigų tarnautojai vs Technikai ir jaunesnieji specialistai: chi tatistika = 0.0073, p-reikšmė = 0.9319, pakoreguota p-reikšmė = 0.9552

Įstaigų tarnautojai vs Nenurodė: chi tatistika = 9.7215, p-reikšmė = 0.0018, pakoreguota p-reikšmė = 0.0030

Įstaigų tarnautojai vs Vadovai: chi tatistika = 4.9457, p-reikšmė = 0.0262, pakoreguota p-reikšmė = 0.0370

Įstaigų tarnautojai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 0.3656, p-reikšmė = 0.5454, pakoreguota p-reikšmė = 0.5953

Įstaigų tarnautojai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 7.0427, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0124

Įstaigų tarnautojai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 1.8853, p-reikšmė = 0.1697, pakoreguota p-reikšmė = 0.2131

Įstaigų tarnautojai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0223, p-reikšmė = 0.8812, pakoreguota p-reikšmė = 0.9141

Technikai ir jaunesnieji specialistai vs Nenurodė: chi tatistika = 28.2874, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Technikai ir jaunesnieji specialistai vs Vadovai: chi tatistika = 5.9330, p-reikšmė = 0.0149, pakoreguota p-reikšmė = 0.0222

Technikai ir jaunesnieji specialistai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 1.5400, p-reikšmė = 0.2146, pakoreguota p-reikšmė = 0.2616

Technikai ir jaunesnieji specialistai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 18.8609, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Technikai ir jaunesnieji specialistai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 4.6264, p-reikšmė = 0.0315, pakoreguota p-reikšmė = 0.0432

Technikai ir jaunesnieji specialistai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0361, p-reikšmė = 0.8493, pakoreguota p-reikšmė = 0.8864

Nenurodė vs Vadovai: chi tatistika = 70.1640, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nenurodė vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 26.9007, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nenurodė vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 1.7358, p-reikšmė = 0.1877, pakoreguota p-reikšmė = 0.2322

Nenurodė vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 11.1430, p-reikšmė = 0.0008, pakoreguota p-reikšmė = 0.0015

Nenurodė vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 19.2568, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vadovai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 16.0496, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0001

Vadovai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 49.4178, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vadovai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 23.1094, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vadovai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 3.5156, p-reikšmė = 0.0608, pakoreguota p-reikšmė = 0.0799

Kvalifikuoti darbininkai ir amatininkai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 11.4013, p-reikšmė = 0.0007, pakoreguota p-reikšmė = 0.0013

Kvalifikuoti darbininkai ir amatininkai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 1.8478, p-reikšmė = 0.1740, pakoreguota p-reikšmė = 0.2169

Kvalifikuoti darbininkai ir amatininkai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 1.0639, p-reikšmė = 0.3023, pakoreguota p-reikšmė = 0.3506

Paslaugų sektoriaus darbuotojai ir pardavėjai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 6.5106, p-reikšmė = 0.0107, pakoreguota p-reikšmė = 0.0163

Paslaugų sektoriaus darbuotojai ir pardavėjai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 15.7706, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0001

Įrenginių ir mašinų operatoriai ir surinkėjai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 3.1710, p-reikšmė = 0.0750, pakoreguota p-reikšmė = 0.0978

Tikrinamas kintamasis: Šiuo metu mokosi
Ne vs Taip: chi tatistika = 4.3971, p-reikšmė = 0.0360, pakoreguota p-reikšmė = 0.0485

Tikrinamas kintamasis: Darbo sutarties tipas
Netaikoma vs Neterminuota sutartis: chi tatistika = 501.3645, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Terminuota sutartis: chi tatistika = 13.0595, p-reikšmė = 0.0003, pakoreguota p-reikšmė = 0.0005

Neterminuota sutartis vs Terminuota sutartis: chi tatistika = 6.1378, p-reikšmė = 0.0132, pakoreguota p-reikšmė = 0.0199

12 Priedas. Šapiro ir Vilko testo rezultatai, kai priklausomas kintamasis – maistas ir nealkoholiniai gėrimai

numeric_column = "Maistas ir nealkoholiniai gėrimai"

Gyvenamoji vieta = 5 didieji miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvenamoji vieta = Kiti miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvenamoji vieta = Kaimas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Regionas = Sostinės regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Regionas = Vidurio ir vakarų Lietuvos regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Vienas gyvenantis asmuo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Kiti namų ūkiai su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Pora be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Pora su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Vienas suaugęs asmuo su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Kiti namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 1 išlaikomu vaiku: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 2 išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 3 ir daugiau išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Senatvės pensininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Samdomieji darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Bedarbiai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Savarankiškai dirbantys asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pensija: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Darbo užmokestis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pašalpa, stipendija ir kitas pragyvenimo šaltinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pajamos iš savo ar šeimos verslo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 60 metų ir vyresni: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 50-59 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 40-49 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = iki 30 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 30-39 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Vidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Aukštesniojo ir aukštojo nesuteikiantis pavidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Aukštesnysis ir aukštasis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Pradinis, nebaigę pradinės mokyklos: Shapiro-Wilk p = 0.0001 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Pagrindinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 2-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 1-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 5-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 4-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 3-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = Butas daugiabučiame name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = Vieno buto namas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = Butas dviejų butų ar sublokuotame name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Būsto savininkas, neturintis išmokėti paskolos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Būsto savininkas, turintis neišmokėtą paskolą: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: Shapiro-Wilk p = 0.0011 ✗ Nenormalu
Būsto valdos statusas = Gyvena nemokėdamas nuomos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Nuomojasi iš būstą mažesne nei rinkos kaina: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas šalto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas šalto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas karšto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas karšto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš šilumos tinklu: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš vietinio šilumos šaltinio: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Nėra stacionaraus šilumos šaltinio: Shapiro-Wilk p = 0.3032 ✓ Normalu
Pagrindinis būsto šildymo būdas = Krosnis, židiny: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 1-2 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 3-4 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 5 ir daugiau: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = <50 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = 50-99 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = 100-199 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = >200 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.7564 ✓ Normalu
Gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinės pilietybės šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinės pilietybės šalis = Kita šalis: Shapiro-Wilk p = 0.0001 ✗ Nenormalu
Tėvo gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Tėvo gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Tėvo gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0002 ✓ Normalu
Lytis = Moteris: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Lytis = Vyras: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Išsituokęs (-usi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Našlys (-ė): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Nevedęs (netekėjusi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Vedęs (ištėkėjusi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Taip, teisiniu pagrindu: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Taip, be teisinio pagrindo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Metaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Specialistai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Įstaigų tarnautojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Technikai ir jaunesnieji specialistai: Shapiro-Wilk p = 0.0008 ✗ Nenormalu
Pagrindinis darbas: profesija = Nenurodė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Vadovai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Kvalifikuoti darbininkai ir amatininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Paslaugų sektoriaus darbuotojai ir pardavėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Įrenginių ir mašinų operatoriai ir surinkėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Šiuo metu mokosi = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Šiuo metu mokosi = Taip: Shapiro-Wilk p = 0.5649 ✓ Normalu
Darbo sutarties tipas = Metaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Darbo sutarties tipas = Neterminuota sutartis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Darbo sutarties tipas = Terminuota sutartis: Shapiro-Wilk p = 0.0543 ✓ Normalu

13 Priedas. Leveno testo rezultatai, kai priklausomas kintamasis – maistas ir nealkoholiniai gėrimai

Dispersijų lygybės testas (Levene):

Gyvenamoji vieta: Levene p = 0.0001 Skirtingos dispersijos

Regionas: Levene p = 0.1638 Vienodos dispersijos

Namų ūkio tipas: Levene p = 0.0000 Skirtingos dispersijos

Išlaikomų vaikų skaičius: Levene p = 0.0000 Skirtingos dispersijos

Socialinė ekonominė grupė: Levene p = 0.0000 Skirtingos dispersijos

Pajamų šaltinis: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio galvos amžius: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio galvos išsilavinimas: Levene p = 0.0000 Skirtingos dispersijos

Pajamų kvintilinė grupė: Levene p = 0.0000 Skirtingos dispersijos

Būsto tipas: Levene p = 0.0000 Skirtingos dispersijos

Būsto valdos statusas: Levene p = 0.0032 Skirtingos dispersijos

Centralizuotas šalto vandens tiekimas: Levene p = 0.1256 Vienodos dispersijos

Centralizuotas karšto vandens tiekimas: Levene p = 0.0000 Skirtingos dispersijos

Pagrindinis būsto šildymo būdas: Levene p = 0.0000 Skirtingos dispersijos

Asmenų skaičius namų ūkyje: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio užimamas naudingasis plotas: Levene p = 0.0000 Skirtingos dispersijos

Motinos gimimo šalis: Levene p = 0.4001 Vienodos dispersijos

Gimimo šalis: Levene p = 0.1796 Vienodos dispersijos

Pagrindinės pilietybės šalis: Levene p = 0.9458 Vienodos dispersijos

Tėvo gimimo šalis: Levene p = 0.0609 Vienodos dispersijos

Lytis: Levene p = 0.0000 Skirtingos dispersijos

Teisinė santuokinė padėtis: Levene p = 0.0000 Skirtingos dispersijos

Gyvena su partneriu: Levene p = 0.0000 Skirtingos dispersijos

Pagrindinis darbas: profesija: Levene p = 0.0000 Skirtingos dispersijos

Šiuo metu mokosi: Levene p = 0.0093 Skirtingos dispersijos

Darbo sutarties tipas: Levene p = 0.0000 Skirtingos dispersijos

14 Priedas. Andersono ir Darlingo testo rezultatai, kai priklausomas kintamasis yra maistui ir gėrimams skirtos išlaidos

Kategorinis kintamasis	Andersono-Darlingo statistika	p-reikšmė	Bent vienos grupės pasiskirstymo forma skiriasi
Gyvenamoji vieta	23,20270132	0,001	Taip
Regionas	4,472108379	0,005	Taip
Namų ūkio tipas	274,6942901	0,001	Taip
Išlaikomų vaikų skaičius	152,358421	0,001	Taip
Socialinė ekonominė grupė	157,4439799	0,001	Taip
Pajamų šaltinis	154,2377628	0,001	Taip
Namų ūkio galvos amžius	94,15015306	0,001	Taip
Namų ūkio galvos išsilavinimas	47,09804284	0,001	Taip
Pajamų kvintilinė grupė	84,61218205	0,001	Taip
Būsto tipas	49,64217153	0,001	Taip
Būsto valdos statusas	18,98092282	0,001	Taip
Centralizuotas šalto vandens tiekimas	7,158687836	0,001	Taip
Centralizuotas karšto vandens tiekimas	62,97275204	0,001	Taip
Pagrindinis būsto šildymo būdas	53,42631696	0,001	Taip
Asmenų skaičius namų ūkyje	264,8247428	0,001	Taip
Namų ūkio užimamas naudingasis plotas	106,875962	0,001	Taip
Motinos gimimo šalis	0,083716805	0,250	Ne
Gimimo šalis	1,574706196	0,072	Taip
Pagrindinės pilietybės šalis	1,393551315	0,086	Taip
Tėvo gimimo šalis	0,10782159	0,250	Ne
Lytis	102,7151692	0,001	Taip
Teisinė santuokinė padėtis	220,5695897	0,001	Taip
Gyvena su partneriu	308,2513007	0,001	Taip
Pagrindinis darbas: profesija	101,6190078	0,001	Taip
Šiuo metu mokosi	2,086892059	0,045	Taip
Darbo sutarties tipas	113,4799034	0,001	Taip

15 Priedas. Moodo medianos testo rezultatai, skirti maisto ir nealkoholinių gėrimų vartojimo išlaidų skirtumams tarp kategorinių grupių vertinti.

Kategorinis kintamasis	χ^2 statistikos reikšmė	p-reikšmė	Požymis yra reikšmingas
Gyvenamoji vieta	35,5961	0,000	Taip
Regionas	1,8732	0,170	Ne
Namų ūkio tipas	716,6756	0,000	Taip
Išlaikomų vaikų skaičius	259,9696	0,000	Taip
Socialinė ekonominė grupė	345,7548	0,000	Taip
Pajamų šaltinis	324,2278	0,000	Taip
Namų ūkio galvos amžius	221,8326	0,000	Taip
Namų ūkio galvos išsilavinimas	111,3406	0,000	Taip
Pajamų kvintilinė grupė	207,1585	0,000	Taip
Būsto tipas	75,5518	0,000	Taip
Būsto valdos statusas	38,9499	0,000	Taip
Centralizuotas šalto vandens tiekimas	7,4106	0,007	Taip
Centralizuotas karšto vandens tiekimas	58,2933	0,000	Taip
Pagrindinis būsto šildymo būdas	95,6065	0,000	Taip
Asmenų skaičius namų ūkyje	371,284	0,000	Taip
Namų ūkio užimamas naudingasis plotas	191,9529	0,000	Taip
Motinos gimimo šalis	2,8808	0,237	Ne
Gimimo šalis	5,4234	0,020	Taip
Pagrindinės pilietybės šalis	4,7338	0,030	Taip
Tėvo gimimo šalis	1,5949	0,451	Ne
Lytis	134,2186	0,000	Taip
Teisinė santuokinė padėtis	467,4096	0,000	Taip
Gyvena su partneriu	526,4904	0,000	Taip
Pagrindinis darbas: profesija	366,031	0,000	Taip
Šiuo metu mokosi	1,9543	0,162	Ne
Darbo sutarties tipas	200,0804	0,000	Taip

16 Priedas. Daugartiniai palyginimai naudojant medianos testą, kai priklausomas kintamasis yra maisto ir nealkoholinių gėrimų vartojimo išlaidas

Tikrinamas kintamasis: Gyvenamoji vieta	
5 didieji miestai vs Kiti miestai: chi tatistika = 0.5379, p-reikšmė = 0.4633, pakoreguota p-reikšmė = 0.5568	

5 didieji miestai vs Kaimas: chi tatistika = 19.6939, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Kiti miestai vs Kaimas: chi tatistika = 28.0169, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Tikrinamas kintamasis: Regionas	
Sostinės regionas vs Vidurio ir vakarų Lietuvos regionas: chi tatistika = 1.8732, p-reikšmė = 0.1711, pakoreguota p-reikšmė = 0.2268	

Tikrinamas kintamasis: Namų ūkio tipas	
Vienas gyvenantis asmuo vs Kiti namų ūkiai su išlaikomais vaikais: chi tatistika = 76.8278, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vienas gyvenantis asmuo vs Pora be išlaikomų vaikų: chi tatistika = 356.1322, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vienas gyvenantis asmuo vs Pora su išlaikomais vaikais: chi tatistika = 332.3079, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vienas gyvenantis asmuo vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 58.3066, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vienas gyvenantis asmuo vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 159.4525, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Kiti namų ūkiai su išlaikomais vaikais vs Pora be išlaikomų vaikų: chi tatistika = 18.4334, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Kiti namų ūkiai su išlaikomais vaikais vs Pora su išlaikomais vaikais: chi tatistika = 0.0005, p-reikšmė = 0.9827, pakoreguota p-reikšmė = 0.9946	

Kiti namų ūkiai su išlaikomais vaikais vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 11.2783, p-reikšmė = 0.0008, pakoreguota p-reikšmė = 0.0015	

Kiti namų ūkiai su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 13.8180, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004	

Pora be išlaikomų vaikų vs Pora su išlaikomais vaikais: chi tatistika = 55.9289, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Pora be išlaikomų vaikų vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 0.1606, p-reikšmė = 0.6886, pakoreguota p-reikšmė = 0.7616	

Pora be išlaikomų vaikų vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 0.1375, p-reikšmė = 0.7108, pakoreguota p-reikšmė = 0.7708	

Pora su išlaikomais vaikais vs Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 24.7769, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Pora su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 33.1905, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vienas suaugęs asmuo su išlaikomais vaikais vs Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 0.3733, p-reikšmė = 0.5412, pakoreguota p-reikšmė = 0.6321	

Tikrinamas kintamasis: Išlaikomų vaikų skaičius	
Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 1 išlaikomu vaiku: chi tatistika = 94.2555, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 114.8695, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Namų ūkiai be išlaikomų vaikų vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 62.2355, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Namų ūkiai su 1 išlaikomu vaiku vs Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 14.6024, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0003	

Namų ūkiai su 1 išlaikomu vaiku vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 32.3727, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Namų ūkiai su 2 išlaikomais vaikais vs Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 8.7469, p-reikšmė = 0.0031, pakoreguota p-reikšmė = 0.0056	

Tikrinamas kintamasis: Socialinė ekonominė grupė	
Senatvės pensininkai vs Samdomieji darbuotojai: chi tatistika = 274.7591, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Senatvės pensininkai vs Bedarbiai: chi tatistika = 4.2131, p-reikšmė = 0.0401, pakoreguota p-reikšmė = 0.0598	

Senatvės pensininkai vs Savarankiškai dirbantys asmenys: chi tatistika = 117.0195, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Samdomieji darbuotojai vs Bedarbiai: chi tatistika = 39.7679, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Samdomieji darbuotojai vs Savarankiškai dirbantys asmenys: chi tatistika = 13.6243, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004	

Bedarbiai vs Savarankiškai dirbantys asmenys: chi tatistika = 42.4109, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Tikrinamas kintamasis: Pajamų šaltinis	
Pensija vs Darbo užmokestis: chi tatistika = 281.9428, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Pensija vs Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 0.3181, p-reikšmė = 0.5727, pakoreguota p-reikšmė = 0.6507	

Pensija vs Pajamos iš savo ar šeimos verslo: chi tatistika = 81.4211, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Darbo užmokestis vs Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 27.0581, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Darbo užmokestis vs Pajamos iš savo ar šeimos verslo: chi tatistika = 7.2900, p-reikšmė = 0.0069, pakoreguota p-reikšmė = 0.0118	

Pašalpa, stipendija ir kitas pragyvenimo šaltinis vs Pajamos iš savo ar šeimos verslo: chi tatistika = 36.7973, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Tikrinamas kintamasis: Namų ūkio galvos išsilavinimas

60 metų ir vyresni vs 50-59 metų: chi tatistika = 83.7768, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni vs 40-49 metų: chi tatistika = 115.8129, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni vs iki 30 metų: chi tatistika = 0.7701, p-reikšmė = 0.3802, pakoreguota p-reikšmė = 0.4694

60 metų ir vyresni vs 30-39 metų: chi tatistika = 73.5551, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-59 metų vs 40-49 metų: chi tatistika = 9.3257, p-reikšmė = 0.0023, pakoreguota p-reikšmė = 0.0041

50-59 metų vs iki 30 metų: chi tatistika = 4.6501, p-reikšmė = 0.0311, pakoreguota p-reikšmė = 0.0476

50-59 metų vs 30-39 metų: chi tatistika = 3.0745, p-reikšmė = 0.0795, pakoreguota p-reikšmė = 0.1145

40-49 metų vs iki 30 metų: chi tatistika = 17.3425, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0001

40-49 metų vs 30-39 metų: chi tatistika = 1.1694, p-reikšmė = 0.2795, pakoreguota p-reikšmė = 0.3510

iki 30 metų vs 30-39 metų: chi tatistika = 13.7922, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004

Tikrinamas kintamasis: Namų ūkio galvos išsilavinimas

Vidurinis vs Aukštesniojo ir aukštojo nesuteikiantis povidurinis: chi tatistika = 0.3536, p-reikšmė = 0.5521, pakoreguota p-reikšmė = 0.6403

Vidurinis vs Aukštesnysis ir aukštasis: chi tatistika = 30.5891, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis vs Pradinis, nebaigę pradinės mokyklos: chi tatistika = 22.9105, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis vs Pagrindinis: chi tatistika = 15.7738, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Aukštesnysis ir aukštasis: chi tatistika = 31.2836, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Pradinis, nebaigę pradinės mokyklos: chi tatistika = 25.9521, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis vs Pagrindinis: chi tatistika = 22.9550, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis vs Pradinis, nebaigę pradinės mokyklos: chi tatistika = 53.1918, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis vs Pagrindinis: chi tatistika = 56.2956, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pradinis, nebaigę pradinės mokyklos vs Pagrindinis: chi tatistika = 2.1456, p-reikšmė = 0.1430, pakoreguota p-reikšmė = 0.1990

Tikrinamas kintamasis: Pajamų kvintilinė grupė

2-a pajamų kvintilinė grupė vs 1-a pajamų kvintilinė grupė: chi tatistika = 2.0844, p-reikšmė = 0.1488, pakoreguota p-reikšmė = 0.2054

2-a pajamų kvintilinė grupė vs 5-a pajamų kvintilinė grupė: chi tatistika = 123.2808, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė vs 4-a pajamų kvintilinė grupė: chi tatistika = 120.3971, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė vs 3-a pajamų kvintilinė grupė: chi tatistika = 54.3196, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė vs 5-a pajamų kvintilinė grupė: chi tatistika = 66.5340, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė vs 4-a pajamų kvintilinė grupė: chi tatistika = 62.0571, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė vs 3-a pajamų kvintilinė grupė: chi tatistika = 22.3066, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

5-a pajamų kvintilinė grupė vs 4-a pajamų kvintilinė grupė: chi tatistika = 0.0108, p-reikšmė = 0.9173, pakoreguota p-reikšmė = 0.9398

5-a pajamų kvintilinė grupė vs 3-a pajamų kvintilinė grupė: chi tatistika = 28.3307, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

4-a pajamų kvintilinė grupė vs 3-a pajamų kvintilinė grupė: chi tatistika = 21.3620, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Būsto tipas

Butas daugiabučiame name vs Vieno buto namas: chi tatistika = 72.3857, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Butas daugiabučiame name vs Butas dviejų butų ar sublokuotame name: chi tatistika = 5.5238, p-reikšmė = 0.0188, pakoreguota p-reikšmė = 0.0304

Butas daugiabučiame name vs Kita gyvenamoji patalpa: chi tatistika = 1.3350, p-reikšmė = 0.2479, pakoreguota p-reikšmė = 0.3161

Vieno buto namas vs Butas dviejų butų ar sublokuotame name: chi tatistika = 0.1452, p-reikšmė = 0.7032, pakoreguota p-reikšmė = 0.7708

Vieno buto namas vs Kita gyvenamoji patalpa: chi tatistika = 1.3358, p-reikšmė = 0.2478, pakoreguota p-reikšmė = 0.3161

Butas dviejų butų ar sublokuotame name vs Kita gyvenamoji patalpa: chi tatistika = 1.3470, p-reikšmė = 0.2458, pakoreguota p-reikšmė = 0.3161

Tikrinamas kintamasis: Būsto valdos statusas
Būsto savininkas, neturintis išmokėti paskolos vs Būsto savininkas, turintis neišmokėtą paskolą: chi tatistika = 30.3304, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, neturintis išmokėti paskolos vs Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 0.1191, p-reikšmė = 0.7300, pakoreguota p-reikšmė = 0.7765

Būsto savininkas, neturintis išmokėti paskolos vs Gyvena nemokėdamas nuomos: chi tatistika = 1.9720, p-reikšmė = 0.1602, pakoreguota p-reikšmė = 0.2161

Būsto savininkas, neturintis išmokėti paskolos vs Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 0.9202, p-reikšmė = 0.3374, pakoreguota p-reikšmė = 0.4205

Būsto savininkas, turintis neišmokėtą paskolą vs Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 4.9469, p-reikšmė = 0.0261, pakoreguota p-reikšmė = 0.0408

Būsto savininkas, turintis neišmokėtą paskolą vs Gyvena nemokėdamas nuomos: chi tatistika = 36.1079, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, turintis neišmokėtą paskolą vs Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 14.6081, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0003

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina vs Gyvena nemokėdamas nuomos: chi tatistika = 0.1400, p-reikšmė = 0.7083, pakoreguota p-reikšmė = 0.7708

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina vs Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 0.1283, p-reikšmė = 0.7202, pakoreguota p-reikšmė = 0.7760

Gyvena nemokėdamas nuomos vs Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Tikrinamas kintamasis: Centralizuotas šalto vandens tiekimas
Taip vs Ne: chi tatistika = 7.4106, p-reikšmė = 0.0065, pakoreguota p-reikšmė = 0.0112

Tikrinamas kintamasis: Centralizuotas karšto vandens tiekimas
Taip vs Ne: chi tatistika = 58.2933, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Pagrindinis būsto šildymo būdas
Centrinis šildymas iš šilumos tinklu vs Centrinis šildymas iš vietinio šilumos šaltinio: chi tatistika = 84.8731, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Centrinis šildymas iš šilumos tinklu vs Nėra stacionaraus šilumos šaltinio: chi tatistika = 0.0639, p-reikšmė = 0.8004, pakoreguota p-reikšmė = 0.8407

Centrinis šildymas iš šilumos tinklu vs Krosnis, židiny: chi tatistika = 0.6604, p-reikšmė = 0.4164, pakoreguota p-reikšmė = 0.5076

Centrinis šildymas iš vietinio šilumos šaltinio vs Nėra stacionaraus šilumos šaltinio: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Centrinis šildymas iš vietinio šilumos šaltinio vs Krosnis, židiny: chi tatistika = 49.9811, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nėra stacionaraus šilumos šaltinio vs Krosnis, židiny: chi tatistika = 0.0673, p-reikšmė = 0.7954, pakoreguota p-reikšmė = 0.8407

Tikrinamas kintamasis: Asmenų skaičius namų ūkyje
1-2 asmenys vs 3-4 asmenys: chi tatistika = 268.1970, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-2 asmenys vs 5 ir daugiau: chi tatistika = 106.8899, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

3-4 asmenys vs 5 ir daugiau: chi tatistika = 29.5767, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Namų ūkio užimamas naudingasis plotas
<50 m² vs 50-99 m²: chi tatistika = 114.8408, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² vs 100-199 m²: chi tatistika = 148.3855, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² vs >200 m²: chi tatistika = 11.3616, p-reikšmė = 0.0007, pakoreguota p-reikšmė = 0.0014

50-99 m² vs 100-199 m²: chi tatistika = 24.6576, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-99 m² vs >200 m²: chi tatistika = 2.0670, p-reikšmė = 0.1505, pakoreguota p-reikšmė = 0.2060

100-199 m² vs >200 m²: chi tatistika = 0.2429, p-reikšmė = 0.6222, pakoreguota p-reikšmė = 0.6973

Tikrinamas kintamasis: Motinos gimimo šalis
Lietuva vs Kita šalis: chi tatistika = 0.5376, p-reikšmė = 0.4634, pakoreguota p-reikšmė = 0.5568

Lietuva vs Nenurodė: chi tatistika = 1.7657, p-reikšmė = 0.1839, pakoreguota p-reikšmė = 0.2418

Kita šalis vs Nenurodė: chi tatistika = 1.9698, p-reikšmė = 0.1605, pakoreguota p-reikšmė = 0.2161

Tikrinamas kintamasis: Gimimo šalis
Lietuva vs Kita šalis: chi tatistika = 5.4234, p-reikšmė = 0.0199, pakoreguota p-reikšmė = 0.0316

Tikrinamas kintamasis: Pagrindinės pilietybės šalis
Lietuva vs Kita šalis: chi tatistika = 4.7338, p-reikšmė = 0.0296, pakoreguota p-reikšmė = 0.0457

Tikrinamas kintamasis: Tėvo gimimo šalis
Lietuva vs Kita šalis: chi tatistika = 1.3785, p-reikšmė = 0.2404, pakoreguota p-reikšmė = 0.3136

Lietuva vs Nenurodė: chi tatistika = 0.0246, p-reikšmė = 0.8755, pakoreguota p-reikšmė = 0.9081

Kita šalis vs Nenurodė: chi tatistika = 0.2626, p-reikšmė = 0.6083, pakoreguota p-reikšmė = 0.6864

Tikrinamas kintamasis: Lytis
Moteris vs Vyras: chi tatistika = 134.2186, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Teisinė santuokinė padėtis
Išsituokęs (-usi) vs Našlys (-ė): chi tatistika = 3.7389, p-reikšmė = 0.0532, pakoreguota p-reikšmė = 0.0772

Išsituokęs (-usi) vs Nevedęs (netekėjusi): chi tatistika = 9.9336, p-reikšmė = 0.0016, pakoreguota p-reikšmė = 0.0030

Išsituokęs (-usi) vs Vedęs (ištekejusi): chi tatistika = 194.2631, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Našlys (-ė) vs Nevedęs (netekėjusi): chi tatistika = 21.1375, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Našlys (-ė) vs Vedęs (ištekejusi): chi tatistika = 347.8943, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nevedęs (netekėjusi) vs Vedęs (ištekejusi): chi tatistika = 99.8755, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Gyvena su partneriu
Ne vs Taip, teisiniu pagrindu: chi tatistika = 503.6821, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Ne vs Taip, be teisinio pagrindo: chi tatistika = 55.4795, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Taip, teisiniu pagrindu vs Taip, be teisinio pagrindo: chi tatistika = 0.4177, p-reikšmė = 0.5181, pakoreguota p-reikšmė = 0.6093

