MYKOLO ROMERIO UNIVERSITETAS

Tatjana Bilevičienė, Steponas Jonušauskas

ATVIROJO KODO PROGRAMŲ TAIKYMAS RINKOS TYRIMUOSE

Vadovėlis

Vilnius
2013

UDK 004.4:339.13(075.8)
Bi238

Recenzavo:

doc. dr. Eugenijus Telešius, Informacinių technologijų institutas
prof. dr. Vitalija Rudzkienė, Mykolo Romerio universitetas

Autorių indėlis:

doc. dr. Steponas Jonušauskas – 4,2 autorinio lanko
doc. dr. Tatjana Bilevičienė – 5,4 autorinio lanko

Mykolo Romerio universiteto Socialinės informatikos fakulteto Informatikos ir programų sistemų katedros 2012 m. spalio 11 d. posėdyje (protokolo Nr. 1INFPSK-16) pritarta leidybai.

Mykolo Romerio universiteto Mokslo programos „Socialinės technologijos“ komiteto 2012 m. lapkričio 5 d. posėdyje (protokolo Nr. MPK2-4) pritarta leidybai.

Mykolo Romerio universiteto Socialinės informatikos fakulteto tarybos 2012 m. lapkričio 8 d. posėdyje (protokolo Nr. 10-55-(1.15 E-20701)) pritarta leidybai.

Mykolo Romerio universiteto Mokslinių-mokomųjų leidinių aprobavimo leidybai komisijos 2012 m. lapkričio 14 d. posėdyje (protokolo Nr. 2L-38) pritarta leidybai.

Visos knygos leidybos teisės saugomos. Ši knyga arba kuri nors jos dalis negali būti dauginama, taisoma arba kitu būdu platinama be leidėjo sutikimo.

ISBN 978-9955-19-513-9 © Mykolo Romerio universitetas, 2013

Turinys

PRATARMĖ 5
1. ATVIROJO KODO PROGRAMŲ TAIKYMAS 8
1.1. Atvirojo kodo programų ypatumai 8
1.2. Statistikos programinių paketų apžvalga 18
1.3. Atvirojo kodo statistinio paketo pasirinkimas 33
2. RINKOS TYRIMŲ ORGANIZAVIMAS 34
2.1. Rinkos tyrimų paskirtis ir reglamentavimas 34
2.2. Statistinio tyrimo projektas 37
2.3. Tyrimo tikslas ir tyrimo modelis 38
2.3.1. Tyrimo problemos ir tikslo formulavimas 38
2.3.2. Koncepcijų žemėlapis 41
2.4. Kokybiniai ir kiekybiniai tyrimai. Tyrimo tipo pasirinkimas 43
2.5. Anketinės apklausos metodas 50
2.5.1. Kintamųjų tipai ir matavimo skalės pasirinkimas 52
2.5.2. Imties sudarymo būdas ir dydis 58
2.5.3. Anketos kodavimo schemos sudarymas. Duomenų kodavimas 65
3. DUOMENŲ SUVEDIMAS Į KOMPIUTERĮ 69
3.1. Pspp programų paketo duomenų redaktorius 69
3.2. Pspp duomenų bazės pildymas 76
3.3. Duomenų įkėlimas iš kitų šaltinių 89
4. PIRMINĖ STATISTINĖ DUOMENŲ ANALIZĖ 94
4.1. Duomenų keitimas ir atranka 94
4.2. Duomenų atranka (filtravimas) 105
4.3. PSPP paketo sintaksės redaktorius 110
4.4. Aprašomoji statistika 112
4.4.1. Vienmatė dažnių analizė 118
4.4.2. Porinės dažnių lentelės 121
4.4.3. Pagrindiniai grafikai: histogramos ir skritulinės diagramos 125
4.4.4. Padėties ir sklaidos charakteristikos 130
5. STATISTINĖS HIPOTEZĖS 144
5.1. Imties skirstiniai. Įverčiai 144
5.2. Statistinių hipotezių tikrinimas 145
6. RYŠIŲ ANALIZĖ 156
6.1. Kintamųjų, išmatuotų nominalioje skalėje arba rangų skalėje su nedideliu kategorijų skaičiumi, statistiniai ryšiai 156
6.1.1. Pirsono χ^2 suderinamumo kriterijus 156
6.1.2. Pirsono χ^2 suderinamumo kriterijaus skaičiavimas (nominalūs kintamieji) 159
6.1.3. Pirsono χ^2 suderinamumo kriterijaus skaičiavimas (ranginiai kintamieji) 164
6.1.4. Kiti χ^2 kriterijaus skaičiavimo būdai 168
6.1.5. Nominaliųjų kintamųjų ryšio matai 169

6.2. Koreliacinė analizė ... 171
6.2.1. Pirsono koreliacijos koeficientas ... 174
6.2.2. Pirsono koreliacijos koeficiento skaičiavimas ... 175
6.2.3. Spirmeno (*Spearman*) koreliacijos koeficientas ... 180
6.2.4. Spirmeno koreliacijos koeficiento skaičiavimas ... 182
6.2.5. Kiti kokybinių kintamųjų ryšio matai ... 185
6.2.6. Kiti tarpusavio ryšio matai ... 188
7. REGRESINĖ ANALIZĖ ... 198
7.1. Tiesinis (paprastas) regresijos modelis ... 198
7.2. Tiesinė regresinė analizė pspp ir spss terpėje ... 201
7.3. Tiesinė regresinė analizė ms excel terpėje ... 204
7.4. Daugialypės tiesinės regresijos modelis ... 208
7.5. Daugialypė tiesinė regresija PSPP ir SPSS terpėje ... 209
7.6. Daugialypė tiesinė regresija *MS EXCEL* terpėje ... 212
8. SKIRTUMŲ ANALIZĖ ... 215
8.1. Hipotezės apie vidurkių lygybę ... 215
8.2. *T* kriterijus vienai imčiai ... 216
8.2.1. *T* kriterijaus vienai imčiai skaičiavimas ... 217
8.3. *T* kriterijus dviem nepriklausomoms imtims ... 219
8.3.1. *T* kriterijaus dviem nepriklausomoms imtims skaičiavimas ... 221
8.4. *T* kriterijaus dviem priklausomoms (porinėms) imtims skaičiavimas pspp ir spss terpėje ... 228
8.5. Dispersinė analizė ... 230
8.6. Vienfaktorinė dispersinė analizė PSPP ir SPSS terpėje ... 235
8.6.1. Kontrastų tikrinimas ... 239
8.7. Skirtumų analizė *MS EXCEL* terpėje ... 242
8.7.1. Vienfaktorinė dispersinė analizė ... 242
8.7.2. *T* kriterijus dviem priklausomoms (porinėms) imtims ... 244
8.7.3. *T* kriterijus dviem nepriklausomoms imtims ... 246
9. FAKTORINĖ ANALIZĖ ... 251
9.1. Faktorinė analizė PSPP ir SPSS terpėje ... 264
10. KLAUSIMINŲ PATIKIMUMO ANALIZĖ ... 274
10.1. Kronbacho ***ALFA*** skaičiavimas ... 276
10.2. Konkordancijos koeficiento skaičiavimas ... 278
11. LOGISTINĖ REGRESIJA IR *ROC* ANALIZĖ ... 280
11.1. Logistinė regresija ... 280
11.1.1. Logistinė regresija SPSS terpėje ... 281
11.2. Operatoriaus charakteringa kreivė ... 284

Literatūra ... 290
1 priedas. χ^2 skirstinio α lygmens kritinės reikšmės ... 293
2 priedas. Stjudento skirstinio α lygmens kritinės reikšmės $t_{\alpha}(n)$ ... 295
3 priedas. Atsakymai ... 297

PRATARMĖ

Globalizavimo procesai, žinių visuomenės plėtra, naujos technologijos keičia šiuolaikinio pasaulio ekonominį ir socialinį modelį. Ekonomikos augimą labiausiai skatina ne paklausa, o perteklinė pasiūla, sparčiai didėja konkurencija. Jau XX a. antrojoje pusėje verslininkai suprato, kad ateityje konkurencinėje kovoje laimės tos įmonės, kurios geriausiai tenkins vartotojų poreikius. Verslo pasaulyje įsivyravo naujoji rinkodara – strategijų ir metodų rinkinys, galintis paskatinti potencialius klientus patikėti paslaugomis ir prekėmis. Norėdami suvokti įmanomas grėsmes ir galimybes, verslo atstovai siekia kuo tikslesnės informacijos apie savo srities rinką, potencialius vartotojus ir konkurentus. Vertinant vartotojo elgesį svarbu suprasti, kad rinkoje veikia skirtingos asmenų grupės su skirtingais interesais. Norint veikti rinkoje, reikia žinoti, koks yra vartotojų skonis, įpročiai, finansinės galimybės. Tik tada galima sėkmingai pasinaudoti savo galimybėmis ir gauti pelną. Rinkos tyrimai – tai kokybinės ir kiekybinės informacijos apie paklausą bei realių ir potencialių prekių bei paslaugų pasiūlą rinkimas, analizė ir gautų apibendrintų išvadų pateikimas taktiniams ar strateginiams komerciniams sprendimams priimti.

Rinkos tyrimų rezultatai yra įmonės steigimo sprendimo priėmimo, naujo rinkos segmento perpratimo, prekybinių plotų praplėtimo, filialo arba atstovybės kitame regione atidarymo, prekių linijos praplėtimo ir t. t. pagrindas Visi tie sprendimai yra siejami su pinigų investavimu, o tai reiškia, kad siejami ir su jų praradimo rizika. Būtent tam, kad būtų galima sumažinti investuojamų lėšų praradimo tikimybę, pagrįsti investicijas ir pateikti labiau apibrėžtus jų grąžinimo laikotarpius, yra vykdomi rinkos tyrimai. Tyrimų rezultatai yra pateikiami kartu su plėtojamo projekto verslo planu. Pastaraisiais dešimtmečiais net ir mažos įmonės atlieka rinkos tyrimus, o gautą informaciją naudoja rinkos dydžiui, konkurencinei aplinkai įvertinti, taip pat naujai rinkos strategijai (produkto, kainos, paskirstymo, pardavimo ir kt.) diegti. Tačiau pastaraisiais metais buvo pastebėtas verslo atstovų dėmesio rinkos tyri-

mams sumažėjimas. Pagrindinė nusivylimo tyrimų rezultatais priežastis – skirtingas verslininkų ir rinkos tyrimo specialistų požiūris į tyrimo procesą. Užsakydami rinkos tyrimą, įmonės atstovai dažnai nežino statistinių tyrimų ypatumų ir galimybių, o tyrėjai ne visada gerai supranta įmonės veiklos specifiką. Todėl verslininkai mano, kad rinkodaros tyrimai daugiausia teikia tik teorinę informaciją, kurią būtų sunku pritaikyti praktiškai, be to, šie tyrimai yra brangūs, todėl jų organizavimas neretai tėra nereikalingas įmonės lėšų švaistymas. Tyrimo reikalingumas ne tik turi būti pagrįstas, suderinti tyrimo tikslai, uždaviniai ir metodai, bet tyrimo proceso metu reikia, kad bendradarbiautų rinkodaros specialistai ir statistiniai tyrimų specialistai. Jie turi kartu aptarti tyrimo rezultatus ir suformuoti išvadas. Tik tokiu atveju rinkos tyrimas turės teigiamą poveikį įmonės veiklai.

Rinkos tyrimo pramonė nustato aukštas profesines ir etikos normas. Veiksmingas rinkos tyrimas priklauso nuo visuomenės pasitikėjimo, kad tyrimas atliekamas sąžiningai, objektyviai, kad jis paremtas pasirengimu bendradarbiauti. Profesiniai praktikos kodeksai, kurie numato rinkos tyrimo projektų atlikimo būdus, skatina šį pasitikėjimą. Pirmą tokį kodeksą 1948 m. paskelbė Europos viešosios nuomonės ir rinkos tyrimų asociacija ESOMAR (*European Society for Opinion and Marketing*). Vėliau buvo parengta daug kitų kodeksų, juos parengė nacionalinės rinkodaros organizacijos, tokios kaip Tarptautiniai prekybos rūmai (ICC – *International Chamber of Commerce*), atstovaujantys tarptautinei rinkodaros bendruomenei. Šiuo metu daugiau kaip 5000 ICC/ESOMAR narių visame pasaulyje laikosi tarptautinio rinkos ir socialinių tyrimų praktikos kodekso reikalavimų.

Rinkos tyrimų procesas apima duomenų rinkimą, analizę ir interpretavimą. Rinkos tyrimų esmę sudaro specialios metodikos, pradedant nuo paprasčiausių ir baigiant naujausiais sudėtingais tyrimo metodais, paremtais sudėtingais matematiniais skaičiavimais, kurių be šiuolaikinės kompiuterinės ir programinės įrangos praktiškai neįmanoma atlikti. Susiduriama ir su statistikos programinių paketų pasirinkimo problema: jų yra daugybė – nuo paprastų iki profesionalių, nuo pigių iki kainuojančių dešimtis tūkstančių eurų, nuo pateiktų atviruoju kodu iki pateiktų vartotojui neprieinama programavimo kalba. Kai statistiniai

tyrimai atliekami savarankiškai, ne visada įmanoma naudoti brangiai kainuojančią specializuotą statistinei informacijai apdoroti skirtą programinę įrangą. Todėl svarbu žinoti kitokių kompiuterinių programų galimybes. Pastaraisiais metais padidėjo atvirojo kodo (*Open Source Definition, OSD*) programų populiarumas. Atvirojo kodo programos platinamos nemokamai, pagal atvirojo kodo licenciją. Tokia programinė įranga dažniausiai pasižymi šiomis savybėmis: nemokama, laisvai prieinami išeities kodai, galima laisvai platinti ir modifikuoti nekeičiant licencijos. Reikalingų atvirojo kodo programų galima ieškoti naudojant įprastas interneto paieškos sistemas. Kita vertus, yra specializuoti portalai, skirti atvirojo kodo programoms saugoti, platinti ir kurti bendradarbiaujant. Didelė šių programų dalis suderinama su *MS Windows* programine įranga, turi panašios išvaizdos vartotojo sąsają. Neblogas statistinių duomenų apdorojimo galimybes turi ir MS Excel programinė įranga.

Autoriai pateikia skaitytojams (nebūtinai tik studentams) vadovėlį, kuriame glaustai, be sudėtingos matematikos bandoma aprašyti, kaip atlikti rinkos tyrimus, kokią statistiką taikyti pateikiant ir interpretuojant gautus rezultatus. Vadovėlyje naudojami studentų tyrimo darbų pavyzdžiai, kuriems spręsti pasirinkti skirtingi programų paketai: atvirojo kodo programų paketas PSPP – supaprastintas patentuoto ir plačiai žinomo statistikos programų paketo SPSS variantas; programų paketas SPSS (*Statistical Package for the Social Science*) (šio paketo licenciją turi nemažai universitetų, jį naudoja ir verslo įmonės), MS Excel. Nagrinėjama programinė įranga knygoje aprašoma tiek, kiek reikia, kad būtų galima išspręsti pagrindinius rinkos tyrimo uždavinius, t. y. vadovėlis nepretenduoja į išsamią PSPP, SPSS, MS Excel programinių paketų apžvalgą, juo labiau į išsamų jų mokymą. Turint omenyje, kad PSPP nerealizuoja visų SPSS taikomų statistinės analizės metodų, autoriai nusprendė atskleisti tik bendras šios programinės įrangos galimybes. Tie patys statistiniai uždaviniai sprendžiami skirtingais būdais, norint parodyti, kad gaunami rezultatai nesiskiria.

Tikimės, kad šis vadovėlis padės skaitytojams parengti tyrimų projektą, surinkti duomenis, juos visapusiškai išanalizuoti ir pateikti vertingą ataskaitą.

1. ATVIROJO KODO PROGRAMŲ TAIKYMAS

1.1. ATVIROJO KODO PROGRAMŲ YPATUMAI

Atvirojo kodo programos šiandien yra tema, prie kurios visi sugrįžta kalbėdami apie programinės įrangos pasirinkimą. Vieniems tai kritikos objektas, kitiems savotiškas šventasis Gralis. Bandysime aptarti atvirojo kodo atsiradimo priežastis, istoriją, teisinę atvirojo kodo dalį, pakalbėsime apie mitus, susijusius su atviruoju kodu.

Bet kokio kūrinio autorius pagal įstatymus turi tam tikras teises į savo kūrinį. Autorius, sukūręs kompiuterinę programą ar jos dalį, nuo jos sukūrimo momento įgauna autoriaus teises į ją: išskirtines teises teikti savo teisių subjektui – kompiuterinei programai ar jos daliai – licencines sąlygas. Licencija yra suprantama kaip „autorių teisių, gretutinių teisių ar *sui generis* teisių subjekto (licenciaro) leidimas, suteikiantis kūrinio, gretutinių teisių ar *sui generis* teisių objekto naudotojui (licenciatui) teisę naudoti kūrinio, gretutinių teisių ar *sui generis* teisių objekto originalą arba jo kopijas (licencijos dalyką) nurodytoje teritorijoje tokiu būdu ir tokiomis sąlygomis, kaip numatyta licencinėje sutartyje“[1]. Tai intelektinė nuosavybė, kuri suprantama kaip neatimama asmens nuosavybė ir gali būti perduodama kam nors kitam lygiai taip pat, kaip ir kitas asmeninis turtas. Autorių teisėms į kompiuterių programą atsirasti pakanka programos sukūrimo fakto, apie turimas autoriaus teises taip pat galima pareikšti pasitelkiant autorių teisių ženklą © arba (c).

Lietuvoje kompiuterinių programų autorių teises saugo Lietuvos Respublikos autorių teisių ir gretutinių teisų įstatymas Nr. VIII-1185, atsakomybė už šių teisių pažeidimą numatyta Lietuvos Respublikos administracinių teisės pažeidimų kodekso 214(10) straipsnyje: „Autorių teisių ir gretutinių teisių pažeidimas“.

[1] Lietuvos Respublikos autorių teisių ir gretutinių teisų įstatymas.

Atvirojo kodo principas – tai laisva autoriaus valia suteikti savo kūriniui, t. y. programai, perpanaudojimo, tobulinimo ir kitas programinio kodo laisves, nustatytas atvirojo kodo licencijoje(-se).

Teisiniu požiūriu atvirasis kodas – tai programiniai produktai, kurių naudojimas, platinimas ir modifikavimas yra nustatytas atvirojo kodo licencijomis bei jų sąlygomis, su kuriomis sutinkantis vartotojas produktu gali disponuoti tiek, kiek jam tai suteikia licencija.

Praėjusio šimtmečio paskutiniaisiais dešimtmečiais sparčiai plečiantis internetui, atvirojo kodo programinė įranga užėmė svarbias pozicijas reguliuojant informacijos srautus, tvarkant elektroninį paštą ar dirbant kasdienius darbus kompiuteriu. Programinės įrangos kūrimo bendrovės stengiasi suderinti savo programas su *Linux* atvirojo kodo operacine sistema, visiškai ar iš dalies atveria savo programų kodus. Naujų kompiuterių pirkėjai planuoja sumažinti išlaidas naujai programinei įrangai pirkti, vietoj įprastinių paketų pasirinkdami *Linux* operacinę sistemą bei atvirojo kodo taikomąsias programas.

1970-ųjų pabaigoje JAV telekomunikacijų gigantui *AT&T* perėmus tuomet pažangiausios operacinės sistemos UNIX nuosavybės teises ir kartu uždraudus savanoriams programuotojams laisvai naudotis jos išeities tekstais, atsirado atvirojo kodo programos kaip tam tikras programuotojų atsakas į besikuriančias programinės įrangos monopolijas. Statistiniai duomenys liudija, kad šiandien didesnę dalį visos programinės įrangos dalies užima uždarojo kodo programinės milžinės *Microsoft* komerciniai produktai. Tačiau atvirojo kodo kompiuterių programos jau yra rimta alternatyva komercinėms privačioms kompiuterių programoms.

Kompiuterio programų autoriai išeities tekstus sukuria ir perkelia į viešus interneto tinklalapius ar kitus kompiuterių tinklus, čia jie faktiškai tampa laisvai prieinami kiekvienam besidominčiam. Šiandien vyrauja kelios skirtingos atvirojo kodo kompiuterių programų teisinės apsaugos koncepcijos, kurios reglamentuoja iš esmės nedaug viena nuo kitos besiskiriančias išeities tekstų pakeitimų, patobulinimų, papildymų ir kitas teises.

Atvirojo kodo programos evoliucionavo iš termino *free software* – laisva programinė įranga. Jos koncepciją sukūrė Richardas Stallmanas,

parinkęs anglų kalba kelias reikšmes turintį apibūdinimą *free*. Anot jo, programinė įranga turi būti „laisva kaip žodžio laisvė, bet ne kaip nemokamas alus“. Taip R. Stallmanas, atvirojo kodo ideologas, mėgino pasakyti, kad iš tiesų atvirojo kodo programinė įranga nėra nemokama. Autorius, programinį kodą padarydamas laisvai prieinamą, padaro programą nemokamą.

Galima įvardyti du svarbius momentus atvirojo kodo licencijų evoliucijoje. Viskas prasidėjo 1970-ųjų pabaigoje JAV. Vienas iš jų – R. Stallmano įkurta organizacija *Free Software Foundation*, ir jo pastangomis sukurta *GNU General Public License* – GPL. Kitas – JAV Kalifornijos Berklio universiteto kurta UNIX operacinės sistemos atmaina, iš kurios atsirado BSD (angl. *Berkeley Software Distribution*) UNIX programų šeima bei BSD genties licencijos.

GPL tipo licencija garantuoja šias teises ir laisves – laisvę platinti programinę įrangą, laisvę gauti jos programinį kodą ir laisvę jį redaguoti/naudoti. GPL licencija į programinės įrangos terminologiją įnešė *copyleft* terminą, reiškiantį neatsiejamą prievolę bet kokiam išvestiniam darbui iš GPL licencijuoto produkto suteikti tos pačios GPL licencijos teises ir pareigas.

Free BSD tipo licencija numato daugelį identiškų sąlygų, kaip ir GPL, tačiau suteikia vartotojui, įdėjusiam savų žinių į esamą programą, teisę ją „sukompiliuoti“, t. y. uždaryti, nesuteikiant kitiems vartotojams „uždarytos“ programos išeities tekstų ir kt.

Atvirojo kodo programinės įrangos kūrimo procesas yra paremtas komunos arba visuomenės principu. Programa tobulinama keičiant, siūlant ir taisant, tai teoriškai gali daryti bet kuris programuotojas savanoris. Pirminis programos variantas yra gerokai patobulinamas ir grįžta pas jos kūrėją/kūrėjus pastebimai atnaujintas. Didesniuose projektuose dažniausiai yra keletas koordinatorių, atsakingų už produkto sukūrimą. Žymiausias tokio atvirojo komandinio darbo pavyzdys – *Linux* operacinė sistema, kurios branduolio kūrimo procesą koordinavo suomių programuotojas Linusas Torvaldsas su keliais bendraminčiais.

Atvirojo kodo programos nuo komercinių uždarojo kodo programų skiriasi dviem aspektais.

Pirmas. Atvirojo kodo programų vartotojai gauna tiek programinį kodą, tiek jau sukompiliuotą (paverstą kompiuteriui suprantamu dvejetainių skaičių kodu) kodą. Programinis kodas – tai programos veikimo mechanizmas, parašytas programavimo kalba, vadinamasis išeities tekstas. Sukompiliuotas kodas, pavirtęs kompiuteriui suprantama kalba, vartotojui jau yra nepasiekiamas, t. y. vartotojas jo tobulinti/keisti negali. Komercinių programų kūrėjai paprastai pateikinėja tik jau sukompiliuotų programų kopijas, nepateikdami išeities tekstų, kurie vartotojams leidžia perskaityti ir suprasti programos veikimo principus, juos koreguoti, taip pat iš jų kurti išvestines programas.

Antras. Atvirasis kodas yra pateikiamas remiantis ne naujai parašyta, bet specifine „viešąja" (t. y. visiems prieinama ir pritaikoma) licencija, tokia kaip GPL. GPL galima suprasti kaip tipinę komercinę sutartį, su kurios sąlygomis vartotojas sutinka išpakuodamas pateiktos programinės įrangos dėžutę arba nuspausdamas kompiuterio pelės klavišą. GPL licencijos sąlygos vartotojui suteikia teisę laisvai naudotis pateikiama programa ir ją tobulinti. Pagrindinė sąlyga, kurią atlikti įpareigoja ši licencija, – tai visiems išvestiniams darbams, sukurtiems pateiktos programos pagrindu, taikyti GPL licenciją bei jos keliamus reikalavimus. Tuo tarpu komercinės licencijos licenciatui suteikia tik naudojimosi teises ir draudžia daryti bet kokius pakeitimus. Komercinių produktų licencijos dažniausiai yra suteikiamos ribotam kompiuterių skaičiui.

Dažnai yra painiojami atvirojo kodo principu sukurti produktai su viešai, t. y. internetu, platinama programine įranga – *shareware* bei *freeware*. Abiem atvejais tai yra licencijos sąlygos, susijusios su specifiniais įkainojimo ir naudojimo kriterijais, bet, kaip rodo praktika, nesuteikia jokios prieigos prie paties programinio kodo, neleidžia daryti patobulinimų ir pan. Šios programos dažniausiai būna bent jau tam tikrą laiko tarpą nemokamos, o jų licencijos vartotoją saisto daugybe įprastoms komercinėms programoms būdingų įsipareigojimų.

Praktikoje atvirojo kodo programos asocijuojasi su nemokamomis programomis ir jų naudojimu *Linux* operacinėje sistemoje. Tačiau toks požiūris yra neteisingas ir neatskleidžia tikrosios atvirojo kodo paskirties – licencijų sąlygomis suteiktos laisvos prieigos prie programos išeities kodo. Atvirojo kodo programos gali bet ne turi būti nemoka-

mos, jos lygiai taip pat, kaip ir komerciniai produktai, yra ginamos įstatymo ir jų naudojimas nėra ribojamas jokia operacine sistema.

Reikėtų atkreipti dėmesį į, mūsų nuomone, svarbų teisinį trūkumą: yra tikimybė padaryti teisės pažeidimą, panaudojant kažkieno autorių teisėmis arba patentais apsaugotą programinį kodą ir perpateiktą kaip atvirąjį kodą. Neįsigilinus į licencines sąlygas (pavyzdžiui, GPL), gali atsirasti nepageidaujama prievolė paviešinti visą programinio produkto išeities kodą, jei nors dalis jo buvo sukurta pagal *copyleft* sąlygas keliančią licenciją. Panašių bėdų gali atsitikti ir dėl didelio „oficialių" atvirojo kodo licencijų skaičiaus: vartotojai rizikuoja nukentėti ir kuo nors įsipareigoti, neįsigilinę į kiekvienos jų sąlygas.

Esama daug atvirojo kodo praktinio pritaikymo formų – įmonės veikimui užtikrinti galima pasirinkti atvirojo kodo principu parašytą ir nemokamą buhalterinės apskaitos, tinklo išteklių valdymo ir kitas programas. Arba tiesiog pasirinkti vieną iš nemokamai internetu platinamų operacinių sistemų – tokių kaip *Caldera Open Linux*, *RedHat Linux*, *Debian/GNU Linux*, *SuSe Linux* paketų kartu su jau pridedamomis įvairiausiomis programomis, jomis aprūpinti visą įmonę ir už tai nieko nemokėti.

Kam kuriamos tokios programos, jeigu vėliau atiduodamos veltui? Ne, veltui niekas nėra dalijama, prisiminkime atvirojo kodo idėją – „laisvas kaip žodis, bet ne kaip nemokamas alus". Atvirojo kodo pasaulis gauna pelną iš tų pačių atvirojo kodo programinių paketų kompaktinių diskų, juos pardavinėjant praktiškai už pagaminimo savikainą su kartu pridedamomis naudojimosi instrukcijomis ir kita pagalba, už kompiuterinių tinklų, kuriuose „sukasi" atvirojo kodo programos, priežiūrą ir tvarkymą, už naujokų apmokymą naudotis nauja aplinka bei programomis.

Mokamų ir nemokamų programų „karas" pamažu blėsta, nes vis daugiau sprendimų veikia debesų kompiuterijos pagrindu. Vartotojai šiandien vis dažniau ne perka licenciją, o moka už tinkle naudojamą programinę įrangą kaip tam tikro lygio paslaugą, arba gauna ją nemokamai mainais už rodomą reklamą. Nemažai vadinamųjų debesų kompiuterijos produktų turi nemokamas apribotų funkcijų versijas. Todėl daugeliui vartotojų, ypač ieškančių programų asmeniniam naudojimui ar smulkiajam verslui, vis svarbiau tampa tai, ar jie galės naudotis šia programine įranga iš bet kur ir bet kada, kokiomis sąlygomis, kiek tai

kainuos per metus, ar saugūs bus jų duomenys. Su interneto paplitimu „atviros inovacijos“ ir „kolektyvinės kūrybos“ koncepcijos aiškiai populiarėja, jų taikymas suteikia pagreitį ir sukuria geresnius rezultatus visiems, todėl net didieji rinkos žaidėjai, ilgai saugoję patentus ir komercines paslaptis, vis plačiau atveria savo programų išeities tekstus ne tik uždaroms partnerių bendruomenėms, bet ir viso pasaulio programų kūrėjų visuomenei.

Vienareikšmiškai sakyti, kad atvirojo kodo programos yra prastesnės ar ne tokios patogios kaip komerciniai produktai, negalima – visko pasitaiko abiem atvejais. Tai ką patarti privačiam ir ypač viešajam sektoriui? Pokyčiai valstybės mastu pareikalautų ne tik išlaidų, bet ir nemažai laiko. Tad kam vargti? Juk sena IT vadovų išmintis sako: „Dar niekas nebuvo išmestas iš darbo už tai, kad pirko pasaulinio lyderio licencijas.“

Šiuo metu yra labai daug mitų apie atvirojo kodo programinę įrangą. Nemažai mitų už atvirąjį kodą ir nemažai mitų prieš. Tačiau blogiausia, kad tie visi mitai gali suformuoti neteisingą supratimą apie atvirojo kodo produktus, o tai ypač aktualu renkantis geriausią programinę įrangą jūsų poreikiams.

Pateiksime keletą dažniausiai apie atvirąjį kodą pasitaikančių *mitų ir neteisingų supratimų*.

Atvirasis kodas yra nemokamas. Atvirojo kodo programinė įranga gali būti be licencijos mokesčio, bet tai ne tas pats, kas sakyti, kad nėra jokių išlaidų ją naudoti. Tikrąją atvirojo kodo kainą sudaro palaikymas (angl. *support*) ir pritaikymas pagal kliento poreikius (angl. *customization*). Jei operacinės sistemos ar duomenų bazės yra didelės sudėtingos programos, joms gali reikėti daug palaikymo – gali tekti samdyti programuotoją arba kitą įmonę, kuri atliktų reikiamus darbus. Paprastesnėms programoms reikia mažiau palaikymo ir dažnai nėra jokio palaikymo mokesčio. Tai, pavyzdžiui, OpenOffice.org ofisinių programų paketas, Firefox naršyklė, Thunderbird pašto programa ir daugelis kitų.

Atvirasis kodas yra Linux. Nors *Linux* operacinė sistema yra geriausiai žinomas atvirojo kodo produktas ir daugelis atvirojo kodo taikomųjų programų yra parašytos pirmiausia (arba tik) *Linux* operacinei sistemai, tačiau atvirasis kodas yra kur kas daugiau negu tik *Linux*. Dalis atvirojo kodo programų yra parašytos tik *Windows* operacinei sistemai,

kitos sėkmingai veikia tiek *Linux*, tiek *Windows* operacinėse sistemose. Todėl naudojant atvirojo kodo programas visai nebūtina dirbti *Linux* operacinėje sistemoje.

Atvirajam kodui nėra palaikymo (angl. *support*). Atvirkščiai, atvirojo kodo programų naujienų grupės yra geriau išvystytos negu komercinių produktų. Esama interneto naujienų grupių ir forumų, kur vartotojai renkasi tam, kad apsikeistų informacija. Dauguma pačių atvirojo kodo kūrėjų palaiko forumus dėl savo vartotojų. Ten galima gauti greitą pagalbą, ypač bendroms programoms ir bendroms problemoms. Kitas palaikymo šaltinis yra kūrėjų internetinės svetainės. Dažnai jose yra žinynų, patarimų ir pastebėtų klaidų ištaisymų, o tai padės jums atsakyti į daugelį klausimų. Taip pat ten rasite vartotojo žinynus, kuriuose bus pateikta informacija, kaip naudotis programa.

Atvirojo kodo programos yra mažiau funkcionalios. Iš tikrųjų daug atvirojo kodo programų funkcionalumu pralenkia ekvivalentiškas komercines programas. Tokie vardai, kaip *Apache, MySQL, PostgreSQL* kalba patys už save.

Suintegruoti atvirąjį kodą ir Windows OS yra labai sunku. Tai *Microsoft* mėgstamiausias mitas. Daug atvirojo kodo programų kuo puikiausiai integruojasi su *Windows OS* ir jo programomis, tokiomis kaip *Excel, Outlook* ir kt. Žinoma, kai kurioms programoms reikalingi tam tikri pakeitimai, kad susiintegruotų, tačiau tie pakeitimai dažniausiai yra nedideli.

Tačiau tai nėra vien tik atvirojo kodo pakeitimai. *Microsoft* programoms taip pat turi būti atlikti tam tikri pakeitimai. Pavyzdžiui, kad senesnės *Microsoft Office* versijos nuskaitytų *Microsoft Office* naujesnių versijų sukurtus failus, reikalinga speciali paslaugų programa, kuri atlieka reikiamus pakeitimus.

Suintegruoti atvirąjį kodą ir Windows OS yra labai paprasta. Taip pat netiesa. Visada atsiras dalykų, kurie neveiks, tačiau daugeliu atvejų bus gana nesudėtinga padaryti, kad atvirasis kodas ir *Windows OS* dirbtų kartu.

Atvirasis kodas yra saugesnis. Blogiukai labiau linkę atakuoti populiarias programas, kurios buvo daugiau komercinės. Kada atvirojo kodo programos tapo labiau žinomos ir naudojamos, jos taip pat buvo pradėtos atakuoti. Visada reikia rūpintis saugumu ir nesvarbu, ar programa yra atvirojo kodo, ar komercinė.

Atvirasis kodas yra mažiau saugesnis. Kai kalbama apie saugumą, pagrindinis atvirojo kodo privalumas yra tas, kad kur kas daugiau akių mato programos kodą. Atvirasis kodas yra laisvai dalijamas ir bet kas gali surasti, paviešinti, netgi ištaisyti klaidas programiniame kode. Daugelis žmonių tai daro ir tai vyksta labai greitai. Komercinių programų atveju visa tai vyksta lėčiau.

Atvirojo kodo programos negali būti išplečiamos (t. y. pritaikomos labai didelėms bendrovėms). Viena iš atvirojo kodo programų charakteristikų yra tai, kad laikui bėgant kūrėjai prideda daugiau savybių ir atnaujinimų atvirojo kodo programoms tam, kad jos būtų pritaikytos labai didelėms bendrovėms.

Atvirojo kodo programas yra sudėtinga naudoti. Pirmosios atvirojo kodo programos buvo sukurtos ekspertų ir orientuotos į programuotojus bei tinklo administratorius, o ne į galutinį vartotoją. Šiandien daugelį atvirojo kodo programų yra labai lengva suinstaliuoti ir naudoti lygiai taip pat, kaip ir geriausias komercines programas. Vis daugiau dėmesio kreipiama į vartotojo sąsają, paslepiant sudėtingumą.

Klausti, ar atvirojo kodo programos jums tinka, yra nekorektiška. Tai tas pats, kas klausti, ar *Microsoft, Oracle* ar kitų žymių bendrovių programos yra tai, ko jums reikia. Reikėtų savęs paklausti, ką jūs norite atlikti, pasiekti? Kai nuspręsite, kokie yra jūsų tikslai, tada galėsite pradėti ieškoti reikiamos programinės įrangos, kuri padės tuos tikslus pasiekti, turint omenyje programos kainą, našumą ir savybes.

Reikalingų atvirojo kodo programų galima ieškoti naudojant įprastas interneto paieškos sistemas. Taip pat yra specializuoti portalai, skirti atvirojo kodo programoms saugoti, platinti ir kurti bendradarbiaujant (dažnai šie portalai vadinami saugyklomis). Šie portalai, be įrankių, skirtų atvirojo kodo programoms kurti bendradarbiaujant, kaupia įvairią statistiką apie projektus: apsilankymų, atsisiuntimų dažnius, klaidų koregavimus ir t. t. Pagal teikiamus rodiklius galima susidaryti nuomonę apie atvirojo kodo kokybę, brandą bei tinkamumą vartotojų poreikiams.

Populiariausios yra šios atvirojo kodo programų saugyklos:

SourceForge. Didžiausia ir plačiausiai žinoma atvirojo kodo projektų saugykla, prižiūrima *VA Software* bendrovės. Ji turi apie pusantro

milijono registruotų vartotojų ir yra sukaupusi daugiau nei 140 tūkstančių projektų. Be programų kūrimo palaikymo priemonių, teikia projekto aktyvumo, srauto statistiką.

GNU Savannah. FSF (angl. *Free Software Foundation*) projektas, skirtas išskirtinai laisvai programinei įrangai kurti ir palaikyti. Naudojama ta pati programinė įranga, kuri taikoma *SourceForge* saugykloje. Taikoma griežtesnė projektų kontrolė, projektų teikėjai turi nurodyti, kuri laisvos programinės įrangos licencija taikoma jų projektui.

BerliOS. Įkurtas FOKUS (Atvirų komunikacijų Fraunhoferio institutas). Daugiausia ši saugykla skirta atvirojo kodo programinei įrangai kurti. Be programų kūrimo palaikymo priemonių, teikia projekto aktyvumo, srauto statistiką.

GForge. Kaip ir *GNU Savannah, GForge* projektas atsirado *VA Software* uždarant pirminius tekstus (*GForge* pagrindas – paskutinė atvira *VA Linux* (tuometinė *VA Software*) sistemos versija). Be programų kūrimo palaikymo priemonių, teikia projekto aktyvumo, srauto statistiką.

Freepository. Išleista kaip atvirojo kodo programinė įranga. Projektams nedaro jokių licencinių apribojimų. Joje sukaupta daugiau nei 12 tūkstančių projektų.

Lietuvos Respublikos teisės aktuose atvirojo kodo apibrėžimas pirmą kartą buvo pateiktas 2002 m. gruodžio 31 d. Lietuvos Respublikos Vyriausybės nutarime „Dėl Elektroninės valdžios koncepcijos patvirtinimo“[2]: atvirasis kodas – laisvai platinami programų tekstai programavimo kalba. Atvirojo kodo programinis produktas apibrėžiamas kaip „sukurtas atvirojo kodo principu programinis produktas, kurį galima modifikuoti laikantis prie jo pridedamų arba jame nurodytų licencijų sąlygų“.

Informacinės visuomenės plėtros komitetas prie Lietuvos Respublikos Vyriausybės jau kelerius metus skatina plačiau naudoti atvirąją programinę įrangą, kuri palaiko atviruosius e. dokumentų standartus. Tokių programų pavyzdžiais galėtų būti lokalizuota naršyklė *Mozila-Firefox*, kalendoriaus programinė įranga *Calendar*, iš dalies lokalizuotas paketas *OpenOffice.org*. Šie lituanizuoti produktai suteikia vartotojams galimybę įsigyti legalią laisvai prieinamą programinę įrangą. Be to, var-

[2] Neteko galios Lietuvos Respublikos Vyriausybės 2009 m. gegužės 13 d. nutarimu Nr. 464.

totojai tampa nebepriklausomi nuo vieno konkretaus gamintojo. Naudojant atvirojo kodo programas ir standartus skatinama programinių produktų įvairovė, užtikrinamas jų sąveikumas, didėja gamintojų konkurencija, todėl mažėja kainos, o tai naudinga vartotojams.

2011 m. gruodžio 8 d. Lietuvos Respublikos Seimas priėmė rezoliuciją „Dėl atvirojo kodo programinės įrangos valstybės ir savivaldybių institucijose įsigijimo ir naudojimo“. Šiame nutarime teigiama, kad šiuo metu Lietuvos valstybės institucijų interneto svetainėse dokumentai ir formos pateikiami plačiai naudojant uždarus nuosavybinius DOC ir XLS formatus, ir tai prieštarauja Lietuvos Respublikos Vyriausybės 2003 m. balandžio 18 d. nutarimui Nr. 480 „Dėl Bendrųjų reikalavimų valstybės institucijų interneto svetainėms patvirtinimo“. Seimas:

- pasisako už tai, kad valstybės, valdymo ir viešojo sektoriaus subjektai pateiktų ir priimtų informaciją naudodami atvirus formatus ir standartus, kad vartotojai ir verslo įmonės galėtų savo nuožiūra nevaržomai naudotis programine įranga;
- remia atvirojo kodo sistemas, jų konkurencingumo didinimą;
- ragina parengti valstybės ir savivaldybių institucijų bei įstaigų perėjimo prie atvirojo kodo programinės įrangos naudojimo strategiją ir nuo 2012 m. 70 procentų sumažinti valstybės ir savivaldybių biudžetų išlaidas uždarojo kodo operacinių sistemų bei biuro programinės įrangos paketams įsigyti, o atliekant viešuosius pirkimus įsigyti kompiuterinę techniką su atvirojo kodo programine įranga.