Tikrinamas kintamasis: Pagrindinis darbas: profesija
Netaikoma vs Specialistai: chi tatistika = 108.0436, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Įstaigų tarnautojai: chi tatistika = 25.1053, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Technikai ir jaunesnieji specialistai: chi tatistika = 28.7666, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Nenurodė: chi tatistika = 20.8438, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Vadovai: chi tatistika = 94.6469, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 123.9600, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 24.2464, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 58.8578, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 48.9421, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai vs Įstaigų tarnautojai: chi tatistika = 0.1235, p-reikšmė = 0.7253, pakoreguota p-reikšmė = 0.7764

Specialistai vs Technikai ir jaunesnieji specialistai: chi tatistika = 0.0450, p-reikšmė = 0.8321, pakoreguota p-reikšmė = 0.8685

Specialistai vs Nenurodė: chi tatistika = 9.1914, p-reikšmė = 0.0024, pakoreguota p-reikšmė = 0.0044

Specialistai vs Vadovai: chi tatistika = 14.2287, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0003

Specialistai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 2.3993, p-reikšmė = 0.1214, pakoreguota p-reikšmė = 0.1718

Specialistai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 8.0541, p-reikšmė = 0.0045, pakoreguota p-reikšmė = 0.0079

Specialistai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.4451, p-reikšmė = 0.5047, pakoreguota p-reikšmė = 0.5977

Specialistai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 13.3238, p-reikšmė = 0.0003, pakoreguota p-reikšmė = 0.0005

Įstaigų tarnautojai vs Technikai ir jaunesnieji specialistai: chi tatistika = 0.0073, p-reikšmė = 0.9319, pakoreguota p-reikšmė = 0.9490

Įstaigų tarnautojai vs Nenurodė: chi tatistika = 4.4952, p-reikšmė = 0.0340, pakoreguota p-reikšmė = 0.0511

Įstaigų tarnautojai vs Vadovai: chi tatistika = 6.1778, p-reikšmė = 0.0129, pakoreguota p-reikšmė = 0.0216

Įstaigų tarnautojai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 1.1844, p-reikšmė = 0.2765, pakoreguota p-reikšmė = 0.3498

Įstaigų tarnautojai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 4.6153, p-reikšmė = 0.0317, pakoreguota p-reikšmė = 0.0481

Įstaigų tarnautojai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.7635, p-reikšmė = 0.3822, pakoreguota p-reikšmė = 0.4694

Įstaigų tarnautojai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 3.7723, p-reikšmė = 0.0521, pakoreguota p-reikšmė = 0.0763

Technikai ir jaunesnieji specialistai vs Nenurodė: chi tatistika = 5.5673, p-reikšmė = 0.0183, pakoreguota p-reikšmė = 0.0300

 Technikai ir jaunesnieji specialistai vs Vadovai: chi tatistika = 8.3349, p-reikšmė = 0.0039, pakoreguota p-reikšmė = 0.0068

 Technikai ir jaunesnieji specialistai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 0.3366, p-reikšmė = 0.5618, pakoreguota p-reikšmė = 0.6470

 Technikai ir jaunesnieji specialistai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 3.9010, p-reikšmė = 0.0483, pakoreguota p-reikšmė = 0.0713

 Technikai ir jaunesnieji specialistai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.4484, p-reikšmė = 0.5031, pakoreguota p-reikšmė = 0.5977

 Technikai ir jaunesnieji specialistai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 5.4422, p-reikšmė = 0.0197, pakoreguota p-reikšmė = 0.0316

 Nenurodė vs Vadovai: chi tatistika = 35.1573, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Nenurodė vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 19.2603, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Nenurodė vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 0.1929, p-reikšmė = 0.6605, pakoreguota p-reikšmė = 0.7354

 Nenurodė vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 2.6447, p-reikšmė = 0.1039, pakoreguota p-reikšmė = 0.1483

 Nenurodė vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 26.0086, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Vadovai vs Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 8.6311, p-reikšmė = 0.0033, pakoreguota p-reikšmė = 0.0059

 Vadovai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 36.8060, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Vadovai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 16.4280, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0001

 Vadovai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0156, p-reikšmė = 0.9005, pakoreguota p-reikšmė = 0.9283

 Kvalifikuoti darbininkai ir amatininkai vs Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 17.6411, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0001

 Kvalifikuoti darbininkai ir amatininkai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 5.3772, p-reikšmė = 0.0204, pakoreguota p-reikšmė = 0.0321

 Kvalifikuoti darbininkai ir amatininkai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 6.9479, p-reikšmė = 0.0084, pakoreguota p-reikšmė = 0.0142

 Paslaugų sektoriaus darbuotojai ir pardavėjai vs Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 2.2373, p-reikšmė = 0.1347, pakoreguota p-reikšmė = 0.1891

 Paslaugų sektoriaus darbuotojai ir pardavėjai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 24.3438, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Įrenginių ir mašinų operatoriai ir surinkėjai vs Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 13.5435, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0005

Tikrinamas kintamasis: Šiuo metu mokosi
 Ne vs Taip: chi tatistika = 1.9543, p-reikšmė = 0.1621, pakoreguota p-reikšmė = 0.2166

Tikrinamas kintamasis: Darbo sutarties tipas
 Netaikoma vs Neterminuota sutartis: chi tatistika = 199.0781, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Netaikoma vs Terminuota sutartis: chi tatistika = 6.1171, p-reikšmė = 0.0134, pakoreguota p-reikšmė = 0.0221

 Neterminuota sutartis vs Terminuota sutartis: chi tatistika = 0.3293, p-reikšmė = 0.5661, pakoreguota p-reikšmė = 0.6475

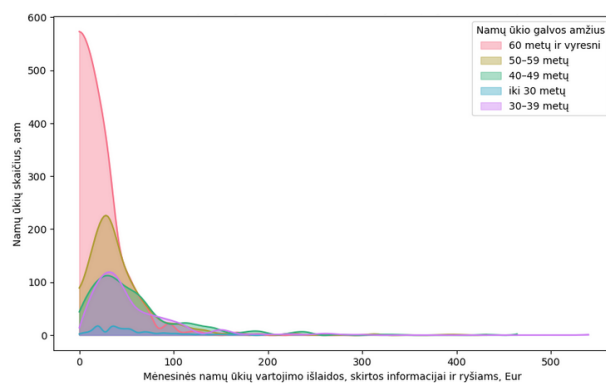
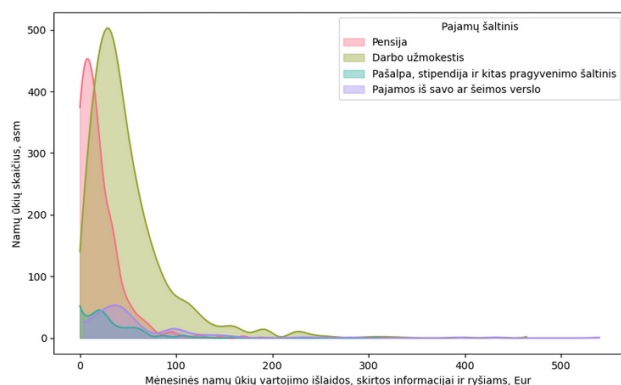
Tikrinamas kintamasis: Šiuo metu mokosi
 Ne ir Taip: chi tatistika = 4.8802, p-reikšmė = 0.0272, pakoreguota p-reikšmė = 0.0420

Tikrinamas kintamasis: Darbo sutarties tipas
 Netaikoma ir Neterminuota sutartis: chi tatistika = 369.3327, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Netaikoma ir Terminuota sutartis: chi tatistika = 20.1588, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

 Neterminuota sutartis ir Terminuota sutartis: chi tatistika = 2.9447, p-reikšmė = 0.0862, pakoreguota p-reikšmė = 0.1251

17 Priedas. Skirstinių tankių grafikai, kai kategorinių kintamųjų grupės yra pasiskirsčiusios pagal informacijai ir ryšiams skirtos išlaidos



18 Priedas. Šapiro ir Vilko testo rezultatai, kai priklausomas kintamasis yra informacijai ir ryšiams skirtos išlaidos

Gyvenamoji vieta = 5 didieji miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvenamoji vieta = Kiti miestai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvenamoji vieta = Kaimas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Regionas = Sostinės regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Regionas = Vidurio ir vakarų Lietuvos regionas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = vienas gyvenantis asmuo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Kiti namų ūkiai su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Pora be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Pora su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Vienas suaugęs asmuo su išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio tipas = Kiti namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 1 išlaikomu vaiku: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 2 išlaikomais vaikais: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Išlaikomų vaikų skaičius = Namų ūkiai su 3 ir daugiau išlaikomų vaikų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Senatvės pensininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Samdomieji darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Bedarbiai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Socialinė ekonominė grupė = Savarankiškai dirbantys asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pensija: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = darbo užmokestis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pašalpa, stipendija ir kitas pragyvenimo šaltinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų šaltinis = Pajamos iš savo ar šeimos verslo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 60 metų ir vyresni: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 50-59 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 40-49 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = iki 30 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos amžius = 30-39 metų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Vidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Aukštesniojo ir aukštojo nesuteikiantis pavidurinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Aukštesnysis ir aukštasis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Pradinis, nebaigęs pradinės mokyklos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio galvos išsilavinimas = Pagrindinis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 2-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 1-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 5-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 4-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pajamų kvintilinė grupė = 3-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = Butas daugiabučiam name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = vieno buto namas: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto tipas = Butas dviejų butų ar sublokuotame name: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Būsto savininkas, neturintis išmokėti paskolos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Būsto savininkas, turintis neišmokėtą paskolą: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: Shapiro-Wilk p = 0.0684 ✓ Normalu
Būsto valdos statusas = Gyvena nemokėdamas nuomos: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Būsto valdos statusas = Nuomojasi šį būstą mažesne nei rinkos kaina: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas šalto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas šalto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas karšto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Centralizuotas karšto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš šilumos tinklų: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš vietinio šilumos šaltinio: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Nėra stacionaraus šilumos šaltinio: Shapiro-Wilk p = 0.0033 ✗ Nenormalu
Pagrindinis būsto šildymo būdas = Krosnis, židiny: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 1-2 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 3-4 asmenys: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Asmenų skaičius namų ūkyje = 5 ir daugiau: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = <50 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = 50-99 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = 100-199 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Namų ūkio užimamas naudingasis plotas = >200 m²: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Motinos gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0011 ✗ Nenormalu
Gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinės pilietybės šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinės pilietybės šalis = Kita šalis: Shapiro-Wilk p = 0.0004 ✗ Nenormalu
Tėvo gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Tėvo gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Tėvo gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Lytis = Moteris: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Lytis = Vyras: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Išituokęs (-usi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Našlys (-ė): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Nevedęs (netekėjusi): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Teisinė santuokinė padėtis = Vedęs (ištekęs): Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Taip, teisiniu pagrindu: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Gyvena su partneriu = Taip, be teisinio pagrindo: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Netaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Specialistai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Įstaigų tarnautojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Technikai ir jaunesnieji specialistai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Nenurodė: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Vadovai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Kvalifikuoti darbininkai ir amatininkai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Paslaugų sektoriaus darbuotojai ir pardavėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Įrenginių ir mašinų operatoriai ir surinkėjai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Pagrindinis darbas: profesija = Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Šiuo metu mokosi = Ne: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Šiuo metu mokosi = Taip: Shapiro-Wilk p = 0.0004 ✗ Nenormalu
Darbo sutarties tipas = Netaikoma: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Darbo sutarties tipas = Neterminuota sutartis: Shapiro-Wilk p = 0.0000 ✗ Nenormalu
Darbo sutarties tipas = Terminuota sutartis: Shapiro-Wilk p = 0.0019 ✗ Nenormalu

19 Priedas. Leveno testo rezultatai, kai priklausomas kintamasis yra informacijai ir ryšiams skirtos išlaidos

Dispersijų lygybės testas (Levene):

Gyvenamoji vieta: Levene p = 0.1294 Vienodos dispersijos

Regionas: Levene p = 0.2628 Vienodos dispersijos

Namų ūkio tipas: Levene p = 0.0000 Skirtingos dispersijos

Išlaikomų vaikų skaičius: Levene p = 0.0000 Skirtingos dispersijos

Socialinė ekonominė grupė: Levene p = 0.0000 Skirtingos dispersijos

Pajamų šaltinis: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio galvos amžius: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio galvos išsilavinimas: Levene p = 0.0000 Skirtingos dispersijos

Pajamų kvintilinė grupė: Levene p = 0.0000 Skirtingos dispersijos

Būsto tipas: Levene p = 0.0001 Skirtingos dispersijos

Būsto valdos statusas: Levene p = 0.0000 Skirtingos dispersijos

Centralizuotas šalto vandens tiekimas: Levene p = 0.1555 Vienodos dispersijos

Centralizuotas karšto vandens tiekimas: Levene p = 0.0004 Skirtingos dispersijos

Pagrindinis būsto šildymo būdas: Levene p = 0.0002 Skirtingos dispersijos

Asmenų skaičius namų ūkyje: Levene p = 0.0000 Skirtingos dispersijos

Namų ūkio užimamas naudingasis plotas: Levene p = 0.0000 Skirtingos dispersijos

Motinos gimimo šalis: Levene p = 0.1488 Vienodos dispersijos

Gimimo šalis: Levene p = 0.0024 Skirtingos dispersijos

Pagrindinės pilietybės šalis: Levene p = 0.8868 Vienodos dispersijos

Tėvo gimimo šalis: Levene p = 0.1187 Vienodos dispersijos

Lytis: Levene p = 0.0000 Skirtingos dispersijos

Teisinė santuokinė padėtis: Levene p = 0.0000 Skirtingos dispersijos

Gyvena su partneriu: Levene p = 0.0000 Skirtingos dispersijos

Pagrindinis darbas: profesija: Levene p = 0.0000 Skirtingos dispersijos

Šiuo metu mokosi: Levene p = 0.0104 Skirtingos dispersijos

Darbo sutarties tipas: Levene p = 0.0000 Skirtingos dispersijos

20 Priedas. Andersono ir Darlingo testo rezultatai kai priklausomas kintamasis yra informacijai ir ryšiams skirtos išlaidos

Gyvenamoji vieta: Anderson-Darling statistika = 17.769306706829223, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Regionas: Anderson-Darling statistika = -0.08638230093801313, p-value = 0.25
Skirtingos pasiskirstymų formos tarp grupių!

Namų ūkio tipas: Anderson-Darling statistika = 346.6463390117104, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Išlaikomų vaikų skaičius: Anderson-Darling statistika = 295.04379108151176, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Socialinė ekonominė grupė: Anderson-Darling statistika = 433.3472247174698, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Pajamų šaltinis: Anderson-Darling statistika = 431.58591700408687, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Namų ūkio galvos amžius: Anderson-Darling statistika = 292.693450662083, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Namų ūkio galvos išsilavinimas: Anderson-Darling statistika = 170.54715847444103, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Pajamų kvintilinė grupė: Anderson-Darling statistika = 119.9303512234279, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Būsto tipas: Anderson-Darling statistika = 5.165012300034979, p-value = 0.0015344234685637768
Skirtingos pasiskirstymų formos tarp grupių!

Būsto valdos statusas: Anderson-Darling statistika = 52.12204951478525, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Centralizuotas šalto vandens tiekimas: Anderson-Darling statistika = 16.434934982509525, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Centralizuotas karšto vandens tiekimas: Anderson-Darling statistika = 9.902043477652485, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Pagrindinis būsto šildymo būdas: Anderson-Darling statistika = 46.31952379404002, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Asmenų skaičius namų ūkyje: Anderson-Darling statistika = 396.48130193023337, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Namų ūkio užimamas naudingasis plotas: Anderson-Darling statistika = 65.8701082187242, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Motinos gimimo šalis: Anderson-Darling statistika = 0.949143731684854, p-value = 0.14611225630097033
Gimimo šalis: Anderson-Darling statistika = 4.4153295620551996, p-value = 0.005658205298543668
Skirtingos pasiskirstymų formos tarp grupių!

Pagrindinės pilietybės šalis: Anderson-Darling statistika = 0.21027487112628418, p-value = 0.25
Tėvo gimimo šalis: Anderson-Darling statistika = 1.2659156523477033, p-value = 0.10361375604824058
Lytis: Anderson-Darling statistika = 106.65741008922863, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Teisinė santuokinė padėtis: Anderson-Darling statistika = 239.6781068379144, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Gyvena su partneriu: Anderson-Darling statistika = 243.30067556719186, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Pagrindinis darbas: profesija: Anderson-Darling statistika = 252.1308755114287, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Šiuo metu mokosi: Anderson-Darling statistika = 8.161790480893705, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

Darbo sutarties tipas: Anderson-Darling statistika = 359.9634546935552, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

21 Priedas. Moodo medianos testo rezultatai, skirti informacijos ir ryšių paslaugų vartojimo išlaidų skirtumams tarp kategorinių požymių grupių vertinti

Kategorinis kintamasis	χ^2 statistikos reikšmė	p-reikšmė	Požymis yra reikšmingas
Gyvenamoji vieta	8,8115	0,0122	Taip
Regionas	0,2020	0,6531	Ne
Namų ūkio tipas	840,0572	0,0000	Taip
Išlaikomų vaikų skaičius	543,2899	0,0000	Taip
Socialinė ekonominė grupė	867,3193	0,0000	Taip
Pajamų šaltinis	857,4629	0,0000	Taip
Namų ūkio galvos amžius	636,6101	0,0000	Taip
Namų ūkio galvos išsilavinimas	295,2076	0,0000	Taip
Pajamų kvintilinė grupė	250,9263	0,0000	Taip
Būsto tipas	5,9552	0,1138	Ne
Būsto valdos statusas	107,7724	0,0000	Taip
Centralizuotas šalto vandens tiekimas	6,2905	0,0121	Taip
Centralizuotas karšto vandens tiekimas	0,9301	0,3348	Ne
Pagrindinis būsto šildymo būdas	67,4721	0,0000	Taip
Asmenų skaičius namų ūkyje	619,0638	0,0000	Taip
Namų ūkio užimamas naudingasis plotas	138,7172	0,0000	Taip
Motinos gimimo šalis	2,5467	0,2799	Ne
Gimimo šalis	5,3786	0,0204	Taip
Pagrindinės pilietybės šalis	0,7670	0,3812	Ne
Tėvo gimimo šalis	3,4247	0,1804	Ne
Lytis	150,8324	0,0000	Taip
Teisinė santuokinė padėtis	480,1647	0,0000	Taip
Gyvena su partneriu	419,9989	0,0000	Taip
Pagrindinis darbas: profesija	872,9468	0,0000	Taip
Šiuo metu mokosi	10,1297	0,0015	Taip
Darbo sutarties tipas	561,364	0,0000	Taip

22 Priedas. Daugartiniai palyginimai naudojant medianos testą, kai priklausomas kintamasis yra informacijai ir ryšiams skirtos išlaidos

Tikrinamas kintamasis: Gyvenamoji vieta
5 didieji miestai ir Kiti miestai: chi tatistika = 7.2255, p-reikšmė = 0.0072, pakoreguota p-reikšmė = 0.0128

5 didieji miestai ir Kaimas: chi tatistika = 5.4471, p-reikšmė = 0.0196, pakoreguota p-reikšmė = 0.0334

Kiti miestai ir Kaimas: chi tatistika = 0.0006, p-reikšmė = 0.9810, pakoreguota p-reikšmė = 0.9989

Tikrinamas kintamasis: Regionas
Sostinės regionas ir Vidurio ir vakarų Lietuvos regionas: chi tatistika = 0.2020, p-reikšmė = 0.6531, pakoreguota p-reikšmė = 0.7691

Tikrinamas kintamasis: Namų ūkio tipas
Vienas gyvenantis asmuo ir Kiti namų ūkiai su išlaikomais vaikais: chi tatistika = 135.2726, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo ir Pora be išlaikomų vaikų: chi tatistika = 248.1231, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo ir Pora su išlaikomais vaikais: chi tatistika = 519.6820, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 126.6433, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas gyvenantis asmuo ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 178.4682, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Kiti namų ūkiai su išlaikomais vaikais ir Pora be išlaikomų vaikų: chi tatistika = 65.9315, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Kiti namų ūkiai su išlaikomais vaikais ir Pora su išlaikomais vaikais: chi tatistika = 0.1211, p-reikšmė = 0.7278, pakoreguota p-reikšmė = 0.8220

Kiti namų ūkiai su išlaikomais vaikais ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 7.2730, p-reikšmė = 0.0070, pakoreguota p-reikšmė = 0.0126

Kiti namų ūkiai su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 35.2198, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora be išlaikomų vaikų ir Pora su išlaikomais vaikais: chi tatistika = 252.5638, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora be išlaikomų vaikų ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 22.3589, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pora be išlaikomų vaikų ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 9.2985, p-reikšmė = 0.0023, pakoreguota p-reikšmė = 0.0044

Pora su išlaikomais vaikais ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 8.3047, p-reikšmė = 0.0040, pakoreguota p-reikšmė = 0.0073

Pora su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 81.9563, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vienas suaugęs asmuo su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 7.3209, p-reikšmė = 0.0060, pakoreguota p-reikšmė = 0.0124

Tikrinamas kintamasis: Išlaikomų vaikų skaičius
Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 1 išlaikomu vaiku: chi tatistika = 252.1983, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 229.6025, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 73.4538, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Namų ūkiai su 1 išlaikomu vaiku ir Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 9.2956, p-reikšmė = 0.0023, pakoreguota p-reikšmė = 0.0044

Namų ūkiai su 2 išlaikomu vaiku ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 1.0081, p-reikšmė = 0.3153, pakoreguota p-reikšmė = 0.4114

Namų ūkiai su 2 išlaikomais vaikais ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 0.4488, p-reikšmė = 0.5029, pakoreguota p-reikšmė = 0.6130

Tikrinamas kintamasis: Socialinė ekonominė grupė
Senatvės pensininkai ir Samdomieji darbuotojai: chi tatistika = 774.8560, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Senatvės pensininkai ir Bedarbiai: chi tatistika = 22.6246, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Senatvės pensininkai ir Savarankiškai dirbantys asmenys: chi tatistika = 216.9592, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Samdomieji darbuotojai ir Bedarbiai: chi tatistika = 104.6133, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Samdomieji darbuotojai ir Savarankiškai dirbantys asmenys: chi tatistika = 8.8999, p-reikšmė = 0.0029, pakoreguota p-reikšmė = 0.0054

Bedarbiai ir Savarankiškai dirbantys asmenys: chi tatistika = 97.7953, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Pajamų šaltinis
Pensija ir Darbo užmokestis: chi tatistika = 812.0417, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pensija ir Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 17.1595, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0001

Pensija ir Pajamos iš savo ar šeimos verslo: chi tatistika = 114.9113, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Darbo užmokestis ir Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 45.9120, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Darbo užmokestis ir Pajamos iš savo ar šeimos verslo: chi tatistika = 2.2899, p-reikšmė = 0.1302, pakoreguota p-reikšmė = 0.1924

Pašalpa, stipendija ir kitas pragyvenimo šaltinis ir Pajamos iš savo ar šeimos verslo: chi tatistika = 47.9747, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Namų ūkio galvos amžius
60 metų ir vyresni ir 50-59 metų: chi tatistika = 225.8244, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni ir 40-49 metų: chi tatistika = 260.8362, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni ir iki 30 metų: chi tatistika = 61.1612, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

60 metų ir vyresni ir 30-39 metų: chi tatistika = 274.7995, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-59 metų ir 40-49 metų: chi tatistika = 31.4557, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-59 metų ir iki 30 metų: chi tatistika = 6.6025, p-reikšmė = 0.0102, pakoreguota p-reikšmė = 0.0177

40-49 metų ir iki 30 metų: chi tatistika = 1.6584, p-reikšmė = 0.1978, pakoreguota p-reikšmė = 0.2730

40-49 metų ir 30-39 metų: chi tatistika = 0.0894, p-reikšmė = 0.7650, pakoreguota p-reikšmė = 0.8538

iki 30 metų ir 30-39 metų: chi tatistika = 0.9974, p-reikšmė = 0.3179, pakoreguota p-reikšmė = 0.4116

Tikrinamas kintamasis: Namų ūkio galvos išsilavinimas
Vidurinis ir Aukštesniojo ir aukštojo nesuteikiantis pavidurinis: chi tatistika = 0.0063, p-reikšmė = 0.9370, pakoreguota p-reikšmė = 0.9670

Vidurinis ir Aukštesnysis ir aukštasis: chi tatistika = 62.9770, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 71.6588, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Vidurinis ir Pagrindinis: chi tatistika = 45.2433, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis pavidurinis ir Aukštesnysis ir aukštasis: chi tatistika = 95.6868, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis pavidurinis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 71.5366, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis pavidurinis ir Pagrindinis: chi tatistika = 45.5773, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 91.7691, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis ir Pagrindinis: chi tatistika = 135.0842, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pradinis, nebaigę pradinės mokyklos ir Pagrindinis: chi tatistika = 9.7865, p-reikšmė = 0.0018, pakoreguota p-reikšmė = 0.0035

Tikrinamas kintamasis: Pajamų kvintilinė grupė
2-a pajamų kvintilinė grupė ir 1-a pajamų kvintilinė grupė: chi tatistika = 18.4386, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė ir 5-a pajamų kvintilinė grupė: chi tatistika = 219.2618, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 119.9420, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 46.4470, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 5-a pajamų kvintilinė grupė: chi tatistika = 76.8884, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 38.4037, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 1.3180, p-reikšmė = 0.2509, pakoreguota p-reikšmė = 0.3380

5-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 9.2118, p-reikšmė = 0.0024, pakoreguota p-reikšmė = 0.0046

5-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 84.0206, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

4-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 33.6629, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Būsto tipas
Butas daugiabučiame name ir Vieno buto namas: chi tatistika = 2.1074, p-reikšmė = 0.1466, pakoreguota p-reikšmė = 0.2110

Butas daugiabučiame name ir Butas dviejų butų ar sublokuotame name: chi tatistika = 1.2450, p-reikšmė = 0.2645, pakoreguota p-reikšmė = 0.3516

Butas daugiabučiame name ir Kita gyvenamoji patalpa: chi tatistika = 1.2199, p-reikšmė = 0.2694, pakoreguota p-reikšmė = 0.3542

Vieno buto namas ir Butas dviejų butų ar sublokuotame name: chi tatistika = 0.4033, p-reikšmė = 0.5254, pakoreguota p-reikšmė = 0.6358

Vieno buto namas ir Kita gyvenamoji patalpa: chi tatistika = 1.3309, p-reikšmė = 0.2487, pakoreguota p-reikšmė = 0.3380

Butas dviejų butų ar sublokuotame name ir Kita gyvenamoji patalpa: chi tatistika = 1.3199, p-reikšmė = 0.2506, pakoreguota p-reikšmė = 0.3380

Tikrinamas kintamasis: Būsto valdos statusas
Būsto savininkas, neturintis išmokėti paskolos ir Būsto savininkas, turintis neišmokėtą paskolą: chi tatistika = 101.3303, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, neturintis išmokėti paskolos ir Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 3.3183, p-reikšmė = 0.0685, pakoreguota p-reikšmė = 0.1059

Būsto savininkas, neturintis išmokėti paskolos ir Gyvena nemokėdamas nuomos: chi tatistika = 0.7848, p-reikšmė = 0.3757, pakoreguota p-reikšmė = 0.4682

Būsto savininkas, neturintis išmokėti paskolos ir Nuomojasi šį būstą mažesne nei rinkos kaina: chi tatistika = 0.1690, p-reikšmė = 0.6810, pakoreguota p-reikšmė = 0.7898

Būsto savininkas, turintis neišmokėtą paskolą ir Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 0.1219, p-reikšmė = 0.7270, pakoreguota p-reikšmė = 0.8220

Būsto savininkas, turintis neišmokėtą paskolą ir Gyvena nemokėdamas nuomos: chi tatistika = 31.5631, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, turintis neišmokėtą paskolą ir Nuomojasi šį būstą mažesne nei rinkos kaina: chi tatistika = 19.6566, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina ir Gyvena nemokėdamas nuomos: chi tatistika = 2.7149, p-reikšmė = 0.0994, pakoreguota p-reikšmė = 0.1496

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina ir Nuomojasi šį būstą mažesne nei rinkos kaina: chi tatistika = 3.4284, p-reikšmė = 0.0641, pakoreguota p-reikšmė = 0.1000

Gyvena nemokėdamas nuomos ir Nuomojasi šį būstą mažesne nei rinkos kaina: chi tatistika = 0.8328, p-reikšmė = 0.3615, pakoreguota p-reikšmė = 0.4556

Tikrinamas kintamasis: Centralizuotas šalto vandens tiekimas
Taip ir Ne: chi tatistika = 6.2905, p-reikšmė = 0.0121, pakoreguota p-reikšmė = 0.0209