Rezoliucijoje pabrėžiama, kad Seimas remiasi kitų Europos valstybių (Italijos, Suomijos) patirtimi ir atvirojo kodo programinės įrangos apibrėžtimi, kuria nustatyta, kad turi būti laisva prieiga prie programinės įrangos išeities kodo, išeities kodo neliečiamumas, laisvas programinės įrangos platinimas, asmenų ir grupių diskriminacijos neleistinumas, panaudojimo sričių ir kitos programinės įrangos apribojimų nebuvimas. Šis atvirojo kodo apibrėžimas identiškas pateiktam *Open Source Initiative*[3] portale apibrėžimui (*The Open Source Definition*).

Statistikos programinių paketų rinkoje randame ir komercinius, ir atvirojo kodo programinius paketus. Rinkdamiesi tarp jų turime opti-

[3] Open Source Initiative. <http://opensource.org/osd>.

maliai apsispręsti funkcionalumo, vartotojo sąsajos patogumo, suderinamumo su kitais programiniais paketais (duomenų sklaidos prasme) ir, aišku, kainos prasme.

1.2. STATISTIKOS PROGRAMINIŲ PAKETŲ APŽVALGA

Statistikiniai skaičiavimai buvo atliekami dar senaisiais laikais. Statistiko darbas visuomet buvo susijęs su skaičiavimais, todėl kompiuterių atsiradimas XX a. padarė didžiulę įtaką teorinės bei praktinės statistikos vystymuisi.

Šiuolaikinė statistika yra neatsiejama nuo kompiuterinės duomenų analizės, padedančios greitai ir veiksmingai spręsti įvairius statistikos uždavinius. Programiniai paketai, kuriuose įdiegti modernūs matematinės statistikos metodai, o daugelis operacijų yra formalizuota, leidžia spręsti taikomuosius uždavinius ne tik gerai įvaldžiusiems tuos metodus, bet ir tik bendrą supratimą apie juos turintiems vartotojams. Tam tik reikia gerai suvokti, kam skirti konkretūs statistiniai metodai, kokia jų taikymo sritis ir kaip interpretuoti gautus rezultatus.

Šiuo metu yra labai daug statistikai skirtų kompiuterinių paketų, o tarp jų kūrėjų vyksta aštri konkurencija. Tarp įvairių paketų randame komercinius paketus ir gana plačiai pradedamus naudoti atvirojo kodo – nemokamus statistikinius paketus. Daugumos populiarių komercinių paketų galimybės yra daugmaž vienodos, o vieno ar kito produkto pasirinkimą dažnai nulemia gamintojo reklaminės veiklos intensyvumas ar aptarnavimo kokybė, kaina, profesiniai interesai arba tiesiog pripratimas.

Kalbant apie kainą, rinkoje yra nemokami ir komerciniai paketai, kuriuos reikia pirkti. Piratinių kopijų naudojimas yra baudžiamas. Prie komercinių paketų priklauso SAS, SPSS, Statistica, Statsoft, S-Plus ir daugybė kitų. Beje, studentams ir dėstytojams (mokymui) skirtos daugelio šių produktų versijos yra gerokai pigesnės. Rinkoje yra ir nemokamų (angl. *free*) statistikos paketų, kurie yra šiek tiek blogesnės kokybės. Čia reikėtų paminėti pasaulinės programuojančių entuziastų bendruomenės GNU operacinės sistemos pagrindu kuriamus R, PSPP produktus, kuriuos pasitelkus šiandien galima spręsti beveik visus statistikos uždavinius.

Svarbus statistikos paketų vertinimo ar klasifikavimo kriterijus yra statistinės analizės komandų vykdymo būdas. Dauguma aukščiau išvardytų paketų turi meniu tipo komandų sistemą. Kiti (programuojami paketai) statistinę analizę atlieka naudojant komandas komandiniame lange. Abu statistikos paketų tipai turi ir privalumų, ir trūkumų. Pirmasis būdas paprastesnis (nereikia mokytis gana greitai užmirštamų komandų), tačiau antrasis statistikos profesionalui teikia daugiau galimybių. Reikėtų žinoti, kad daugelis paketų turi abi galimybes: galima dirbti naudojantis grafine vartotojo sąsaja arba programuoti komandiniame faile (pvz., SPSS, PSPP).

Pateiksime 2012 m. statistikos paketų lyginamąsias lenteles (žr. 1.1 ir 1.2 lenteles).

1.1 lentelė. 2012 m. statistikos paketų palyginimas

Produktas	**Produkto kūrėjas**	**Paskutinė versija**	**Kaina (USD)**	**Atvirasis kodas**	**Programos licencija**	**Sąsaja**
ADaMSoft	Marco Scarno	2012 05 05	Nemokamas	Taip	GNU GPL	CLI/ GUI
Analyse-it	Analyse-it		185–495	Ne	Nuosavybės teisės	GUI
ASReml	VSN International	2009 m. spalis	>150	Ne	Nuosavybės teisės	CLI
BMDP	Statistical Solutions		1095	Ne	Nuosavybės teisės	
Dataplot	Alan Heckert	2005 m. kovas	Nemokamas	Taip	Viešas	CLI/ GUI
Epi Info	Centers for Disease Control and Prevention	2011 01 26	Nemokamas	Taip	Viešas	CLI/ GUI
EViews	IHS	2011 m. lapkritis	Student.: 40 Akad.: 425 Komerc.: 1075	Ne	Nuosavybės teisės	CLI/ GUI
GAUSS	Aptech systems	2011 m. spalis		Ne	Nuosavybės teisės	CLI/ GUI

1.1 lentelės tęsinys

GenStat	VSN International	2011 m. liepa	>190	Ne	Nuosavybės teisės	CLI/ GUI
GraphPad Prism	GraphPad Software, Inc.	2009 m. vasaris	595	Ne	Nuosavybės teisės	GUI
gretl	The gretl Team	2012 05 01	Nemokamas	Taip	GNU GPL	CLI/ GUI
JMP	SAS Institute	2010 m. spalis	Komerc.: 1895 Student.: 29,95/49,95	Ne	Nuosavybės teisės	GUI/ CLI
Maple	Maplesoft	2012 0328	Komerc.: 2275 Student.: 99	Ne	Nuosavybės teisės	CLI/ GUI
Mathematica	Wolfram Research	8.0.4 2011 m. spalis	Profesional.: 2495 Mokymui: 1095 Student.: 140 Student. metams: 69,95 Asmeniui: 295	Ne	Nuosavybės teisės	CLI/ GUI
MATLAB	The MathWorks	Du kartus per metus	Priklauso nuo daugelio aplinkybių	Ne	Nuosavybės teisės	CLI/ GUI
MedCalc	MedCalc Software bvba	2010 11 14	395	Ne	Nuosavybės teisės	GUI
Minitab	Minitab Inc.	2010 05 18	Neribotam laikui: 895–1395 Metinė licenc.: 542 Akad.: 29,99/49,99/ 99,99	Ne	Nuosavybės teisės	CLI/ GUI
NCSS	NCSS, LLC	2007 12 01	Akad.: 495 Komerc.: 795	Ne	Nuosavybės teisės	GUI

1.1 lentelės tęsinys

NMath Stats	CenterSpace Software	2009 m. lapkritis	1295	Ne	Nuosavybės teisės	CLI
NumXL	Spider Financial	2009 m. spalis	Komerc.: 150 Student.: 50	Ne	Nuosavybės teisės	GUI
OpenEpi	A. Dean, K. Sullivan, M. Soe	2011 06 23	Nemokamas	Taip	GNU GPL	GUI
Origin	OriginLab	2011 m. lapkritis	Student.: 50 Akad.: 550 Komerc.: 1095	Ne	Nuosavybės teisės	GUI
Ox programming language	OxMetrics, J.A. Doornik	2011 m. rugpjūtis	Nemokamas akad. vartot.	Ne	Nuosavybės teisės	CLI
OxMetrics	OxMetrics, J.A. Doornik	2011 m. rugpjūtis	Student. ir akad.: licencijuota Komerc.: >500	Ne	Nuosavybės teisės	CLI/ GUI
Primer	Primer-E	2007 m. vasaris	500–1000	Ne	Nuosavybės teisės	GUI
PSPP	GNU Project	2012 02 04	Nemokamas	Taip	GNU GPL	CLI/ GUI
QI Macros	KnowWare International, Inc.	2012 09 14	Student: 99 / Industry: 199	Ne	Nuosavybės teisės	CLI/ GUI
R	R Foundation	2011 04 13	Nemokamas	Taip	GNU GPL	CLI/ GUI
RATS	Estima	2010 10 01	500	Ne	Nuosavybės teisės	CLI/ GUI
Revolution Analytics	Norman Nie	2007	Akad. versija – nemokama Profesionalams mokama licencija	Taip	Nuosavybės teisės	CLI/ GUI
ROOT	ROOT Analysis Framework	2010 06 30	Nemokamas	Taip	GNU GPL	GUI

1.1 lentelės tęsinys

SAGE	>100 kūrėjų pasaulyje	4.5.2 2010 m. rugpjūtis	Nemokamas	Taip	GNU GPL	CLI &GUI
Salstat	Alan J. Salmoni, Mark Livingstone	2004	Nemokamas	Taip	GNU GPL	CLI &GUI
SAS	SAS Institute	2011 m. gruodis	~6000 komplektas PC ~28000 Windows serveriui pirmuosius metus už BASE, STAT, GRAPH ir ACCESS modulius Atskiri moduliai licencijuoti atskirai Už tolesnius metus kaina mažėja dvigubai	Ne	Nuosavybės teisės	CLI/ GUI
SHAZAM	SHAZAM Analytics Ltd	2012 m. birželis	1 studentui 49 Standartas 249 Profesionalus paketas 495 Akad. padaliniui: standartas 1600 Profesionalus paketas 2100	Ne	Nuosavybės teisės	CLI/ GUI
SigmaXL	SigmaXL Inc.	2011 01 18	Neribota licencija 249	Ne	Nuosavybės teisės	GUI

1.1 lentelės pabaiga

SimStat	Provalis Research	2010	Neribota licencija 255–695	Ne	Nuosavybės teisės	CLI/ GUI
SOCR	UCLA	2008 10 28	Nemokamas	Taip	LGPL	GUI
SOFA Statistics	Grant Paton-Simpson	2010 m. balandis	Nemokamas	Taip	AGPL	GUI
SPlus	Insightful Inc.	2010	Metams 2399	Ne	Nuosavybės teisės	CLI
SPSS	IBM	2011	4975	Ne	Nuosavybės teisės	CLI/ GUI
Stata	StataCorp	2011 m. liepa	Akad. nuo 595 Gamybai 1195	Ne	Nuosavybės teisės	CLI/ GUI
Statgraphics	StatPoint	2009 m. spalis	695–1495	Ne	Nuosavybės teisės	GUI
STATISTICA	StatSoft	2010 m. lapkritis	>695	Ne	Nuosavybės teisės	GUI
StatPlus	AnalystSoft	2007 01 07	150	Ne	Nuosavybės teisės	GUI
SYSTAT	Systat Software Inc.	2007 02 21	1299	Ne	Nuosavybės teisės	CLI/ GUI
TSP	TSP International	2009 m. rugsėjis	Student.: 120 Akad.: 600 Komerc.: 1200	Ne	Nuosavybės teisės	CLI
UNISTAT	Unistat Ltd	2011 04 08	895, 495, 300	Ne	Nuosavybės teisės	GUI, Excel
Winpepi	J. H. Abramson	2008 m. birželis	Nemokamas	Ne	Nuosavybės teisės	GUI
WPS	World Programming	2012 m. vasaris		Ne	Nuosavybės teisės	CLI/ GUI
XploRe	MD*Tech	2006		Ne	Nuosavybės teisės	GUI

Pastaba. CLI – komandinė vartotojo sąsaja, GUI – grafinė vartotojo sąsaja

1.2 lentelė. Statistikos programų palaikomos operacinės sistemos

Produktas	Windows	Mac OS	Linux	BSD*	Unix
ADaMSoft	Taip	Taip	Taip	Taip	Taip
Analyse-it	Taip	Ne	Ne	Ne	Ne
BMDP	Taip				
Dataplot	Taip	Taip	Taip	Taip	Taip
Epi Info	Taip	Ne	Ne	Ne	Ne
EViews	Taip	Taip	Ne	Ne	Ne
GAUSS	Taip	Taip	Taip	Ne	Taip
GraphPad Prism	Taip	Taip	Ne	Ne	Ne
gretl	Taip	Taip	Taip	Ne	Ne
JMP	Taip	Taip	Taip	Ne	Ne
JHepWork	Taip	Taip	Taip	Taip	Taip
Maple	Taip	Taip	Taip	?	Taip
Matlab	Taip	Taip	Taip	Ne	Ne
Mathematica	Taip	Taip	Taip	Ne	Taip
MedCalc	Taip	Ne	Ne	Ne	Ne
Minitab	Taip	Apribota	Ne	Ne	Ne
NCSS	Taip	Ne	Ne	Ne	Ne
NMath Stats	Taip	Ne	Ne	Ne	Ne
NumXL	Taip	Ne	Ne	Ne	Ne
OpenEpi	Taip	Taip	Taip	Taip	Taip
Origin	Taip	Ne	Ne	Ne	Ne
Primer	Taip	Ne	Ne	Ne	Ne
PSPP	Taip	Taip	Taip	Taip	Taip
QI Macros	Taip	Taip	Ne	Ne	Ne
R Commander	Taip	Taip	Taip	Taip	Taip
R	Taip	Taip	Taip	Taip	Taip
RATS	Taip	Taip	Taip	Ne	Taip
RKWard	Taip	Ne	Taip	Ne	Taip
ROOT	Taip	Taip	Taip	Taip	Taip
Sage	Dalinai	Taip	Taip	Ne	Taip
Salstat	Taip	Taip	Taip	Taip	Taip
SAS	Taip	Apribota	Taip	Ne	Taip

1.2 lentelės pabaiga

SHAZAM	Taip	Ne	Ne	Ne	Ne
SigmaXL	Taip	Ne	Ne	Ne	Ne
SimStat	Taip				
SOCR	Taip	Taip	Taip	Taip	Taip
SOFA Statistics	Taip	Taip	Taip	Taip	Taip
SPlus	Taip	Ne	Taip	Ne	Taip
SPSS	Taip	Taip	Taip	Ne	Ne
Stata	Taip	Taip	Taip	Ne	Taip
Statgraphics	Taip	Ne	Ne	Ne	Ne
STATISTICA	Taip	Ne	Ne	Ne	Ne
StatPlus	Taip	Taip	Ne	Ne	Ne
SYSTAT	Taip	Ne	Ne	Ne	Ne
TSP	Taip	Taip	Taip	Taip	Taip
UNISTAT	Taip	Ne	Ne	Ne	Ne
The Unscrambler	Taip	Ne	Ne	Ne	Ne
Winpepi	Taip	Ne	Ne	Ne	Ne
WPS	Taip	Taip	Taip	Ne	Taip
XploRe	Taip	Ne	Taip	Ne	Taip

* BSD (angl. *Berkeley Software Distribution*) – operacinė sistema, sukurta UNIX OS pagrindu. Šiuo metu plačiausiai naudojamos BSD tipo operacinės sistemos yra kilusios iš 386BSD: FreeBSD, DragonFly BSD, Mac OS X, NetBSD, OpenBSD.

STATISTICA

STATISTICA yra matematinės statistikos programų paketas, sprendžiantis statistikos problemas. Jo kūrėja – amerikiečių bendrovė *StatSoft, Inc* – pasaulinė lyderė programinės įrangos duomenų analizės srityje – viena didžiausių statistikos ir analizės programinės įrangos gamintojų pasaulyje.

StatSoft, Inc. yra pasaulinis duomenų analizės, duomenų valdymo, duomenų atvaizdavimo, duomenų gavybos (angl. *data mining*), prognozinės analizės ir kokybės kontrolės programinės įrangos tiekėja verslui ir asmeninių kompiuterių turėtojams.

Bendrovė įsteigta 1984 m. kaip grupės universitetų profesorių ir mokslininkų bendrija. Pirmoji produkcija buvo meniu valdomų statis-

tikinių procedūrų bibliotekos, skirtos mikrokompiuterių platformoms Apple II, CP/M, Commodore ir PC DOS 1.0.

Šiandien *StatSoft, Inc*. yra privati korporacija , kurios buveinė yra Tulsoje (Oklahoma, JAV). Korporacija turi filialus 24 pasaulio šalyse. *StatSoft, Inc* produktai yra naudojami universitetuose, įmonėse ir valstybinėse įstaigose visame pasaulyje.

Programinis paketas apima daug duomenų analizės, tvarkymo, atvaizdavimo ir duomenų gavybos procedūrų, taip pat įvairius prognozių modeliavimo, klasterizacijos, duomenų vaizdavimo ir tyrimo metodus. Papildomos galimybės atsiranda integruojantis su *R* – nemokamu atvirojo kodo programų paketu.

Pradedant STATISTICA 9, kuri pasirodė 2009 m. gegužę, siūlomos 32 bitų ir 64 bitų versijos. STATISTICA 10 pasirodė 2010 m. lapkritį. Programa veikia *Microsoft Windows* aplinkoje.

SPSS

IBM SPSS Statistics – tai specializuota statistinė programinė įranga, leidžianti vartotojams atlikti visą duomenų analizės procesą:

- įkelti duomenis iš įvairių šaltinių;
- parengti duomenis (pvz., atlikti pataisas, sukurti naujus kintamuosius, užkoduoti kategorijas ir praleistas reikšmes, sujungti duomenis, išrinkti imtį, sugrupuoti kintamuosius ir t. t.);
- visapusiškai išanalizuoti duomenis tiksliais statistiniais metodais ir gauti reikšmingus rezultatus bei statistinius kriterijus;
- lengvai ir suprantamai pateikti gautus rezultatus grafikais bei analitinėmis lentelėmis;
- eksportuoti rezultatus įvairias formatais.

Šia programa naudojasi daugiau nei 300 000 įmonių visame pasaulyje, tarp kurių yra ir 95 proc. organizacijų iš 1000 pelningiausių *Fortune* sąrašo įmonių. Lietuvoje šia programine įranga naudojasi daugiau nei 200 įmonių bei organizacijų.

IBM SPSS Statistics programinę įrangą naudoja akademinės bei mokslo institucijos, siekdamos parengti įvairių sričių specialistus ir atlikti reikšmingus mokslinius tyrimus, viešojo sektoriaus įstaigos, siekdamos geriau suprasti visuomenės poreikius ir veiksmingai juos tenkin-

ti mažiausiomis sąnaudomis, bei verslo bendrovės, siekdamos suprasti ir optimizuoti savo bendravimo su klientais bei verslo procesus ir įgyti konkurencinį pranašumą.

IBM SPSS Statistics programinę įrangą sudaro 15 savarankiškų modulių, tarp kurių yra ir bazinis modulis *IBM SPSS Statistics Base*, aprėpiantis dažniausiai naudojamus analizės metodus, todėl šis modulis rekomenduojamas visiems vartotojams. Kiti moduliai apima specifines analizės rūšis (pvz., *Regressions, Advanced Statistics*) arba yra taikomi konkrečiose verslo srityse (pvz., *IBM SPSS Direct Marketing, IBM SPSS Custom Tables*).

Nuo 18 versijos visi *IBM SPSS Statistics* moduliai gali dirbti atskirai be *IBM SPSS Statistics Base* instaliacijos, nes visos duomenų įkėlimo, parengimo bei grafikų kūrimo galimybės yra įtrauktos į kiekvieną modulį.

Nuo 2009 m. balandžio įsigaliojo nauji SPSS produktų pavadinimai su prierašu *PASW* (angl. *Predictive Analytics Software* – prognozinės analitikos programinė įranga). Tačiau 2009 m. spalį *SPSS Inc*. tapus IBM dalimi, vietoj PASW atsirado IBM SPSS. Pačių produktų sudėtis liko nepakitusi.