Tikrinamas kintamasis: Centralizuotas karšto vandens tiekimas
Taip ir Ne: chi tatistika = 0.9301, p-reikšmė = 0.3348, pakoreguota p-reikšmė = 0.4269

Tikrinamas kintamasis: Pagrindinis būsto šildymo būdas
Centrinis šildymas iš šilumos tinklu ir Centrinis šildymas iš vietinio šilumos šaltinio: chi tatistika = 20.9929, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Centrinis šildymas iš šilumos tinklu ir Nėra stacionaraus šilumos šaltinio: chi tatistika = 0.6929, p-reikšmė = 0.4052, pakoreguota p-reikšmė = 0.4976

Centrinis šildymas iš vietinio šilumos šaltinio ir Krosnis, židiny: chi tatistika = 61.5115, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Mėra stacionaraus šilumos šaltinio ir Krosnis, židiny: chi tatistika = 3.2511, p-reikšmė = 0.0714, pakoreguota p-reikšmė = 0.1094

Tikrinamas kintamasis: Asmenų skaičius namų ūkyje

1-2 asmenys ir 3-4 asmenys: chi tatistika = 488.5366, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-2 asmenys ir 5 ir daugiau: chi tatistika = 117.9367, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

3-4 asmenys ir 5 ir daugiau: chi tatistika = 12.0277, p-reikšmė = 0.0005, pakoreguota p-reikšmė = 0.0011

Tikrinamas kintamasis: Namų ūkio užimamas naudingasis plotas

<50 m² ir 50-99 m²: chi tatistika = 36.2921, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² ir 100-199 m²: chi tatistika = 120.9312, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² ir >200 m²: chi tatistika = 14.6794, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0003

50-99 m² ir 100-199 m²: chi tatistika = 61.9502, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-99 m² ir >200 m²: chi tatistika = 7.7437, p-reikšmė = 0.0054, pakoreguota p-reikšmė = 0.0099

100-199 m² ir >200 m²: chi tatistika = 0.0561, p-reikšmė = 0.8128, pakoreguota p-reikšmė = 0.8872

Tikrinamas kintamasis: Motinos gimimo šalis

Lietuva ir Kita šalis: chi tatistika = 0.2804, p-reikšmė = 0.5964, pakoreguota p-reikšmė = 0.7166

Lietuva ir Nenurodė: chi tatistika = 1.7035, p-reikšmė = 0.1918, pakoreguota p-reikšmė = 0.2670

Kita šalis ir Nenurodė: chi tatistika = 0.9454, p-reikšmė = 0.3309, pakoreguota p-reikšmė = 0.4251

Tikrinamas kintamasis: Gimimo šalis

Lietuva ir Kita šalis: chi tatistika = 5.3786, p-reikšmė = 0.0204, pakoreguota p-reikšmė = 0.0344

Tikrinamas kintamasis: Pagrindinės pilietybės šalis

Lietuva ir Kita šalis: chi tatistika = 0.7670, p-reikšmė = 0.3812, pakoreguota p-reikšmė = 0.4715

Tikrinamas kintamasis: Tėvo gimimo šalis

Lietuva ir Kita šalis: chi tatistika = 0.0805, p-reikšmė = 0.7766, pakoreguota p-reikšmė = 0.8589

Lietuva ir Nenurodė: chi tatistika = 2.7965, p-reikšmė = 0.0945, pakoreguota p-reikšmė = 0.1434

Kita šalis ir Nenurodė: chi tatistika = 2.1302, p-reikšmė = 0.1444, pakoreguota p-reikšmė = 0.2101

Tikrinamas kintamasis: Lytis

Moteris ir Vyras: chi tatistika = 150.8324, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Teisinė santuokinė padėtis

Išsituokęs (-usi) ir Našlys (-ė): chi tatistika = 59.0504, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Išsituokęs (-usi) ir Nevedęs (netekėjusi): chi tatistika = 15.2659, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002

Išsituokęs (-usi) ir Vedęs (ištekejusi): chi tatistika = 144.7944, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Našlys (-ė) ir Nevedęs (netekėjusi): chi tatistika = 127.7110, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Našlys (-ė) ir Vedęs (ištekejusi): chi tatistika = 436.6640, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nevedęs (netekėjusi) ir Vedęs (ištekejusi): chi tatistika = 19.6865, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Gyvena su partneriu

Ne ir Taip, teisiniu pagrindu: chi tatistika = 406.9971, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Ne ir Taip, be teisinio pagrindo: chi tatistika = 38.8355, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Taip, teisiniu pagrindu ir Taip, be teisinio pagrindo: chi tatistika = 0.1867, p-reikšmė = 0.6657, pakoreguota p-reikšmė = 0.7774

Tikrinamas kintamasis: Pagrindinis darbas: profesija

Netaikoma ir Specialistai: chi tatistika = 350.4824, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Įstaigų tarnautojai: chi tatistika = 68.8343, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Technikai ir jaunesnieji specialistai: chi tatistika = 100.4472, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Nenurodė: chi tatistika = 87.3357, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Vadovai: chi tatistika = 163.0463, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 231.0358, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 107.4857, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 164.4684, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 35.4308, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai ir Įstaigų tarnautojai: chi tatistika = 0.0220, p-reikšmė = 0.8822, pakoreguota p-reikšmė = 0.9272

Specialistai ir Technikai ir jaunesnieji specialistai: chi tatistika = 0.0038, p-reikšmė = 0.9509, pakoreguota p-reikšmė = 0.9743

Specialistai ir Technikai ir Jaunesnieji specialistai: chi tatistika = 0.0038, p-reikšmė = 0.9509, pakoreguota p-reikšmė = 0.9743

Specialistai ir Nenurodė: chi tatistika = 20.3457, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Specialistai ir Vadovai: chi tatistika = 2.6387, p-reikšmė = 0.1043, pakoreguota p-reikšmė = 0.1555

Specialistai ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 0.0217, p-reikšmė = 0.8828, pakoreguota p-reikšmė = 0.9272

Specialistai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 4.5834, p-reikšmė = 0.0323, pakoreguota p-reikšmė = 0.0534

Specialistai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.2727, p-reikšmė = 0.6015, pakoreguota p-reikšmė = 0.7175

Specialistai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0401, p-reikšmė = 0.8412, pakoreguota p-reikšmė = 0.9063

Įstaigų tarnautojai ir Technikai ir Jaunesnieji specialistai: chi tatistika = 0.0374, p-reikšmė = 0.8467, pakoreguota p-reikšmė = 0.9064

Įstaigų tarnautojai ir Nenurodė: chi tatistika = 7.0240, p-reikšmė = 0.0080, pakoreguota p-reikšmė = 0.0141

Įstaigų tarnautojai ir Vadovai: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Įstaigų tarnautojai ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 0.0654, p-reikšmė = 0.7981, pakoreguota p-reikšmė = 0.8769

Įstaigų tarnautojai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 1.9323, p-reikšmė = 0.1645, pakoreguota p-reikšmė = 0.2328

Įstaigų tarnautojai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.1402, p-reikšmė = 0.7081, pakoreguota p-reikšmė = 0.8155

Įstaigų tarnautojai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.2009, p-reikšmė = 0.6540, pakoreguota p-reikšmė = 0.7691

Technikai ir Jaunesnieji specialistai ir Nenurodė: chi tatistika = 4.8739, p-reikšmė = 0.0273, pakoreguota p-reikšmė = 0.0455

Technikai ir Jaunesnieji specialistai ir Vadovai: chi tatistika = 1.8191, p-reikšmė = 0.1774, pakoreguota p-reikšmė = 0.2490

Technikai ir Jaunesnieji specialistai ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Technikai ir Jaunesnieji specialistai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 1.2411, p-reikšmė = 0.2653, pakoreguota p-reikšmė = 0.3516

Technikai ir Jaunesnieji specialistai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000

Technikai ir Jaunesnieji specialistai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0060, p-reikšmė = 0.9380, pakoreguota p-reikšmė = 0.9670

Nenurodė ir Vadovai: chi tatistika = 40.7483, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nenurodė ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 20.6865, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nenurodė ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 1.9355, p-reikšmė = 0.1642, pakoreguota p-reikšmė = 0.2328

Nenurodė ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 20.5721, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nenurodė ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 4.3173, p-reikšmė = 0.0377, pakoreguota p-reikšmė = 0.0612

Nenurodė ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 4.3173, p-reikšmė = 0.0377, pakoreguota p-reikšmė = 0.0612

Vadovai ir Kvalifikuoti darbininkai ir amatininkai: chi tatistika = 4.4465, p-reikšmė = 0.0350, pakoreguota p-reikšmė = 0.0573

Vadovai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 14.1048, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004

Vadovai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 3.4796, p-reikšmė = 0.0621, pakoreguota p-reikšmė = 0.0979

Vadovai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 2.1271, p-reikšmė = 0.1447, pakoreguota p-reikšmė = 0.2101

Kvalifikuoti darbininkai ir amatininkai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi tatistika = 3.8414, p-reikšmė = 0.0500, pakoreguota p-reikšmė = 0.0795

Kvalifikuoti darbininkai ir amatininkai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 0.0296, p-reikšmė = 0.8635, pakoreguota p-reikšmė = 0.9185

Kvalifikuoti darbininkai ir amatininkai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0078, p-reikšmė = 0.7669, pakoreguota p-reikšmė = 0.8538

Paslaugų sektoriaus darbuotojai ir pardavėjai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi tatistika = 4.0953, p-reikšmė = 0.0430, pakoreguota p-reikšmė = 0.0691

Paslaugų sektoriaus darbuotojai ir pardavėjai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.8280, p-reikšmė = 0.3628, pakoreguota p-reikšmė = 0.4556

Įrenginių ir mašinų operatoriai ir surinkėjai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi tatistika = 0.0141, p-reikšmė = 0.9055, pakoreguota p-reikšmė = 0.9451

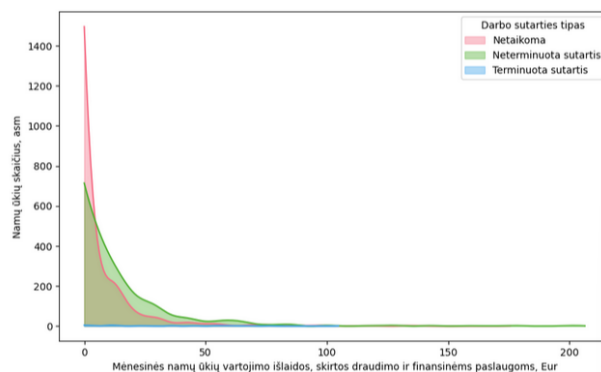
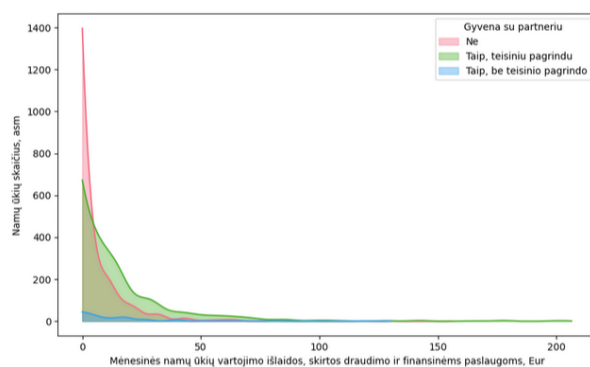
Tikrinamas kintamasis: Šiuo metu mokosi
Ne ir Taip: chi tatistika = 10.1297, p-reikšmė = 0.0015, pakoreguota p-reikšmė = 0.0029

Tikrinamas kintamasis: Darbo sutarties tipas
Netaikoma ir Neterminuota sutartis: chi tatistika = 554.6093, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Netaikoma ir Terminuota sutartis: chi tatistika = 13.3154, p-reikšmė = 0.0003, pakoreguota p-reikšmė = 0.0005

Neterminuota sutartis ir Terminuota sutartis: chi tatistika = 0.1205, p-reikšmė = 0.7284, pakoreguota p-reikšmė = 0.8220

23 Priedas. Skirstinių tankių grafikai, kai priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos



24 Priedas. Šapiro ir Vilko testo rezultatai, kai priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos

Gyvenamoji vieta = 5 didieji miestai: Shapiro-Wilk p = 0.0000	✗	Normalu
Gyvenamoji vieta = Kiti miestai: Shapiro-Wilk p = 0.0000	✗	Normalu
Gyvenamoji vieta = Kaimas: Shapiro-Wilk p = 0.0000	✗	Normalu
Regionas = Sostinės regionas: Shapiro-Wilk p = 0.0000	✗	Normalu
Regionas = Vidurio ir vakarų Lietuvos regionas: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Vienas gyvenantis asmuo: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Kiti namų ūkiai su išlaikomais vaikais: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Pora be išlaikomų vaikų: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Pora su išlaikomais vaikais: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Vienas suaugęs asmuo su išlaikomais vaikais: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio tipas = Kiti namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000	✗	Normalu
Išlaikomų vaikų skaičius = Namų ūkiai be išlaikomų vaikų: Shapiro-Wilk p = 0.0000	✗	Normalu
Išlaikomų vaikų skaičius = Namų ūkiai su 1 išlaikomu vaiku: Shapiro-Wilk p = 0.0000	✗	Normalu
Išlaikomų vaikų skaičius = Namų ūkiai su 2 išlaikomais vaikais: Shapiro-Wilk p = 0.0000	✗	Normalu
Išlaikomų vaikų skaičius = Namų ūkiai su 3 ir daugiau išlaikomų vaikų: Shapiro-Wilk p = 0.0000	✗	Normalu
Socialinė ekonominė grupė = Senatvės pensininkai: Shapiro-Wilk p = 0.0000	✗	Normalu
Socialinė ekonominė grupė = Samdomieji darbuotojai: Shapiro-Wilk p = 0.0000	✗	Normalu
Socialinė ekonominė grupė = Bedarbiai: Shapiro-Wilk p = 0.0000	✗	Normalu
Socialinė ekonominė grupė = Savarankiškai dirbantys asmenys: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų šaltinis = Pensija: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų šaltinis = Darbo užmokestis: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų šaltinis = Pašalpa, stipendija ir kitas pragyvenimo šaltinis: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų šaltinis = Pajamos iš savo ar šeimos verslo: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos amžius = 60 metų ir vyresni: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos amžius = 50-59 metų: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos amžius = 40-49 metų: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos amžius = iki 30 metų: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos amžius = 30-39 metų: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos išsilavinimas = Vidurinis: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos išsilavinimas = Aukštesniojo ir aukštojo nesuteikiantis povidurinis: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos išsilavinimas = Aukštesnysis ir aukštasis: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos išsilavinimas = Pradinis, nebaigę pradinės mokyklos: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio galvos išsilavinimas = Pagrindinis: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų kvintilinė grupė = 2-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų kvintilinė grupė = 1-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų kvintilinė grupė = 5-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų kvintilinė grupė = 4-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000	✗	Normalu
Pajamų kvintilinė grupė = 3-a pajamų kvintilinė grupė: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto tipas = Butas daugiabučiam name: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto tipas = Vieno buto namas: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto tipas = Butas dviejų butų ar sublokuotame name: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto valdos statusas = Būsto savininkas, neturintis išmokėti paskolos: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto valdos statusas = Būsto savininkas, turintis neišmokėtą paskolą: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto valdos statusas = Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto valdos statusas = Gyvena nemokėdamas nuomos: Shapiro-Wilk p = 0.0000	✗	Normalu
Būsto valdos statusas = Nuomojasi šį būstą mažesne nei rinkos kaina: Shapiro-Wilk p = 0.0000	✗	Normalu
Centralizuotas šalto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000	✗	Normalu
Centralizuotas šalto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000	✗	Normalu
Centralizuotas karšto vandens tiekimas = Taip: Shapiro-Wilk p = 0.0000	✗	Normalu
Centralizuotas karšto vandens tiekimas = Ne: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš šilumos tinklų: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis būsto šildymo būdas = Centrinis šildymas iš vietinio šilumos šaltinio: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis būsto šildymo būdas = Nėra stacionaraus šilumos šaltinio: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis būsto šildymo būdas = Krosnis, židiny: Shapiro-Wilk p = 0.0000	✗	Normalu
Asmenų skaičius namų ūkyje = 1-2 asmenys: Shapiro-Wilk p = 0.0000	✗	Normalu
Asmenų skaičius namų ūkyje = 3-4 asmenys: Shapiro-Wilk p = 0.0000	✗	Normalu
Asmenų skaičius namų ūkyje = 5 ir daugiau: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio užimamas naudingasis plotas = <50 m²: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio užimamas naudingasis plotas = 50-99 m²: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio užimamas naudingasis plotas = 100-199 m²: Shapiro-Wilk p = 0.0000	✗	Normalu
Namų ūkio užimamas naudingasis plotas = >200 m²: Shapiro-Wilk p = 0.0002	✗	Normalu
Motinos gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000	✗	Normalu
Motinos gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000	✗	Normalu
Motinos gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0000	✗	Normalu
Gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000	✗	Normalu
Gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinės pilietybės šalis = Lietuva: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinės pilietybės šalis = Kita šalis: Shapiro-Wilk p = 0.0000	✗	Normalu
Tėvo gimimo šalis = Lietuva: Shapiro-Wilk p = 0.0000	✗	Normalu
Tėvo gimimo šalis = Kita šalis: Shapiro-Wilk p = 0.0000	✗	Normalu
Tėvo gimimo šalis = Nenurodė: Shapiro-Wilk p = 0.0000	✗	Normalu
Lytis = Moteris: Shapiro-Wilk p = 0.0000	✗	Normalu
Lytis = Vyras: Shapiro-Wilk p = 0.0000	✗	Normalu
Teisinė santuokinė padėtis = Išituokęs (-usi): Shapiro-Wilk p = 0.0000	✗	Normalu
Teisinė santuokinė padėtis = Našlys (-ė): Shapiro-Wilk p = 0.0000	✗	Normalu
Teisinė santuokinė padėtis = Nevedęs (netekėjusi): Shapiro-Wilk p = 0.0000	✗	Normalu
Teisinė santuokinė padėtis = Vedęs (ištekejusi): Shapiro-Wilk p = 0.0000	✗	Normalu
Gyvena su partneriu = Ne: Shapiro-Wilk p = 0.0000	✗	Normalu
Gyvena su partneriu = Taip, teisiniu pagrindu: Shapiro-Wilk p = 0.0000	✗	Normalu
Gyvena su partneriu = Taip, be teisinio pagrindo: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Netaikoma: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Specialistai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Ištaigų tarnautojai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Technikai ir jaunesnieji specialistai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Nenurodė: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Vadovai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Kvalifikuoti darbininkai ir amatininkai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Paslaugų sektoriaus darbuotojai ir pardavėjai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Įprastinių ir mašinų operatoriai ir surinkėjai: Shapiro-Wilk p = 0.0000	✗	Normalu
Pagrindinis darbas: profesija = Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: Shapiro-Wilk p = 0.0000	✗	Normalu
Šiuo metu mokosi = Ne: Shapiro-Wilk p = 0.0000	✗	Normalu
Šiuo metu mokosi = Taip: Shapiro-Wilk p = 0.0011	✗	Normalu
Darbo sutarties tipas = Netaikoma: Shapiro-Wilk p = 0.0000	✗	Normalu
Darbo sutarties tipas = Neterminuota sutartis: Shapiro-Wilk p = 0.0000	✗	Normalu
Darbo sutarties tipas = Terminuota sutartis: Shapiro-Wilk p = 0.0175	✗	Normalu

25 Priedas. Leveno testo rezultatai, kai priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos

Dispersijų lygybės testas (Levene):

Gyvenamoji vieta: Levene p = 0.0000 Skirtingos dispersijos
Regionas: Levene p = 0.0020 Skirtingos dispersijos
Namų ūkio tipas: Levene p = 0.0000 Skirtingos dispersijos
Išlaikomų vaikų skaičius: Levene p = 0.0000 Skirtingos dispersijos
Socialinė ekonominė grupė: Levene p = 0.0000 Skirtingos dispersijos
Pajamų šaltinis: Levene p = 0.0000 Skirtingos dispersijos
Namų ūkio galvos amžius: Levene p = 0.0000 Skirtingos dispersijos
Namų ūkio galvos išsilavinimas: Levene p = 0.0000 Skirtingos dispersijos
Pajamų kvintilinė grupė: Levene p = 0.0000 Skirtingos dispersijos
Būsto tipas: Levene p = 0.0000 Skirtingos dispersijos
Būsto valdos statusas: Levene p = 0.0000 Skirtingos dispersijos
Centralizuotas šalto vandens tiekimas: Levene p = 0.0328 Skirtingos dispersijos
Centralizuotas karšto vandens tiekimas: Levene p = 0.0000 Skirtingos dispersijos
Pagrindinis būsto šildymo būdas: Levene p = 0.0000 Skirtingos dispersijos
Asmenų skaičius namų ūkyje: Levene p = 0.0000 Skirtingos dispersijos
Namų ūkio užimamas naudingasis plotas: Levene p = 0.0000 Skirtingos dispersijos
Motinos gimimo šalis: Levene p = 0.1552 Vienodos dispersijos
Gimimo šalis: Levene p = 0.0175 Skirtingos dispersijos
Pagrindinės pilietybės šalis: Levene p = 0.2463 Vienodos dispersijos
Tėvo gimimo šalis: Levene p = 0.0426 Skirtingos dispersijos
Lytis: Levene p = 0.0000 Skirtingos dispersijos
Teisinė santuokinė padėtis: Levene p = 0.0000 Skirtingos dispersijos
Gyvena su partneriu: Levene p = 0.0000 Skirtingos dispersijos
Pagrindinis darbas: profesija: Levene p = 0.0000 Skirtingos dispersijos
Šiuo metu mokosi: Levene p = 0.3039 Vienodos dispersijos
Darbo sutarties tipas: Levene p = 0.0000 Skirtingos dispersijos

26 Priedas. Andersono ir Darlingo testo rezultatai kai priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos

Gyvenamoji vieta: Anderson-Darling statistika = 25.659943685457, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Regionas: Anderson-Darling statistika = 6.756461317377376, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Namų ūkio tipas: Anderson-Darling statistika = 228.45410401241233, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Išlaikomų vaikų skaičius: Anderson-Darling statistika = 128.8429804138085, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Socialinė ekonominė grupė: Anderson-Darling statistika = 323.10176754307327, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Pajamų šaltinis: Anderson-Darling statistika = 318.2604190415055, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Namų ūkio galvos amžius: Anderson-Darling statistika = 164.954612435276, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Namų ūkio galvos išsilavinimas: Anderson-Darling statistika = 146.1474414961845, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Pajamų kvintilinė grupė: Anderson-Darling statistika = 144.5037856746223, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Būsto tipas: Anderson-Darling statistika = 88.18923305860754, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Būsto valdos statusas: Anderson-Darling statistika = 69.74787575120278, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Centralizuotas šalto vandens tiekimas: Anderson-Darling statistika = 7.028100855152678, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Centralizuotas karšto vandens tiekimas: Anderson-Darling statistika = 81.05451658778215, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Pagrindinis būsto šildymo būdas: Anderson-Darling statistika = 104.47973709221756, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Asmenų skaičius namų ūkyje: Anderson-Darling statistika = 215.63477540172173, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Namų ūkio užimamas naudingasis plotas: Anderson-Darling statistika = 203.3787355346096, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Motinos gimimo šalis: Anderson-Darling statistika = 2.610833092731109, p-value = 0.024139825867299625
Skirtingos pasiskirstymų formos tarp grupių!
Gimimo šalis: Anderson-Darling statistika = 12.415224258143821, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Pagrindinės pilietybės šalis: Anderson-Darling statistika = -1.0064913448324702, p-value = 0.25
Tėvo gimimo šalis: Anderson-Darling statistika = 3.4402501437551996, p-value = 0.009849577529076559
Skirtingos pasiskirstymų formos tarp grupių!
Lytis: Anderson-Darling statistika = 151.5145949940202, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Teisinė santuokinė padėtis: Anderson-Darling statistika = 231.70198210920853, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Gyvena su partneriu: Anderson-Darling statistika = 294.5815279113811, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Pagrindinis darbas: profesija: Anderson-Darling statistika = 217.39493232931997, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!
Šiuo metu mokosi: Anderson-Darling statistika = 4.507687325855196, p-value = 0.0052330089391711505
Skirtingos pasiskirstymų formos tarp grupių!
Darbo sutarties tipas: Anderson-Darling statistika = 267.03218593637575, p-value = 0.001
Skirtingos pasiskirstymų formos tarp grupių!

27 Priedas. Moodo medianos testo rezultatai, skirti draudimo ir finansinių paslaugų vartojimo išlaidų skirtumams tarp kategorinių požymių grupių vertinti.