R

R yra GNU laisvai platinama kalba bei statistinių skaičiavimų ir grafikos aplinka, kuri suteikia didelę statistikos ir grafinių metodų įvairovę: tiesinis ir netiesinis modeliavimas, statistiniai testai, laiko eilučių analizė, klasifikavimas, klasterizacija ir t. t. Tai atvirojo kodo (angl. *open source*) programinės įrangos paketas, skirtas statistiniams skaičiavimams ir grafiniams jų atvaizdavimams. Autoriai *Ross Ihaka* ir *Robert Gentleman.* Projekto pradžia 1983 m. Buvo siekta sukurti į *UNIX* panašią atvirojo kodo operacinę sistemą. GNU yra rekursinė santrumpa: **G**NU **N**ėra **U**NIX(angl. ***G****NU* ***N****ot* ***U****nix*). Šios sistemos variantai dabar žinomi *Linux* vardu.

R programų paketas yra jaunesnysis komercinio *S-Plus* paketo „brolis“. Šių dviejų kalbų sintaksė praktiškai ta pati, nors programiniai interpretavimo principai skiriasi. Dauguma *S* kalba parašytų programų veikia ir *R* aplinkoje. *R* yra paplitusi kaip komandinių eilučių (visos užduotys įvedamos naudojant komandas) produktas. Yra ir kelios grafinės vartotojo sąsajos.

R iš esmės yra programavimo kalba su specializuota (statistikos reikmėms skirta) aplinka. *R* funkcijos yra sujungtos į bibliotekas, kurios gali būti prijungtos prie darbinės srities arba, kai nebereikalingos, atjungtos.

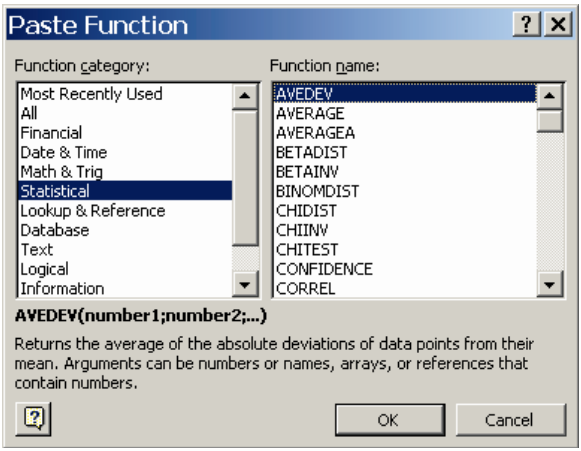
R projekto adresas internete yra http://www.r-project.org, tačiau diegiant siūloma kreiptis į vieną iš CRAN (angl. *Central R Archive Network*) veidrodinių kopijų. Ten yra paskutinė R versija (2012 m. pabaigoje – R2.15.2).

MS Excel

Kompiuterinės programos, skirtos lentelėms kurti ir tvarkyti, vadinamos skaičiuoklėmis. Lietuvoje labiausiai paplitusi skaičiuoklė *Microsoft Excel*.

Visos *MS Excel* versijos turi didelį kiekį statistinių funkcijų, kurios gali būti labai paprastai panaudotos duomenų analizei.

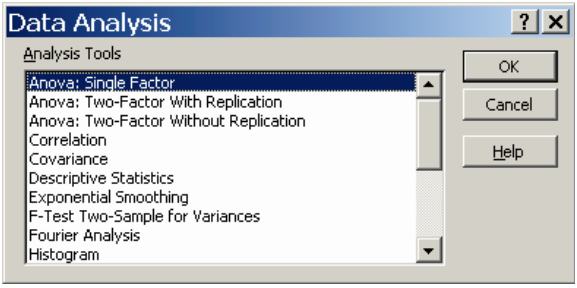
- Pasirinkite *Insert* → *Function…* → *Statistical* (žr. 1.1 pav.)



1.1 pav. *MS Excel* langas *Paste Function*. Statistinių funkcijų įterpimas

MS Excel turi statistikams naudingą įrankį – duomenų analizės įrankių rinkinį *Data Analysis*. Senesnėse *Excel* versijose (iki MS Excel 7):

- Pasirinkite *Tools* → *Data Analysis…*. Atsiveria langas *Data Analysis* (žr. 1.2 pav.).

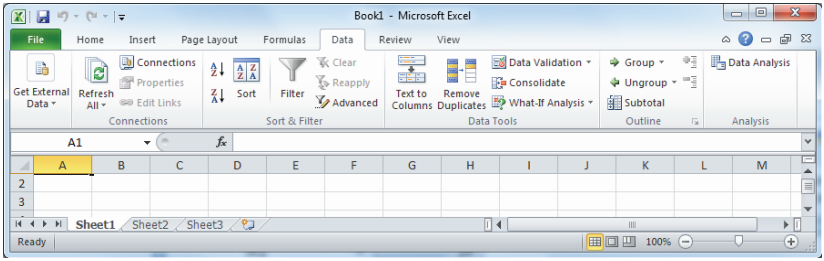


1.2 pav. *MS Excel* langas *Data Analysis*. Įrankių duomenų analizei pasirinkimas

- Pasirinkite norimus analizės įrankius. Spragtelėkite mygtuką *OK*.

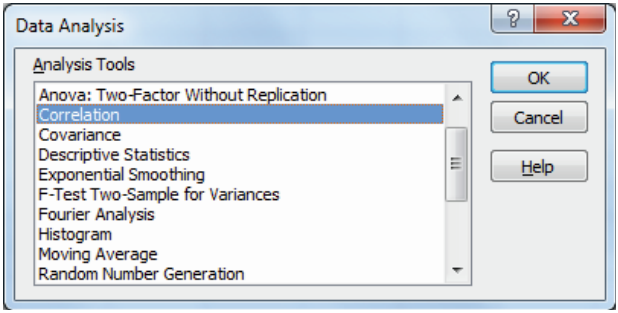
Naujesnėms *Excel* versijoms duomenų analizės įrankiai pasirenkami kiek kitaip (pavyzdyje MS Excel 10):

- *Data* → *Data Analysis* (žr. 1.3 pav.).



1.3 pav. Duomenų analizės pasirinkimas *MS Excel 10*

- Atsiveria langas *Data Analysis* (žr. 1.4).

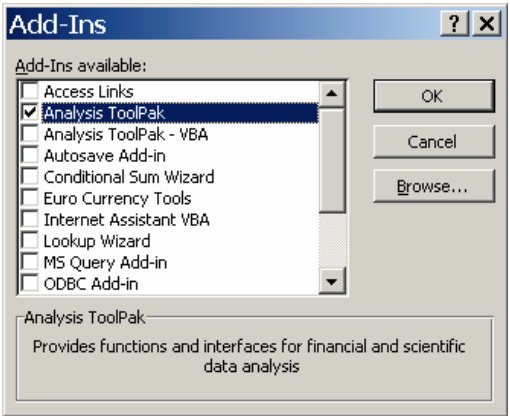


1.4 pav. Duomenų analizės įrankių pasirinkimas *MS Excel 10*

- Pasirinkite norimus analizės įrankius. Spragtelėkite mygtuką *OK*.

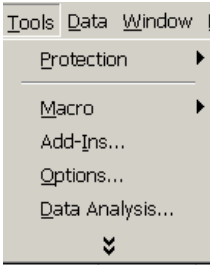
Paprastai, naujai įkėlus *Excel* skaičiuoklę, įrankius duomenų analizei tenka papildomai įkelti. Senesnėms *Excel* versijoms:

- Pasirinkite *Tools* → *Add-Ins…* Atveriamas langas *Add-Ins*. Pažymėkite varnele *Analysis ToolPak* (žr. 1.5 pav.). Paspauskite *OK*.



1.5 pav. *MS Excel* langas *Add-Ins*. Įrankių paketo duomenų analizei pasirinkimas

- Meniu *Tools* atsiranda komanda *Data Analysis…* (žr. 1.6 pav.)



1.6 pav. Komanda *Data Analysis…* meniu *Tools*

Naujesnėms *Excel* versijoms (pavyzdyje *MS Excel 10*):

- Pasirinkite *File* → *Options* → *Add-Ins*.

- Atsivėrusiame lange pasirinkite *Analysis ToolPak* ir spragtelėkite mygtuką *Go*.
- Grįžtate į *Add-Ins* langelį. Pažymėkite varnele *Analysis ToolPak* ir spragtelėkite mygtuką *OK*.
- *Data* meniu atsiranda *Data Analysis*.

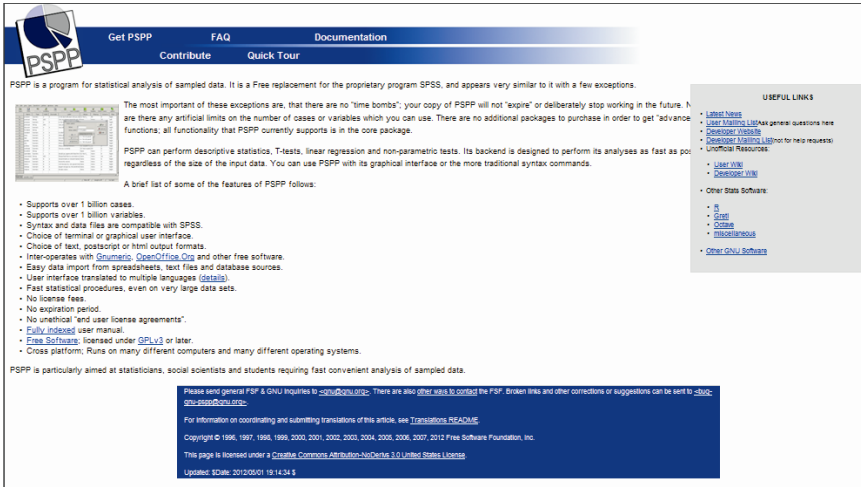
PSPP programų paketo apžvalga

PSPP yra laisvai platinamas programų paketas, skirtas statistikams, mokslininkams ir studentams, greitai ir paprastai statistinių duomenų analizei atlikti. Ši programų paketas – supaprastintas patentuoto ir plačiai žinomo statistikos programų paketo SPSS (angl. *Statistical Package for the Social Statistics*) variantas.

PSPP neturi jokio oficialaus akronimo (akronimas – termino, kurį sudaro keli sudėtiniai žodžiai, santrumpa, panaudojant tų žodžių pirmąsias (kartais – ir kitas) raides). Anglų kalba siūlomi:

- Perfect Statistics Professionally Presented;
- Probabilities Sometimes Prevent Problems;
- People Should Prefer PSPP.

Oficialus internetinis adresas http://www.gnu.org/software/pspp/



1.7. pav. Oficialus PSPP interneto puslapis

Kaip rašoma programų paketo pristatyme, PSPP neturi „laiko bombų“, turima kopija nenustos galioti ar programa ateityje sąmoningai nesustos. Nėra jokių dirbtinių apribojimų nagrinėjamų įrašų (atvejų, respondentų atsakymų) ar kintamųjų skaičiui. Taip pat nereikia pirkti jokių papildomų programinių paketų – viskas yra PSPP branduolyje.

PSPP gali atlikti aprašomąją statistiką, T-testus, tiesinę regresiją ir neparametrinius testus. PSPP turi grafinę vartotojo sąsają, taip pat ją galima naudoti komandiniu pavidalu.

Trumpas PSPP savybių sąrašas:

- Palaiko daugiau kaip 1 milijardą stebėjimo duomenų;
- Palaiko daugiau kaip 1 milijardą kintamųjų;
- Sintaksės ir duomenų failai (rinkmenos) suderinami su SPSS;
- Galima rinktis komandinę ar grafinę vartotojo sąsają;
- Galima rezultatus pateikti ASCII, PostScript ar HTMLl formatu;
- Suderinama su Gnumeric, OpenOffice.Org ir kita laisva programine įranga;
- Lengvai importuojami duomenys iš skaičiuoklių, tekstinių failų ir duomenų bazių;
- Vartotojo sąsaja išversta į daugelį kalbų (lietuvių taip pat);
- Statistinės procedūros net ir su daug duomenų veikia greitai;
- Jokių licencijos mokesčių;
- Jokio baigtinio veikimo periodo;
- Jokių vartotojo sutarčių;
- Visiškai indeksuotas vartotojo vadovas;
- Laisva programinė įranga, licencijuota pagal GPLv3 ar vėlesnę versiją;
- Veikia įvairiuose kompiuteriuose ir su įvairiomis operacinėmis sistemomis.

1.3. ATVIROJO KODO STATISTINIO PAKETO PASIRINKIMAS

Statistinės analizės metodai šiuo metu dažnai taikomi socialiniuose ir ekonominiuose tyrimuose. Įmonės ir organizacijos ne tik renka savo veiklos parametrų statistiką, bet ir organizuoja tyrimus, skirtus jų veiklos analizei. Paplito statistiniai rinkos tyrimų metodai, kurie leidžia įvertinti vartotojų poreikius ir įmonės veiklos efektyvumą, atlikti rinkos segmentavimą, prognozuoti veiklos rezultatus.

Atlikta statistikos programinių paketų apžvalga rodo, kad, nepaisant šios programinės įrangos įvairovės, geresni programų paketai brangiai kainuoja, oficialiai platinamos tik licencijuotos jų kopijos. Ne kiekvienas vartotojas pajėgia įsigyti tokią programinę įrangą. Akademinės įstaigos (universitetai) gali tam tikras programas taikyti nemokamai arba su nuolaida, tai tinka studijų procesui. Tačiau kitos organizacijos tokių galimybių neturi, ypač kalbant apie smulkiojo ir vidutinio verslo atstovus. Nemažai kainuoja ir specializuotoje firmoje užsakyti statistinius tyrimus. Todėl verta atkreipti dėmesį į atviro kodo programinę įrangą.

Knygos autoriai siūlo rinktis PSPP statistikos programų paketą – atvirojo kodo, nemokamą, funkcionalumo prasme labai artimą komerciniam statistikos paketui IBM SPSS. Kaip dar vieną PSPP paketo pasirinkimo argumentą pabrėžtumėme tai, kad vartotojo sąsaja leidžia pasirinkti ir lietuvių kalbą, ko neturi daugelis atvirojo kodo bei komercinių paketų. Šių paketų panašumas padeda vartotojui lengvai persiorientuoti nuo vieno paketo taikymo į kitą, naudojami vienodo formato duomenų failai. SPSS paketas taikomas kaip PSPP pagrindas, numatyta PSPP terpėje įgyvendinti daugumą SPSS galimybių. Vadovėlyje pateikiamas šių programų palyginimas.

Aišku, PSPP programinį paketą dar galima tobulinti – reikėtų tobulinti duomenų įvedimą/apsikeitimą iš įvairių programinių produktų, grafinį rezultatų pateikimą. Atvirojo kodo bendruomenės entuziastų darbas nuteikia optimistiškai – kasmet atsiranda naujų PSPP versijų su esminiais patobulinimais.

2. RINKOS TYRIMŲ ORGANIZAVIMAS

2.1. RINKOS TYRIMŲ PASKIRTIS IR REGLAMENTAVIMAS

Visa organizuota verslo veikla yra skirta klientų poreikiams tenkinti. Pagrindinis rinkos analizės tikslas yra nagrinėti šių poreikių bei jais išreiškiamos paklausos prigimtį ir nustatyti, kas yra klientai. Rinkos analizė yra bendras terminas, apibrėžiantis metodus ir priemones, naudojamas šiam tikslui siekti.

Šiandien vartotojas gali rinktis iš daugelio produktų, pagamintų įvairiose šalyse. Vartotojo pasirinkimą dažniausiai lemia įvairūs motyvai, kuriuos aktyviai formuoja reklama. Naujausia pardavimų vadybos strategija – ne tik gaminti, bet kartu ir auginti savo prekių ar paslaugų vartotoją. Marketingas (rinkodara) – tai poreikių išsiaiškinimo ir jiems tenkinti reikalingų sprendimų priėmimo bei įgyvendinimo procesas, padedantis siekti žmogaus ar organizacijos tikslų.

Yra daug rinkos apibrėžimų. Ekonomikos požiūriu *rinka* – tai pirkėjų ir pardavėjų susitikimo vieta, kur prekių mainų proceso metu veikia pasiūlos ir paklausos pusiausvyrą reguliuojantis mechanizmas, išreikštas kaina. Marketingo požiūriu *rinka* – tai visuma esamų ir potencialių pirkėjų, siekiančių patenkinti savo poreikius pakeičiant ką nors vertinga su prekes ar paslaugas siūlančiais pardavėjais. Pagrindiniai rinkos dalyviai yra vartotojai (pirkėjai) ir gamintojai (pardavėjai). Rinka – tai yra prekių (įskaitant darbo jėgą, piniginius išteklius) mainų sistema, kurios funkcionavimas priklauso nuo pasiūlos ir paklausos santykio, esant gamintojų konkurencijai ir atitinkamai kainodaros sistemai. Rinkos segmentavimas – tai rinkos skaidymas į dalis, kuriose vienodai ar panašiai reaguojama į marketingo (rinkodaros) veiksmus. Sėkmingo rinkos segmentavimo pagrindas yra geras rinkos tyrimas.

Svarbiausias *rinkos tyrimo* tikslas – suprasti vartotojo elgesį ir motyvus, nulemiančius vienokį ar kitokį elgesį ir apsisprendimą pirkti. Žmogaus elgesį lemia jo individualios savybės, socialinė aplinka ir

veiksniai, susiję su socialine aplinka, be kurių neįmanoma suprasti vartojimo ar pasirinkimo motyvų (mokymosi procesas, įsitraukimas, noras rizikuoti). *Rinkos tyrimai* padeda įvertinti, ką žmonės mąsto, jaučia ir daro siekdami patenkinti savo norus, troškimus ir pomėgius, kuriuos sieja su produktų ir paslaugų pirkimu.

Rinkos analizę reikia atlikti pradedant verslą arba pristatant naują produktą. Dažnai analizė atliekama verslo strategijos veiksmingumui vertinti. Rinkos tyrimų procesas apima: duomenų rinkimą, analizę ir interpretavimą. Ši informacija reikalinga tam, kad būtų numatomi ateities įvykiai, vartotojų elgesys. Analizė leidžia sumažinti riziką, suprasti galimybes ir grėsmes, nustatyti pardavimų apimtis ir suplanuoti marketingo (pardavimų) strategiją ir planus. Rinkos analizė gali būti kokybinė (pavyzdžiui, imant interviu ir dirbant su stebimomis grupėmis) ir kiekybinė (klientų tyrimai ir pardavimų informacija).

Marketingo (rinkodaros) tyrimai yra įmonės išorės veiksnių (rinkos) ir vidaus veiksnių informacijos rinkimas, analizė, išvadų ir siūlymų rengimas. *Rinkos tyrimai* – tai yra sisteminių duomenų apie konkurentus, produktus, kainas, užimamą rinkos dalį, makrofaktorius, tendencijas kaupimas ir analizavimas, siekiant suformuoti marketingo strategiją. Rinkos tyrimų rezultatai yra įmonės steigimo sprendimo priėmimo, naujo rinkos segmento perpratimo, prekybinių plotų praplėtimo, filialo arba atstovybės kitame regione atidarymo, prekių linijos išplėtimo ir t. t. pagrindas. Visi tie sprendimai yra siejami su pinigų investavimu, o tai reiškia, kad siejami ir su jų praradimo rizika. Būtent tam, kad būtų galima sumažinti investuojamų santaupų praradimo tikimybę, pagrįsti investicijas ir pateikti labiau apibrėžtus jų grąžinimo laikotarpius, yra vykdomi rinkos tyrimai. Tyrimų rezultatai gali būti pateikiami kartu su plėtojamo projekto verslo planu.

Marketingo tyrimai skirstomi į keletą grupių. Bendra grupavimo sistema dar nėra griežta, tuos pačius elementus galima atrasti keliose grupėse, tačiau dažniausiai skiriamos šešios pagrindinės grupės su pogrupiais:

1. Vartotojų tyrimai – pateikia informaciją apie rinkos segmentus, vartotojų charakteristikas ir vartotojų motyvaciją.
2. Reklamos tyrimai – pateikia informaciją apie veiksmingiausius reklamos bendrovių tipus ir geriausius reklamos projektus.

3. Produkto tyrimai – padeda ištirti, kaip potencialūs vartotojai priima naujas produkto idėjas.
4. Paskirstymo (distribucijos) tyrimai – skirti įvertinti veiksmingiausius distribucijos kanalus ir tinkamiausius distributorius.
5. Pardavimų tyrimai – leidžia įvertinti prekės pateikimo parduotuvėje ar internetiniame portale patrauklumą, pardavėjų komunikacijos lygį, pritaikyti tinkamiausias pardavimų ir prekių išdėstymo formas.
6. Marketingo aplinkos tyrimai – padeda įvertinti politinius, ekonominius, socialinius ir technologinius veiksnius, kurie gali lemti marketingo sprendimų priėmimą.

Veiksmingas rinkos tyrimas priklauso nuo visuomenės pasitikėjimo, kad tyrimas atliekamas sąžiningai, objektyviai, be šiurkštaus įsibrovimo ar nenaudos respondentams ir kad jis paremtas jų pasirengimu bendradarbiauti. Profesiniai praktikos kodeksai, numatantys rinkos tyrimo projektų atlikimo būdus, skatina šį pasitikėjimą. Teisiškai rinkos tyrimus reglamentuoja statistinį tyrimų teisės aktai. ESOMAR ir Tarptautiniai prekybos rūmai skatina taikyti ICC/ESOMAR tarptautinį Rinkos tyrimo ir Socialinių tyrimų Praktikos kodeksą, kuris yra pasaulinis bendrai taikytinas kodeksas, nurodantis, kaip turi būti atliekami rinkos tyrimai visame pasaulyje. Tarptautinio standarto ISO 20252 normos rinkos, nuomonės ir socialiniams tyrimams buvo patvirtintos ir imtos taikyti 2006 m. Lietuvoje statistiniai tyrimai vykdomi vadovaujantis Lietuvos ir Europos Sąjungos teisės aktais:

1. Lietuvos Respublikos statistikos įstatymo pakeitimo įstatymu. 1999 m. gruodžio 23 d. Nr. VIII-1511.
2. Europos Parlamento ir Tarybos reglamentu (EB) Nr. 295/2008, 2008 m. kovo 11 d. dėl verslo struktūros statistikos (nauja redakcija).

Prieš atliekant tyrimą ir pasirenkant tyrimų metodą būtina surinkti visą informaciją apie sukurtus teorinius modelius, kuriuos būtų galima panaudoti atliekant tyrimą, ir anksčiau atliktus tyrimus.

Įvertinus anksčiau atliktus (ar neatliktus) tyrimus, pradiniame tyrimo etape pasirenkama viena iš trijų galimų tyrimo alternatyvų (tyrimo tipų):

- paaiškinamasis tyrimas, kai reikia nustatyti problemos dydį ir nežinomi jokie modeliai, nuo kurių galima pradėti tyrimą;

- anksčiau sukurtų modelių atnaujinimas ir išplėtojimas;
- tyrimas, skirtas iškeltoms hipotezėms patikrinti.

Tarp šių sričių nėra tiksliai nubrėžtos ribos, tačiau dažniausiai rinkos tyrimų tikslu tampa žinių apie rinką (vartotojus, konkurentus, tiekimus, apskritai rinkos sąlygas) surinkimas.

2.2. STATISTINIO TYRIMO PROJEKTAS

Atliekant statistinį tyrimą, visais atvejais būtinas tyrimo projektas. Projekte pateikiama tyrimo struktūra, kuri sieja visus tyrimo elementus. Įmonė (organizacija) rinkos tyrimą gali atlikti savarankiškai arba užsakyti rinkos tyrimų bendrovei. Šiuo atveju parengti rinkos tyrimo projektą – svarbus ir atsakingas darbas, reikalingas kliento ir tyrėjo abipusių pastangų bei bendradarbiavimo. Vykdytojui svarbu kuo daugiau žinoti apie situaciją, kurią klientas nori pakeisti ar išspręsti, taip pat svarbu aptarti ir suderinti daugybę tyrimo projekto aspektų bei niuansų. Klientas turi pateikti tyrėjui trumpą apžvalgą apie problemos, kuriai spręsti reikalingas tyrimas, kilmę, ir nurodyti, ką jis tikisi sužinoti, taip pat kokie veiksmai bus atliekami atsižvelgus į tyrimo rezultatus. Pasitelkęs šią informaciją vykdytojas rengia tyrimo projektą, kurio metu turi būti surinkti reikalingi ir praktiškai naudingi duomenys. Aiškus ir detalus susitarimas dėl įvairių projekto aspektų padės užtikrinti tyrimo kokybę, veiksmingumą, gerokai sumažins klaidų, nesusipratimų tikimybę ir galiausiai leis taupyti laiką bei išteklius.

Tyrimo projektą sudaro šie pagrindiniai elementai:

- tyrimo problema;
- tikslas ir uždaviniai;
- metodika;
- tyrimo procedūros parinkimas;
- matavimo skalės parinkimas;
- imties sudarymo būdas ir dydis;
- biudžetas ir laiko diagrama;
- duomenų kodavimas ir analizė;
- ataskaitos parengimas.

Neatsižvelgiant į tai, kokia yra tyrimo apimtis ir kiti veiksniai, tyrimo projekte turi atsispindėti trys pagrindiniai statistinio tyrimo tarpsniai: duomenų surinkimas, duomenų suvedimas ir jų analizė bei ataskaitos parengimas.

2.3. TYRIMO TIKSLAS IR TYRIMO MODELIS

2.3.1. Tyrimo problemos ir tikslo formulavimas

Rinkos tyrimo planavimą reikia pradėti nuo nagrinėjamos problemos apibrėžimo ir tyrimo tikslų formulavimo. Tiksliai apibrėžti problemą daugeliu atveju apsunkina dalykinėms situacijoms būdingas neapibrėžtumas ir problemų neaiškumas. Taip pat dažnai problemų ir jų sprendimų būdų būna keletas, todėl tyrėjui būna nelengva užduotis, ką pasirinkti. Taigi prieš pradedant tyrimą svarbiausias uždavinys – tiksliai ir aiškiai suformuluoti problemą ir nustatyti, kokia informacija būtina jai spręsti. Paprastai surinkti ir išanalizuoti duomenis dažniausiai patikima statistikams, tačiau suformuluoti problemas gali tik rinkos procesus gerai išmanantis vadybininkas.

Tyrimo tikslas – tai rezultatas, kurį norima gauti, naudojant tam tikrus tyrimo metodus. Pavyzdžiui: „Nustatyti produkto, kurį pasiūlysime artimiausiu metu, paklausias ir nepaklausias savybes.“ Tyrimo uždaviniuose apibrėžiama, kokius konkrečius žingsnius reikia atlikti, norint pasiekti užsibrėžtą tikslą.

Tinkamai suformuluotas tikslas visada prasideda veiksmažodžiu, reiškiančiu veiksmą. Apibrėžiant tyrimo tikslą, vartotini tipiniai veiksmažodžiai. Pavyzdžiui, problemą arba galimybes galima *nustatyti*, reklamos priemones arba koncepcijas galima *parinkti* ir pan. Tyrimo tikslas turi būti suformuluotas trumpai, aiškiai ir konkrečiai. Taip suformuluotas tikslas turi privalumų: aiškiai matoma, ką norima pasiekti (taupomas laikas, pinigai ir kiti ištekliai); dažnai paaiškėja, kad suformuluotas tikslas atspindi tik dalį to, ką buvo norima pasiekti.

Kiekvienai tyrimo stadijai būdingi saviti tikslai. Iš 2.1 lentelės matyti, kokie yra tinkamiausi tyrimo tikslą apibūdinantys veiksmažodžiai.

2.1 lentelė. Sprendimų priėmimų stadijos, tyrimo tikslai ir tyrimo metodai

Stadija	Tikslas	Pagrindiniai tyrimo metodai	Pagalbiniai metodai
Situacijos įvertinimas. Kas vyksta?	Nustatyti, išsiaiškinti, atsekti	Antriniai tyrimai, vizitai pas vartotojus	Fokus grupės, apklausos
Alternatyvų suformulavimas. Kokios galimybės?	Sukurti, ištirti, aprašyti	Fokus grupės, vizitai pas vartotojus	Antriniai tyrimai
Alternatyvos parinkimas. Kiek mes gausime? Kas geriau?	Įvertinti, testuoti, išrinkti, nustatyti pirmenybes	Eksperimentai, apklausos, parinkimo modeliai, naudingumo testai	Antriniai tyrimai
Sėkmės įvertinimas. Kaip gerai mes tai atlikome?	Išmatuoti, pažingsniui pereiti	Apklausos, antriniai tyrimai	Vizitai pas vartotojus

Tyrimo tikslai gali būti:

- *žvalgybiniai*, t. y. aiškinantys problemą, o galbūt ir padedantys suformuluoti pradines hipotezes;
- *aprašomieji*, t. y. padedantys aprašyti ir išsiaiškinti konkrečius reiškinius, pavyzdžiui, kiek žmonių naudojasi kuria nors paslaugos rūšimi ir kiek jų yra girdėję apie užsakovo firmą;
- *eksperimentiniai*, t. y. tikrinantys hipotezes apie priežasčių ir pasekmių ryšius, pavyzdžiui, kad paslaugų perkėlimas į internetą 10 proc. padidintų norinčių ją gauti skaičių.

Kaip matyti iš 2.1 lentelės, tyrimo tikslams pasiekti reikia naudotis tam tikrais tyrimo metodais. Kiekvienas tyrimo metodas gali būti pagrindinis vienose stadijose, siekiant tam tikro tikslo, bet gali būti pagalbinis kitais atvejais. Naudojant šią lentelę, galima greitai nustatyti, kokiu atveju verta parinkti vieną ar kitą metodą. Šią lentelę galima panaudoti planuojant tyrimus viso sprendimo priėmimo proceso metu.

Nustačius tyrimo tikslus, svarbu apibrėžti *informacijos šaltinius*. Būtina išsiaiškinti, kokia reikalinga informacija ir kaip veiksmingai surinkti duomenis.

Tyrėjas gali naudotis pirminiais arba antriniais duomenimis, arba ir vienais, ir kitais vienu metu.

Antriniai duomenys. Tai tie duomenys, kurie buvo renkami kažkuriam tikslui, nesusijusiam su dabartine problema, t. y. duomenys, kurie jau buvo surinkti anksčiau ir kuriuos galima iš dalies panaudoti sprendžiant dabartinę problemą. Kadangi antriniai tyrimai remiasi jau turima informacija, jie dar vadinami *kabinetiniais tyrimais* (angl. *desk research*). Antriniai duomenys gali būti *vidiniai ir išoriniai*. Plačiausiai naudojami *išoriniai antriniai duomenys* – gyventojų surašymai ir valstybės institucijų registruojama statistika, taip pat visiems prieinama informacija, pateikiama duomenų bazėse, laikraščiuose ir žurnaluose. *Vidinių antrinių duomenų* galima rasti savo įmonės buhalterijoje, vidinio įmonės audito ataskaitose, taip pat vartotojų pateikiamuose pasiūlymuose ir skunduose.

Pirminiai duomenys. Tais atvejais, kai nėra tinkamų antrinių duomenų arba jie neišsamūs, atliekami tyrimai ir renkami pirminiai duomenys (lot. – *ad hoc* – specialus, šiam tikslui). Tai duomenys, kurie renkami konkrečiai problemai spręsti. Daugeliu rinkos tyrimų atvejų renkami išoriniai pirminiai duomenys. Deja, dažnai renkant šiuos duomenis sugalvojama keletas klausimų ir suieškomi asmenys, kuriuos patogiausia apklausti. Taip surinkti duomenys dažniausiai pasirodo esantys labai mažai naudingi, o neretai dar ir klaidinantys. Todėl reikėtų sudaryti atskirą pirminių duomenų rinkimo planą, kuriame būtų numatytas imties dydis, apklausos metodas ir planuojami statistinio tyrimo metodai.

Beveik visada antrinius duomenis naudoti yra pigiau ir rezultatas gaunamas greičiau, nei atliekant savo tyrimus. Praktiškai kiekviename projekte pirmiausia aiškinamasi, kokie antriniai tyrimai yra prienami ir ką jais remiantis galima išsiaiškinti.

2.2 lentelė. Pirminių ir antrinių duomenų palyginimas

	Pirminiai duomenys	**Antriniai duomenys**
Rinkimo tikslas	Tyrimo problemai spręsti	Kitiems tikslams
Rinkimo procesas	Reikalauja nemažai pastangų	Greitas, lengvas
Rinkimo kaina	Didelė	Nedidelė
Rinkimo laikas	Ilgas	Trumpas

Nustatant problemas ir tikslus labai svarbų vaidmenį gali suvaidinti tinkamo tiriamo nagrinėjamo reiškinio, proceso modelio tipo pasirinkimas.

2.3.2. Koncepcijų žemėlapis

Modeliai iš esmės yra abstrakčios konstrukcijos, kuriomis mėginama pavaizduoti tik kai kurias realaus pasaulio savybes. Modelio sąvokos apibrėžimus galima rasti ne tik moksliniuose, bet ir populiariuose šaltiniuose. Pavyzdžiui, tarptautinių žodžių žodyne teigiama, kad „modelis – originalo atvaizdas, tapatus pasirinktu struktūros lygmeniu arba pasirinktomis funkcijomis". Klausimas, ar modeliai yra empiriškai teisingi, neturi prasmės, nes to nė negali būti. Reikėtų kalbėti ne apie modelio empirinį teisingumą, bet apie nagrinėjamo modelio tinkamumą konkrečiu empiriniu atveju.

Dažniausiai skiriami penki modelių tipai:

- verbalinis (žodinis);
- grafinis;
- statistinis;
- matematinis;
- konceptualus (sąvokinis).

Konceptualus modelis yra vaizdinis metodas (diagrama), vaizduojantis priežastinius ryšius tarp veiksnių (faktorių), kurie, tyrėjų manymu, yra reikšmingi nagrinėjamai problemai ar objektui. Geras modelis turėtų aiškiai ir pagrįstai susieti objektą su jį tiesiogiai veikiančiais faktoriais (tiesioginėmis grėsmėmis ar galimybėmis), netiesioginiais faktoriais ir strateginėmis veiklomis, kurios daro įtaką šiems faktoriams. Konceptualus modelis turi būti lydimas tekstinio aprašymo, kuris sudarytą diagramą paaiškina žodžiais.

Konceptualaus modelio tikslas – atskleisti, kaip įvairios veiklos susijusios viena su kita, kaip jos logiškai yra išdėstytos ir sutvarkytos. Konceptualiuose modeliuose konceptams (abstrakčioms idėjoms ar bendriems principams, išreiškiamiems kaip trumpos frazės ar tekstas) jungti tarpusavyje taikomi koncepcijų žemėlapių metodai.

Statistinio tyrimo procesas tampa suprantamesnis ir lengvesnis, kai pradedama nuo konceptualaus modelio, vartojančio žmonėms lengvai suprantamas sąvokas (koncepcijas) ir kalbą.

Sąvokų žemėlapis (arba koncepcijų žemėlapis) tai yra hierarchinė diagrama, naudojama grafiniu būdu atvaizduoti žinias. Sąvokos išdės-

tomos nuo pačios svarbiausios, pagrindinės iki pačių specifiškiausių ir tarpusavyje sujungiamos linijomis arba strėlėmis.

Braižant koncepcijų žemėlapį, reikia išskirti pagrindinę koncepciją (nagrinėjamą problemą). Tolesnis žingsnis – surašomos pagrindinės sąvokos (veiksniai), kurios bus vartojamos šiame žemėlapyje (dažniausiai pakanka 10–20 sąvokų ir jas reikia išdėstyti hierarchiškai). Gera hierarchinė struktūra prasideda nuo bendros, daugelį reikšmių aprėpiančios sąvokos. Bendriausia sąvoka (veiksnys) yra užrašoma lapo viršuje arba centre, vėliau, remiantis jų hierarchiškumu, išdėstomos likusios sąvokos (veiksniai), jas jungiant tokia tvarka ir tokiais jungiamaisiais žodžiais, kad susidarytų teisingi, tarpusavyje sujungti teiginiai.

Pagrindinės koncepcijų žemėlapio braižymo taisyklės:

- koncepcijų žemėlapis visada turi hierarchiją;
- koncepcijos (veiksniai) sujungiamos rodyklėmis. Rodyklė rodo, kuria kryptimi skaitomas teiginys;
- rodyklės įvardijamos vienu ar keliais žodžiais. Taip sukuriamas prasminis teiginys;
- konkreti koncepcija (veiksnys) žemėlapyje gali būti pažymėta tik vieną kartą;
- kuo daugiau ryšių turi koncepcija (veiksnys), tuo ji svarbesnė.

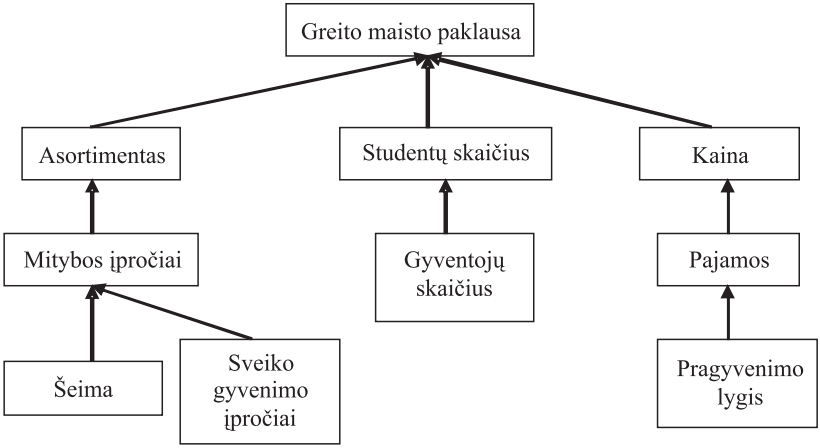
Reikia patikslinti koncepcijų žemėlapio aprašyme vartojamas sąvokas. Koncepcijos (ar esybės, angl. *entities*) yra specifinis konceptualaus modeliavimo terminas. Lotyniškai *entitas* – egzistuojantis daiktas ir atitinka nagrinėjamą objektą ar sistemą. Konceptualus modelis įvertina, kokią įtaką tam tikros koncepcijos turi kitų koncepcijų būsenai, t. y. priežastinius konceptų ryšius. Taikant daugiamačius statistinius metodus koncepcijų ryšiui vertinti, sąvoka *koncepcija* pervardijama į sąvoką *veiksnys (faktorius).* Žemesniu lygmeniu terminas *veiksnys (faktorius)* kartais suprantamas daug siauriau ir pakeičiamas terminu *kintamasis*. Terminu *kintamasis* norima pasakyti, kad koncepcija yra išmatuojama ir jai priskiriamos skaitinės reikšmės. Su kintamuoju susijusios skaitinės reikšmės keičiasi nagrinėjant atskirus atvejus.

Konceptualaus modelio kūrimo schema apima šiuos etapus:

- nagrinėjamo reiškinio apibrėžimą;
- koncepcijų įvardijimą ir žemėlapio sudarymą;

- norimų patikrinti teorinių teiginių prioritetų nustatymą;
- tinkamų tyrimo ir matavimo metodų parinkimą;
- duomenų rinkimo strategijos detalizavimą;
- priimtinų duomenų analizės metodų parinkimą.

Konceptualaus modelio žemėlapio pranašumas yra tai, kad jis vizualus ir jame lengva pamatyti grėsmes, galimybes bei trūkumas – kad dažnai modelis tampa labai sudėtingas ir nevaldomas (siekiant to išvengti stengiamasi vadovautis taisykle, kad viename puslapyje negali būti daugiau kaip 25 diagramos elementai).