Kategorinis kintamasis	χ^2 statistikos reikšmė	p-reikšmė	Požymis yra reikšmingas
Gyvenamoji vieta	42,9500	0,0000	Taip
Regionas	2,5114	0,113	Ne
Namų ūkio tipas	498,2127	0,0000	Taip
Išlaikomų vaikų skaičius	218,4466	0,0000	Taip
Socialinė ekonominė grupė	573,1127	0,0000	Taip
Pajamų šaltinis	581,7603	0,0000	Taip
Namų ūkio galvos amžius	341,1052	0,0000	Taip
Namų ūkio galvos išsilavinimas	241,8934	0,0000	Taip
Pajamų kvintilinė grupė	280,8074	0,0000	Taip
Būsto tipas	145,1574	0,0000	Taip
Būsto valdos statusas	131,1982	0,0000	Taip
Centralizuotas šalto vandens tiekimas	7,4106	0,0065	Taip
Centralizuotas karšto vandens tiekimas	92,9801	0,0000	Taip
Pagrindinis būsto šildymo būdas	208,8342	0,0000	Taip
Asmenų skaičius namų ūkyje	305,8853	0,0000	Taip
Namų ūkio užimamas naudingasis plotas	324,7483	0,0000	Taip
Motinos gimimo šalis	4,1764	0,1239	Ne
Gimimo šalis	8,3916	0,0038	Taip
Pagrindinės pilietybės šalis	0,0000	1,0000	Ne
Tėvo gimimo šalis	8,7578	0,0125	Taip
Lytis	170,268	0,0000	Taip
Teisinė santuokinė padėtis	411,2959	0,0000	Taip
Gyvena su partneriu	430,4937	0,0000	Taip
Pagrindinis darbas: profesija	667,4551	0,0000	Taip
Šiuo metu mokosi	1,9543	0,1621	Ne
Darbo sutarties tipas	392,2796	0,0000	Taip

28 Priedas. Daugkartiniai palyginimai naudojant medianos testą, kai priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos

Tikrinamas kintamasis: Gyvenamoji vieta	
5 didieji miestai ir Kiti miestai: chi tatistika = 2.2775, p-reikšmė = 0.1313, pakoreguota p-reikšmė = 0.1713	
5 didieji miestai ir Kaimas: chi tatistika = 22.3277, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Kiti miestai ir Kaimas: chi tatistika = 39.5326, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Tikrinamas kintamasis: Regionas	
Sostinės regionas ir Vidurio ir vakarų Lietuvos regionas: chi tatistika = 2.5114, p-reikšmė = 0.1130, pakoreguota p-reikšmė = 0.1498	
Tikrinamas kintamasis: Namų ūkio tipas	
Vienas gyvenantis asmuo ir kiti namų ūkiai su išlaikomais vaikais: chi tatistika = 55.4072, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vienas gyvenantis asmuo ir Pora be išlaikomų vaikų: chi tatistika = 253.7193, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vienas gyvenantis asmuo ir Pora su išlaikomais vaikais: chi tatistika = 340.5749, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vienas gyvenantis asmuo ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 64.6183, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vienas gyvenantis asmuo ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 87.8534, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Kiti namų ūkiai su išlaikomais vaikais ir Pora be išlaikomų vaikų: chi tatistika = 5.0045, p-reikšmė = 0.0253, pakoreguota p-reikšmė = 0.0377	
Kiti namų ūkiai su išlaikomais vaikais ir Pora su išlaikomais vaikais: chi tatistika = 5.6718, p-reikšmė = 0.0172, pakoreguota p-reikšmė = 0.0262	
Kiti namų ūkiai su išlaikomais vaikais ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 6.1323, p-reikšmė = 0.0133, pakoreguota p-reikšmė = 0.0205	
Kiti namų ūkiai su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 5.8748, p-reikšmė = 0.0154, pakoreguota p-reikšmė = 0.0235	
Pora be išlaikomų vaikų ir Pora su išlaikomais vaikais: chi tatistika = 90.1373, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Pora be išlaikomų vaikų ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 1.0433, p-reikšmė = 0.3071, pakoreguota p-reikšmė = 0.3637	
Pora be išlaikomų vaikų ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 0.2713, p-reikšmė = 0.6025, pakoreguota p-reikšmė = 0.6576	
Pora su išlaikomais vaikais ir Vienas suaugęs asmuo su išlaikomais vaikais: chi tatistika = 44.5986, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Pora su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 68.7818, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vienas suaugęs asmuo su išlaikomais vaikais ir Kiti namų ūkiai be išlaikomų vaikų: chi tatistika = 0.0290, p-reikšmė = 0.8649, pakoreguota p-reikšmė = 0.8861	
Tikrinamas kintamasis: Išlaikomų vaikų skaičius	
Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 1 išlaikomu vaiku: chi tatistika = 90.0664, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 125.9943, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Namų ūkiai be išlaikomų vaikų ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 21.9412, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Namų ūkiai su 1 išlaikomu vaiku ir Namų ūkiai su 2 išlaikomais vaikais: chi tatistika = 8.4756, p-reikšmė = 0.0036, pakoreguota p-reikšmė = 0.0062	
Namų ūkiai su 1 išlaikomu vaiku ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 0.6099, p-reikšmė = 0.4348, pakoreguota p-reikšmė = 0.4974	
Namų ūkiai su 2 išlaikomais vaikais ir Namų ūkiai su 3 ir daugiau išlaikomų vaikų: chi tatistika = 0.1033, p-reikšmė = 0.7480, pakoreguota p-reikšmė = 0.7956	
Tikrinamas kintamasis: Socialinė ekonominė grupė	
Senatvės pensininkai ir Samdomieji darbuotojai: chi tatistika = 475.5312, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Senatvės pensininkai ir Bedarbiai: chi tatistika = 1.5263, p-reikšmė = 0.2167, pakoreguota p-reikšmė = 0.2680	
Senatvės pensininkai ir Savarankiškai dirbantys asmenys: chi tatistika = 153.5974, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Samdomieji darbuotojai ir Bedarbiai: chi tatistika = 89.0145, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Samdomieji darbuotojai ir Savarankiškai dirbantys asmenys: chi tatistika = 7.6973, p-reikšmė = 0.0055, pakoreguota p-reikšmė = 0.0091	
Bedarbiai ir Savarankiškai dirbantys asmenys: chi tatistika = 104.2594, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Tikrinamas kintamasis: Pajamų šaltinis	
Pensija ir Darbo užmokestis: chi tatistika = 511.5468, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Pensija ir Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 2.0305, p-reikšmė = 0.1542, pakoreguota p-reikšmė = 0.1980	
Pensija ir Pajamos iš savo ar šeimos verslo: chi tatistika = 137.6017, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Darbo užmokestis ir Pašalpa, stipendija ir kitas pragyvenimo šaltinis: chi tatistika = 51.8063, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Darbo užmokestis ir Pajamos iš savo ar šeimos verslo: chi tatistika = 10.8900, p-reikšmė = 0.0010, pakoreguota p-reikšmė = 0.0018	
Pašalpa, stipendija ir kitas pragyvenimo šaltinis ir Pajamos iš savo ar šeimos verslo: chi tatistika = 67.3212, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Tikrinamas kintamasis: Namų ūkio galvos amžius	
60 metų ir vyresni ir 50-59 metų: chi tatistika = 143.5359, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
60 metų ir vyresni ir 40-49 metų: chi tatistika = 159.5961, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
60 metų ir vyresni ir iki 30 metų: chi tatistika = 9.9010, p-reikšmė = 0.0017, pakoreguota p-reikšmė = 0.0030	
60 metų ir vyresni ir 30-39 metų: chi tatistika = 154.6375, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
50-59 metų ir 40-49 metų: chi tatistika = 7.4470, p-reikšmė = 0.0064, pakoreguota p-reikšmė = 0.0102	
50-59 metų ir iki 30 metų: chi tatistika = 4.5406, p-reikšmė = 0.0331, pakoreguota p-reikšmė = 0.0480	
50-59 metų ir 30-39 metų: chi tatistika = 9.5087, p-reikšmė = 0.0020, pakoreguota p-reikšmė = 0.0036	
40-49 metų ir iki 30 metų: chi tatistika = 12.9352, p-reikšmė = 0.0003, pakoreguota p-reikšmė = 0.0006	
40-49 metų ir 30-39 metų: chi tatistika = 0.1246, p-reikšmė = 0.7240, pakoreguota p-reikšmė = 0.7766	
iki 30 metų ir 30-39 metų: chi tatistika = 13.7922, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004	
Tikrinamas kintamasis: Namų ūkio galvos išsilavinimas	
Vidurinis ir Aukštesniojo ir aukštojo nesuteikiantis pavidurinis: chi tatistika = 2.8060, p-reikšmė = 0.0939, pakoreguota p-reikšmė = 0.1284	
Vidurinis ir Aukštesnysis ir aukštasis: chi tatistika = 86.3327, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vidurinis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 32.9911, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	
Vidurinis ir Pagrindinis: chi tatistika = 26.6412, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000	

Vidurinis ir Pagrindinis: chi tatistika = 26.6412, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis ir Aukštesnysis ir aukštasis: chi tatistika = 74.5441, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 36.4540, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesniojo ir aukštojo nesuteikiantis povidurinis ir Pagrindinis: chi tatistika = 47.2093, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis ir Pradinis, nebaigę pradinės mokyklos: chi tatistika = 67.7192, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Aukštesnysis ir aukštasis ir Pagrindinis: chi tatistika = 141.1160, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Pradinis, nebaigę pradinės mokyklos ir Pagrindinis: chi tatistika = 9.1885, p-reikšmė = 0.0024, pakoreguota p-reikšmė = 0.0043

Tikrinamas kintamasis: Pajamų kvintilinė grupė

2-a pajamų kvintilinė grupė ir 1-a pajamų kvintilinė grupė: chi tatistika = 0.0420, p-reikšmė = 0.8376, pakoreguota p-reikšmė = 0.8778

2-a pajamų kvintilinė grupė ir 5-a pajamų kvintilinė grupė: chi tatistika = 204.6708, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 95.0082, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

2-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 35.2853, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 5-a pajamų kvintilinė grupė: chi tatistika = 159.9796, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 83.5040, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 27.8187, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

5-a pajamų kvintilinė grupė ir 4-a pajamų kvintilinė grupė: chi tatistika = 13.2451, p-reikšmė = 0.0003, pakoreguota p-reikšmė = 0.0005

5-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 69.9595, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

4-a pajamų kvintilinė grupė ir 3-a pajamų kvintilinė grupė: chi tatistika = 18.9183, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Būsto tipas

Butas daugiabučiam name ir Vieno buto namas: chi tatistika = 138.5876, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Butas daugiabučiam name ir Butas dviejų butų ar sublokuotame name: chi tatistika = 7.4945, p-reikšmė = 0.0062, pakoreguota p-reikšmė = 0.0181

Butas daugiabučiam name ir Kita gyvenamoji patalpa: chi tatistika = 1.1771, p-reikšmė = 0.2779, pakoreguota p-reikšmė = 0.3339

Vieno buto namas ir Butas dviejų butų ar sublokuotame name: chi tatistika = 6.2559, p-reikšmė = 0.0124, pakoreguota p-reikšmė = 0.0195

Vieno buto namas ir Kita gyvenamoji patalpa: chi tatistika = 1.2821, p-reikšmė = 0.2575, pakoreguota p-reikšmė = 0.3116

Butas dviejų butų ar sublokuotame name ir Kita gyvenamoji patalpa: chi tatistika = 1.3470, p-reikšmė = 0.2458, pakoreguota p-reikšmė = 0.2996

Tikrinamas kintamasis: Būsto valdovė statusas

Būsto savininkas, neturintis išmokėti paskolos ir Būsto savininkas, turintis neišmokėtą paskolą: chi tatistika = 100.0341, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, neturintis išmokėti paskolos ir Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 0.1233, p-reikšmė = 0.7255, pakoreguota p-reikšmė = 0.7766

Būsto savininkas, neturintis išmokėti paskolos ir Gyvena nemokėdamas nuomos: chi tatistika = 13.8659, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004

Būsto savininkas, neturintis išmokėti paskolos ir Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 8.0857, p-reikšmė = 0.0045, pakoreguota p-reikšmė = 0.0075

Būsto savininkas, turintis neišmokėtą paskolą ir Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina: chi tatistika = 4.7199, p-reikšmė = 0.0298, pakoreguota p-reikšmė = 0.0441

Būsto savininkas, turintis neišmokėtą paskolą ir Gyvena nemokėdamas nuomos: chi tatistika = 100.2824, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Būsto savininkas, turintis neišmokėtą paskolą ir Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 39.2898, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina ir Gyvena nemokėdamas nuomos: chi tatistika = 2.5511, p-reikšmė = 0.1102, pakoreguota p-reikšmė = 0.1473

Nuomininkas ar subnuomininkas, mokantis nuomą rinkos kaina ir Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 4.0815, p-reikšmė = 0.0434, pakoreguota p-reikšmė = 0.0619

Gyvena nemokėdamas nuomos ir Nuomojasi ši būstą mažesne nei rinkos kaina: chi tatistika = 1.0727, p-reikšmė = 0.3003, pakoreguota p-reikšmė = 0.3583

Tikrinamas kintamasis: Centralizuotas šalto vandens tiekimas

Taip ir Ne: chi tatistika = 7.4106, p-reikšmė = 0.0065, pakoreguota p-reikšmė = 0.0103

Tikrinamas kintamasis: Centralizuotas karšto vandens tiekimas

Taip ir Ne: chi tatistika = 92.9801, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Tikrinamas kintamasis: Pagrindinis būsto šildymo būdas

Centrinis šildymas iš šilumos tinklu ir Centrinis šildymas iš vietinio šilumos šaltinio: chi tatistika = 185.6823, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Centrinis šildymas iš šilumos tinklu ir Nėra stacionaraus šilumos šaltinio: chi tatistika = 0.7241, p-reikšmė = 0.3948, pakoreguota p-reikšmė = 0.4547

Centrinis šildymas iš šilumos tinklu ir Krosnis, židiny: chi tatistika = 4.5268, p-reikšmė = 0.0334, pakoreguota p-reikšmė = 0.0480

Centrinis šildymas iš vietinio šilumos šaltinio ir Nėra stacionaraus šilumos šaltinio: chi tatistika = 1.5789, p-reikšmė = 0.2089, pakoreguota p-reikšmė = 0.2604

Centrinis šildymas iš vietinio šilumos šaltinio ir Krosnis, židiny: chi tatistika = 96.4317, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

Nėra stacionaraus šilumos šaltinio ir Krosnis, židiny: chi tatistika = 0.8155, p-reikšmė = 0.3665, pakoreguota p-reikšmė = 0.4250

Tikrinamas kintamasis: Asmenų skaičius namų ūkyje

1-2 asmenys ir 3-4 asmenys: chi tatistika = 271.2489, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

1-2 asmenys ir 5 ir daugiau: chi tatistika = 47.9738, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

3-4 asmenys ir 5 ir daugiau: chi tatistika = 1.9390, p-reikšmė = 0.1638, pakoreguota p-reikšmė = 0.2072

Tikrinamas kintamasis: Namų ūkio užimamas naudingasis plotas

<50 m² ir 50-99 m²: chi tatistika = 100.4254, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² ir 100-199 m²: chi tatistika = 264.1065, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

<50 m² ir >200 m²: chi tatistika = 22.1414, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-99 m² ir 100-199 m²: chi tatistika = 136.7309, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

50-99 m² ir >200 m²: chi tatistika = 21.0039, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000

100-199 m² ir >200 m²: chi tatistika = 4.5603, p-reikšmė = 0.0327, pakoreguota p-reikšmė = 0.0479

```
Tikrinamas kintamasis: Motinos gimimo šalis
Lietuva ir Kita šalis: chi statistika = 3.7910, p-reikšmė = 0.0515, pakoreguota p-reikšmė = 0.0723
-----
Lietuva ir Nenurodė: chi statistika = 0.0385, p-reikšmė = 0.8445, pakoreguota p-reikšmė = 0.8778
-----
Kita šalis ir Nenurodė: chi statistika = 0.1271, p-reikšmė = 0.7215, pakoreguota p-reikšmė = 0.7766
-----

Tikrinamas kintamasis: Gimimo šalis
Lietuva ir Kita šalis: chi statistika = 8.3916, p-reikšmė = 0.0038, pakoreguota p-reikšmė = 0.0064
-----

Tikrinamas kintamasis: Pagrindinės pilietybės šalis
Lietuva ir Kita šalis: chi statistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000
-----

Tikrinamas kintamasis: Tėvo gimimo šalis
Lietuva ir Kita šalis: chi statistika = 7.8919, p-reikšmė = 0.0050, pakoreguota p-reikšmė = 0.0083
-----
Lietuva ir Nenurodė: chi statistika = 0.0761, p-reikšmė = 0.7826, pakoreguota p-reikšmė = 0.8272
-----
Kita šalis ir Nenurodė: chi statistika = 0.0025, p-reikšmė = 0.9605, pakoreguota p-reikšmė = 0.9721
-----

Tikrinamas kintamasis: Lytis
Motėris ir Vyras: chi statistika = 170.2680, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----

Tikrinamas kintamasis: Teisinė santuokinė padėtis
Išituokęs (-usi) ir Našlys (-ė): chi statistika = 24.1145, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Išituokęs (-usi) ir Nevedęs (netekėjusi): chi statistika = 1.7351, p-reikšmė = 0.1878, pakoreguota p-reikšmė = 0.2358
-----
Išituokęs (-usi) ir Vedęs (ištekęjusi): chi statistika = 164.1097, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Našlys (-ė) ir Nevedęs (netekėjusi): chi statistika = 36.9679, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Našlys (-ė) ir Vedęs (ištekęjusi): chi statistika = 302.9318, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Nevedęs (netekėjusi) ir Vedęs (ištekęjusi): chi statistika = 79.9996, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----

Tikrinamas kintamasis: Gyvena su partneriu
Ne ir Taip, teisinio pagrindu: chi statistika = 410.6620, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Ne ir Taip, be teisinio pagrindo: chi statistika = 58.6138, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Taip, teisinio pagrindu ir Taip, be teisinio pagrindo: chi statistika = 0.0376, p-reikšmė = 0.8463, pakoreguota p-reikšmė = 0.8778
-----

Tikrinamas kintamasis: Pagrindinis darbas: profesija
Netaikoma ir Specialistai: chi statistika = 301.5263, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Įstaigų tarnautojai: chi statistika = 34.4386, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Technikai ir jaunesnieji specialistai: chi statistika = 58.8446, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----

-----
Netaikoma ir Nenurodė: chi statistika = 34.8958, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Vadovai: chi statistika = 136.6145, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 157.7672, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 67.3250, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 115.8522, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 42.9690, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Specialistai ir Įstaigų tarnautojai: chi statistika = 0.3431, p-reikšmė = 0.5581, pakoreguota p-reikšmė = 0.6131
-----
Specialistai ir Technikai ir jaunesnieji specialistai: chi statistika = 0.8977, p-reikšmė = 0.3434, pakoreguota p-reikšmė = 0.4010
-----
Specialistai ir Nenurodė: chi statistika = 63.3888, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Specialistai ir Vadovai: chi statistika = 15.5216, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002
-----
Specialistai ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 2.6558, p-reikšmė = 0.1032, pakoreguota p-reikšmė = 0.1389
-----
Specialistai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 14.1256, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0003
-----
Specialistai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 2.9067, p-reikšmė = 0.0882, pakoreguota p-reikšmė = 0.1217
-----
Specialistai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 0.0206, p-reikšmė = 0.8859, pakoreguota p-reikšmė = 0.9021
-----
Įstaigų tarnautojai ir Technikai ir jaunesnieji specialistai: chi statistika = 0.0000, p-reikšmė = 1.0000, pakoreguota p-reikšmė = 1.0000
-----
Įstaigų tarnautojai ir Nenurodė: chi statistika = 34.9478, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0002
-----
Įstaigų tarnautojai ir Vadovai: chi statistika = 9.0528, p-reikšmė = 0.0026, pakoreguota p-reikšmė = 0.0046
-----
Įstaigų tarnautojai ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 0.4219, p-reikšmė = 0.5160, pakoreguota p-reikšmė = 0.5783
-----
Įstaigų tarnautojai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 1.4441, p-reikšmė = 0.2295, pakoreguota p-reikšmė = 0.2818
-----
Įstaigų tarnautojai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 0.3895, p-reikšmė = 0.5325, pakoreguota p-reikšmė = 0.5800
-----
Įstaigų tarnautojai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 0.5580, p-reikšmė = 0.4551, pakoreguota p-reikšmė = 0.5170
-----
Technikai ir jaunesnieji specialistai ir Nenurodė: chi statistika = 24.6518, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Technikai ir jaunesnieji specialistai ir Vadovai: chi statistika = 16.1216, p-reikšmė = 0.0001, pakoreguota p-reikšmė = 0.0001
-----
Technikai ir jaunesnieji specialistai ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 0.4010, p-reikšmė = 0.5261, pakoreguota p-reikšmė = 0.5858
-----
Technikai ir jaunesnieji specialistai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 3.0970, p-reikšmė = 0.0784, pakoreguota p-reikšmė = 0.1092
-----
Technikai ir jaunesnieji specialistai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 0.5463, p-reikšmė = 0.4598, pakoreguota p-reikšmė = 0.5189
-----
Technikai ir jaunesnieji specialistai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 0.9873, p-reikšmė = 0.3204, pakoreguota p-reikšmė = 0.3768
-----
Nenurodė ir Vadovai: chi statistika = 83.3446, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Nenurodė ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 29.3463, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Nenurodė ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 13.4397, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0005
-----
Nenurodė ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 24.6477, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----

-----
Nenurodė ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 24.6422, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Nenurodė ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 22.0083, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Vadovai ir Kvalifikuoti darbininkai ir amatininkai: chi statistika = 15.5785, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Vadovai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 44.1505, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Vadovai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 33.9541, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Vadovai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 5.6006, p-reikšmė = 0.0175, pakoreguota p-reikšmė = 0.0264
-----
Kvalifikuoti darbininkai ir amatininkai ir Paslaugų sektoriaus darbuotojai ir pardavėjai: chi statistika = 3.8408, p-reikšmė = 0.0500, pakoreguota p-reikšmė = 0.0708
-----
Kvalifikuoti darbininkai ir amatininkai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 0.0206, p-reikšmė = 0.8835, pakoreguota p-reikšmė = 0.8861
-----
Kvalifikuoti darbininkai ir amatininkai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 2.2197, p-reikšmė = 0.1363, pakoreguota p-reikšmė = 0.1764
-----
Paslaugų sektoriaus darbuotojai ir pardavėjai ir Įrenginių ir mašinų operatoriai ir surinkėjai: chi statistika = 2.7947, p-reikšmė = 0.0046, pakoreguota p-reikšmė = 0.1284
-----
Paslaugų sektoriaus darbuotojai ir pardavėjai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 7.4542, p-reikšmė = 0.0063, pakoreguota p-reikšmė = 0.0182
-----
Įrenginių ir mašinų operatoriai ir surinkėjai ir Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbuotojai: chi statistika = 2.3817, p-reikšmė = 0.1228, pakoreguota p-reikšmė = 0.1614
-----

-----
Tikrinamas kintamasis: Šiuo metu mokosi
Ne ir Taip: chi statistika = 1.9543, p-reikšmė = 0.1621, pakoreguota p-reikšmė = 0.2067
-----

-----
Tikrinamas kintamasis: Darbo sutarties tipas
Netaikoma ir Neterminuota sutartis: chi statistika = 302.3534, p-reikšmė = 0.0000, pakoreguota p-reikšmė = 0.0000
-----
Netaikoma ir Terminuota sutartis: chi statistika = 13.6077, p-reikšmė = 0.0002, pakoreguota p-reikšmė = 0.0004
-----
Neterminuota sutartis ir Terminuota sutartis: chi statistika = 6.1650, p-reikšmė = 0.0130, pakoreguota p-reikšmė = 0.0203
-----
```

29 Priedas. Bendrų vartojimo išlaidų aprašomoji statistika pagal kategorinius kintamuosius

Lietuvos namų ūkių pasiskirstymo pagal gyvenamąją vietą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q ₁)	Trečiasis kvantilis (Q ₃)	Maksimali reikšmė (Eur)
5 didieji miestai	1507	825,1	658,4	622,3	56,2	399,0	1051,4	4491,2
Kaimas	1594	804,7	646,9	628,6	83,6	388,9	1032,0	7099,3
Kiti miestai	1233	737,7	578,1	586,3	64,0	359,7	912,3	7984,8

Lietuvos namų ūkių pasiskirstymo pagal regioną ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q ₁)	Trečiasis kvantilis (Q ₃)	Maksimali reikšmė (Eur)
Sostinės regionas	1003	824,5	613,8	707,8	56,2	359,5	1054,4	7099,3
Vidurio ir vakarų Lietuvos regionas	3331	783,2	636,6	584,7	64,0	389,4	1009,5	7984,8

Lietuvos namų ūkių pasiskirstymo pagal namų ūkio tipą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q ₁)	Trečiasis kvantilis (Q ₃)	Maksimali reikšmė (Eur)
Kiti namų ūkiai be išlaikomų vaikų	500	866,1	718,6	673,3	106,2	459,6	1048,5	7984,8
Kiti namų ūkiai su išlaikomais vaikais	153	1174,7	1002,4	674	229,2	704,1	1418,7	3891,8
Pora be išlaikomų vaikų	1242	859,1	714,9	580,4	82,2	480,7	1067,1	6031,3
Pora su išlaikomais vaikais	498	1334,9	1177,8	753	184,8	824	1618,5	5474,2
Vienas gyvenantis asmuo	1763	521,8	422,8	391,5	56,2	282	644,5	7099,3
Vienas suaugęs asmuo su išlaikomais vaikais	178	962,2	824,4	607,3	174,8	529	1189,4	4083,2

Lietuvos namų ūkių pasiskirstymo pagal išlaikomų vaikų skaičių ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupė	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q ₁)	Trečiasis kvantilis (Q ₃)	Maksimali reikšmė (Eur)
Namų ūkiai be išlaikomų vaikų	3505	690,4	553,2	538,3	56,2	349,1	867,9	7984,8
Namų ūkiai su 1 išlaikomu vaiku	446	1107,5	957,7	687,9	174,8	645,4	1365,7	5474,2

Namų ūkiai su 2 išlaikomais vaikais	285	1323,4	1165,3	740,5	184,8	780,9	1589,4	4475,2
Namų ūkiai su 3 ir daugiau išlaikomų vaikų	98	1476,1	1321	742,1	305	1001,8	1793,8	4491,2

Lietuvos namų ūkių pasiskirstymo pagal socialinę – ekonominę grupę ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupė	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvantilis (Q3)	Maksimali reikšmė (Eur)
Bedarbiai	352	627,8	502,7	452,5	56,2	326,1	786,4	2581,1
Samdomieji darbuotojai	1939	1005,6	844	674,7	106,4	566,8	1270,7	7984,8
Savarankiškai dirbantys asmenys	276	1222,8	1018,5	845,1	171,5	712,9	1445,8	6031,3
Senatvės pensininkai	1767	524,9	436,8	349,8	64	285,4	659,1	3744,5

Lietuvos namų ūkių pasiskirstymo pagal pajamų šaltinį ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvantilis (Q3)	Maksimali reikšmė (Eur)
Darbo užmokestis	2013	1011,5	854,8	660,7	106,4	567,8	1276,3	7099,3
Pajamos iš savo ar šeimos verslo	195	1230	1000,6	992,7	176,8	692,3	1462,7	7984,8
Pašalpa, stipendija ir kitas pragyvenimo šaltinis	225	644	486,8	495,4	56,2	317,5	830,8	2920,2
Pensija	1901	533,9	441,8	363,8	64	290,2	668,2	3744,5

Lietuvos namų ūkių pasiskirstymo pagal namų ūkio galvos išsilavinimo lygį ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvantilis (Q3)	Maksimali reikšmė (Eur)
Aukštesniojo ir aukštojo nesuteikiantis povidurinis	1299	718,6	589,2	545,9	92,7	374,7	890,7	7984,8
Aukštesnysis ir aukštasis	1590	989,4	797,4	712	56,2	492,3	1266,3	7099,3
Pagrindinis	355	531,2	413,2	415	83,6	261,3	655,5	3744,5
Pradinis, nebaigę pradinės mokyklos	123	395,6	328,4	267,2	82,2	211,4	498	1547,9
Vidurinis	967	715,4	587,1	520,4	87,1	354,2	924,9	5416,4

Lietuvos namų ūkių pasiskirstymo pagal namų ūkių būsto tipą bei valdos statusą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Būsto tipas	Butas daugiabučiame name	2425	744,2	587,2	551,2	56,2	365	961,6	5474,2
	Butas dviejų butų ar sublokuotame name	293	905,1	691,5	791,4	105,2	360,7	1109,3	4544,2
	Kita gyvenamoji patalpa	3	177,8	171,4	14,1	168	169,7	182,6	193,9
	Vieno buto namas	1613	846,4	691,0	662,4	64	418,1	1089,6	7984,8
Būsto valdos statusas	Būsto savininkas, neturintis išmokėti paskolos	3828	762,2	613,3	578,3	56,2	373,2	973,9	7984,8
	Būsto savininkas, turintis neišmokėtą paskolą	206	1421,2	1182,6	941,4	125,9	777,8	1788,5	7099,3
	Gyvena nemokėdamas nuomos	213	705,5	563,4	531,2	85,5	341,5	919,8	4083,2
	Nuomininkas ar subnuomininkas, mokantis nuomą	33	1215,1	920,1	768,9	263,9	584,4	1795,7	3117,8
	Nuomojasi šį būstą mažesne nei rinkos kaina	54	645,1	551,1	380	148,7	390,4	758,9	1811,6

Lietuvos namų ūkių pasiskirstymo pagal karšto bei šalto vandens tiekimą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Centralizuotas šalto vandens tiekimas	Ne	877	787,2	619,6	670,2	56,2	358,9	1023,1	7099,3
	Taip	3457	794,1	632,1	601	82,2	389,8	1015,6	7984,8
Centralizuotas karšto vandens tiekimas	Ne	2276	833,8	672,8	657,2	56,2	401,1	1067,9	7984,8
	Taip	2058	747,3	585,5	562,6	82,2	364,6	950	5474,2

Lietuvos namų ūkių pasiskirstymo pagal pagrindinį būsto šildymo būdą bei namų ūkio asmenų skaičių ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Pagrindinis būsto	Centrinis šildymas iš viešinio	1526	917,5	747,4	704,2	85,5	478,6	1132,2	7984,8

šildymo būdas	šilumos šaltinio								
	Centrinis šildymas iš šilumos tinklu	2157	750,4	595,1	556,8	82,2	368,6	959,4	5474,2
	Krosnis, židinys	635	633,3	453,3	514,7	64	281	787,8	3891,8
	Nėra stacionarus šilumos šaltinio	16	929,1	835,9	592,3	56,2	509,6	1194,6	2116,7

Lietuvos namų ūkių pasiskirstymo pagal lytį ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Lytis	Moterys	2349	675,7	532,4	496,4	56,2	338,9	862,7	4083,2
	Vyras	1985	931,3	753,1	707,4	82,2	476,6	1168	7984,8

Lietuvos namų ūkių pasiskirstymo pagal motinos, tėvo, namų ūkio gimimo šalis, pagrindinės pilietybės šalis ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Motinos gimimo šalis	Kita šalis	287	782,9	561,6	578,2	105,2	368,3	1053,7	3188,8
	Lietuva	4019	793,3	631,4	619,4	56,2	385	1015,2	7984,8
	Nenurodė	28	814,2	810,5	419,4	171,5	482,4	1026,4	1697,8
Gimimo šalis	Kita šalis	336	739,7	547,1	578,3	105,2	354,2	936,3	4544,2
	Lietuva	3998	797,2	636,9	618,4	56,2	386,3	1020,5	7984,8
Pagrindinės pilietybės šalis	Kita šalis	36	1037,5	956,8	610,4	131,7	637,8	1376,7	2864
	Lietuva	4298	790,7	626,7	615,3	56,2	381,6	1013,5	7984,8
Tėvo gimimo šalis	Kita šalis	313	764,7	578,0	544,3	105,2	381	1000	3903,1
	Lietuva	3979	795,4	633,4	622,5	56,2	383,2	1017,9	7984,8
	Nenurodė	42	749,5	703,6	431,8	171,5	396,8	1012,4	1795,7

Lietuvos namų ūkių pasiskirstymo pagal teisinę santuokinę padėtį, gyvenimą su partneriu ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Mediana (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Teisinė santuokinė padėtis	Išsituokęs (-usi)	832	634,3	489,4	483,3	56,2	313,5	794,5	3903,1
	Našlys (-ė)	796	483,5	403,5	316,2	87,7	273,1	596,3	2542,7
	Nevedęs (netekėjęs)	585	733,2	597,4	587,1	64	354,8	940,1	7099,3

	Vedęs (ištekejusi)	2121	987,4	825,8	682,5	82,2	540,8	1233,4	7984,8
Gyvena su partneriu	Ne	2321	596,4	471	454,2	56,2	308,6	747,1	7099,3
	Taip, be teisinio pagrindo	171	1027,4	859,2	707,1	139,1	549,5	1278,5	3828
	Taip, teisiniu pagrindu	1842	1018,4	860,7	693,5	82,2	570,2	1267,7	7984,8