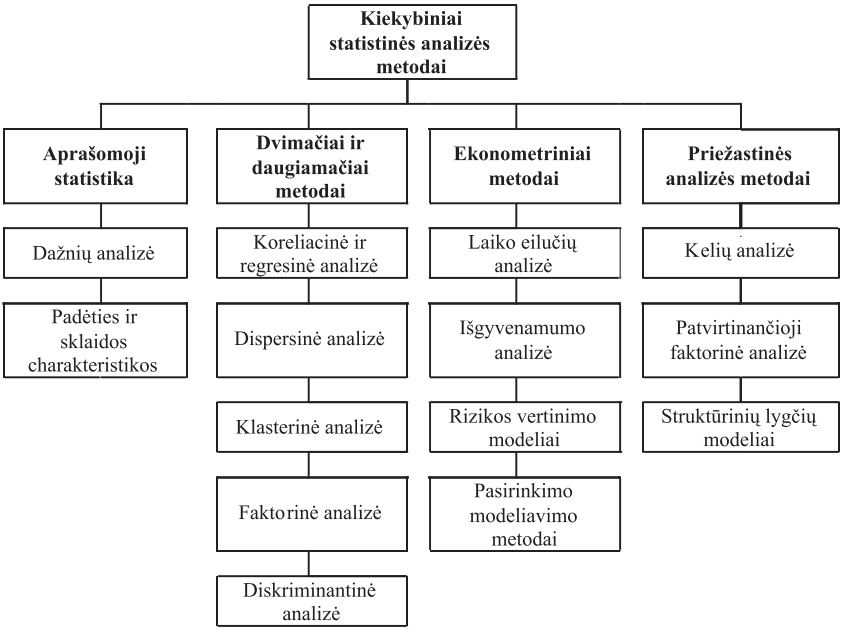
2.1 pav. Greito maisto paklausos koncepcijų žemėlapis

2.4. KOKYBINIAI IR KIEKYBINIAI TYRIMAI. TYRIMO TIPO PASIRINKIMAS

Visi metodai, taikomi tyrimo procesui, skirstomi į dvi dideles klases: *kiekybinius* ir *kokybinius*. *Kiekybiniai* metodai taip vadinami dėl to, kad jų tikslas – gauti kiekybinę informaciją apie didelį tyrimo objektų (vartotojų, įmonių, tiekėjų) skaičių. Kiekybiniai tyrimai yra susiję su išmatuojamais (objektyviais) dalykais, skaičiais ir logika. Tradicinis empirinis kiekybinis tyrimų metodas yra *apklausa*. Jos tikslas – iš nedidelės populiacijos dalies (imties) gauti išvadas apie visos populiacijos elgesį.

Remiantis moksliniais metodais sudaroma imtis, respondentai atsako į klausimus ir nustatomas jų atsakymų dažnis bei kitos statistinės atsakymų charakteristikos. Jeigu imtis statistiškai patikima, iš jos gauti rezultatai taikomi visai populiacijai. Visa populiacija dažniausiai tiriama tik atliekant gyventojų surašymą. *Kiekybinių tyrimų* pagrindą sudaro griežti statistiniai modeliai ir didelės imtys. Tai leidžia ne tik įvertinti prielaidas ar nuomones, bet ir nustatyti tikslias kiekybines (skaitines) tiriamų rodiklių reikšmes. Empirinių duomenų įvairovė ir kiekybinių metodų gausa leidžia tyrėjui rinktis norimą metodą iš gausios kiekybinių metodų klasės (žr. 2.2 pav.).

Kiekybiniai tyrimai yra pagrindinis instrumentas, kurį pasitelkus gaunama informacija, reikalinga planuojant ir priimant sprendimą. Kiekybiniai metodai taikomi vertinant gamybos apimtis, rentabilumą, formuojant kainą, produkto parametrus, ieškant nišų rinkoje ir daugeliu kitų atvejų. Šis instrumentas dažniausiai taikomas tada, kai būtinos hipotezės apie gamybą ar vartotojų elgesį jau suformuluotos taikant *kokybinius metodus*.



2.2 pav. Pagrindiniai kiekybiniai statistinės analizės metodai

Kiekybinių metodų privalumas tas, kad jie sumažina neteisingų sprendimų priėmimo riziką ir netikslių planavimo parametrų parinkimą. Įsitikinimas, kad ir be tyrimų situacija rinkoje visai aiški, dažnai nulemia netikslius, mažai efektyvius veiksmus ir veiksmai būna panašūs į bandymų bei klaidų metodo taikymą.

Ilgą laiką tyrėjai praktikavo tik kiekybinių duomenų analizę. Taip atsitiko todėl, kad socialiniai mokslai (psichologija, ekonomika, sociologija) savo tyrimo metodus perėmė iš fizinių mokslų, skirtų fizinių objektų analizei ir nepritaikytų sudėtingam žmonių elgesiui vertinti. Tačiau kiekybiniai tyrimai nepritaikyti žmonių elgesiui, jų skoniui, suvokimui ir produkto vertinimui paaiškinti. Tiems trūkumams kompensuoti buvo pasitelkti *kokybiniai* tyrimo metodai. Kokybiniai metodai nagrinėja žodžius, paveikslus ir įsivaizduojamus (subjektyvius) dalykus.

Kokybiniai tyrimo metodai grindžiami visai kitais principais nei kiekybiniai metodai. Kokybiniais jie vadinami todėl, kad aiškinant ir interpretuojant sąvokas ir problemas naudojami ne skaičiai, o žodžiai. Kokybinio tyrimo pagrindas yra subjektyvaus patyrimo prasmės suvokimas ir interpretavimas, o ne statistiniai matavimai. Kokybiniai metodai remiasi prielaida, kad skirtingi žmonės skirtingai suvokia realybę, be to, to paties asmens suvokimas skiriasi neatsižvelgiant į laiką, vietą ir kitas aplinkybes. Kokybiniai metodai dėmesį kreipia į subjekto požiūrio perteikimą, subjekto suteiktos prasmės atskleidimą suvokiant situacijos kontekstą.

Dažnai kokybiniai tyrimai yra kiekybinių tyrimų preliudija. Jie tinka žvalgybiniams tyrimams, problemoms nustatyti, priežasčių analizei ir naujų idėjų paieškai, o ne savybių ar nuomonių kiekybinių pasiskirstymui įvertinti. Kokybiniai tyrimai yra greiti ir pigesni nei kokybiniai. Atliekant kokybinius tyrimus apsiribojama mažu respondentų skaičiumi, todėl ***kokybinių tyrimo metodų rezultatai negali būti apibendrinti visai populiacijai***. Tačiau kokybiniai tyrimai suteikia daug vertingos informacijos, padedančios išsiaiškinti ir tiksliai apibrėžti problemas ir todėl juos naudoja daugelis tyrėjų. Kokybiniams tyrimams taikomi kokybiniai grupinių apklausų metodai, stebėjimo metodai, dokumentų analizė, projekciniai ir asociatyviniai metodai. Daugelį kokybinių metodų sukūrė psichologai, todėl šiems metodams taikyti reikalingos specialios žinios. Tokių metodų taikymo galimybės rinkodaroje ir socialiniuose tyrimuose ribotos.

Kokybinius tyrimo metodus rekomenduojama taikyti atvejais, kai būtina:

- sudaryti giluminį vartojimo modelį, suprasti vartotojo elgseną ir veiksnius, nulemiančius vartotojo pasirinkimą;
- suprasti vartotojo sprendimo pirkti priėmimo procesą;
- aprašyti vartotojo požiūrį į produktą, įmonę ir prekės ženklą;
- įvertinti pasitenkinimo esamais produktais lygį.

Labai svarbus kokybinių tyrimų vaidmuo kuriant naujus produktus. Šiuo atveju kokybiniai tyrimai leidžia:

- suprasti, ar rinkoje yra niša naujam produktui;
- įvertinti naujo produkto prekinį ženklą (brendą): pavadinimą, pakuotę, simbolį ir kt.;
- suprasti požiūrį į naują produktą;
- nustatyti ir patikslinti pozicionavimo strategiją.

Kokybiniai tyrimai apima labai skirtingas tyrimo metodikas, kurios skirstomos į keletą stambių grupių:

- **Grupinės diskusijos (fokus grupės)**. Fokus grupių metodas naudingas tyrimo pradžioje, kai tiriama situacija ir siūlomos idėjos. Fokus grupes dažniausiai sudaro 6–12 asmenų grupės. Pašnekesys trunka 1,5–2 valandas. Jis dažnai vyksta specialioje patalpoje, kuri įrengta taip, kad ją galima stebėti iš vienpusio veidrodžio ir naudoti garso ir vaizdo įrašus. Grupinei diskusijai pagal sudarytą planą vadovauja patyręs vadovas (interviuotojas). Dalyviai vienas su kitu susiduria tiesiogiai. Grupinės diskusijos metu tarp nepažįstamų dalyvių vyksta du pagrindiniai procesai: grupės jungimasis ir skirstymasis. Pradžioje respondentai stengiasi sukurti vieną grupę, bet ilgainiui jį suskyla į atskiras grupeles. Šios dvi priešingos tendencijos rinkodaros tyrimuose duoda labai gerų rezultatų. Jungimosi etape išryškėja bendros produkto savybės. Skilimo etape atsiskleidžia produkto trūkumai ir skirtumai nuo kitų produktų.
- **Giluminiai (asmeniniai) interviu**. Tai ilgai trunkantis interviu, kai respondentui pateikiami atvirojo tipo klausimai. Dažniausiai kalbamasi dviese, bet pokalbis gali vykti ir triadomis (vienas interviuotojas – du respondentai). Galima sakyti, kad giluminio interviu metodas yra priešingas anketavimui: res-

pondentui nepateikiami klausimai, o leidžiama pasisakyti laisvai, netaupant laiko. Tam reikalingas specialiai apmokytas interviuotojas, žinantis, ką jis gali daryti ir ko jokiu būdu daryti negalima. Idealiu atveju respondentas iš viso neturi žinoti, ką būtent norima išsiaiškinti, apie kokią prekę (prekės ženklą ar firmą) kalbama.

- **Pusiau struktūrizuoti (formalizuoti) interviu**. Kitaip nei giluminis interviu, šiam metodui būdinga vidinė struktūra, tačiau respondentui leidžiama netrukdomai reikšti savo mintis. Pusiau struktūrizuotas interviu taikomas apklausiant specialistus, nes jie nelinkę kalbėti nežinoma tema. Kaip ir giluminiame interviu, interviuotojas privalo laikytis taip, kad neturėtų jokios įtakos respondentui. Interviu metu naudojami ruošinukai, į kuriuos surašytos pagrindinės temos, etapai, svarbiausi momentai, kad nebūtų praleista kokia nors svarbi problema. Pusiau struktūrizuotas interviu dažniausiai trunka trumpiau nei giluminis interviu, nes pokalbis vyksta laikantis iš anksto sudaryto plano. Čia ypač svarbu stengtis, kad pagalbinės interviuotojo frazės nebūtų panašios į atsakymus.
- **Stebėjimai**. Stebėjimas – tai toks pradinių duomenų rinkimo būdas, kai tyrėjas tiesiogiai stebi žmones ir aplinką. Stebėjimai vykdomi įvairiais tikslais. Jie gali reikalingi kaip pradinių hipotezių formulavimo informacijos šaltinis, tikrinant kitais metodais tiriamą objektą. Tyrėjas gali stabtelėti vietose, susijusiose su nagrinėjama problema. Pavyzdžiui, stebėti gyventojų elgesį vykdomų akcijų metu. Stebėjimas yra daug laiko reikalaujantis metodas. Stebėjimo rezultatų aprašo apiforminimas kartais užima dukart daugiau laiko nei pats stebėjimas. Pagal supančią aplinką skiriami lauko sąlygomis vykdomi stebėjimai, kai procesai stebimi natūralioje aplinkoje (parduotuvėje, prie vitrinų ar lauko skydų) ir laboratoriniai, kai vykdomi dirbtinai sukurtoje aplinkoje. Stebėjimų rezultatai fiksuojami garso ar vaizdo technikos priemonėmis.
- **Projekciniai metodai**. Projekciniai metodai perimti iš klinikinės psichologijos. Šie metodai paremti projekcijos idėja, t. y. priskyrimu kitiems idėjų, minčių, nuomonių, norų, motyvų, ribojančių veiksnių, kurių žmogus nemano esant savais ir nuo

kurių nesąmoningai atsiriboja. Apklausos metu prašoma paaiškinti ar išsakyti savo nuomonę apie nagrinėjamą situaciją. Projekciniai metodai gali būti kelių pavidalų:

Tematinis apercepcinis testas. Vietoje paprasto klausimo respondentui pateikiama viena ar kelios situacijos ir prašoma jas pakomentuoti. Pavyzdžiui, žinoma, kad pirkėjas, vertindamas produkto kokybę, dažnai turi omenyje jo kainą, nors žodžiais tvirtina priešingai.

Sakinio užbaigimas. Taikant šį metodą respondentui pateikiama nežinomos frazės, turinčios tiesioginį ryšį su nagrinėjama tema, pradžia ir prašoma pabaigti sakinį. Šis metodas dažnai taikomas tiriant prekės įvaizdį.

Piešinio užbaigimas. Respondentui duodamas piešinėlis, panašus į komiksą, kuriame vaizduojamas personažų monologas arba dialogas nagrinėjama tema. Respondentas turi užpildyti tuščias vietas, paliktas veikėjų pasisakymams.

- **Asociatyviniai metodai**. Bendras šių metodų tikslas – nustatyti, kokias asociacijas sukelia prekė, prekinis ženklas, įmonė ir t. t. Iš asociatyvinių metodų dažniausiai taikomi šie:

Laisvoji asociacija: respondentui siūloma laisvai ir spontaniškai išreikšti žodžiais, kokias idėjas sukelia jam vienas ar kitas daiktas arba problema.

Žodinė asociacija: respondentui pateikiama keletas požymių (atsparus, greitai dūžtantis, minkštas, greitas, lėtas ir pan.) ir siūloma susieti kiekvieną iš jų su tiriama preke, prekiniu ženklu ir pan.

Vaizdinė asociacija: šis metodas gali būti taikomas visoms respondentų grupėms, neišskiriant vaikų ir užsieniečių. Tiriant šiuo metodu nekalbant rodomas tiriamas objektas (prekė, modelis, prekinis ženklas), o paskui – lengvai atpažįstami stereotipiniai pirkėjų vaizdai (darbininkas, ūkininkas, sekretorė, sportininkas, tarnautojas, verslininkas, namų šeimininkė). Paskutinis etapas – sujungti šiuos vaizdus su tiriamu daiktu.

Kiniško portreto tipo asociacija. Šis metodas taikomas kūrybinėms grupėms, bet tinka ir rinkos tyrimams tais atvejais, kai tiriamas prekės, prekinio ženklo, įmonės įspūdis. Respondento prašoma susieti objektą su gyvūnu, mitologiniu personažu, gėle ir t. t. Tuo tikslu dažnai naudojami šio tipo sakiniai: „Jeigu (prekės, modelio, prekinio ženklo pavadinimas) būtų gyvūnas, tai šis gyvūnas būtų …“, „Jeigu (prekės, modelio, prekinio ženklo pavadinimas) būtų mitologinis personažas, tai būtų…“

- **Eksperimentai**. Eksperimento tikslas – patikrinti, kokioms sąlygoms esant dominantis stimulas sukelia pageidautiną reakciją. Eksperimentas retai kada naudojamas tyrimo pradžioje. Dažniausiai juo tikslinamos alternatyvų pasirinkimas. Pavyzdžiui, tipiniai eksperimento metodikos taikymo atvejai yra optimalios reklamos varianto, produkto dizaino parinkimas. Tuo tikslu testuojamos atskiros gyventojų grupės.

Laboratoriniai eksperimentai vykdomi esant tam tikroms dirbtinėms sąlygoms, kurios padeda išvengti šalutinių veiksnių įtakos. Pavyzdžiui, vertinant pirkėjų reakciją į įvairius reklamos tipus, galima pakviesti tam tikros lyties, amžiaus, socialinės padėties pirkėjus. Pastaraisiais metais laboratoriniams eksperimentams atlikti vis plačiau taikomi kompiuteriai. Sukurtos kompiuterių programos, kurios leidžia vartotojui „vaikščioti" po parduotuvę ir rinktis prekes. Kompiuteryje sumontuota technika ne tik registruoja pirkinius, bet ir matuoja žmonių reakciją į atskirus elementus: kainą, dizaino spalvą ir formą, prekių išdėstymą.

Lauko eksperimentai atliekami natūraliomis sąlygomis: parduotuvėse, vartotojų namuose. Nors šių eksperimentų rezultatai turėtų būti patikimesni nei laboratorinių, jiems didelę įtaką turi šalutiniai veiksniai, jie ilgiau užtunka ir brangiau kainuoja. Lauko sąlygomis atliekami įvairių tipų eksperimentai vadinami „bandomąja rinkodara" ar rinkos testavimu. Firmos lauko eksperimentai gali būti atliekami iškart keliose parduotuvėse ar net skirtinguose miestuose.

- **Ekspertinis vertinimas**. Taikant ekspertinio vertinimo metodą atliekami šie žingsniai:
  - sudaromas ekspertinio tyrimo planas;
  - parenkamas ekspertų apklausos metodas ir parengiami būtini dokumentai;
  - atrenkami ekspertai;
  - atliekama ekspertinio vertinimo procedūra;
  - analizuojami ekspertinio vertinimo duomenys, pašalinamos klaidos ir prieštaravimai;
  - rezultatai interpretuojami;
  - atlekiamas oficialus ekspertinio vertinimo rezultatų ir išvadų pateikimas.

Rinkos tyrimuose taikomi ir kiekybiniai, ir kokybiniai tyrimo metodai, taip pat jų deriniai, nes šie tyrimo metodai dažnai papildo vienas kitą.

Pastaba. Nereikia painioti kiekybinių ir kokybinių tyrimo metodų su kiekybiniais ir kokybiniais kintamaisiais. Ir kokybiniai, ir kiekybiniai tyrimai gali naudoti visų tipų kintamuosius, skiriasi tik tyrimo metodika ir duomenų apdorojimo metodai.

2.5. ANKETINĖS APKLAUSOS METODAS

Atliekantys apklausą tyrėjai skirstomi į dvi grupes: atsakingieji tyrėjai ir tyrimo vykdytojai arba interviuotojai (apklausėjai).

Apklausai atlikti taikomos trys pagrindinės formos:

- pašnekesys;
- interviu;
- anketavimas.

Dažniausiai naudojamos kelios anketinės apklausos rūšys:

- anketinė apklausa el. paštu arba paštu;
- telefoninė apklausa;
- apklausa per žiniasklaidą;
- ekspertų apklausa, kai atrenkama žmonių grupė (ekspertai), turinti kurios nors srities žinių.

Ekspertai nagrinėja problemą ir kiekybiškai ar kokybiškai ją įvertina.

Dažnai (bet nebūtinai!) anketose būna šios dalys: įvadinė, padėties, pagrindinė ir baigiamoji.

Įvadinėje anketos dalyje formuluojamas kreipimasis į respondentą, pateikiamos apklausos sąlygos ir atsakymų teikimo taisyklės.

Padėties dalį sudaro klausimai apie socialinę demografinę respondento padėtį. Šie klausimai skirti analizės rezultatų interpretavimui palengvinti.

Pagrindinę dalį sudaro klausimai apie nagrinėjamą temą.

Baigiamojoje dalyje respondentui gali būti suteikta galimybė laisvai pasisakyti apie tiriamą problemą.

Padėties ir baigiamoji dalys nėra privalomos, jos gali būti praleistos. Įvadinė dalis gali būti pakeista žodiniu paaiškinimu.

Pagrindinės anketos dalies klausimai skirstomi į tris grupes:

- dichotominiai (klausimai, į kuriuos galima atsakyti tik „taip“ arba „ne“. Kartais jie patikslinami papildomu klausimu: „jei taip, tai kodėl?“);
- atvirieji (numato laisvos formos atsakymą. Pateikiant šio tipo klausimus galima gauti išsamią informaciją, tačiau atsakymai sunkiai analizuojami ir formalizuojami);
- uždarieji (numato keletą parengtų atsakymų. Išrenkamas tas, kuris atitinka respondento nuomonę. Rekomenduotini ne mažiau kaip keturi atsakymų variantai).

Jeigu viešosios nuomonės tyrimai yra skelbiami viešai, prie jų turi būti pridėta oficiali tyrimo suvestinė, o joje turi būti nurodyta:

- tyrimą atlikusios organizacijos pavadinimas;
- duomenų rinkimo laikas;
- apibrėžta tiriamoji populiacija;
- naudotas atrankos metodas, imties dydis ir apklausos vieta. Turi būti nurodyta, kuris regionas buvo tiriamas, kaip jis buvo suskirstytas smulkiau ir kaip imtis išdėstyta geografiškai;
- apklausos būdas ir sutikusių dalyvauti tyrime respondentų procentas. Jei šis rodiklis yra mažas, reikėtų išsiaiškinti nenoro dalyvauti tyrime priežastis;
- norint išvengti galimų nesusipratimų, turėtų būti paskelbtas teiktų klausimų sąrašas;
- jeigu pateikiami rezultatai labai skiriasi nuo pradinių surinktų duomenų, turi būti nurodyta, ar buvo taikyti statistiniai svorių metodai bei kitos statistinės procedūros.