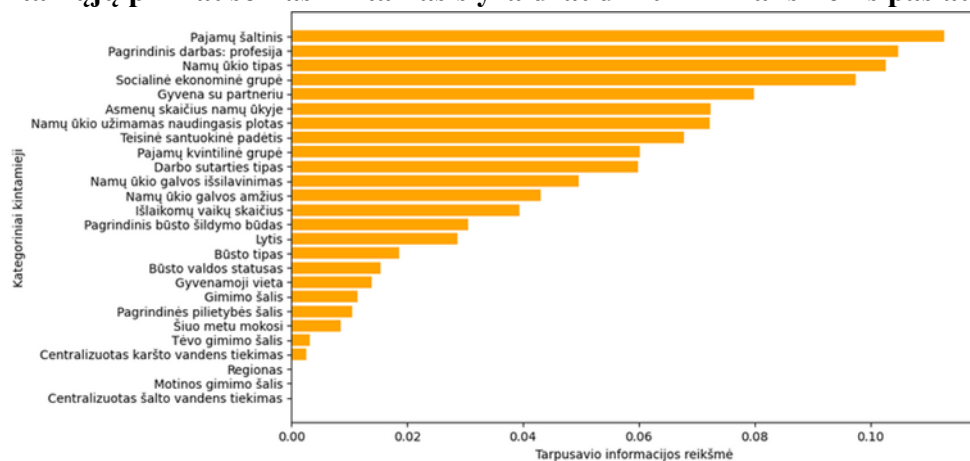
Lietuvos namų ūkių pasiskirstymo pagal pagrindinį darbą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Grupė	N	Vidurkis (Eur)	Medianas (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Kvalifikuoti darbininkai ir amatininkai	368	1025,3	892,6	617,9	125,9	625,9	1278,3	5259
Kvalifikuoti žemės, miškų ir žuvininkystės ūkio darbininkai	96	1188,9	996,1	938,2	176,8	635,1	1455,7	6031,3
Nenurodė	274	743,4	639,6	452,2	106,4	414,9	937,2	2920,2
Netaikoma	2119	541,9	446,4	370,7	56,2	295,0	677,3	3744,5
Paslaugų sektoriaus darbuotojai ir pardavėjai	246	807,9	695,5	494,8	113,4	454	1041,6	4041,5
Specialistai	550	1113,3	940,8	733,4	150	637,2	1410,6	7099,3
Technikai ir jaunesnieji specialistai	133	1096,9	962,3	572,2	195	653,5	1485,9	2756,4
Vadovai	192	1476,2	1185,6	1030,5	171,5	780,8	1888,9	7984,8
Įrenginių ir mašinų operatoriai ir surinkėjai	272	938,6	816,7	521,7	190,8	574,1	1175,9	3331,0
Įstaigų tarnautojai	84	1149,1	938,3	792,4	259	608,5	1391,6	4401,2

Lietuvos namų ūkių pasiskirstymo pagal mokymosi statusą, darbo sutarties tipą ir bendras mėnesines vartojimo išlaidas 2021 m. pagrindinė aprašomoji statistika

Kategorinis kintamasis	Grupės	N	Vidurkis (Eur)	Medianas (Eur)	Standartinis nuokrypis (Eur)	Minimali reikšmė (Eur)	Pirmasis kvartilis (Q1)	Trečiasis kvartilis (Q3)	Maksimali reikšmė (Eur)
Šiuo metu mokosi	Ne	4301	791,1	627,3	609,3	56,2	381,7	1016,2	7984,8
	Taip	33	1005,5	788,8	1172,7	290,5	486,7	1055,8	7099,3
Darbo sutarties tipas	Netaikoma	2395	620,4	479,7	500,9	56,2	309,5	768	6031,3
	Neterminuota sutartis	1911	1001,0	840,0	672,0	106,4	562,8	1261,4	7984,8
	Terminuota sutartis	28	1318,0	1152,4	791,4	162,5	835,6	1756,8	3188,8

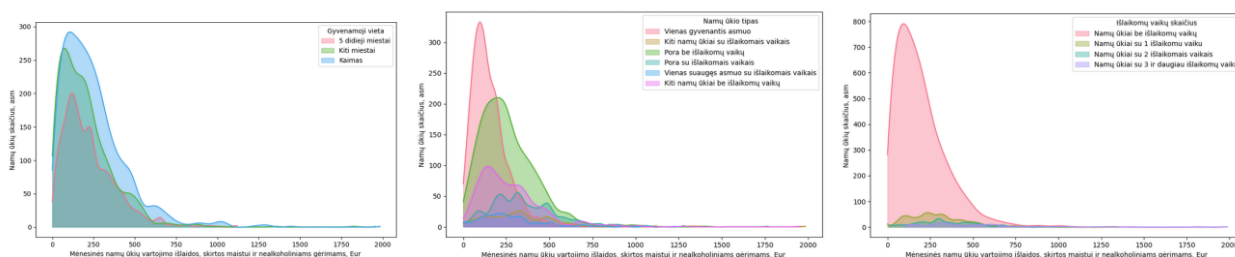
30 Priedas. Tarpusavio informacijos metodo panaudojimas požymių atrankai, kai kategorinių kintamųjų priklausomas kintamasis yra draudimui ir finansinėms paslaugoms skirtos išlaidos



31 Priedas. Maisto ir nealkoholinių gėrimų, informacijos ir ryšių bei draudimo ir finansinių paslaugų vartojimo išlaidas veikiančių požymių identifikavimas

Maisto ir nealkoholinių gėrimų vartojimo išlaidas veikiančių požymių identifikavimas

Kaip ir bendrųjų išlaidų atveju, siekiant korektiškai taikyti dispersinės analizės Fišerio statistikos testą analizuojamiems duomenims, t. y., kategorinių kintamųjų grupių pasiskirstymui pagal maistui ir nealkoholiniams gėrimams skirtos išlaidos, turi būti patikrintos normalumo ir dispersijų panašumo prielaidos. Prielaidų neatitikimas matyti ir vizualiai, kai, remiantis tankio grafikais, didžiosios daugumos kategorinių kintamųjų pasiskirstymai tarp grupių pagal išlaidas, skirtas maistui ir nealkoholiniams gėrimams, yra susikcentravę dešinėje pusėje, t. y., duomenys yra asimetriški, artimi nulinei reikšmei bei turi ilgą uodegą, kas neatitinka įprasto normaliojo pasiskirstymo varpo formos.



Keletas maisto ir nealkoholinių gėrimų vartojimo išlaidų tankių grafikų pagal kategorinių kintamųjų grupes

Skirstinių uodegos ir skirtingas jų plotis rodo, kad duomenų dispersija tarp grupių skiriasi, kas gali lemti dispersijų homogeniškumo prielaidos pažeidimą. Atliktas papildomas Šapiro ir Vilko testas (žr. 12 priedą) parodė, kad maistui ir nealkoholiniams gėrimams skirtų išlaidų pasiskirstymas pagal daugumos kategorinių požymių lygius, neatitiko normalumo prielaidos, nes Šapiro ir Vilko testo p -reikšmė buvo lygi 0,000, kas yra mažiau už 5% reikšmingumo lygmenį. Normalumo prielaidą tenkino tik namų ūkių, kurie neturėjo stacionaraus šilumos šaltinio (p -reikšmė = 0,303), nenurodė savo motinos bei tėvo gimimo šalies (atitinkamai p -reikšmės 0,756 ir 0,08), apklausos metu mokėsi (p -reikšmė = 0,564) ir dirbo pagal terminuotą darbo sutartį (p -reikšmė = 0,054), grupės. Verta pabrėžti, kad šiose grupėse taip pat buvo mažai stebinių, o tai galėjo paveikti testo rezultatus. Kadangi yra siekiama atlikti bendrą požymių atranką, yra laikomasi prielaidos, kad visos stebimos grupės bendrai tenkintų taikomų metodų sąlygas. Tuo tarpu patikrinus dispersijų homogeniškumo prielaidą naudojant statistinį Leveno testą (žr. 13 priedą), buvo patvirtinta, kad grupių dispersijos nėra panašios. Leveno testo rezultatai parodė, kad tarp didžiosios daugumos kategorinių kintamųjų grupių bent viena pora turėjo statistiškai reikšmingai skirtingas dispersijas, nes Leveno testo statistikos p -reikšmė buvo mažesnė už pasirinktą 5 % reikšmingumo lygmenį. Dviejų svarbiausių prielaidų pažeidimas reiškia, kad dispersinės analizės metodo taikymas būtų netinkamas.

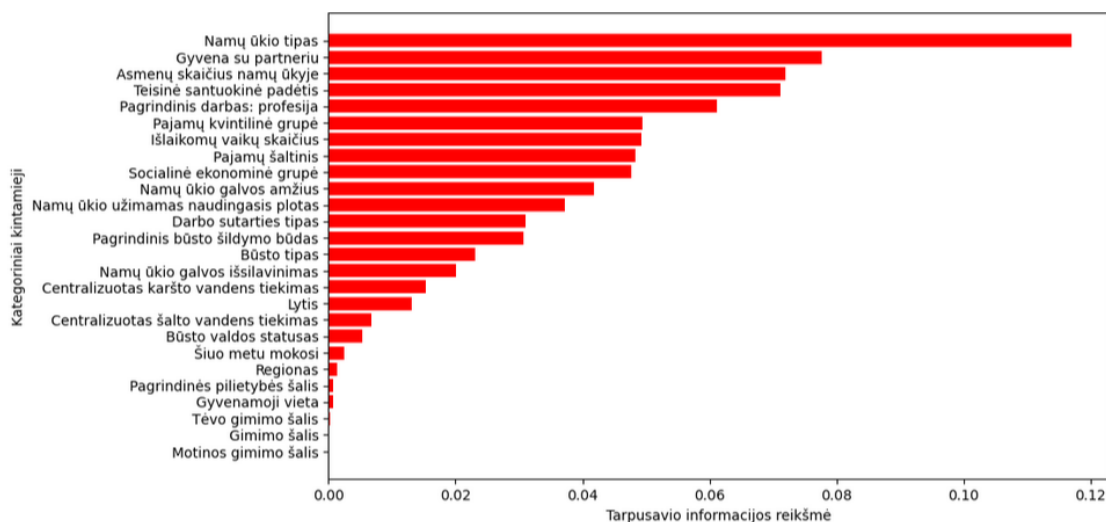
Siekiant požymių atrankai pritaikyti Kruskalio ir Voliso testą, prieš tai, naudojant k -imčių Andersono ir Darlingo testą (žr. 14 priedą), buvo patikrinta neparametrinių testų taikymui reikalinga grupių pasiskirstymų formų panašumo prielaida. Andersono ir Darlingo testu buvo tikrinta nulinė hipotezė, teigianti, kad visų imčių skirstiniai yra panašūs. Buvo priimta išvada, kad skirstiniai nėra panašios formos, nes daugumos kategorinių kintamųjų bent vienos grupės duomenų skirstinys skyrėsi nuo kitų grupių skirstinių, pagal gautas Andersono ir Darlingo statistikos p -reikšmes, išskyrus „tėvo gimimo

šalis“ ir „motinos gimimo šalis“ kintamųjų grupės, buvo mažesnės už pasirinktą 5 % reikšmingumo lygmenį. Kruskalio ir Voliso testo taikymas šiuo atveju būtų taip pat nepagrįstas.

Reikšmingų požymių, skirtų maistui ir nealkoholiniams gėrimams išlaidų dydžio nustatymui, atrankai buvo panaudotas kitas neparametrinis metodas – Moodo medianos testas. Taikant testą buvo formuojama nulinė hipotezė, teigianti, jog skirtingų kategorijų grupių medianos pagal vartojimo išlaidų dydį, skirtą maistui ir nealkoholiniams gėrimams, statistiškai reikšmingai nesiskiria. Remiantis testo rezultatais (žr. 15 priedą), galima teigti, kad visų kategorinių kintamųjų, išskyrus „regionas“, „tėvo gimimo šalis“, „motinos gimimo šalis“ ir „šiuo metu mokosi“, medianos tarp grupių statistiškai reikšmingai skyrėsi, nes gautos χ^2 statistikos p-reikšmės buvo mažesnės už 5 % reikšmingumo lygį. Verta paminėti, kad kuo didesnė yra χ^2 statistikos reikšmė, tuo yra labiau tikėtina, kad kategorinio kintamojo grupės tikrai skiriasi pagal išlaidų medianą, o ne tik atsitiktinai, kas kartu rodo ir to požymio didesnę poveikį maisto ir nealkoholinių gėrimų vartojimo išlaidoms.

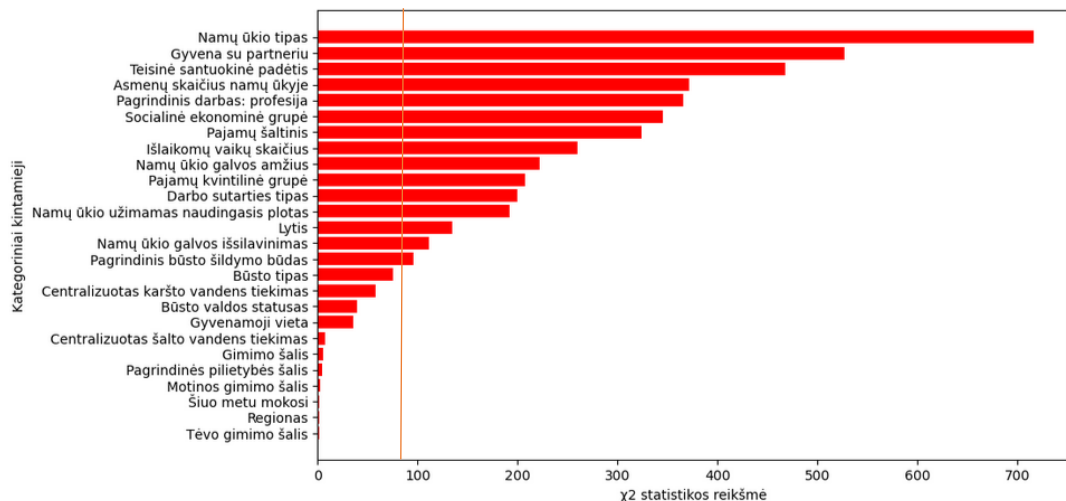
Daugybiniai poriniai palyginimai taip pat buvo atlikti naudojant porinius Moodo medianos testus. Verta paminėti, kad atliekant daug testų vienu metu p-reikšmės buvo taikoma Benjamino ir Hochbergo korekcija (žr. 16 priedą). Įdomu tai, kad Moodo medianos poriniai testai pagal maisto ir alkoholinių gėrimų vartojimo išlaidas parodė, jog namų ūkiai, kurie gyvena poroje ir tie, kurie gyvena vieni, bet su išlaikomais vaikais, maistui ir nealkoholiniams gėrimams pagal medianą išleidžia labai panašią dalį savo biudžeto, t. y., statistiškai reikšmingai tarpusavyje nesiskiria (χ^2 statistikos pakoreguota p-reikšmė didesnė nei 5 % reikšmingumo lygmuo). Statistiškai reikšmingų skirtumų pagal maistui ir nealkoholiniams gėrimams išleidžiamas sumas nebuvo nustatyta ir tarp namų ūkių, kurie šildosi naudodami šilumos tinklų paslaugas ir tų, kurie turi nuosavą krosnį ar židinį.

TI metodo rezultatai interpretuojami taip, kad kuo didesnė yra TI mato reikšmė, tuo svarbesnis priklausomam kintamajam yra ir požymis.



Tarpusavio informacijos metodo panaudojimas požymių atrankai, kai vertinamos maisto ir nealkoholinių gėrimų vartojimo išlaidos

Remiantis TI mato rezultatais, galima teigti, kad namų ūkių tipas, t. y., šeimyninė padėtis, yra svarbiausias kategorinis kintamasis turintis įtakos maisto ir nealkoholinių gėrimų vartojimui. Palyginimui, buvo vizualizuotos ir χ^2 statistikos reikšmės kurių dydžiai turėtų padėti identifikuoti svarbiausius vartojimo išlaidų pokyčius veikiančius požymius.



Moodo medianos testo panaudojimas požymių atrankai, kai vertinamos maisto ir nealkoholinių gėrimų vartojimo išlaidos

Abu metodai sutinka, kad pagal aukščiausius reikšmių (tiek TI, tiek χ^2 statistikos) dydžius svarbiausias maisto ir nealkoholinių gėrimų vartojimo išlaidas veikiantis kintamasis yra namų ūkio tipas (vienas gyvenantis asmuo, vaikus auginantis asmuo, poros su ar be išlaikomų vaikų ir kt.). Taip pat labai svarbus yra ir namų ūkyje esantis asmenų kiekis, gyvenimo su ar be partnerio statusas. TI metodas nurodo, kad vertinant namų ūkių išlaidas maistui ir nealkoholiniams gėrimams yra svarbu įvertinti ir asmenų santuokinę padėtį bei profesiją, kurios yra panašaus svarbumo kaip ir asmenų skaičius namų ūkyje, nors medianos testo rezultatai šiuos požymius vertina kaip mažiau aktualius. Abu metodai leidžia priimti panašias išvadas ir dėl mažiausiai svarbių kintamųjų, kurie yra susiję su asmens ar jo tėvų kilme, nes abiejų metodų reikšmės šiais atvejais yra artimos nuliui. Galutinės požymių atrankos išvados buvo pateiktos naudojant Moodo medianos testo rezultatus dėl anksčiau paminėtos galimos priklausomybės pervertinimo rizika tarpusavio informacijos metodo taikymo atveju.

Remiantis Moodo medianos testo rezultatais, nustatyta, kad mažiausiai informacijos apie namų ūkių išlaidas, skirtas maistui ir nealkoholiniams gėrimams, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 80 (žr. 15 priedą) t. y., „tėvo gimimo šalis“, „motinos gimimo šalis“, „gimimo šalis“, „regionas“, „šiuo metu mokosi“, „pagrindinės pilietybės šalis“, „centralizuotas šalto vandens tiekimas“, „centralizuotas karšto vandens tiekimas“, „būsto tipas“, „gyvenamoji vieta“, „būsto valdos statusas“. Dėl šios priežasties išvardyti kintamieji, kuriant modelius, turėtų būti vertinami kaip pertekliniai ir šalinami iš tolimesnės analizės.

Informacijos ir ryšių išlaidas veikiančių požymių identifikavimas

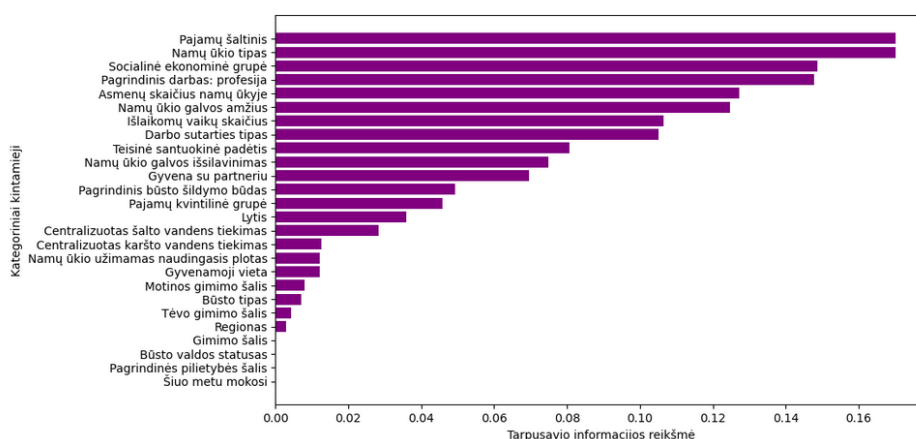
Prieš taikant įvairius statistinius testus mėnesinės namų ūkių išlaidos, skirtos informacijai ir ryšiams, buvo taip pat įvertintos vizualiniu būdu. Iš tankio grafikų (žr. 17 priedą) buvo pastebėta, kad namų ūkių išlaidų ryšiams ir informacijai pasiskirstymas, vertinant jį pagal namų ūkių pajamų šaltinį, amžių ir kitus likusius panašiai išsidėsčiusius kategorinius kintamuosius, tarp grupių pasižymi dešiniąja asimetrija, ilgomis teigiamomis uodegomis, skirtingu modalumu ir plačia dispersija. Toks pasiskirstymas, kaip ir kitų išlaidų kategorijų atveju, lėmė, jog ir šį kartą dispersinės analizės metodai nebuvo taikyti dėl normalumo prielaidos pažeidimo (Šapiro ir Vilko testas, žr. 18 priedą) ir statistškai reikšmingų dispersijų skirtumų tarp grupių (Leveno testas, žr. 19 priedą). Kruskalio ir Voliso testas taip pat nebuvo taikytas, nes Andersono ir Darlingo testo rezultatai (žr. 20 priedą), parodė, kad

didžiosios daugumos kategorinių kintamųjų bent vienos grupės skirstinys, pagal informacijai ir ryšiams skirtos išlaidos, statistiškai reikšmingai skyrėsi tarp lyginamųjų grupių. Dėl šių priežasčių požymių atrankai buvo taikytas TI matas ir Moodo medianos testas, kuriuo yra testuojama nulinė hipotezė, teigianti, jog kategorinių kintamųjų grupių medianos pagal informacijai ir ryšiams skiriamas išlaidas statistiškai reikšmingai nesiskiria.

Moodo medianos testo rezultatai (žr. 21 priedą) rodo, kad 20 iš 26 kategorinių kintamųjų išlaidų, skirtų informacijai ir ryšiams, medianos tarp grupių statistiškai reikšmingai skyrėsi, nes gautos χ^2 statistikos p-reikšmės buvo mažesnės už 5 % reikšmingumo lygį. Verta paminėti, kad kuo didesnė yra χ^2 statistikos reikšmė, tuo labiau tikėtina, jog kategorinio kintamojo grupės tikrai skiriasi pagal išlaidų medianą, o ne tik atsitiktinai, kas kartu rodo ir to požymio didesnę poveikį išlaidoms, skiriamoms informacijai ir ryšiams.

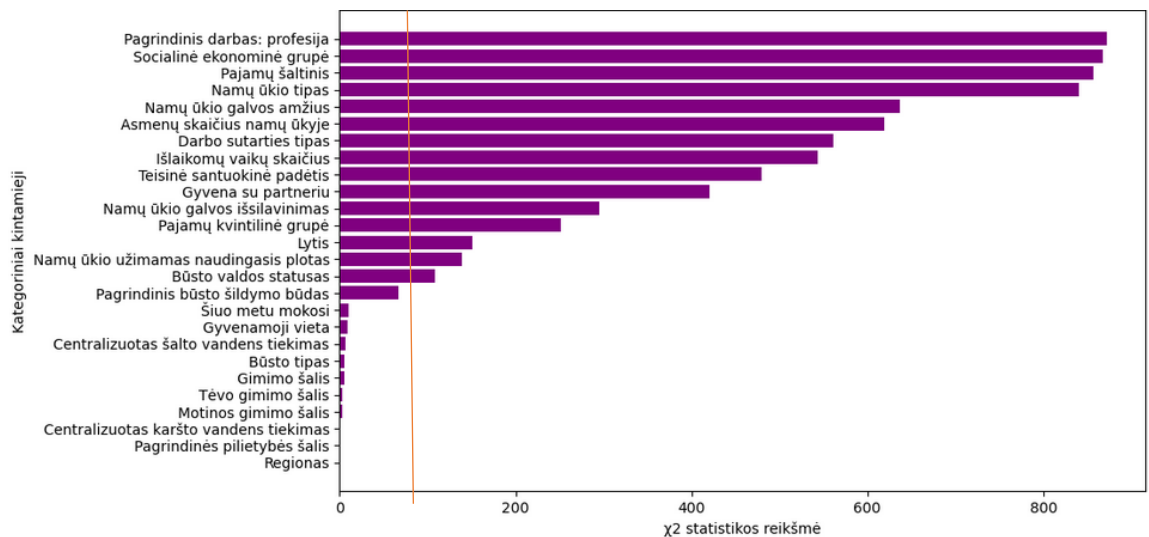
Kaip ir ankstesnėse išlaidų analizėse, daugkartiniai poriniai palyginimai su reikšmingais požymiais ir šįkart buvo atlikti naudojant porinius Moodo medianos testus. Įdomu tai, kad (žr. 22 priedą), vertinant informacijai ir ryšiams skirtų išlaidų medianas, namų ūkiai, kurie gyvena kaime, statistiškai reikšmingai nesiskiria (χ^2 statistikos p-reikšmė $> 0,05$) nuo kituose miestuose gyvenančiųjų, tai gali rodyti intensyviai Lietuvos pastangas plėsti ryšių infrastruktūrą ir užtikrinti plačią paslaugų prieigą.

TI metodo rezultatai interpretuojami taip, kad kuo didesnė yra TI reikšmė, tuo daugiau informacijos požymis teikia apie namų ūkių informacijai ir ryšiams skiriamų išlaidų dydį.



Tarpusavio informacijos metodo panaudojimas požymių atrankai, kai vertinamos informacijos ir ryšių paslaugų vartojimo išlaidos

Remiantis TI didžiausia reikšme galima teigti, kad svarbiausias požymis, paaiškinantis namų ūkių mėnesinių išlaidų, skiriamų informacijai ir ryšiams, pokyčius, yra namų ūkių pajamų šaltinis. Palyginimui buvo vizualizuotos ir χ^2 statistikos reikšmės, kurių dydžiai turėtų padėti identifikuoti svarbiausius vartojimo išlaidų pokyčius veikiančius požymius.



Moodo medianos testo panaudojimas požymių atrankai, kai vertinamos informacijos ir ryšių paslaugų vartojimo išlaidos

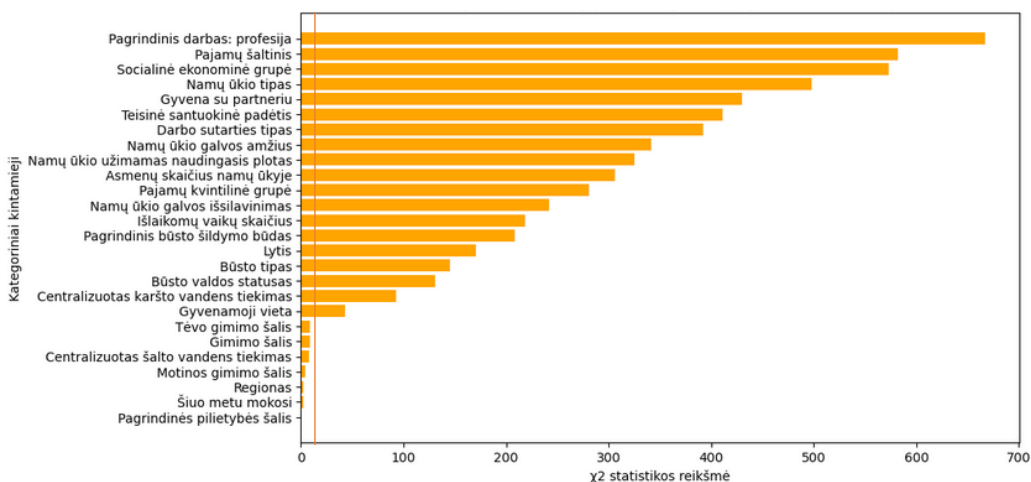
Remiantis Moodo medianos testo rezultatais galima matyti, kad namų ūkių informacijai ir ryšiams skiriamas išlaidas labiausiai turėtų veikti pasirinktos gyventojų profesijos ir priklausymas atitinkamai socialinei-ekonominėi grupei (aukščiausios χ^2 statistikos reikšmės). Abu metodai sutinka, kad informacijos ir ryšių išlaidas minimaliai veikia namų ūkio galvos gyvenamasis regionas, dabartinio mokymosi statusas, būsto tipas, motinos gimimo šalis, pagrindinės pilietybės šalis, tėvo gimimo šalis, gyvenamoji vieta, būsto valdos statusas ir pasirinkimas karštą vandenį gauti centralizuotu būdu (tiek χ^2 statistikos, tiek TI reikšmės yra žemos). Visgi svarbesniu buvo pripažintas Moodo medianos testas, kurio rezultatais remiantis, buvo nustatyta, jog mažiausiai informacijos apie namų ūkių išlaidas, skirtas informacijai ir ryšiams, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 70 (žr. 21 priedą) t. y., „regionas“, „pagrindinės pilietybės šalis“, „centralizuotas karšto vandens tiekimas“, „motinos gimimo šalis“, „tėvo gimimo šalis“, „gimimo šalis“, „būsto tipas“, „centralizuotas šalto vandens tiekimas“, „gyvenamoji vieta“, „šiuo metu mokosi“, „būsto tipas“, „pagrindinis būsto šildymo būdas“. Mažos svarbos kintamieji tolesniame ryšio tarp namų ūkių charakteristikų ir išlaidų, skirtų ryšiams bei informacijai, tyrime bus vertinami kaip rekomenduojami pašalinti.

Išlaidas draudimui ir finansinėms paslaugoms veikiančių požymių identifikavimas

Prieš taikant įvairius statistinius testus mėnesinės namų ūkių išlaidos, skirtos draudimui ir finansinėms paslaugoms, buvo įvertintos vizualiniu būdu. Iš tankio grafikų (žr. 23 priedą) matyti, kad namų ūkių išlaidų draudimui ir finansinėms paslaugoms pasiskirstymas, vertinant jį pagal namų ūkio gyvenimo su ar be partnerio statusą, darbo sutarties tipą ir kitus likusius panašiai išsidėsčiusius kategorinius kintamuosius, kaip ir ankstesnių išlaidų analizių atvejais, tarp grupių pasižymi dešiniąja asimetrija, ilgomis teigiamomis uodegomis, skirtingu duomenų taškų išsibarstymu.

Taip pat kaip ir ankstesniais atvejais, dispersinės analizės metodai nebuvo taikyti dėl normalumo prielaidos pažeidimo (Šapiro ir Vilko testas, žr. 24 priedą) ir dispersijų homogeniškumo prielaidos (Leveno testas, žr. 25 priedą) pažeidimo. Kruskalio ir Voliso testas taip pat nebuvo taikytas, nes Andersono ir Darlingo testo rezultatai (žr. 26 priedą), parodė, kad didžiosios daugumos kategorinių kintamųjų bent vienos grupės skirstinys, pagal draudimui ir finansinėms paslaugoms skirtas išlaidas, statistiškai reikšmingai skyrėsi tarp lyginamųjų grupių. Taip pat dėl galimos priklausomybių pervertinimo rizikos, TI mato rezultatai turėtų būti vertinami atsargiai. Tai reiškia, kad požymių

atranka buvo vykdoma taikant tik Moodo medianos testą, kuriuo yra testuojama nulinė hipotezė, teigianti, jog skirtingų kategorinių kintamųjų grupių medianos pagal išlaidas draudimui ir finansinėms paslaugoms statistiškai reikšmingai nesiskiria. Moodo medianos testo rezultatai parodė (žr. 27 priedą), kad didžioji dalis kategorinių kintamųjų išlaidų, skirtų draudimui ir finansinėms paslaugoms, medianos tarp grupių statistiškai reikšmingai skyrėsi, nes gautos χ^2 statistikos p-reikšmės buvo mažesnės už 5 % reikšmingumo lygį. Verta paminėti, kad aukštesnė χ^2 statistikos reikšmė rodo požymio didesnę svarbumą draudimui ir finansinėms paslaugoms skirtoms išlaidoms.



Moodo medianos testo panaudojimas požymių atrankai, kai vertinamos draudimo ir finansinių paslaugų vartojimo išlaidos

Kaip ir ankstesnėse išlaidų analizėse, daugkartiniai poriniai palyginimai su reikšmingais požymiais ir šįkart buvo atlikti naudojant porinius Moodo medianos testus (žr. 28 priedą). Įdomu tai, kad vertinant draudimui ir finansinėms paslaugoms skiriamų išlaidų medianas tarp vieno asmens su išlaikomais vaikais ir poros asmenų be išlaikomų vaikų, nebuvo atrastas statistiškai reikšmingas skirtumas (χ^2 statistikos pakoreguota p-reikšmė $> 0,05$).

Pagal (žr. 27 priedą) rezultatus vizualizuotos χ^2 statistikos reikšmės, turėtų padėti identifikuoti svarbiausius vartojimo išlaidų pokyčius veikiančius požymius. Remiantis Moodo medianos testo rezultatais galima matyti, kad namų ūkių draudimui ir finansinėms paslaugoms skiriamas išlaidas labiausiai turėtų veikti jų pasirinkta profesija (aukščiausia χ^2 statistikos reikšmė). Pastebėtina, kad nors ir skirtingu eiliškumu, tačiau abu metodai (žr. grafiką ir 30 priedą), t. y., tiek medianos testas, tiek TI matas, sutinka, jog penki svarbiausi požymiai turintys įtakos namų ūkių draudimo ir finansinių paslaugų vartojimui yra profesija, pajamų šaltinis, socialinė-ekonominė grupė, namų ūkio tipas ir gyvenimo su partneriu statusas.

Remiantis Moodo medianos testo rezultatais, kurie dažnai sutapo su taikytu ϕ_K koreliacijos koeficientu gautais rezultatais, nustatyta, jog mažiausiai informacijos apie namų ūkių išlaidas, skirtas draudimui ir finansinėms paslaugoms, suteiktų veiksniai, kurių χ^2 statistikos reikšmė yra mažesnė už 9 (žr. 27 priedą ir paveikslą), t. y., „pagrindinės pilietybės šalis“, „šiuo metu mokosi“, „regionas“, „motinos gimimo šalis“, „centralizuotas šalto vandens tiekimas“, „gimimo šalis“ ir „tėvo gimimo šalis“. Tolimesnėje analizėje, t. y., kuriant modelius, vertinančius draudimo ir finansinių paslaugų išlaidų priklausomybę nuo Lietuvos namų ūkių charakteristikų, išvardinti požymiai turėtų būti laikomi pertekliniais.

32 Priedas. Daugialypės tiesinės regresijos modelių kūrimo bandymai, kai priklausomas kintamasis – bendros vartojimo išlaidos

Išvesti kintamieji?	Vinorizacija?	LOG?	F statistika (prob)	VIF ismeti?	Otimizavimas	Modelio vertinimas	Radj (train)	MAE	RMSE	R2	MEDAE	Normalumas	Valitas	Multiko	
NE	NE	NE		0 NE	OLS	Test MAE: 116.7948 RMSE: 482.1311 R²: 0.3559 MedianAE: 211.1511		0,467	316	482	0,35	211.15	Jarque-Bera: stat=13.244, p=0.001	LM Statistic: 429.239 LM-Test p-value: 0.047	EGZISTUOJA
Taip	Ne	Taip		0 NE	OLS	Test MAE: 116.4890 RMSE: 479.4515 R²: 0.3635 MedianAE: 213.3964		0,461	316	479	0,36	213	Jarque-Bera: stat=12.440, p=0.002	LM-Test p-value: 0.047	EGZISTUOJA
Taip	Ne	Taip		0 Taip	OLS	Test MAE: 116.9105 RMSE: 506.8190 R²: 0.2688 MedianAE: 220.8105		0,4	336	506	0,28	220	Jarque-Bera: stat=10.863, p=0.004,	LM-Test p-value: 0.005	Nera
Taip	Taip (5%)	Taip		0	Taip	Test MAE: 128.4271 RMSE: 523.6822 R²: 0.3396 MedianAE: 215.3841		0,464	328	523	0,3396	215	Jarque-Bera: stat=7.420, p=0.024	M-Test p-value: 0.08	Nera

33 Priedas. Mašininio mokymosi algoritmų modelio kūrimo bandymai, kai priklausomas kintamasis – bendros vartojimo išlaidos

Priklausomas kintamasis	Otimizavimas	Parametru tinktelis	Otimizavimo funkcija	Iteracijų skaičius	Geriausi parametrai	Modelio vertinimas	MAE	RMSE	R2	MEDAE
Visiems	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 248835.11751648383 RFECF: 499.8120861621372 R²: 0.4882786927754885 MEDAE: 212.37139999999929</pre>	329	499	0,308	212
1gyv	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 107.8121484925349 RFECF: 18976.812770773 R²: 0.2647081197110441 R²: 0.1449230077213484 MEDAE: 186.2660090000000</pre>	187	288	0,21	130
1gyv	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 119.1130151048608 RFECF: 7222.4728002422 RFECF: 108.649635111028 R²: 0.13841112020048 RFECF: 118.13088480070</pre>	174	281	0,25	118
Visiems	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 334.8158345357910 RFECF: 258732.9182215215 RFECF: 586.0378045402979 R²: 0.288927521384286 RFECF: 216.3427687508767</pre>	334	508	0,28	216
Visiems	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 119.1130151048608 RFECF: 7222.4728002422 RFECF: 108.649635111028 R²: 0.13841112020048 RFECF: 118.13088480070</pre>	324	466	0,305	217
Visiems	Nera	Nera	criterion='squared_error' (by default)	by default	?	<pre>RF: 108.7243108800007 RFECF: 187809.4011333205 RFECF: 419.138027352629 R²: 0.110441279098128 RFECF: 219.1181187187187</pre>	308	432	0,31	219
1gyv	Bajeso	<pre>search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) }</pre>	r2	20 142]	'max_depth', 48, 'min_samples_leaf', 14, 'min_samples_split', 18, 'n_estimators', 142]	169,856 246	0,24	118		

Visiems	Bajeso	<pre> search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) } </pre>	reg:absoluterror/r2	30	[[('colsample_bytree', 0.3), ('learning_rate', 0.04270780496837394), ('max_depth', 2), ('n_estimators', 2000), ('reg_alpha', 1e-08), ('reg_lambda', 1e-08), ('subsample', 1.0)]	MAE: 283.847769346475 MSE: 167111.118507974805 RFECF: 488.79226816870536 R²: 0.3869389603372544 MEDAE: 195.2153143866392	283	408	0,38	195
Visiems	Bajeso	<pre> search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) } </pre>	reg:absoluterror/r2	30	0.819580974464558), ('learning_rate', 0.0015498371417640521), ('max_depth', 16), ('n_estimators', 2176), ('reg_alpha', 7.018185837401145), ('reg_lambda', 0.028397516492231722), ('subsample',	MAE: 281.56578351702007 MSE: 171689.20272401464 RFECF: 414.3539582579183 R²: 0.3701817273155473 MEDAE: 189.37685742121866	281	414	0,37	189
Visiems	Bajeso	<pre> search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) } </pre>	reg:absoluterror/neg_root_mean_squared_error	60	0.9349013850658064), ('learning_rate', 0.006274236904479497), ('max_depth', 3), ('n_estimators', 1800), ('reg_alpha', 5.000000000000001), ('reg_lambda', 1e-09), ('subsample', 0.6)	MAE: 279.67877103986935 MSE: 185188.848596444805 RFECF: 486.44418161318849 R²: 0.3910880148892761 MEDAE: 189.116477976513622	279	406	0,39	189
Visiems	Bajeso	<pre> search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) } </pre>	MAE/neg_root_mean_squared_error	30	0.5976006297583061)	MAE: 297.20991551856186 MSE: 281128.1879258180813 RFECF: 488.64672587782785 R²: 0.358170521809280073 MEDAE: 199.48992788895949	297	448	0,35	199

Visiems	Bajeso	<pre> search_space = { 'criterion': Integer(0, 100), 'max_depth': Integer(1, 10), 'min_samples_split': Integer(1, 20), 'min_samples_leaf': Integer(1, 20) } </pre>			Geriausi hiperparametrai kvantiliui 0.25: OrderedDict({'colsample_bylevel', 1.0), ('depth', 5), ('iterations', 600), ('l2_leaf_reg', 1.0), ('learning_rate', 0.03187461249757041), ('subsample', 1.0)); Geriausi hiperparametrai kvantiliui 0.5: OrderedDict({'colsample_bylevel', 0.8820207917706828), ('depth', 4), ('iterations', 580), ('l2_leaf_reg', 2.0691862961261718), ('learning_rate', 0.04678945087112739), ('subsample', 0.8242355897456615)); Geriausi hiperparametrai kvantiliui 0.75: OrderedDict({'colsample_bylevel', 0.8007260614509487), ('depth', 4), ('iterations', 522), ('l2_leaf_reg', 1.7970255745729233), ('learning_rate', 0.037902192781288875), ('subsample', 80.0.7824144546506252))	Kvantilis: 0.25 Pinball Loss: 102.7572 Pseudo R²: 0.2160 Kvantilis: 0.5 Pinball Loss: 152.6557 Pseudo R²: 0.2467 Kvantilis: 0.75 Pinball Loss: 145.5839 Pseudo R²: 0.2738
---------	--------	--	--	--	---	--

Įvertinti kintamieji	Log?	Vincitizacija?	Modelis	Priklausomas kintamasis	Otimizuojamas parametru tinktelis	SCORING	Otimizavimo funkcija	Iteracijų skaičius	Geriausi parametrai	Modelio vertinimas	MAE	RMSE	R2	MEDAE
Ne	Ne	Ne	RF	bendros išlaidos	r1 = RandomForestRegressor(random_state=42)	r2	Numatyti	Numatyti	<pre>bootstrap: True 'crg_alpha': 0.0, 'criterion': 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None, bootstrap: True 'crg_alpha': 0.0, 'criterion': 'squared_error', 'max_depth': None, 'max_features': 1.0, 'max_leaf_nodes': None,</pre>					

NE	TAP	Ne	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_depth, 10), (max_features', 0.4372003183162620), (min_samples_leaf, 2), (min_samples_split, 10), (n_estimators, 300) }</pre>	329	525	0.33	217	
TAP	NE	Ne	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_features', 0.4259879893983204), (min_samples_leaf, 2), (min_samples_split, 20), (n_estimators, 377) }</pre>	316	479	0.36	213	
TAP	NE	Ne	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	r2	<pre>{(max_features', 0.4937982386102801), (min_samples_leaf, 5), (min_samples_split, 15), (n_estimators, 375) }</pre>	315	479	0.363	210	
TAP	TAP	Ne	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	r2	<pre>{(max_features', 0.50267227081244), (min_samples_leaf, 3), (n_estimators, 222) }</pre>	308	477	0.369	205.76	
TAP	TAP	Ne	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_depth, 10), (max_features', 0.50269460798746), (min_samples_leaf, 4), (n_estimators, 383) }</pre>	274	372	0.4	202	
NE	NE	TAP (2 proc)	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_features', 0.4481293440574813), (min_samples_leaf, 4), (min_samples_split, 5), (n_estimators, 100) }</pre>	278	372	0.4	208	
TAP	TAP	TAP (2 proc)	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_depth, 17), (max_features', 0.4481293440574813), (min_samples_leaf, 3), (min_samples_split, 10) }</pre>	279	373	0.4	207	
TAP	NE	TAP (2 proc)	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_absolute_error	<pre>{(max_depth, 30), (max_features', 0.354105934532276), (min_samples_leaf, 2), (n_estimators, 30) }</pre>	293	424	0.36	196	
TAP	TAP	TAP (2 proc)	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	neg_mean_squared_error	<pre>{(max_depth, 30), (max_features', 0.40226922367968), (min_samples_leaf, 4), (n_estimators, 30) }</pre>	291	414	0.39	203	
TAP	TAP	TAP (2 proc)	RF	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') }</pre>	r2	<pre>{(max_features', 0.584468670103078), (min_samples_leaf, 2), (min_samples_split, 30), (n_estimators, 279) }</pre>	290	411	0.39	206	
NE	NE	NE	XGBOOST	islands	<pre>model = XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', # GPU acceleration predictor='gpu_predictor', # GPU prediction random_state=0, verbosity=0)</pre>	r	nera	nera	334	508	0.28	216
NE	TAP	NE	XGBOOST	islands	<pre>model = XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', # GPU acceleration predictor='gpu_predictor', # GPU prediction random_state=0, verbosity=0)</pre>	r	nera	nera	330	525	0.23	200
NE	NE	NE	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(max_depth, 145), (max_features', 0.51815150277716), (gamma, 0.81782001837319), (learning_rate, 0.016770816323626), (max_depth, 4), (min_child_weight, 5), (n_estimators, 145) }</pre>	328	509	0.33	231	
NE	NE	NE	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(colsample_bytree', 0.5922473581556799), (gamma, 0.87051372660621), (learning_rate', 0.0285197078471529), (max_depth, 5), (min_child_weight, 7), (n_estimators, 213), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5), (colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	318	508	0.34	206	
NE	NE	NE	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	313	512	0.33	196	
NE	TAP	NE	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	307	484	0.35	200	
TAP	TAP	NE	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	307	475	0.37	208	
TAP	TAP	TAP (2 proc)	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_absolute_error	<pre>{(colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	291	414	0.39	204	
TAP	TAP	TAP (2 proc)	XGBOOST	islands	<pre>search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='log-uniform'), 'colsample_bytree': Real(0.1, 1.0), 'reg_alpha': Real(1e-6, 1.0, prior='log-uniform'), 'reg_lambda': Real(1e-6, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0, 1.0), }</pre>	neg_mean_squared_error	<pre>{(colsample_bytree', 0.830535744065641), (gamma, 1.0), (learning_rate', 0.011718042246737), (gamma, 1.0), (learning_rate', 0.0151734792648793), (max_depth, 5), (min_child_weight, 8), (reg_alpha, 1e-06), (reg_lambda, 1e-06), (subsample, 0.5) }</pre>	291	410	0.404	206	

Taip	Taip	Taip (2 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_median_absolute_error		reg_absolute_error	<pre>(colsample_bytree', 0.5258349614816475), (gamma', 0.306841005550057), (learning_rate', 0.02057913719615543), (max_depth', 10), (min_child_weight', 1), (n_estimators', 259), (reg_alpha', 0.000544168985160891), (reg_lambda', 0.005620355727276), (subsample', 2.0.345662021762743)) (gamma', 1.0), (learning_rate', 0.02871957979504032), (max_depth', 2), (min_child_weight', 1), (n_estimators', 500), (reg_alpha', 1e-05), (reg_lambda', 2.18624557766777e-06), (subsample', 150 0.345662021762743))</pre>	288	419	0.37	198
Taip	Taip	Taip (2 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_root_mean_squared_error		reg_absolute_error	<pre>(gamma', 0.0999129414040294), (learning_rate', 0.0059574684053254), (max_depth', 20), (min_child_weight', 8), (n_estimators', 779), (reg_alpha', 1e-05), (subsample', 150 0.3511331771434714)</pre>	294	413	0.3958	209
taip	Taip	Taip (2 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_mean_squared_error		reg_absolute_error	<pre>(gamma', 0.000276361667303), (learning_rate', 0.60276361667303), (gamma', 2.0), (learning_rate', 0.0082053608223329), (max_depth', 13), (colsample_bytree', 0.6321066914076459), (gamma', 1.258482431814097), (learning_rate', 0.00836140540675845), (max_depth', 3), (min_child_weight', 8), (n_estimators', 1000), (reg_alpha', 0.15971463611985274), (reg_lambda', 3.347813451523055e-06), (subsample', 150 0.44881812162431)</pre>	290	411	0.4	205
Taip	NE	Taip (2 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_mean_squared_error		reg_absolute_error	<pre>(colsample_bytree', 0.6321066914076459), (gamma', 1.258482431814097), (learning_rate', 0.00836140540675845), (max_depth', 3), (min_child_weight', 8), (n_estimators', 1000), (reg_alpha', 0.15971463611985274), (reg_lambda', 3.347813451523055e-06), (subsample', 150 0.44881812162431)</pre>	293	422	0.36	199
NE	taip	Taip (2 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_mean_squared_error		reg_absolute_error	<pre>(colsample_bytree', 0.6321066914076459), (gamma', 1.258482431814097), (learning_rate', 0.00836140540675845), (max_depth', 3), (min_child_weight', 8), (n_estimators', 1000), (reg_alpha', 0.15971463611985274), (reg_lambda', 3.347813451523055e-06), (subsample', 150 0.44881812162431)</pre>	293	413	0.39	206
Taip	taip	Taip (3 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_mean_squared_error		reg_absolute_error	<pre>(colsample_bytree', 0.6321066914076459), (gamma', 1.258482431814097), (learning_rate', 0.00836140540675845), (max_depth', 3), (min_child_weight', 8), (n_estimators', 1000), (reg_alpha', 0.15971463611985274), (reg_lambda', 3.347813451523055e-06), (subsample', 150 0.44881812162431)</pre>	285	394	0.4	210
Taip	taip	Taip (3 proc)	RF	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "min_samples_split": Integer(2, 10), "min_samples_leaf": Integer(2, 10), "max_features": Real(0.1, 0.999, prior="uniform") }</pre>	reg_mean_absolute_error		reg_absolute_error	<pre>(max_depth', 16), (max_features', 0.3858300812448705), (min_samples_leaf', 3), (min_samples_split', 17), (n_estimators', 357), (max_depth', 27), (max_features', 0.422748887934543), (min_samples_leaf', 6), (min_samples_split', 16), (n_estimators', 188), (gamma', 2.0), (learning_rate', 0.0010172132676625), (max_depth', 13), (min_child_weight', 8), (n_estimators', 811), (reg_alpha', 150 0.00844358829473177)</pre>	283	394	0.4	195
Taip	taip	Taip (3 proc)	RF	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "min_samples_split": Integer(2, 10), "min_samples_leaf": Integer(2, 10), "max_features": Real(0.1, 0.999, prior="uniform") }</pre>	reg_mean_absolute_error		reg_absolute_error	<pre>(max_depth', 16), (max_features', 0.3858300812448705), (min_samples_leaf', 3), (min_samples_split', 17), (n_estimators', 357), (max_depth', 27), (max_features', 0.422748887934543), (min_samples_leaf', 6), (min_samples_split', 16), (n_estimators', 188), (gamma', 2.0), (learning_rate', 0.0010172132676625), (max_depth', 13), (min_child_weight', 8), (n_estimators', 811), (reg_alpha', 150 0.00844358829473177)</pre>	285	395	0.39	202
Taip	taip	Taip (3 proc)	XGBBOOST	bendros islaides	<pre>search_space = { "n_estimators": Integer(10, 500), "max_depth": Integer(2, 10), "learning_rate": Real(0.5, 0.9, prior="log-uniform"), "subsample": Real(0.5, 1.0), "colsample_bytree": Real(0.5, 1.0), "reg_alpha": Real(0.1, 0.9, prior="log-uniform"), "reg_lambda": Real(0.1, 0.9, prior="log-uniform"), "min_child_weight": Integer(1, 8), "gamma": Real(0.1, 0.9), }</pre>	reg_mean_absolute_error		reg_absolute_error	<pre>(colsample_bytree', 0.6321066914076459), (gamma', 1.258482431814097), (learning_rate', 0.00836140540675845), (max_depth', 3), (min_child_weight', 8), (n_estimators', 1000), (reg_alpha', 0.15971463611985274), (reg_lambda', 3.347813451523055e-06), (subsample', 150 0.44881812162431)</pre>	284.39	391.01	0.4	201.71

--- TRAIN ---
MAE: 245.2627 | RMSE: 348.6228 | R²: 0.5390 | MEDAE: 161.9425

34 Priedas. Modelių kūrimo bandymai, kai priklausomas kintamasis – maistas ir nealkoholiniai gėrimai

Išmesti kintamieji	Vinorizacija?	LOG?	F statistika (prob)	VIF išmeti?	Modelis	Optimizavimas	Modelio vertinimas	Radj (train)	MAE	RMSE	R2	MEDAE	Normalumas	Varitas	Multiko				
Taip	Ne	Taip	5.73e-211	NE	Tiesinė	OLS	Test RMSE: 186.2117 RMSE: 153.6299 R²: 0.2613 MedianMAE: 76.1764	0,25	106	153	0,26	78	Jarque-Bera: stat=17436.447, p=0.0000	LM-Test p-value: 0.769	EGZISTUOJA				
Taip	Ne	Taip	0	Taip	Tiesinė	OLS	Test RMSE: 111.8771 RMSE: 164.3628 R²: 0.1548 MedianMAE: 85.0301	0,178	113	164	0,15	85	Jarque-Bera: stat=17922.780, p=0.0000	LM-Test p-value: 0.000	Nera				
Taip	Taip (5 proc)	Taip	7.60e-223	Taip	Tiesinė	OLS	Test RMSE: 107.8838 RMSE: 151.6160 R²: 0.2154 MedianMAE: 84.1339	0,27	107	151	0,25	81	Jarque-Bera: stat=14013.269, p=0.0000	LM-Statistic: 371.121/LM-Test p-value: 0.000	Nera				
NE	NE	NE	0.00	NE	Tiesinė	OLS	Test RMSE: 187.1422 RMSE: 155.8370 R²: 0.2498 MedianMAE: 77.1881	0.265	107	155	0,24	77	Jarque-Bera: stat=18262.407, p=0.0000	LM-Statistic: 316.078/LM-Test p-value: 0.001	EGZISTUOJA				
Išmesti kintamieji	Log?	Vinorizacija?	Modelis	Priklausomas kintamasis	Optimizavimas	Optimizavimo metrika	Optimizavimo funkcija	Iteracijų skaičius	Geriausi parametrai	Modelio vertinimas	MAE	RMSE	R2	MEDAE					
Ne	Ne	Ne	RF	bendros išlaidos	rf = RandomForestRegressor(random_state=42)	reg_mean_absolute_error	1	7	7	7	110	159	0,2	84					
Ne	Taip	Ne	RF	bendros išlaidos	rf = RandomForestRegressor(random_state=42)	reg_mean_absolute_error	1	7	7	7	109	162	0,17	76,63					
Taip	Ne	Ne	RF	bendros išlaidos	rf = RandomForestRegressor(random_state=42)	reg_mean_absolute_error	1	7	7	7	112	161	0,18	84,64					
Taip	Taip	Ne	RF	bendros išlaidos	rf = RandomForestRegressor(random_state=42)	reg_mean_absolute_error	1	7	7	7	111	165	0,14	75,08					
NE	Ne	Ne	RF	bendros išlaidos	<pre>model = RandomForestRegressor(objective='reg:squarederror', tree_method='hist', predictor='auto', random_state=42, verbosity=0, n_estimators=250, max_depth=13, min_samples_leaf=6, min_samples_split=10, min_child_weight=1, gamma=0.001, colsample_bytree=0.8, subsample=0.8, bootstrap=True, warm_start=True, n_jobs=-1) rf = model rf.fit(X_train, y_train) y_pred = rf.predict(X_test) rmse = np.sqrt(mean_squared_error(y_test, y_pred))</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	100
Taip	Taip	taip (2 proc)	RF	bendros išlaidos	<pre>model = RandomForestRegressor(objective='reg:squarederror', tree_method='hist', predictor='auto', random_state=42, verbosity=0, n_estimators=250, max_depth=13, min_samples_leaf=6, min_samples_split=10, min_child_weight=1, gamma=0.001, colsample_bytree=0.8, subsample=0.8, bootstrap=True, warm_start=True, n_jobs=-1) rf = model rf.fit(X_train, y_train) y_pred = rf.predict(X_test) rmse = np.sqrt(mean_squared_error(y_test, y_pred))</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	100
NE	NE	NE	XGBOOST	bendros išlaidos	<pre>model = XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) rmse = np.sqrt(mean_squared_error(y_test, y_pred))</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	100
NE	Taip	NE	XGBOOST	bendros išlaidos	<pre>model = XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) rmse = np.sqrt(mean_squared_error(y_test, y_pred))</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	100
NE	NE	NE	XGBOOST	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	100
Taip	Taip	Taip (2 proc)	XGBOOST	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	
Taip	Taip	Taip (2 proc)	XGBOOST	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	
Taip	Taip	Taip (3 proc)	RF	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	
Ne	Ne	Taip (3 proc)	RF	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	
TAIP	Ne	Taip (3 proc)	RF	bendros išlaidos	<pre>param_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(1, 16), 'learning_rate': Real(0.0, 0.3, prior='log-uniform'), 'subsample': Real(0.5, 1.0), 'colsample_bytree': Real(0.5, 1.0), 'reg_alpha': Real(0.1, 1.0, prior='log-uniform'), 'reg_lambda': Real(0.1, 1.0, prior='log-uniform'), 'min_child_weight': Integer(1, 10), 'gamma': Real(0.1, 1.0), }</pre>	reg_mean_absolute_error	50	reg_absolterror	100	100	100	100	100	100	100	100	100	100	

TAP	TAP	Tap (3 proc)	RF	bendros istados	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_mean_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	98.49	129.99	0.29	75.45
Ne	Ne	Tap (3 proc)	XGBOOST	bendros istados	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_mean_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	98	130.9	0.28	73.94
Ne	Tap	Tap (3 proc)	XGBOOST	bendros istados	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_mean_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	98.3	129.7	0.29	75.87
TAP	ne	Tap (3 proc)	XGBOOST	bendros istados	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_mean_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	97.43	130.37	0.28	72.8
TAP	TAP	Tap (3 proc)	XGBOOST	bendros istados	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_mean_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	reg_absolute_error	<pre>{ "n_estimators": 100, "max_depth": 30, "learning_rate": 0.01, "subsample": 0.5, "colsample_bytree": 0.5, "reg_alpha": 1e-06, "reg_lambda": 1e-06, "min_child_weight": 1, "gamma": 0 }</pre>	97.79	128.27	0.3101	75.38

```
→ Smearing factor: 1.0649
--- TRAIN ---
MAE: 96.0391 | RMSE: 125.6973 | R2: 0.3561 | MEDAE: 74.3989

--- TEST ---
MAE: 97.7892 | RMSE: 128.2742 | R2: 0.3101 | MEDAE: 75.3981
```