Be abejo, iškyla apklausų patikimumo problema. Yra trys svarbiausi veiksniai, kurie gali sukelti sistemines paklaidas:

- praktikoje nepatikrintos apklausos procedūros;
- nepakankamos pastangos įtikinti respondentus dalyvauti apklausoje ir todėl didelis atsisakiusių dalyvauti apklausose procentas;
- nerūpestingai atlikta apklausa ir nepakankama kokybės kontrolė.

2.5.1. Kintamųjų tipai ir matavimo skalės pasirinkimas

Požymiai – grupavimo pagrindas duomenų visumai suskirstyti į pogrupius. Jų parinkimas priklauso nuo grupavimo tikslo ir preliminarinės analizės. Skiriami kokybiniai (atributiniai) ir kiekybiniai požymiai.

Parinkus duomenų visumos nariams tam tikrus grupavimo požymius, juos tenka matuoti. Kiekvieno požymio matuojamos reikšmės sudaro duomenų aibę. Tuos požymius vadinsime *kintamaisiais*. Duomenų visumos ar dalies aibė – tai atitinkamų kintamųjų reikšmių aibė. Kintamieji, kaip ir požymiai, skirstomi į *kokybinius* ir *kiekybinius*.

Su kintamųjų tipu siejasi matavimo skalės. Kokybinių kintamųjų reikšmės gali būti išmatuotos *nominaliojoje* (*pavadinimų*) arba *ranginėje* skalėje, kiekybinių – *metrinėje* (intervalų ar santykių) skalėje. Matavimų skalė lemia duomenų analizės metodo parinkimą.

Pavadinimų (nominalioji) skalė (angl. *nominal measurement scale*). Ši skalė taikoma tada, kai imties objektus galima tik klasifikuoti, t. y. priskirti prie vienos ar kitos grupės. Imties nariams arba jų grupėms priskiriami kodai arba kiti simboliai. Grupės numeruojamos laisvai, o jų numeriai turi ne skaičių, bet tam tikrų grupių (vardų) prasmę. Kintamieji, kurie matuojami pavadinimų skalėje, vadinami *nominaliaisiais kintamaisiais*. Šių kintamųjų pavyzdžiai: lytis, socialinis statusas, tautybė, kraujo grupė, studentų numeris mokomųjų dalykų žurnale, komandos žaidėjo numeris, telefono numeris, akademinės grupės numeris ir t. t.

Nominalieji kintamieji tarpusavyje nepalyginami, jų reikšmių aritmetinis vidurkis neturi prasmės.

Dichotominė matavimų skalė (angl. *dichotomous measurement scale*) – taikant šią skalę, požymių reikšmės yra priskiriamos prie vienos iš dviejų galimų kategorijų. Pavyzdžiui, gyvas arba miręs, rūko – nerūko ir kt.

Ranginė (tvarkos) skalė (angl. *ordinal measurement scale*) Ši skalė taikoma tada, kai galima nustatyti objektų tiriamojo požymio skirtumus ir pagal tai išrikiuoti objektus į eilę. Kintamieji, kurie matuojami rangų skalėje, yra vadinami *ranginiais kintamaisiais*. Šių kintamųjų pavyzdžiai: konkurso dalyviams pagal parodytus rezultatus skiriamos vietos, pedagoginiai ir moksliniai vardai (dėstytojas, docentas, daktaras, habilituotas daktaras ir t. t.). Prie klasikinių ranginės skalės kintamųjų priskiriami kintamieji, gauti siejant duomenis į klases, pavyzdžiui, pa-

jamas skirstant intervalais: iki 500 Lt; nuo 501 iki 1000 Lt; nuo 1001 iki 2000 Lt; per 2000 Lt.

Rangų skalėjė išmatuoti dydžiai gali būti lyginami tik eiliškumui nustatyti. Tai nėra kiekybinės charkteristikos: pirmą vietą krose užėmęs sportininkas nebėga tris kartus greičiau už trečioje vietoje likusį bėgiką. Vartojant šią skalę galima teigti, kad vieno objekto požymio reikšmė didesnė, lygi ar mažesnė už kito objekto, nors neįmanoma nustatyti, kiek ji didesnė ar mažesnė. Idealioje ranginėje skalėje sutampančių verčių neturėtų būti. Praktikoje sutampančių rangų būna. Tokiais atvejais taikant ranginius metodus neretai tenka priimti specialias pataisas. Mažiau tikslesnis šios skalės variantas – balinė matavimo skalė (pvz., žinių įvertinimas balais, ligos stadija, dangos korozinis pažeistumas ir pan.). Joje sutampančių verčių būna daug.

Jei ranginę skalę sudaro 15 ir daugiau pozicijų, kintamąjį galima nagrinėti kaip kiekybinį.

Ranginiai ir nominalieji kintamieji dar vadinami *kategoriniais*.

Intervalų skalė (angl. *interval measurement scale*). Ši skalė naudojama, kai imties objektai yra skaitiniai, juos galima įvertinti kiekybiškai. Nulis ir intervalo ilgis parenkami laisvai, dviejų šios skalės intervalų santykis nepriklauso nei nuo numatomų vienetų, nei nuo nulinio taško. *Intervalinių kintamųjų* pavyzdžiai: temperatūra, kalendorinis laikas (nulinis taškas – Kristaus gimimas, Romos įkūrimas, laiko intervalai pagal Saulę, Mėnulį ir kt.), gabumų, intelekto testų rezultatai.

Intervalų skalėje nurodomi kiekybiniai kintamųjų reikšmių skirtumai, išreikšti matavimo vienetais (metrais, sekundėmis, laipsniais, taškais ir pan.). Šie skirtumai gali būti tarp atskirų intervalų arba nuo kurio nors pasirinkto atskaitos taško, t. y. nulinė reikšmė dar nereiškia, kad tiriamasis požymis visai nepasireiškia, o tik tai, kad jis nesiskiria nuo sąlyginio atskaitos nulio. Duomenis intervalų skalėje (intervalinius kintamuosius) galima apdoroti visais be apribojimų statistikos metodais.

Santykių skalė (angl. *ratio measurement scale*). Ši skalė skiriasi nuo intervalų skalės tik tuo, kad joje yra apibrėžtas nulinis taškas (ataskaitos pradžia). Santykių skalės kintamųjų pavyzdžiai: kaina, laikas, svoris, amžius ir kt. Skaičių, gautų matuojant požymius, santykis parodo kiekybinį matuojamojo požymio santykį, kuris npriklauso nuo matavimo vienetų.

Atliekant statistinę analizę, skirtumai tarp intervalinės ir santykių skalės kintamųjų paprastai laikomi nereikšmingais, ir šių kintamųjų analizei taikomi tie patys metodai. SPSS taikymo praktikoje skirtumas tarp kintamųjų, matuojamų pagal intervalų arba santykių skalę, yra neesminis. Todėl toliau nagrinėsime matuojamus pagal intervalų skalę (intervalinius) kintamuosius.

Teorijoje pateikiami keturi pagrindiniai matavimo skalių tipai:

- lygių intervalų, arba Thurstone'o skalė;
- suminė, arba Likerto skalė;
- kaupiamoji, arba Guttmano skalė;
- semantinė diferencialinė skalė.

Plačiau panagrinėsime šias matavimo skales.

Lygių intervalų, arba Thurstone'o skalė. 1929 m. L. Thurstone'as pasiūlė ranginės skalės intervalus, sudalytus lygiais intervalais, vertinti kaip iš tiesų lygius intervalus ir priskirti juos ne kokybinei ranginei skalei, o kiekybinei intervalinei skalei.

Ši skalė išliko kaip metodologinis pavyzdys ir turi teorinę vertę, bet tyrimuose naudojama retai.

Sudarant lygių intervalų skalę, kaip ir daugeliu kitų atvejų, pradedama nuo didelės aibės formuluočių (teiginių) apie tiriamą požymį. Šios formuluotės turi būti panašios gramatinės struktūros. Kiekviena formuluotė gali būti teiginys, su kuriuo galima sutikti arba nesutikti. Pavyzdžiui, nagrinėjant, kokios savybės būtinos, norint tapti lyderiu, galima formuluoti šias savybes (teiginius):

- bendras kultūros lygis;
- sugebėjimas mokytis;
- logika;
- savikritika;
- atsakomybė;
- energija ir aktyvumas;
- tikslo siekimas ir kryptingumas;
- savitvarda;
- ištvermingumas;
- asmenybės branda;
- padorumas;
- sugebėjimas bendrauti;

- pakantumas kitų nuomonei;
- lankstus elgesys;
- sugebėjimas sukurti palankų įvaizdį.

Svarbiausia Thurstone'o technikos procedūra yra teiginių ekspertinis išrūšiavimas pagal tai, ar teiginys atžvilgiu tiriamo požymio yra palankus, ar nepalankus. Pavyzdžiui, teiginys „Aš nenoriu, kad gėjai gyventų mano kaimynystėje" kintamojo „tolerancija" atžvilgiu būtų žemai.

Rikiavimui naudojama kelių ekspertų nuomonės. Kiekvienas ekspertas teiginius išrikiuoja ir suteikia jiems numerius. Formuojant skalę, pirmenybė teikiama tiems teiginiams, dėl kurių ekspertų nuomonės sutapo.

Suminė, arba Likerto skalė. Šią skalę pasiūlė Rensis Likertas. Tęsdamas Thurstone'o darbus, 1932 m. jis pasiūlė metodą, kaip išreikšti skalę tokiu būdu, kad vieno klausimo taškus būtų galima sudėti su kitų klausimų taškais ir taip gauti bendrus (suminius) taškus.

Likertas sudarė standartinę 5 taškų skalę pavartodamas būdvardžius (vadinamuosius inkarus – angl. *anchors*), įgyjančius reikšmes nuo „visai sutinku" iki „visai nesutinku" su atitinkamomis skaitinės reikšmės nuo 5 iki 1.

Likerto skalė yra viena iš dažniausiai naudojamų priverstinio pasirinkimo (angl. *forced-choice*) skalių. Ji yra plačiai apklausose naudojama skalė, ypač matuojant nuostatas, įsitikinimus ir nuomones.

Pagrindinės šios skalės konstravimo idėjos yra šios:

a) nagrinėjamas klausimas pateikiamas kaip konstatuojamasis sakinys;
b) tada pateikiama keletas atsakymo variantų arba pasirinkimų, kurie galėtų parodyti sutikimo ar pritarimo šiam sakiniui mastą.

Pagal šią skalę respondentai išreiškia savo požiūrį į pateiktus teiginius pasirinkdami vieną iš jiems pateiktų (iš anksto paruoštų) skalės reikšmių.

Likerto pasiūlyta skalė tapo dažniausiai naudojama (ir plūstama) matavimo skale.

Pagal Likerto reikalavimus sudaryta 5 taškų matavimo skalė privalo turėti du teigiamus taškus, neutralų tašką ir du neigiamus taškus.

Standartinė 5 taškų Likerto skalė reiškia:

5 = giežtai teigiamas požiūris;

4 = šiek tiek teigiamas požiūris;
3 = neapsisprendęs (neutralus);
2 = šiek tiek neigiamas požiūris;
1 = griežtai neigiamas požiūris.

Likerto skalės „klausimai“ yra pateikiami tvirtinimo forma, su kuriais respondentas gali sutikti arba nesutikti.

Paklausus dešimt klausimų galima gauti didžiausią suminį balą 50 (kai pasirinkti 5 kiekvieno punkto taškai). Taigi, nagrinėjant nuomonę pasitelkus klausimyną, šiuo atveju 50 reikštų didžiausią galimą nuomonių sumą, o taškai – mažiau nei 50 atitiktų mažesnes nuomonių sumas.

1 pavyzdys.

Teiginys. Reguliari techninė priežiūra yra automobilio saugaus eksploatavimo garantas.

1 – Visiškai nesutinku.
2 – Nuosaikiai nesutinku.
3 – Vienodai sutinku ir nesutinku.
4 – Nuosaikiai sutinku.
5 – Visiškai sutinku.

2 pavyzdys.

Teiginys. Nuo plaukų spalvos priklauso žmogaus charakteris.

1 – Visiškai nesutinku.
2 – Nuosaikiai nesutinku.
3 – Vienodai sutinku ir nesutinku.
4 – Nuosaikiai sutinku.
5 – Visiškai sutinku.

Kartais bandoma sudaryti Likerto skalę kaip klausimą, pavyzdžiui: „Ar nuo plaukų spalvos priklauso žmogaus charakteris?“ Tačiau negalima „sutikti“ ar „nesutikti“su klausimu! Tam, kad būtų galima išmatuoti sutikimo ar nesutikimo intensyvumą, pateikiamame dalyke turi būti kažkas, į ką reikėtų sureaguoti, būtent, deklaratyvus teiginys.

Kitas natūraliai kylantis klausimas – kiek atsakymų variantų reikia pateikti? Dažniausiai taikoma 5 ar 6 pasirinkimo alternatyvų skalė. Nėra jokios griežtos taisyklės, ar turi būti lyginis, ar nelyginis pasirinkimo alternatyvų skaičius. Sprendimas priklauso nuo to, kas yra tiriama, taip pat nuo ankstesnių šios srities darbų apimties. Tai viena iš daugelio problemų, kuri gali būti patikrinta atliekant pilotinį tyrimą.

Visada iškyla vidurinio taško problema. Klausimas – ar turi būti vidurinysis, neutralus ar „balansuojantis“ taškas? Pirmiausia reikėtų nutarti, kaip šis vidurinis taškas bus pavadintas. Galimi tokie vidurinio taško variantai:

a) neutralu;

b) nei sutinku, nei nesutinku (skatina negalvoti);

c) vienodai sutinku ir nesutinku.

Kai kas pasakytų, kad visi šie pažymėjimai yra ekvivalentiški. Tačiau šio tipo literatūroje teigiama, kad šie žymėjimai turi subtilius, bet aiškius skirtumus, kuriuos jie turėtų sukelti respondentų galvose. Pavyzdžiui, R. DeVellis (1991) teigia, kad:

žymė b) leidžia numanyti apatišką nesidomėjimą, o likusios žymės reiškia griežtą, net lygų potraukį ir sutikimui, ir nesutikimui;

žymės a) ir c) sukelia aktyvų mąstymą, o b) turėtų patikti tam, kas visai nenori galvoti apie pateikiamą problemą.

Vidurinio taško problemos esmę tikriausiai būtų galima apibrėžti klausimu; „Ar vidurinio taško pateikimas nesudarys sąlygų respondentui „išsisukti“? Tuo tarpu tyrėjai visada nori žinoti požiūrių intensyvumą.

Atlikus gausius tyrimus, buvo nustatyta:

- kai pateikiamas vidurinis taškas, jį pasirenka apie 20 procentų apklausiamųjų. Tačiau, jei šios žymės nėra, jos neprašoma ir nepasigendama;
- vidurinio taško buvimas/nebuvimas turi labai nedidelį poveikį kitų kategorijų atsakymų pasiskirstymui.

Guttmano skalė. Matuojant Guttmano skale teiginiai išdėstomi tokia tvarka, kad respondentui sutikus su bet kuriuo iš eilėje išdėstytų teiginių, logiškai jis turėtų sutikti ir su visais prieš tai buvusiais teiginiais.

Semantinė diferencialinė skalė. Šį skalių sudarymo metodą, panašų į Likerto, sukūrė Ch. Osgoodas apie 1950 m.

Semantinė diferencialinė skalė telkiasi apie tiriamą objektą (stimulą), kuriuo gali būti asmuo, reiškinys, pozicija ir pan.

Konstruojant semantinę diferencialinę skalę privalu:

1. Nustatyti stimulą.
2. Kitu žingsniu išvardyti kelias būdvardžių poras (inkarus). Kaip kraštinės reikšmės (inkarai) apgalvotai naudojamos priešingos reikšmės (antonimai).

3. Respondentas pasirenka linijos atkarpą, kuri geriausiai atitinka jo nuomonę apie stimulą.

Semantinės diferencialinės skalės pavyzdys:

AUTOMOBILIŲ PARDAVĖJAI

Garbingi	__	__	__	__	__	__	__	__	__	__	Negarbingi
Ramūs	__	__	__	__	__	__	__	__	__	__	Triukšmingi

Šiame pavyzdyje stimulas yra automobilių pardavėjai.

Šis stimulas vertinamas panaudojant būdvardžių, apimančių visą galimą nuomonių įvairovę, poras.

Praktiškai vienos ar dviejų porų būdvardžių neužtenka, kad būtų galima apibūdinti stimulą, todėl šis pavyzdys yra neišsamus.

Lieka neaišku, kiek galimų pasirinkimo linijų (brūkšnių) reikia pateikti tarp būdvardžių poros. Kaip paprastai apklausose, griežtų taisyklių nėra. Dažniausiai naudojama nuo 7 iki 9 pasirinkimo linijų.

Kartais naudojama mažiau pasirinkimų ir jie numeruojami.

Šiuo atveju visiems skalės taškams iš anksto suteikiamos reikšmės: +2 +1 0 –1 –2.

Sudarant semantinę diferencialinę skalę priešingos reikšmės neturi priešdėlio „ne“, pavyzdžiui:

Naudingas				**Žalingas**
+2	+1	0	–1	–2

2.5.2. Imties sudarymo būdas ir dydis

Kasdienėje praktikoje mes susiduriame su daugybe tvirtinimų ar paneigimų, kuriuose minima tam tikra skaitinė arba kiekybinė informacija.

Pavyzdžiai.

- Daugiau kaip pusė mūsų šalies gyventojų tiki geresne Lietuvos ateitimi.
- Per 60 proc. Lietuvos gyventojų pasitiki spauda, o daugiau kaip 70 proc. – Bažnyčia.
- X politiko reitingas per paskutinius 3 mėnesius krito 10 proc.

Tokius teiginius vadiname statistiniais teiginiais. Jie paremti duomenimis, kurie gauti statistinio stebėjimo metu. Renkant duomenis visuomet reikia ištirti tam tikrų objektų ar individų grupę, kuri statistikoje vadinama *statistine visuma, generaline aibe arba populiacija*. Dažnai yra sudėtinga tiksliai apibūdinti populiaciją, nustatyti jos dydį, o visa tai padarius įsitikinama, kad visuma tiriamų objektų yra labai didelė. Vykdyti ištisinį tyrimą, rinkti duomenis apie kiekvieną objektą ar asmenį dažnai yra neįmanoma arba tam reikia didelių laiko ir materialinių sąnaudų. Todėl dažniausiai renkami duomenys tik iš dalies visumos narių. Dalis populiacijos objektų arba asmenų, apie kuriuos renkame mus dominančius duomenis, vadiname *imtimi*.

Pagrindinės priežastys, dėl kurių tiriama ne visa populiacija, o atrenkama imtis:

- sutaupoma pinigų;
- sutaupoma laiko;
- kai kurie populiacijos nariai gali būti neprieinami;
- tiriant kai kuriuos elementus gali tekti sunaikinti;
- tikslumas – ištirti mažesnį kiekį elementų iš kruopščiai atrinktos imties galima tiksliau.

Metodas, kuriuo atrenkama tinkama imtis, turint tikslą nustatyti visai populiacijai būdingus bruožus ar ypatybes, vadinamas *atranka*.

Esminė imties sudarymo problema – parinkimas imties, kuri atstovautų visai generalinei aibei (populiacijai) – būtų *reprezentatyvi*. Pagrindinė imties reprezentatyvumo užtikrinimo sąlyga yra jos *atsitiktinis* parinkimas iš populiacijos. Užtikrindami tokį imties parinkimą, mes galėsime, remdamiesi imties tyrimo rezultatais, daryti išvadas apie visą populiaciją.

Imtys sudaromos įvairiais būdais (žr. 2.3 pav.), tačiau pagrindiniai imčių sudarymo būdai yra du: *tikimybinis* ir *netikimybinis*. Atrenkant tikimybiniu būdu, kiekvieno elemento tikimybė patekti į imtį yra žinoma, o sudarant netikimybines imtis, daug lemia atsitiktinumas.

Trumpai apibūdinsime kiekvieną *tikimybinės imties* sudarymo atvejį.

Sudarant *paprastąją atsitiktinę* imtį (*atsitiktinės atrankos metodas*), visų populiacijos elementų galimybės priklausyti imčiai yra vienodos. Tai užtikrinama naudojant kompiuterinį atsitiktinių skaičių generatorių arba, sunumeravus visus elementus, traukiant juos burtų

būdu. Šis būdas naudotinas esant nedideliam populiacijos narių skaičiui. Didelėse populiacijose priskirti kiekvienam individui žymę ir paskui atlikti atsitiktinį traukimą nėra paprasta.