35 Priedas. Modelių kūrimo bandymai, kai priklausomas kintamasis – informacija ir ryšiai

Įvesti kintamieji	Vinorizacija?	LOG?	F statistika (prob)	VIF išmeti?	Modelis	Optimizavimas	Modelio vertinimas	Radj (train)	MAE	RMSE	R2	MEDAE	Normalumas	Valtas	Multiko
Taip	Ne	Taip		0 NE	Tiesinė	OLS	Test MAE: 20.7629 RMSE: 30.8752 R²: 0.2861 MedLMSE: 14.1348	0,4	20	30	0,24	14,13	Jarque-Bera: stat=2050.543, p=0.000	LM-Test p-value: 0.248	EGZISTUOJA
Taip	Ne	Taip	4.57e-312	Taip	Tiesinė	OLS	Test MAE: 21.4312 RMSE: 31.3637 R²: 0.2208 MedLMSE: 15.0326	0,35	21	31	0,22	15	Jarque-Bera: stat=2030.120, p=0.000	LM-Test p-value: 0.001	Nera
Taip	Taip (5 proc)	Taip		0 Taip	Tiesinė	OLS	Test MAE: 23.0452 RMSE: 32.2706 R²: 0.2528 MedLMSE: 14.6272	0,38	23.04	39.27	0,25	14,62	Jarque-Bera: stat=2292.794, p=0.000	LM Statistic: 203.004LM-Test p-value: 0.080	Nera
NE	NE	NE	0.00	NE	Tiesinė	OLS	Test MAE: 20.9326 RMSE: 31.0183 R²: 0.2391 MedLMSE: 14.4631	0.404	20.93	31.01	0,23	14,46	Jarque-Bera: stat=2004.434, p=0.000	LM-Test p-value: 0.000	EGZISTUOJA
Įvesti kintamieji?	Log?	Vinorizacija?	Modelis	Priklausomas kintamasis	Optimizavimas	Parametrai tinkamumas	Optimizavimo funkcija	Iteracijų skaičius	Optimalūs parametrai	Modelio vertinimas	MAE	RMSE	R2	MEDAE	
Ne	Ne	Ne	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	21	32	0.14	14	
Ne	Taip	Ne	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	19	31	0.22	13	
Taip	Ne	Ne	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	22	33	0.12	14	
Taip	Taip	Ne	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	20	31	0.19	13,56	
NE	Ne	Ne	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	neg_mean_absolute_error	50	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 16, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 22.28451510604879 RMSE: 23.865709969093 R²: 0.27856674866989 R²: 0.27750484459119815 RMSEP: 15.2403897157316	22,22	38,7	0,27	15,23	
Taip	Taip	Taip (2 proc)	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	r2	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 16, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 18.56750575519251 RMSE: 20.3389273242473 R²: 0.34289326089484 RMSEP: 17.90186615837395	18,46	26,13	0,34	12,99	
NE	NE	NE	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	?	nera	nera	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "verbosity": "objective":	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	22	33	0.1	14,72	
NE	Taip	NE	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	?	nera	nera	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": None, "verbosity": "objective":	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	21	32,59	0.15	13,09	
Taip	Taip	Taip (2 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 13, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	20,96	33,48	0,29	15	
Taip	Taip	Taip (2 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_root_mean_squared_error	50	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 13, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	18	26	0.34	13	
Taip	Taip	Taip (3 proc)	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	neg_mean_absolute_error	300	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 11, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17	24,6	0.35	13	
Ne	Ne	Taip (3 proc)	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	neg_mean_absolute_error	100	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 11, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17	24	0.36	13	
Taip	Ne	Taip (3 proc)	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 11, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	18	24,4	0.36	13,6	
Taip	Taip	Taip (3 proc)	RF	bendros islaidos	rf = RandomForestRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 11, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17,9	24,58	0.35	13,2	
Ne	Ne	Taip (3 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 2.0, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17,4	24,82	0.34	12,42	
Ne	Taip	Taip (3 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 20, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17,69	24,47	0.376	13,88	
Taip	ne	Taip (3 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 2.0, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17,47	24,9	0.33	12,51	
Taip	Taip	Taip (3 proc)	XGBOOST	bendros islaidos	xgb = XGBRegressor(random_state=42)	neg_mean_absolute_error	150	100	boosting: True, "ccp_alpha": 0.0, "criterion": "squared_error", "max_depth": 22, "max_features": "sqrt", "min_samples_leaf": 10, "min_samples_split": 10, "n_estimators": 100, "oob_score": True, "random_state": 42, "verbose": 0, "warm_start": False	MAE: 17.3256 RMSE: 24.5082 R²: 0.4200 MEDAE: 11.8332	17,8	24,42	0.3634	13,57	

→ Smearing factor: 1.1758

--- TRAIN ---

MAE: 17.3256 | RMSE: 24.5082 | R²: 0.4200 | MEDAE: 11.8332

--- TEST ---

MAE: 17.6942 | RMSE: 24.1793 | R²: 0.3760 | MEDAE: 13.0195

36 Priedas. Modelių kūrimo bandymai, kai priklausomas kintamasis – draudimas ir finansinės paslaugos

Išmesti kintamieji	Vinorizacija?	LOG?	F statistika (prob)	VIF išmet?	Modelis	Optimizavimas	Modelio vertinimas	Radi (train)	MAE	RMSE	R2	MEDAE	Normalumas	Valitas	Multiko		
Taip	Ne	Taip		0 NE	Tiesinė	OLS	Test RMSE: 30.5336 RMSE: 19.4989 R²: 0.2793 Mediane: 5.4335	0,35	10	19	0,27	5,4	Jarque-Bera: stat=29.000, p=0.000	LM-Test p-value: 0.000	EGZISTUOJA		
Taip	Ne	Taip	1.31e-294	Taip	Tiesinė	OLS	Test RMSE: 30.1182 RMSE: 19.1599 R²: 0.2604 Mediane: 5.4889	0,34	10,31	19,15	0,26	5,6	Jarque-Bera: stat=29.000, p=0.000	LM-Test p-value: 0.000	Nera		
Taip	Taip (5 proc)	Taip	1.19e-311	Taip	Tiesinė	OLS	Test RMSE: 30.5923 RMSE: 20.1181 R²: 0.3342 Mediane: 5.3565	0,35	10,55	20,32	0,1342	5,35	Jarque-Bera: stat=29.000, p=0.000	LM-Test p-value: 0.000	Nera		
NE	NE	NE	2.80e-310	NE	Tiesinė	OLS	Test RMSE: 30.1398 RMSE: 19.6149 R²: 0.2792 Mediane: 5.5307	0.359	10,13	18,9	0,27	5,5	Jarque-Bera: stat=29.157, p=0.000	LM-Test p-value: 0.000	EGZISTUOJA		
Išmesti kintamieji	Log?	Vinorizacija?	Modelis	Optimizavimas	Parametrai tinktis	Optimizavimo funkcija	Iteracijų skaičius	Gerisusių parametrų	Modelio vertinimas	MAE	RMSE	R2	MEDAE				
Ne	Ne	Ne	RF	rf	RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	MAE: 10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,6	19	0,24	6,13				
Ne	Taip	Ne	RF	rf	RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	MAE: 10.260610151781453 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,2	20	0,12	4,71				
Taip	Ne	Ne	RF	rf	RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	MAE: 10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	9	19	0,19	3,9				
Taip	Taip	Ne	RF	rf	RandomForestRegressor(random_state=42)	Numatyti	GP	Numatyti	MAE: 10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,31	20,85	0,12	4,82				
NE	Ne	Ne	RF		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') } rf = RandomForestRegressor(random_state=42) rf.fit(X_train, y_train) y_pred = rf.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_mean_absolute_error	50	(n_estimators=186)	MAE: 10.48/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,3	18,78	0,2006	5,78				
Taip	Taip	taip (2 proc)	RF		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') } rf = RandomForestRegressor(random_state=42) rf.fit(X_train, y_train) y_pred = rf.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	r2	150	(n_estimators=500)	MAE: 8.701804519250306 RMSE: 10.48/10.66474976/5.9819 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	8,7	13,74	0,27	5,18				
NE	NE	NE	XGBOOST		model = xgb.XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	?	nera	nera	MAE: 10.48/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	11	20	0,16	6,16				
NE	Taip	NE	XGBOOST		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'learning_rate': Real(0, 0.5, prior='log uniform'), 'colsample_bytree': Real(0.1, 1.0), 'colsample_bynode': Real(0.1, 1.0), 'reg_alpha': Real(0, 1.0, prior='log uniform'), 'reg_lambda': Real(0, 1.0, prior='log uniform'), 'min_child_weight': Integer(1, 8), 'gamma': Real(0, 1.0) } model = xgb.XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	?	nera	nera	MAE: 10.48/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,75	21	0,05	5				
NE	NE	NE	XGBOOST		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'learning_rate': Real(0, 0.5, prior='log uniform'), 'colsample_bytree': Real(0.1, 1.0), 'colsample_bynode': Real(0.1, 1.0), 'reg_alpha': Real(0, 1.0, prior='log uniform'), 'reg_lambda': Real(0, 1.0, prior='log uniform'), 'min_child_weight': Integer(1, 8), 'gamma': Real(0, 1.0) } model = xgb.XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_mean_absolute_error	150	(n_estimators=140)	MAE: 10.48/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	10,07	18,13	0,25	6,34				
Taip	Taip	Taip (2 proc)	XGBOOST		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'learning_rate': Real(0, 0.5, prior='log uniform'), 'colsample_bytree': Real(0.1, 1.0), 'colsample_bynode': Real(0.1, 1.0), 'reg_alpha': Real(0, 1.0, prior='log uniform'), 'reg_lambda': Real(0, 1.0, prior='log uniform'), 'min_child_weight': Integer(1, 8), 'gamma': Real(0, 1.0) } model = xgb.XGBRegressor(objective='reg:squarederror', tree_method='gpu_hist', predictor='gpu_predictor', random_state=42, verbosity=0) model.fit(X_train, y_train) y_pred = model.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_root_mean_squared_error	neg_absoluteerror	150	(n_estimators=0.5)	MAE: 10.48/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	8,8	13	0,28	5,6			
Taip	Taip	Taip (3 proc)	RF		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') } rf = RandomForestRegressor(random_state=42) rf.fit(X_train, y_train) y_pred = rf.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_mean_absolute_error	300	(n_estimators=180)	MAE: 8.41422556/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	8,41	12,84	0,285	5,1				
Ne	Ne	Taip (3 proc)	RF		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') } rf = RandomForestRegressor(random_state=42) rf.fit(X_train, y_train) y_pred = rf.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_mean_absolute_error	100	(n_estimators=800)	MAE: 8.41422556/10.66474976/5.9819 RMSE: 17.4135418799974 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	8,6	12,64	0,307	5,56				
Taip	Ne	Taip (3 proc)	RF		search_space = { 'n_estimators': Integer(10, 500), 'max_depth': Integer(2, 30), 'min_samples_split': Integer(2, 30), 'min_samples_leaf': Integer(2, 30), 'max_features': Real(0.1, 0.999, prior='uniform') } rf = RandomForestRegressor(random_state=42) rf.fit(X_train, y_train) y_pred = rf.predict(X_test) mae = mean_absolute_error(y_test, y_pred) rmse = mean_squared_error(y_test, y_pred) r2 = r2_score(y_test, y_pred) medae = median_absolute_error(y_test, y_pred)	neg_mean_absolute_error	150	(n_estimators=656)	MAE: 8.597974716112073 RMSE: 10.48/10.66474976/5.9819 R²: 19.34258045866618 RMSE: 20.791988063245982 R²: 0.1290293364083346 MEDAE: 4.27258910674264	8,59	12,62	0,309	5,63				

Smearing factor: 7.210751444659868e+17
TRAINING MAE: 7.2620
TRAINING MSE: 109.8641
TRAINING RMSE: 10.4816
TRAINING R²: 0.4542
TRAINING MEDAE: 4.8260
MAE: 8.597974716112073
MSE: 159.4325993677814
RMSE: 12.626662241771632
R²: 0.3090874461037473
MEDAE 5.632748008667291

143

37 Priedas. Maisto ir nealkoholinių gėrimų, informacijos ir ryšių, draudimo ir finansinių paslaugų vartojimo išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimas

Maisto vartojimo išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimas

Anksčiau buvo nustatyta (žr. 4 priedą), kad maksimali maisto ir nealkoholinių gėrimų vartojimo išlaidų suma tarp Lietuvos gyventojų siekė 2 tūkst. Eur per mėnesį. Taikant skirtingas viršutines vinzorizacijos metodo ribas, buvo apribotos ir skirtingos vartojimo išlaidų sumos: taikant 1 % ribą, maksimali išlaidų suma buvo apribota iki 893,15 Eur, 2 % – iki 730,39 Eur, 3 % – iki 664,88 Eur, 4 % – iki 614,02 Eur, ir 5 % – iki 581,07 Eur. Maisto ir nealkoholinių gėrimų vartojimo išlaidų priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimui atlikti buvo palyginti trys sukurti modeliai.

Tyrime sukurtų regresijos modelių palyginimas, kai priklausomas kintamasis yra maisto ir nealkoholinių gėrimų išlaidos

	Daugialypė tiesinė regresija		Atsitiktinių miškų regresija		Ekstremalaus gradientinio stiprinimo algoritmas	
	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas
Nereikšmingi požymiai išmesti	Ne	Taip	Ne	Taip	Ne	Taip
Taikyta vinzorizacija/proc.	Ne	Taip/5%	Ne	Taip/3%	Ne	Taip/3%
Logarimtuotos išlaidos	Ne	Taip	Ne	Ne	Ne	Taip
MAE, Eur.	107,14	107,88	110,43	98,1	118,92	97,78
RMSE, Eur.	155,03	151,61	159,60	128,37	173,54	128,27
R²	0,24	0,25	0,20	0,30	0,1	0,31
MedAE, Eur.	77,10	81,13	84,11	78,02	81,24	75,39

Kiekvieno metodo atveju buvo pateiktas pradinis modelio variantas, t. y., su neapdorotu priklausomu kintamuoju, neišmestais nereikšmingais kintamaisiais, tai leido įvertinti pradinę modelio kokybę ir ją vėliau palyginti su optimizuotu modeliu. Geriausias konkretaus metodo modelio variantas buvo tas, kuris po duomenų paruošimo (išskirčių tvarkymo, logaritmavimo, nereikšmingų požymių pašalinimo (žr. 31 priedą) ir modelio optimizavimo (hiperparametrų derinimo) pasiekė aukščiausias kokybės metrikų reikšmes tarp visų tuo metodu bandytų variantų (žr. 34 priedą).

Kaip ir bendrųjų išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimo atveju, daugialypės tiesinės regresijos modelio galutinis variantas, vertinant pagal kokybės metrikas pasirodė prasčiau nei pradinis variantas, tačiau jo atveju buvo pašalinta multikolinearumo problema, dėl kurios buvo išmesti kai kurie aiškinamieji kintamieji, kurie galėjo būti svarbūs. Verta paminėti, kad sukurtas modelio liekanos neatitiko kitų esminių klasikinei regresijai taikomų prielaidų (Vaito testo p-reikšmė = 0,00; *Jarque-Bera* testo p-reikšmė = 0,00). Vertinat atsitiktinių miškų modelius matyti, kad pradinis modelio variantas pagal kokybės metrikas buvo prastesnis už daugialypės tiesinės regresijos modelį, tačiau atlikus duomenų paruošimą ir hiperparametrų optimizavimą, kurio metu buvo minimizuojama vidutinė absoliuti paklaida, atsitiktinių miškų modelio rezultatai pagerėjo, t. y., modelio aiškinamoji galia (R^2) išaugo apie 10 proc. punktų, o klaidų metrikų (MAE, RMSE, MedAE) reikšmės buvo artimesnės nuliui. Pastebėtina, kad ekstremalaus gradientinio stiprinimo algoritmo metodu gautas pradinis modelio variantas buvo blogiausiai duomenis aprašantis modelis iš visų trijų metodų, tačiau atlikus papildomą duomenų paruošimą ir modelių optimizavimą, buvo sukurtas geriausias, tarp visų

taikytų metodų, regresijos modelis maisto ir nealkoholinių gėrimų vartojimo išlaidų poveikio nuo mikrolygmens veiksmų vertinimui. Regresijos modelis, paremtas ekstremalaus gradientinio stiprinimo algoritmu, pasižymėjo aukščiausiu determinacijos koeficientu tarp visų taikytų metodų, t. y., $R^2 = 0,31$, kai atsitiktinių miškų buvo $R^2 = 0,30$, o tiesinės regresijos – $R^2 = 0,25$, tai reiškia, kad geriausias modelis gali paaiškinti iki 31 % Lietuvos namų ūkių maisto ir nealkoholinių gėrimų vartojimo išlaidų variacijos. Kitos klaidos taip pat mažesnės, pavyzdžiui, vidutinė absoliuti paklaida ekstremalaus gradientinio stiprinimo algoritmo atveju siekė 97,78 Eur, kai atsitiktinių miškų – 98,1 Eur, ir 107,14 Eur – daugialypės tiesinės regresijos atveju. Galima teigti, kad išlaidų prognozės nuo tikrųjų išlaidų sumų vidutiniškai skiriasi per 97,78 eurus. Remiantis metrika, galima teigti, kad 50 % visų stebėtų namų ūkių prognozių paklaidos neviršijo 75,39 Eur. Mažesnė modelių paaiškinamoji galia, lyginant su bendrųjų vartojimo išlaidų vertinimui sukurtu modeliu, gali būti paaiškinama tuo, kad tyrime naudoti požymiai buvo dar silpniau susiję su maisto ir nealkoholinių gėrimų išlaidų kategorija (ϕ_K koreliacijos koeficientų matrica). Tyrimo metu geriausiu pripažinto modelio optimalūs hiperparametrai, rasti naudojant Bajeso optimizavimo algoritmą.

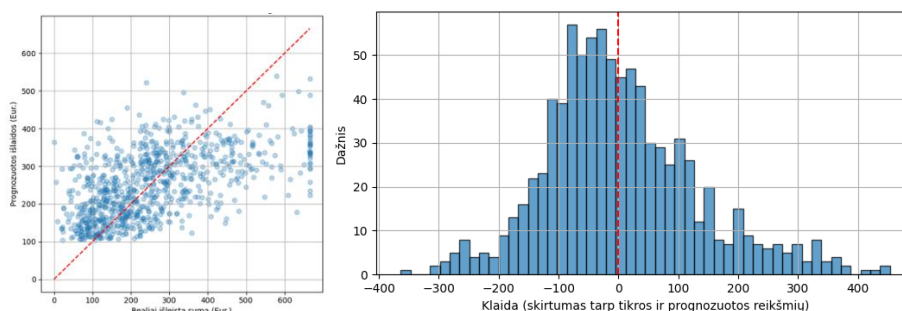
Regresijos modelio, paremto ekstremalaus gradientinio stiprinimo algoritmu, paieškos intervalai ir optimalus derinys, kai priklausomas kintamasis yra maisto ir nealkoholinių gėrimų išlaidos

Hiperparametrai Bajeso optimizavimui	Paieškos intervalas	Geriausias hiperparametrų derinys
Medžių skaičius	[10 ; 1000]	248
Didžiausias medžio gylis	[2; 30]	2,00
Mokymosi greitis	[1e-4 ; 1,0]	0,04
Dalinė duomenų imtis	[0,3 ; 1,0]	0,45
Kintamųjų dalis, naudojama kiekvieno medžio kūrimui	[0,3 ; 1,0]	1,00
L1 reguliavimo koeficientas svoriams	[1e-6; 5,0]	4,99
L2 reguliavimo koeficientas svoriams	[1e-6 ; 5,0]	1,82
Minimali instancijų svorių suma, reikalinga mazge	[2 ; 20]	2,00
Minimalus nuostolio funkcijos sumažėjimas, reikalingas, kad įvyktų mazgo padalijimas	[1e-6 ; 2,0]	1e-06

Hiperparametrų optimizavimo metu buvo naudojama MAE metrika, o nuostolio funkcija – absoliutinė paklaida, kuri buvo pasirinkta vietoje numatytosios kvadratinės paklaidos. Naudotas maksimalus iteracijų skaičius buvo 150. Geriausias hiperparametrų derinys rodo, kad modelis optimaliausiai veikė su labai sekliais medžiais (didžiausias medžio gylis = 2), tai gali reikšti, duomenyse nebuvo atrasta sudėtingų sąveikų. Taip pat buvo naudoti visi į modelį patekę požymiai (kintamųjų dalis, naudojama kiekvieno medžio kūrimui = 1), o tai rodo, kad ankstesnis neinformatyvių požymių pašalinimas buvo efektyvus. Nedidelė minimali instancijų svorių suma, reikalinga mazge, t. y., 2, ir minimalus nuostolio funkcijos sumažėjimas, reikalingas, kad įvyktų mazgo padalijimas, t. y., 1e-06, gali rodyti, jog modelis jautriai reagavo į smulkius pokyčius, tačiau tai buvo kontroliuojama naudojant stiprią reguliaciją.

Kadangi regresijos modelis, sukurtas naudojant ekstremalaus gradientinio stiprinimo algoritmą, pasižymėjo aukščiausią prognozavimo kokybę, t. y., aukščiausiu R^2 ir žemiausiomis MAE, RMSE bei MedAE reikšmėmis, jis buvo pasirinktas kaip pagrindinis tyrimo įrankis vertinant mikrolygmens veiksmų poveikį maisto ir nealkoholinių gėrimų vartojimo išlaidoms. Vertinant modelio kokybę pagal metrikas, apskaičiuotas naudojant mokymo imties duomenis, buvo gauta, kad MAE = 96,03 Eur, RMSE = 125,70 Eur, $R^2 = 0,35$, MedAE = 74,40 Eur, kai naudojant testavimo imtį – MAE = 97,78 Eur, RMSE = 128,27 Eur, $R^2 = 0,31$, MedAE = 75,39 Eur. Nedideli kokybės metrikų skirtumai

tarp mokymo ir testavimo duomenų rodo, kad sukurtas modelis nebuvo permokytas. Modelio kokybei įvertinti taip pat buvo pateikti diagnostikos grafikai.



Ekstremalaus gradientinio stiprinimo regresijos modelio diagnostika prognozuojant maisto ir nealkoholinių gėrimų vartojimo išlaidas

Matyti panaši situacija kaip ir modelio, sukurto bendrosioms vartojimo išlaidoms vertinti, atveju, t. y., modeliu yra pervertinamos mažų išlaidų sumos (iki 400 Eur), nes sklaidos diagrama rodo, kad taškai yra virš raudonos tiesės ir rodo, kad modeliu yra pervertinamos mažos išlaidų sumos. Tuo tarpu, ypač didelės vartojimo sumos (didesnės už 500 Eur) yra nuvertinamos, nes dauguma taškų yra susitelkę žemiau idealios linijos. Remiantis paklaidų histograma galima teigti, kad didžioji dalis paklaidų yra susitelkusi intervale nuo – 90 Eur, iki + 30 Eur, kai ypač dažni atvejai pasitaiko neigiamoje pusėje, ir reiškia, kad modelis yra labiau linkęs nuvertinti tikrąsias vartojimo išlaidas. Nepaisant modelio polinkio nuvertinti didesnes išlaidas, modeliu gautos prognozės išlieka pakankamai patikimos, ypač vertinant vidutines vartojimo išlaidų sumas.

Siekiant įvertinti mikrolygmens veiksnių bei jų sąveikų poveikį modelio prognozuojamoms išlaidoms, buvo naudotas SHAP metodas, paaiškinantis ekstremalaus gradientinio stiprinimo algoritmu gautas prognozes. SHAP analizės suminė diagrama padėjo identifikuoti svarbiausius veiksnius, remiantis jų vidutiniu indėliu į modelio prognozę, jį lyginant su vidutine prognoze, ir leido nustatyti kiekvieno veiksnio poveikio kryptį.

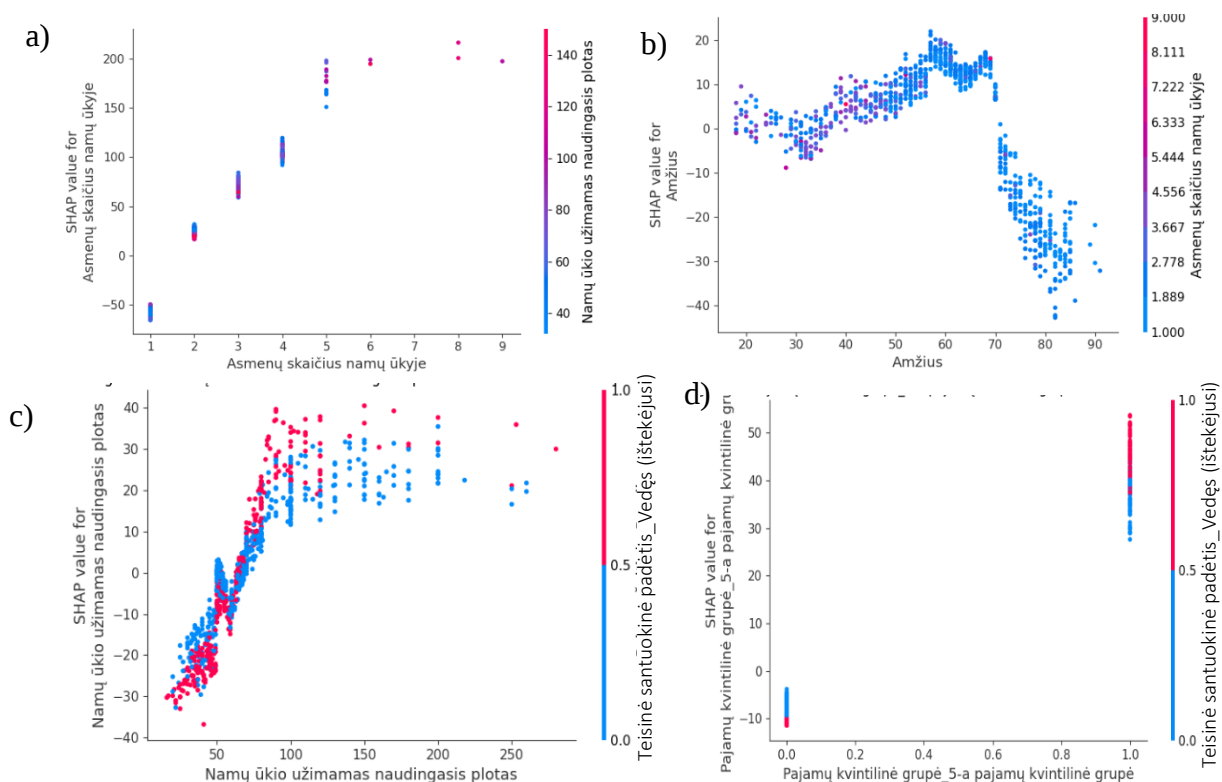


Požymių poveikis maisto ir nealkoholinių gėrimų vartojimo išlaidoms, nustatytas naudojant SHAP metodą, pritaikytą pagal ekstremalaus gradientinio stiprinimo algoritmu sukurtą regresijos modelį

Vidutinė prognozuota mėnesinė maisto ir nealkoholinių gėrimų vartojimo išlaidų suma siekė 233,58 Eur. Remiantis suminės diagramos informacija, prognozuojant šią išlaidų kategoriją svarbiausi veiksniai buvo asmenų skaičius namų ūkyje, užimamas naudingasis plotas, priklausymas konkrečiai pajamų kvintilinei grupei bei namų ūkio galvos amžius. Visų išvardintų veiksnių poveikio kryptis yra tokia pati, kaip ir bendrųjų vartojimo išlaidų atveju, pavyzdžiui, namų ūkiai, kurie priklauso pajamų 5-ai grupei, pagal prognozavimo modelį, yra linkę maisto ir nealkoholinių gėrimų vartojimui išleisti vidutiniškai apie 90 Eur., daugiau, juos lyginant su namų ūkiais, esančiais 1-oje kvintilinėje grupėje. Tai rodo teigiamos krypties poveikį. Vienas iš svarbesnių, ir nuo bendrųjų vartojimo išlaidų prognozės modelio besiskiriančių veiksnių, yra namų ūkio teisinė santuokinė padėtis, kuri rodo, jog santuokoje gyvenantys namų ūkiai teigiamai veikia maisto ir gėrimų vartojimo išlaidas (teigiamos SHAP reikšmės). Vedę (ištekęsios) gyventojai prie tiriamos išlaidų kategorijos prognozuojamos sumos vidutiniškai prisideda apie 40 Eur, lyginant juos su išsituokusiais namų ūkiais, kurie yra referencinė grupė.

Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas. Šios diagramos padeda suprasti veiksnių įtaką modeliui atliktoms išlaidų prognozėms, kartu įvertinant ir sąveiką su kitu veiksnio. Nevisiškai tiesinis ryšys tarp asmenų skaičiaus namų ūkyje ir vartojimo išlaidų buvo taip pat nustatytas ir maisto bei nealkoholinių gėrimų kategorijoje. Grafike (žr. a) matyti kad SHAP reikšmės pasiekia daugiau nei 200 Eur, kai namų ūkį sudaro 6 ir daugiau asmenų, o tuo tarpu asmenys, gyvenantys vieni mažuose būstuose (užimančiuose apie 40 m² naudingojo ploto), vidutinės išlaidas gali mažinti daugiau nei per 50 Eur nuo vidutinės visų namų ūkių išleidžiamos sumos. Kitas grafikas (žr. b) taip pat rodo jau anksčiau nustatytą netiesinį ryšį tarp vartojimo išlaidų ir namų ūkio galvos amžiaus. Tiriant maisto išlaidų kategoriją matyti, kad teigiamos SHAP reikšmės svyruoja apie 10 Eur, kai namų ūkio galvos amžius yra tarp 20 – 50 metų ir asmenų skaičius namų ūkyje įprastai yra didesnis nei 1 (violetiniai taškai). Ypač didelės teigiamos SHAP reikšmės, rodančios didesnes nei vidutinės maistui skirtos išlaidos (vidutiniškai didesnes nei 10 Eur), yra

matomos namų ūkiuose, kurių galvos amžius siekia 50 – 60 metų, net ir esant mažesniai asmenų skaičiui. Tuo tarpu vyresni nei 70 metų asmenys, dažniausiai gyvenantys vieni (mėlyni taškai), neigiamai veikia maisto vartojimo išlaidas.



SHAP priklausomybės diagramos, vertinančios kintamųjų tarpusavio sąveikas prognozuojamų maisto ir nealkoholinių gėrimų vartojimo išlaidų atžvilgiu

Priklausomybės diagrama, rodanti namų ūkio užimamo naudingąjo ploto poveikį prognozuojamomis maisto ir nealkoholinių gėrimų išlaidomis (žr., c) rodo, kad santuokoje gyvenantys asmenys yra labiau linkę gyventi didesnę naudingąją plotą turinčiuose būstuose, ir tai kartu prisideda prie vidutiniškai per 40 Eur., didesnių išlaidų nei vidutinė suma. Maisto išlaidų sumos didėjimą teigiamai veikia ir faktas, jog namų ūkiai yra susituokę bei priklauso 5-ai pajamų kvintilinei grupei (žr. d), nes šios kombinacijos teigiamos SHAP reikšmės dažnai yra tarp 40 – 50 Eur, tačiau jei asmenys yra susituokę, bet nepriklauso aukščiausias pajamas uždirbančių asmenų grupei, jų vidutinė išlaidų suma nuo vidurkio gali mažėti per daugiau nei 10 Eur.

Informacijos ir ryšių išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimas

Anksčiau buvo nustatyta (žr. 4 priedą), kad informacijos ir ryšių paslaugų vartojimo išlaidų suma tarp Lietuvos gyventojų siekė 558,32 Eur. per mėnesį. Taikant 3 % viršutinę vinzorizacijos metodo ribą, maksimali išlaidų suma buvo apribota iki 136,29 Eur. Informacijos ir ryšių paslaugų vartojimo išlaidų priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimui atlikti buvo palyginti trys sukurti regresiniai modeliai (žr. lent.). Kiekvieno metodo atveju buvo pateiktas pradinis modelio variantas, t. y., su neapdorotu priklausomu kintamuoju, neišmestais nereikšmingais kintamaisiais, tai leido įvertinti pradinę modelio kokybę ir ją vėliau palyginti su optimizuotu modeliu. Geriausias konkretaus metodo modelio variantas buvo tas, kuris po duomenų paruošimo (išskirčių tvarkymo, logaritmvimo, nereikšmingų požymių pašalinimo (žr. 31 priedą) ir modelio optimizavimo

(hiperparametrų derinimo) pasiekė aukščiausias kokybės metrikų reikšmes tarp visų tuo metodu bandytų variantų (žr. 35 priedą).

Tyrime sukurtų regresijos modelių palyginimas, kai priklausomas kintamasis yra informacijos ir ryšių paslaugų vartojimo išlaidos

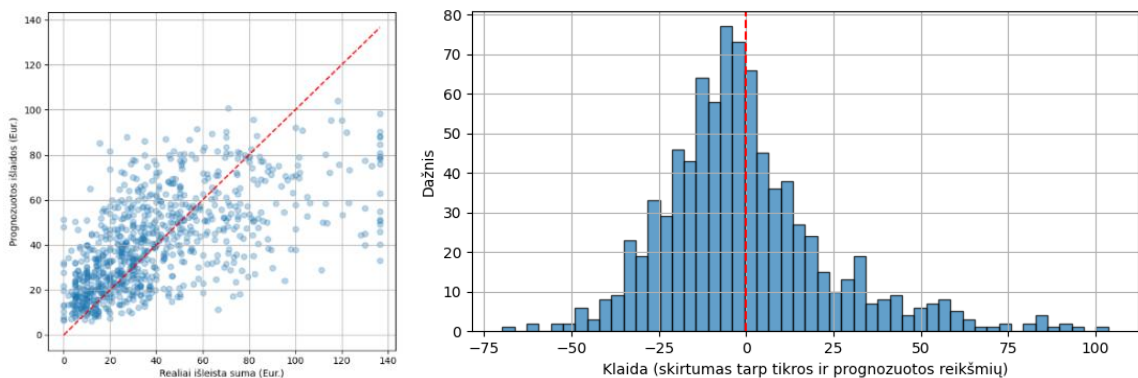
	Daugialypė tiesinė regresija		Atsitiktinių miškų regresija		Ekstremalaus gradientinio stiprinimo algoritmas	
	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas
Nereikšmingi požymiai išmesti	Ne	Taip	Ne	Ne	Ne	Ne
Taikyta vinzorizacija/proc.	Ne	Taip/5%	Ne	Taip/3%	Ne	Taip/3%
Logarimtuotos išlaidos	Ne	Taip	Ne	Ne	Ne	Taip
MAE, Eur.	20,93	23,04	21,73	17,93	22,76	17,69
RMSE, Eur.	31,01	39,27	32,94	24,31	33,73	24,17
R²	0,23	0,25	0,14	0,36	0,1	0,38
MedAE, Eur.	14,46	14,62	14,78	13,33	14,72	13,02

Daugialypės tiesinės regresijos galutinis modelio variantas, vertinant pagal paklaidų metrikas, pasirodė prastesnis nei pradinis, dėl siekio tenkinti tiesinės regresijos taikymui reikalingas prielaidas. Vertinant informacijos ir ryšių paslaugų vartojimo išlaidas, geriausiu modeliu buvo pripažintas ekstremalaus gradientinio stiprinimo galutinis variantas ($R^2 = 0,38$), kuris aplenkė tiek atsitiktinių miškų ($R^2 = 0,36$), tiek daugialypės tiesinės regresijos ($R^2 = 0,25$) modelius. Tai reiškia, kad geriausias modelis gali paaiškinti iki 38 % informacijos ir ryšių paslaugų vartojimo išlaidų dispersijos. Vertinant modelio kokybę pagal klaidų metrikas, ekstremalaus gradientinio stiprinimo modelis taip pat buvo panašesnis – jo MAE, RMSE ir MedAE reikšmės, lyginant su kitais sukurtais modeliais, buvo arčiausiai nulio, kai vidutinė absoliuti paklaida siekė 17,69 Eur, o atsitiktinių miškų modelio atveju – 17,93 Eur ir 23,04 Eur – daugialypės tiesinės regresijos atveju. Galima teigti, kad išlaidų prognozės nuo tikrųjų išlaidų sumų vidutiniškai skiriasi per 17,69 eurus.

Tyrimo metu geriausiu pripažinto modelio optimalūs hiperparametrai buvo rasti naudojant Bajeso optimizavimo algoritmą. Hiperparametrų optimizavimo metu buvo naudojama MAE metrika, nuostolio funkcija – absoliutinė paklaida, kuri buvo pasirinkta vietoje numatytosios kvadratinės paklaidos ir maksimalus iteracijų skaičius buvo nustatytas iki 150. Geriausio ekstremalaus gradientinio stiprinimo modelio hiperparametrų derinys buvo toks: 715 medžiai, didžiausias medžio gylis – 30, mokymosi greitis – 10^{-6} , dalinė duomenų imtos – 0,3, kintamųjų dalis, naudojama kiekvieno medžio kūrimui – 0,3, L1 reguliavimo koeficientas svoriams – 10^{-6} , L2 reguliavimo koeficientas svoriams – 1,07, minimali instancijų svorių suma, reikalinga mazge – 20, minimalus nuostolio funkcijos sumažėjimas, reikalingas, kad įvyktų mazgo padalijimas – 2,0. Vertinant modelio kokybę pagal metrikas, apskaičiuotas naudojant mokymo imties duomenis, buvo gauta, kad MAE = 17,33 Eur, RMSE = 24,51 Eur, $R^2 = 0,42$, MedAE = 11,83 Eur. Nedideli kokybės metrikų skirtumai tarp mokymo ir testavimo duomenų rodo, kad sukurtas modelis nebuvo permokytas.

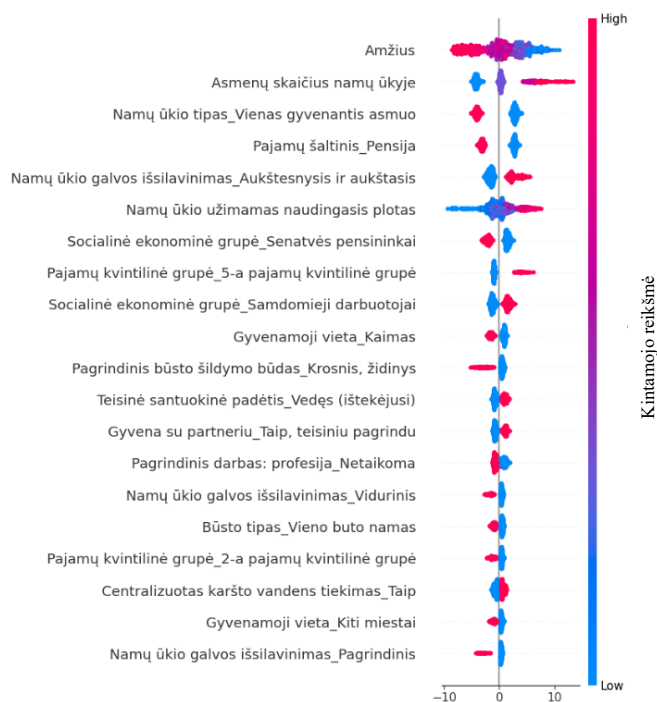
Modelio, sukurto taikant ekstremalaus gradientinio stiprinimo algoritmą, kokybei įvertinti taip pat buvo pateikti diagnostikos grafikai. Matyti panaši situacija kaip ir ankstesnių regresinių modelių atveju, t. y., rezultatai pasižymi didele dispersija bei modelis ypač didelės vartojimo sumas (didesnės už 80 Eur) yra linkęs nepakankamai įvertinti, nes dauguma taškų yra susitelkę žemiau idealios linijos. Remiantis paklaidų histograma galima teigti, kad didžioji dalis paklaidų yra susitelkusi intervale nuo

– 15 Eur, iki + 5 Eur, kai ypač dažni atvejai pasitaiko neigiamoje pusėje, ir reiškia, kad modelis yra labiau linkęs nuvertinti tikrąsias vartojimo išlaidas.



Ekstremalaus gradientinio stiprinimo algoritmu sukurto regresijos modelio diagnostika prognozuojant informacijos ir ryšių paslaugų vartojimo išlaidas

SHAP analizės suminė diagrama, sukurta remiantis ekstremalaus gradientinio stiprinimo algoritmu paremto regresijos modelio rezultatais, padėjo identifikuoti svarbiausius veiksnius, remiantis jų vidutiniu indėliu į modelio prognozę, jį lyginant su vidutine prognoze. Vidutinė prognozuota mėnesinė informacijos ir ryšių paslaugų vartojimo išlaidų suma siekė 34,74 Eur.



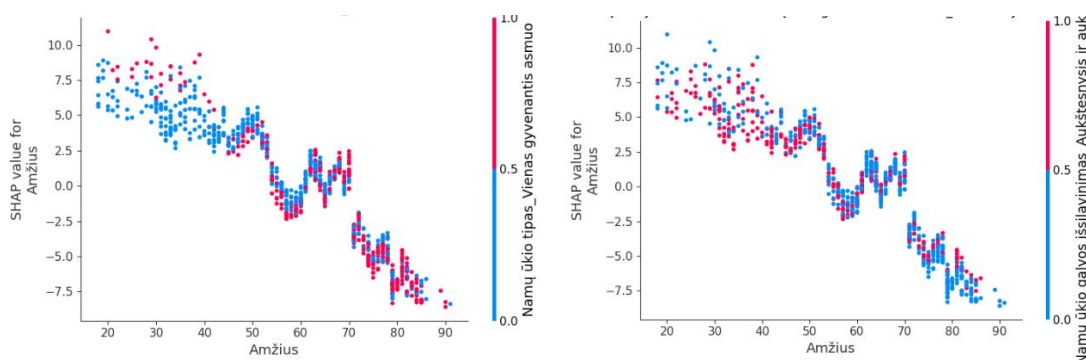
Informacijos ir ryšių paslaugų vartojimo išlaidos (Eur) pagal SHAP

Požymių poveikis informacijos ir ryšių vartojimo išlaidoms, nustatytas naudojant SHAP metodą, pritaikytą pagal ekstremalaus gradientinio stiprinimo algoritmą sukurta regresijos modelį

Remiantis SHAP suminės diagramos informacija, galima teigti, kad informacijos ir ryšių vartojimo išlaidas stipriausiai veikia namų ūkio galvos amžius, kurio poveikį galima vertinti kaip atvirkštinį, t. y., kuo asmuo yra jaunesnis, tuo jo prognozuojamos išlaidos yra didesnės, ir gali viršyti bendrą vidurkį daugiau nei 10 Eur. Trečias pagal svarbą kintamasis yra vieno gyvenančio asmens fiktyvusis požymis, kurio buvimas yra susijęs su mažesnėmis vartojimo išlaidomis, skirtomis informacijai ir ryšiams. Tai reiškia, kad tokio tipo namų ūkiai šiai išlaidų kategorijai išleidžia vidutiniškai iki 10 Eur

mažiau, nei kiti namų ūkiai be išlaikomų vaikų. Pensijos, kaip pagrindinio pajamų šaltinio, poveikis vartojimo išlaidoms yra panašus į poveikį, kuris buvo nustatytas vieno gyvenančio asmens atveju, tai gali būti paaiškinama tuo, kad pensinio amžiaus asmenys dažnai gyvena vieni. Aukštesnįjį ar aukštąjį išsilavinimą turinčių asmenų poveikis informacijai ir ryšiams skirtoms išlaidoms yra teigiamas, ir rodo, kad tokį išsilavinimo lygį turintys asmenys, prognozuojant išlaidas, nuo vidurkio prisidėtų per 5 Eur daugiau, lyginant su asmenimis, kurių išsilavinimas yra aukštesniojo ir aukštojo nesuteikiantis pavidurinis.

Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas. Šios diagramos padeda suprasti veiksnių įtaką modeliui atliktoms išlaidų prognozėms, kartu įvertinant ir sąveiką su kitu veiksnio.



SHAP priklausomybės diagramos, vertinančios svarbių kintamųjų tarpusavio sąveikas prognozuojamų informacijos ir ryšių paslaugų vartojimo išlaidų atžvilgiu

Matyti, kad didžiausios teigiamos SHAP reikšmės, t. y., prognozuojamos informacijai ir ryšiams skirtos išlaidos, viršijančios vidurkį daugiau nei 10 Eur pasitaiko tarp jaunų (apie 20 – 40 metų) ir vieno gyvenančių (raudoni taškai) asmenų. Bendrai nustatyta, kad jauni ir vidutinio amžiaus asmenys (20 – 50 metų), negyvenantys vieni (mėlyni taškai), yra susiję su teigiamu poveikiu informacijos ir ryšių paslaugų vartojimo išlaidoms. Tuo tarpu namų ūkiai, kurie yra vyresni nei 60 metų asmenys ir dažniausiai gyvena vieni, yra susiję su informacijos vartojimo išlaidų mažinimu, nes tokiu atveju prognozuojamos išlaidos yra mažesnės nei vidutinės. Taip pat nustatyta, kad jaunesni nei 50 metų asmenys ir turintys aukštesnįjį ar aukštąjį išsilavinimą, pasižymi teigiamomis SHAP reikšmėmis, tai rodo teigiamą poveikį informacijos ir ryšių vartojimo išlaidoms. Verta paminėti, kad didžiausias teigiamas atotrūkis nuo vidurkio (per daugiau nei 10 Eur) yra pastebimas būtent jaunų (iki 30 metų) ir neturinčių aukštojo ar aukštesniojo išsilavinimo namų ūkių. Verta pabrėžti, kad nors prognozuojamų išlaidų dydis mažėja priklausomai nuo amžiaus augimo, tačiau net ir vyresnio amžiaus grupėse (nuo 70 metų) aukštesnįjį ar aukštąjį išsilavinimą turintys asmenys ir toliau išsiskiria didesnėmis informacijos ir ryšių paslaugoms skirtomis išlaidomis, juos lyginant su gyventojais, turinčiais aukštesniojo ir aukštojo nesuteikiantį pavidurinį išsilavinimo lygį (raudoni taškai virš mėlynų), o tai rodo teigiamą išsilavinimo poveikį informacijos ir ryšių paslaugų vartojimo išlaidoms, nors gyventojų amžius ir didėja.

Finansinių paslaugų išlaidų priklausomybės nuo mikrolygmens veiksnių vertinimas

Anksčiau buvo nustatyta (žr. 4 priedą), kad draudimo ir finansinių paslaugų vartojimo išlaidų suma tarp Lietuvos gyventojų siekė 213,45 Eur per mėnesį. Taikant 3 % viršutinę vintorizacijos metodo ribą, maksimali išlaidų suma buvo apribota iki 59,79 Eur. Draudimo ir finansinių paslaugų vartojimo išlaidų priklausomybės nuo Lietuvos namų ūkių charakteristikų vertinimui atlikti buvo palyginti trys

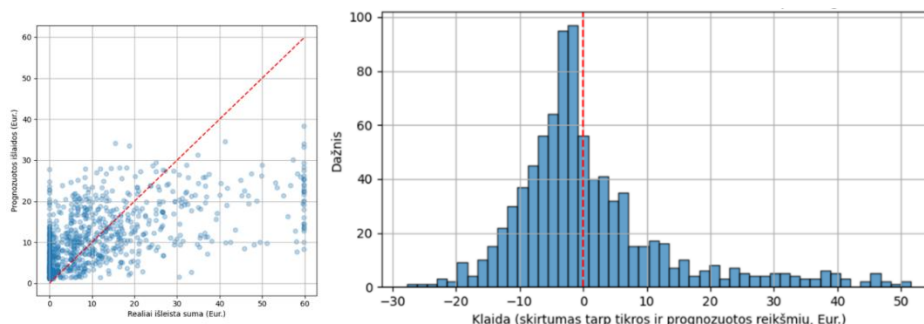
sukurti regresijos modeliai (žr. lent.). Kiekvieno metodo atveju buvo pateiktas pradinis modelio variantas, t. y., su neapdorotu priklausomu kintamuoju, neišmestais nereikšmingais kintamaisiais, tai leido įvertinti pradinę modelio kokybę ir ją vėliau palyginti su optimizuotu modeliu. Geriausias konkretaus metodo modelio variantas buvo tas, kuris po duomenų paruošimo (išskirčių tvarkymo, logaritmvimo, nereikšmingų požymių pašalinimo (žr. 31 priedą) ir modelio optimizavimo (hiperparametrų derinimo) pasiekė aukščiausias kokybės metrikų reikšmes tarp visų tuo metodu bandytų variantų (žr. 36 priedą).

Tyrimo sukurtų regresijos modelių palyginimas, kai priklausomas kintamasis yra draudimo ir finansinių paslaugų vartojimo išlaidos

	Daugialypė tiesinė regresija		Atsitiktinių miškų regresija		Ekstremalaus gradientinio stiprinimo algoritmas	
	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas	Pradinis variantas	Galutinis variantas
Nereikšmingi požymiai išmesti	Ne	Taip	Ne	Taip	Ne	Ne
Taikyta vinzorizacija/proc.	Ne	Ne	Ne	Taip/3%	Ne	Taip/3%
Logarimtuotos išlaidos	Ne	Taip	Ne	Ne	Ne	Taip
MAE, Eur.	10,13	10,31	10,66	8,59	11,23	8,88
RMSE, Eur.	18,91	19,16	19,34	12,62	20,36	12,71
R²	0,27	0,26	0,25	0,31	0,16	0,30
MedAE, Eur.	5,53	5,43	6,14	5,63	6,16	6,03

Daugialypės tiesinės regresijos galutinis modelio variantas, vertinant pagal kokybės metrikas, pasirodė prastesnis nei pradinis, dėl siekio tenkinti tiesinės regresijos taikymui reikalingas prielaidas. Vertinant draudimo ir finansinių paslaugų vartojimo išlaidas, geriausiu modeliu buvo pripažintas atsitiktinių miškų regresijos galutinis variantas ($R^2 = 0,31$), kuris aplenkė tiek ekstremalaus gradientinio stiprinimo ($R^2 = 0,30$), tiek daugialypės tiesinės regresijos ($R^2 = 0,26$) modelius. Tai reiškia, kad geriausias modelis gali paaiškinti iki 31 % draudimo ir finansinių paslaugų išlaidų dispersijos. Vertinant modelio kokybę pagal klaidų metrikas, atsitiktinių miškų modelis taip pat yra panašesnis – jo MAE, RMSE ir MedAE reikšmės tarp visų modelių buvo arčiausiai nulio, kai vidutinė absoliuti paklaida siekė 8,59 Eur, o ekstremalaus gradientinio algoritmo atveju – 8,88 Eur ir 10,13 Eur – daugialypės tiesinės regresijos atveju. Galima teigti, kad 50 % modelių atliktų prognozių nuo tikrųjų draudimui ir finansinėms paslaugoms skiriamų išlaidų sumų nukrypsta per mažiau nei 5,63 Eur.

Tyrimo metu geriausiu pripažinto modelio optimalūs hiperparametrai buvo rasti naudojant Bajeso optimizavimo algoritmą. Hiperparametrų optimizavimo metu buvo naudojama MAE metrika ir maksimalus iteracijų skaičius buvo nustatytas iki 150. Geriausio atsitiktinių miškų regresijos modelio hiperparametrų derinys buvo toks: 656 medžiai, 5 stebinių skaičius lapo mazge, didžiausias medžio gylis – 13, minimalus stebinių skaičius mazgo padalinimui – 33, o didžiausia požymių dalis, svarstoma kiekviename mazge padalinimui – 0,34. Vertinant modelio kokybę pagal metrikas, apskaičiuotas naudojant mokymo imties duomenis, buvo gauta, kad MAE = 7,26 Eur, RMSE = 10,48 Eur, $R^2 = 0,45$, MedAE = 4,83 Eur. Pakankamai nedideli kokybės metrikų skirtumai tarp mokymo ir testavimo duomenų rodo, kad sukurtas modelis nebuvo permokytas. Modelio, sukurtas taikant atsitiktinį miškų algoritmą, kokybei įvertinti taip pat buvo pateikti diagnostikos grafikai.



Atsitiktinių miškų regresijos modelio diagnostika prognozuojant draudimo ir finansinių paslaugų vartojimo išlaidas

Ir šiuo atveju matyti panaši situacija kaip ir ankstesnių regresinių modelių atveju, t. y., rezultatai pasižymi didele dispersija bei modelis dideles vartojimo išlaidų sumas (didesnės už 30 Eur) yra linkęs nepakankamai įvertinti, nes dauguma taškų yra susitelkę žemiau idealios linijos. Remiantis paklaidų histograma galima teigti, kad didžioji dalis paklaidų yra intervale nuo 0 Eur, iki – 5 Eur, kai dažniausiai pasitaiko neigiamos paklaidos, o tai reiškia, kad modelis yra labiau linkęs nuvertinti tikrąsias vartojimo išlaidas, ypač mažesnių sumų (0 – 10 Eur) atveju.

SHAP analizės suminė diagrama, sukurta remiantis atsitiktinių miškų regresijos modelio rezultatais, padėjo identifikuoti svarbiausius veiksnius, remiantis jų vidutiniu indėliu į modelio prognozę, jį lyginant su vidutine prognoze. Vidutinė prognozuota mėnesinė draudimo ir finansinių paslaugų vartojimo išlaidų suma siekė 11,11 Eur.

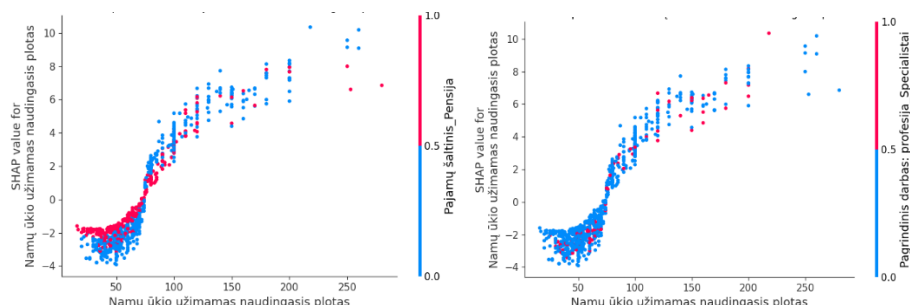


Draudimo ir finansinių paslaugų vartojimo išlaidos (Eur) pagal SHAP reikšmę

Požymių poveikis draudimo ir finansinių paslaugų vartojimo išlaidoms, nustatytas naudojant SHAP metodą, pritaikytą pagal atsitiktinių miškų algoritmą sukurta regresijos modelį

Remiantis SHAP suminės diagramos informacija, matyti, kad didžiausią poveikį prognozuojamoms draudimo ir finansinių paslaugų išlaidoms turi namų ūkio užimamas naudingasis plotas, pensija kaip pagrindinis pajamų šaltinis bei priklausymas 5-ai pajamų kvintilinei grupei, kai didesnis užimamas naudingasis plotas ir priklausymas 5-ai kvintilinei grupei turi teigiamą poveikį, nes gali padidinti išlaidas daugiau nei per 10 Eur nuo vidurkio, o gaunamos pajamos iš pensijos – atvirkštinį, t. y., pensininkai, juos lyginant su asmenimis, kurių pagrindinis pajamų šaltinis yra darbo užmokestis, yra

linkę išleisti mažiau. Svarbų poveikį finansinių paslaugų vartojimui taip pat turi ir didesnis asmenų skaičius namų ūkyje bei aukštesnis išsilavinimas, pagal kuriuos prognozuojant vidutinė išlaidų suma didėtų apie 5 eurus. Svarbiausių kintamųjų tarpusavio sąveikos buvo tirtos naudojant priklausomybės diagramas. Šios diagramos padeda suprasti veiksnių įtaką modeliui atliktoms išlaidų prognozėms, kartu įvertinant ir sąveiką su kitu veiksnio.



SHAP priklausomybės diagramos, vertinančios svarbių kintamųjų tarpusavio sąveikas prognozuojamų draudimo ir finansinių paslaugų vartojimo išlaidų atžvilgiu

Priklausomybės diagramos atskleidžia tarp draudimo ir finansinių paslaugų vartojimo išlaidų bei namų ūkių užimamo naudingojo ploto egzistuojantį netiesinį ryšį. Daugiausiai namų ūkių, kurių pajamų šaltinis yra pensija, gyvena būstuose, turinčiuose iki 100 m² naudingojo ploto, ir kartu šie požymiai lemia neigiamas SHAP reikšmes, tai reiškia, kad prognozuojamos draudimo ir finansinių paslaugų vartojimo išlaidos, esant tokioms sąlygoms būtų mažesnės nei visų namų ūkių vidurkis. Taip pat matyti, kad specialistų profesiją turintys asmenys, juos lyginant su kvalifikuotais darbininkais ar amatininkais, yra labiau linkę išleisti didesnę sumą draudimo ir finansinių paslaugų kategorijai, ypač jei jie gyvena apie 100 – 180 m² naudingojo ploto būstuose. Visgi kitų profesijų atstovai, gyvenantys dar didesnio ploto būstuose, gali viršyti bendrą prognozuojamą išlaidų vidurkį dar labiau, t. y., daugiau nei 10 Eur.



Lietuvos
mokslo
taryba

PADĖKA

Juditai Pilipauskaitei

*Už dalyvavimą studentų vasaros mokslinės praktikos
konferencijoje „Mokslo Vasara“ 2024 m. rugpjūčio 28 d.*

Lietuvos mokslo tarybos
pirmininkas

Dr. Gintaras Valinčius