Sisteminės imties sudarymo principas – atsižvelgiant į populiacijos ir būsimos imties dydį parenkamas išrinkimo žingsnis, populiacijos elementai išrikiuojami į eilę, iš pirmųjų pasirenkamas vienas, o toliau pasirinktu žingsniu atrenkami visi kiti elementai. Jei populiacija didelė, šis atrankos būdas artimas atsitiktinei atrankai.

Sluoksninės imties atveju visa populiacija suskirstoma į sluoksnius (mažesnes aibes), kurių kiekviena yra vienalytė tiriamojo požymio atžvilgiu. Kiekvienam sluoksniui taikomas paprastosios atsitiktinės imties sudarymo metodas. Tokiu būdu galima atlikti visos populiacijos ir atskirų sluoksnių tyrimą. Teoriškai įrodoma, kad taikant sluoksninės imties sudarymo schemą, gaunami tikslesni rezultatai nei paprastosios atsitiktinės imties atveju.

Lizdinės (klasterinės) imties atveju visa populiacija suskirstoma pagal tam tikrą požymį į panašias grupes – lizdus (klasterius). Iš visų lizdų aibės paprastosios atsitiktinės imties būdu atrenkama dalis. Pagal tiriamąjį požymį išskirtos grupės (lizdai) turi būti kiek galima panašūs vienas į kitą. Dažniausiai naudojama lizdinės imties forma – teritorinė imtis, kurioje lizdai sudaromi panaudojant žemėlapius. Šiuo atveju tirti objektus yra paprasčiau ir pigiau, nes nereikia turėti informacijos apie smulkiausius imties vienetus, lengviau vykdyti duomenų rinkimą.

Visi aptarti tikimybiniai imčių sudarymo atvejai sudaro galimybę surinktų statistinių duomenų analizei taikyti tikimybių teoriją.

Netikimybinės (neatsitiktinės) imtys pasižymi subjektyvumu, todėl yra mažiau reprezentatyvios. Šių imčių atveju neužtikrinamos vienodos galimybės kiekvienam populiacijos elementui patekti į imtį. Gautiems tiriamo požymio įverčiams nėra matematinio jų įvertinimo pagrindimo. Pagrindinis šių imčių privalumas – paprastas, prieinamas jų sudarymas, kuris esant ir nedideliam imties tūriui (imties elementų skaičiui), kai tyrimą vykdo gerai žinantys tyrimo objektus asmenys, duoda pakankamai patikimus rezultatus.

Netikimybinių imčių sudarymo būdai.

Ekspertinės imties atveju elementai į imtį traukiami atsižvelgiant į ekspertų nuomonę. Imtis pasiteisina ir duoda gerus rezultatus, kai jos

apimtis nedidelė. Populiacijos vienetų savybės yra smarkiai kintančios, todėl tokią imtį turi sudaryti atitinkamos srities geras specialistas.

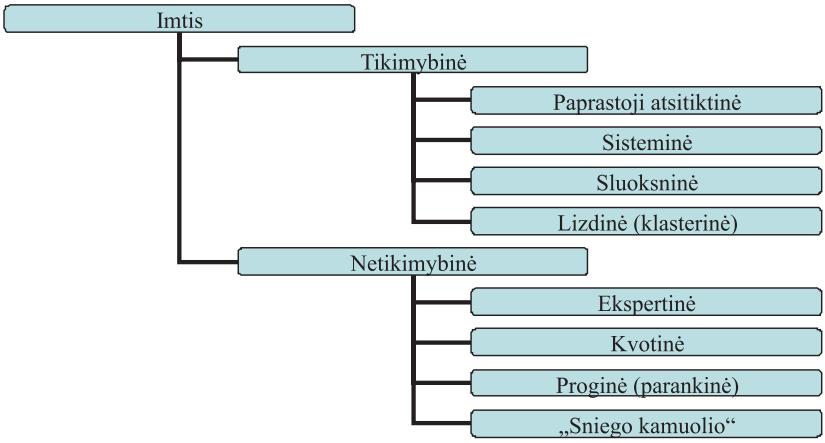
Kvotinė atranka gaunama tada, kai populiacijos pogrupiams nustatomos jų atstovavimo kvotos. Toliau kiekvienoje kvotoje atsitiktinai pasirenkant elementus yra užtikrinamas didesnis imties reprezentatyvumas. Ši atranka panaši į sluoksniuotą, tačiau respondentai (kvotos) pasirenkami neatsitiktinai. Kvotos būdas naudingas, kai nereikia didelio tikslumo, o pakanka bendro vaizdo. Šis būdas gali būti taikomas operatyviai tiriant visuomenės nuomonę, prekių rinką.

Proginės (parankios) imties atveju į imtį įtraukiami pirmi pasitaikę populiacijos elementai. Tai mažai reprezentatyvi, pigi, paprastai vykdoma atranka.

„Sniego kamuolio“ atranka yra tada, kai atsižvelgiant į gautą informaciją keičiamas respondentų skaičius (dažniausiai didinamas).

Imtys neapima visos populiacijos, todėl jos elementams apskaičiuotas parametras skiriasi nuo visos populiacijos atitinkamo parametro. Gaunama *imties paklaida*, kurią dažniausiai lemia atrankos klaidos. Dažniausios atrankos klaidų priežastys:

Atsitiktinumas (priklauso nuo imties didumo). Pavyzdžiui, išmatavus 100 moterų ūgį, visos buvo aukštesnės nei 170 cm.



2.3 pav. Imčių parinkimo atvejai

Tendencingumas (sisteminės paklaidos). Nulemia „matavimo įrankio“ netobulumas. Pavyzdžiui, vykdant apklausą internetu, bus su-

žinota nuomonė tik tų respondentų, kurie turi internetą. Sistemingos paklaidos dažnai daromos prognozuojant rinkimų rezultatus.

Imties paklaidai didelį poveikį turi ir *atsakymo lygmuo* – atsakiusiųjų ir visų parinktų respondentų santykis.

Atsakymų matavimo klaidos:

Apklausos poveikis. Tas pats respondentas skirtingų apklausų metu gali pateikti skirtingus atsakymus. Pavyzdžiui, kai kurie respondentai linkę „pasijauninti", kiti nurodo daugiau metų nei turi.

Atsakančiojo (respondento) poveikis. Kai kurie respondentai pateikia neteisingus atsakymus, norėdami padaryti įspūdį tyrėjams. Pavyzdžiui, ūkininkai vyras ir žmona dažnai pateikia skirtingus duomenis apie gautą derlių.

Nepasitikėjimas tyrimo tikslu. Pavyzdys: klausimas apie pajamas. Jei klausia oficialus asmuo – pajamos paprastai būną mažesnės.

Tendencinga apklausa. Tai vyksta, kai tyrėjai yra suinteresuoti. Pavyzdžiui, nuo trąšų gamintojų priklausantys tyrėjai gali parinkti geresnius sklypus, norėdami įrodyti trąšų poveikį derliui.

Imties dydžio nustatymas yra vienas iš svarbiausių tyrimo projekto uždavinių, kartu ir vienas iš sudėtingesnių. Pirma, reikia apibrėžti populiaciją. Antra, reikia išspręsti keletą įvairių problemų: turimas tyrimo finansavimas gali turėti lemiamos įtakos imties dydžiui; imties dydis priklauso nuo būsimos analizės tipo; nuo norimo pasiekti įvertinimo tikslumo; nuo tiriamų kintamųjų skaičiaus; nuo tiriamos populiacijos vienalytiškumo ir t. t.

Matematiškai imties dydis yra funkcija, priklausanti nuo kelių argumentų: norimo pasiekti įverčių tikslumo, populiacijos sklaidos ir norimo patikimumo.

Socialiniuose ir rinkos tyrimuose paprastai naudojamas 95 proc. ar rečiau 99 proc. *patikimumo (pasikliovimo) lygmuo*.

Imties dydį galima skaičiuoti pagal formulę

$$n = \frac{z^2 * s^2}{\Delta^2}, \qquad (2.1)$$

čia: Δ – tyrimo tikslumas (paklaida) procentais; z – konstanta, apibrėžiama pagal pasirinktą patikimumo lygmens reikšmę (patikimumo lygmenų konstantų z reikšmės pateiktos 2.3 lentelėje); s – standartinis nuokrypis

$$s = \sqrt{\Delta(100 - \Delta)} \tag{2.2}$$

2.3 lentelė. Patikimumo lygmenų konstantos

Patikimumo lygmuo	**Konstanta** z
60 %	±0,84
70 %	±1,03
80 %	±1,29
85 %	±1,44
90 %	±1,64
95 %	±1,96
97 %	±2,18
99 %	±2,58
99,7 %	±2,58

Pavyzdys. Nustatysime imties dydį, kai norime patikrinti, ar tikrai 35 proc. asmenų padaro eismo avarijas būdami neblaivūs. Šiuo atveju dispersija:

$$s^2 = 35 * (100 - 35) = 2275. \tag{2.3}$$

Matavimo paklaida $\pm$5 %, konstanta z = 1,96, kai patikimumo lygmuo 95 %. Tada

$$n = \frac{1{,}96^2 * 2275}{25} \approx 350. \tag{2.4}$$

Šiuo atveju imties dydis yra 350 respondentų.

Imties statistinė paklaida skaičiuojama pagal formulę

$$\Delta_{x\%} = \pm z\sqrt{\frac{pq}{n}}\,, \tag{2.5}$$

čia z – statistinė konstanta atitinkamam patikimumo lygmeniui; $p = q$ = 50 % – tiriamojo požymio pasirodymo/nepasirodymo tikimybė (mūsų atveju respondento patekimas/nepatekimas į imtį; atsitiktinėms imtims tikimybė lygi ½, t. y. 50 %); n – imties dydis – apklaustųjų skaičius.

1 pavyzdys. Apklausta 1000 respondentų. Patikimumo lygmuo 95 %. Imties statistinė paklaida:

$$\Delta = \pm 1{,}96 * \sqrt{\frac{50 * 50}{1000}} = \pm 1{,}96 * \sqrt{2{,}5} = \pm 3{,}1\% . \qquad (2.6)$$

Apskaičiuota statistinė paklaida taikoma visų imties viduje naudojamų kintamųjų reikšmėms. Pavyzdžiui, jei apklausos metu nustatėme, kad 17 proc. respondentų mokosi labai gerais ir puikiais pažymiais, tai ši reikšmė kinta nuo 13,9 proc. (17 % – 3,1 %) iki 20,1 proc. (17 % + 3,1 %).

2 pavyzdys. Kai populiacijos dydis žinomas, imties dydžiui nustatyti galima naudoti formulę [Schwarze, 1993]

$$n = \frac{N * 1{,}96^2 * p * q}{\varepsilon^2 * (N-1) + 1{,}96^2 * p * q}; \qquad (2.7)$$

čia: N – populiacijos dydis, ε – pageidaujamas tikslumas, paprastai ε = 5 %, $p = q$ = 50 % – tiriamojo požymio pasirodymo/nepasirodymo tikimybė (mūsų atveju respondento patekimas/nepatekimas į imtį; atsitiktinėms imtims tikimybė lygi ½, t. y. 50 proc.); 1,96 – atitinka standartizuoto normaliojo skirstinio 95 procentų patikimumo lygmenį (žr. 1.3 lentelę).

Tipinėje mokykloje mokosi apie 1200 mokinių. Norime ištirti, kiek mokinių reikėtų apklausti, norint nustatyti, kiek mokinių gauna dienpinigius. Patikimumo lygmuo 95 proc., pageidaujamas tikslumas 5 proc. Pagal duotą formulę randame n = 291.

3 pavyzdys. Kartais reikia nustatyti paklaidos dydį, kai imtis žinoma. Mūsų pavyzdyje reikėjo apklausti 291 mokinį, tačiau pavyko apklausti tik 200. Kokiu tikslumu pavyko apklausa?

Pagal mūsų formulę randame

$$\varepsilon = \sqrt{\frac{N * 1{,}96^2 * p * q}{n * (N-1) + 1{,}96^2 * p * q}} = 6{,}8\%. \qquad (2.8)$$

Taigi, jei nustatytume, kad 70 proc. apklaustųjų yra būdinga tam tikra savybė su 95 proc. patikimumu, galime sakyti, kad šia savybe pasižymi nuo 63,2 iki 76,8 proc. populiacijos.

Imties dydžiui skaičiuoti internete yra daug įvairių skaičiuoklių. Jas nesunkiai galime rasti interneto naršyklėje atlikę paiešką, pavyzdžiui „imties skaičiuoklė“ arba „imties dydžio skaičiuoklė“.

2.5.3. Anketos kodavimo schemos sudarymas. Duomenų kodavimas

Apklausos anketų rezultatai tolesnei analizei atlikti turi būti suvesti į kompiuterinę duomenų bazę. Anketos klausimai ir atsakymų variantai paprastai formalizuojami apibrėžiant kintamuosius duomenų bazėse (mūsų atveju PSPP). Visa tai atvaizduojama anketos kodavimo schemoje – lentelėje, kurioje įrašyti anketos klausimai, kintamųjų vardai ir atsakymų kintamųjų kodai. 2.4 lentelėje matome apklausos anketos apie kavos pirkimą kodavimo schemą.

2.4 lentelė. Anketos kodavimo schema

Anketos klausimas	**Kintamojo vardas duomenų bazėje**	**Kintamojo tipas**	**Atsakymų variantai**	**Atsakymo kodas**
1. Jūsų lytis?	*Lytis*	Nominalusis	Vyras Moteris	1 2
2. Jūsų amžius?	*Amžius*	Ranginis*	15–25 m. 26–35 m. 36–45 m. Daugiau kaip 45 m.	1 2 3 4
3. Jūsų mėnesinės pajamos?	*Pajamos*	Ranginis	Iki 800 Lt 801–1000 Lt 1001–1500 Lt 1501–2000 Lt Daugiau kaip 2000 Lt	1 2 3 4 5
4. Kokio dydžio pakuotę perkate? _____	*Pakuotės dydis*	Kiekybinis	Neribojama	Papildomai koduoti nereikia
5. Įvertinkite pagal 10 balų skalę šias kavos rūšis: Paulig Nescafe Jacobs Eduscho Folgers Merrild Elite Ar Mani Tshibo	 *R_A_vertin* *R_B_vertin* *R_C_vertin* *R_D_vertin* *R_E_vertin* *R_F_vertin* *R_G_vertin* *R_H_vertin* *R_I_vertin*	Ranginiai	Kiekvienam kintamajam nuo 1 iki 10	Papildomai koduoti nereikia

2.4 lentelės tęsinys

6. Kokias kavos rūšis dar žinote? _____	*Kitos_rūšys*	Nominalusis	Neribojama	Papildomai koduoti nereikia
7. Kur perkate kavą?	*Pirkimo_vieta_PC* *Pirkimo_vieta_SP* *Pirkimo_vieta_K*	Nominalusis	Prekybos centre Specializuotoje parduotuvėje Kitur	1 2 3

Pastaba. Šioje apklausoje kintamasis *Amžius* yra ranginio tipo. Paprastai amžius suprantamas kaip santykių skalės kiekybinis kintamasis. Ranginė skalė aprėpia mažiau informacijos nei santykių. Apklausos vykdytojai, tirdami priklausomybę nuo amžiaus grupių, turi būti atsargūs, nes respondentai gali priklausyti tik vienai amžiaus grupei, pavyzdžiui, dieninių studijų pirmo kurso studentai.

Skirtingo tipo klausimai yra koduojami trijų tipų kintamaisiais:

Uždarieji klausimai su vieno varianto atsakymais koduojami vienu kintamuoju (*Lytis, Amžius, Pakuotės_dydis*). Kintamojo kitimo skalės tipas gali būti bet koks (atitinkamai pavyzdyje: vardinė, ranginė, intervalinė).

Uždarieji klausimai su daugiavariančiais atsakymais koduojami keliais vienvariančiais kintamaisiais (*Pirkimo_vieta_PC, Pirkimo_vieta_SP, Pirkimo_vieta_K*). Kintamųjų skalė – tik vardinė.

Atvirieji klausimai su bet kuriuo skaičiumi atsakymų variantų koduojami vienu kintamuoju. Kintamojo skalė gali būti intervalinė (skaitmeniniams kintamiesiems *Pakuotės_dydis*) arba vardinė (nominaliniams kintamiesiems *Kitos_rūšys*).

Kiekybiniai kintamieji papildomai nekoduojami, nes jau yra išreikšti skaičiais.Tačiau jei kintamasis yra kokybinis, prieš pradedant statistinę analizę tokį kintamąjį reikia koduoti: tekstinė informacija keičiama į skaitmeninę. Pavyzdžiui: kokybinį kintamąjį *Lytis*, kurio galimi atsakymai yra „Moteris“ arba „Vyras“ galima pakeisti skaičiais 2 ir 1, kur 2 reiškia kad stebėjimas atitinka įrašą „Moteris“, o 1 – atitinkamai „Vyras“.

Ranginiai kintamieji išreikšti skaičiais/skaičiumi (pavyzdžiui, įvertinimas pagal dešimtbalę skalę arba užimamą vietą) irgi papildomai ne-

koduojami. Jei ranginis kintamasis turi tekstinę reikšmę („gerai“, „labai gerai“, ... arba „30–40 m.“, „41–50 m.“ ...), tai šios reikšmės koduojamos.

Tam, kad būtų galima išvengti klaidų, nustatant ranginių kintamųjų ryšį, koduojant ranginius kintamuosius, reikia, kad rangavimas būtų vieną kryptimi – nuo didžiausio rango iki mažiausio, arba atvirkščiai (žr. 2.5 lentelę).

2.5 lentelė. Ranginių kintamųjų kodavimas

Anketos klausimas	Kintamasis	Atsakymo variantai	Teisingas kodavimas
Jūsų pajamos	*Pajamos*	iki 100 Lt 101–500 Lt 501–1000 Lt 1001–2000 Lt	1 2 3 4
Jūsų išsilavinimas	*Išsilavinimas*	Aukštasis Vidurinis Pradinis	3 2 1

Sunkiau teisingai atlikti kintamųjų kodavimą, kai viename klausime nurodoma keletas atsakymo variantų. Pavyzdžiui, „Sužymėkite svarbos tvarka (nuo 1 – svarbiausia iki 3 – mažiausiai svarbu) jogurto savybes: „Skonis“, „Kaina“, „Ekologiškumas“.

Šiuo atveju kiekvienas respondentas nurodo kiekvienos savybės svarbą. Atsakymai įvedami kaip atskiri kintamieji: *Skonis*, *Kaina*, *Ekologiškumas*. Tačiau šiuo atveju rangavimo skalė atvirkštinė, nei parodyta 2.5 lentelėje. Todėl reikia atlikti perkodavimą: 1 vieta – 3, 2 vieta – 2, 3 vieta – 1.

Nominalių kintamųjų kodavimo tvarka neturi reikšmės, tačiau patariama kintamuosius *taip/ne* koduoti: *taip* – 1, *ne* – 0.

Reikia neužmiršti, kad koduojant kokybinius kintamuosius, kintamiesiems suteikiami skaitmeniniai kodai, o ne reikšmės: 2.5 lentelėje kintamojo *Pajamos* kodai niekaip neatspindi tikrųjų pajamų reikšmių. Kodavimas reikalingas suformuoti pirminės informacijos duomenų bazei (iš anketų atsakymų), kuri suvedama į kompiuterį ir analizuojama statistiniais ir kognityviniais metodais.

KONTROLINIAI KLAUSIMAI

2.1. Kokie yra pagrindiniai tyrimo projekto elementai?

2.2. Kokius žinote tiriamų procesų aprašymo modelius?

2.3. Kokios dalys įvardijamos apklausos anketoje?

2.4. Nuo ko priklauso atsakymų matavimo klaidos?

2.5. Kokias kintamųjų matavimų skales žinote?

2.6. Ką vadiname populiacija ir imtimi? Ar visada populiacija ir imtis skiriasi?

2.7. Kokie pagrindiniai imčių sudarymo tikslai?

2.8. Koks patikimumo lygmuo dažniausiai taikomas socialiniuose tyrimuose?

2.9. Kaip vedant duomenis į kompiuterį koduojami kiekybiniai kintamieji?

UŽDUOTYS

2.1. 2010 m. Lietuvoje buvo apie 370 tūkst. registruotų bedarbių. Reikia ištirti, su kokiomis pagrindinėmis problemomis jie susiduria, norėdami gauti pašalpas. Kiek bedarbių reikėtų apklausti, norint gauti 95 proc. patikimumo lygmens ir 5 proc. tikslumo atsakymus?

3. DUOMENŲ SUVEDIMAS Į KOMPIUTERĮ

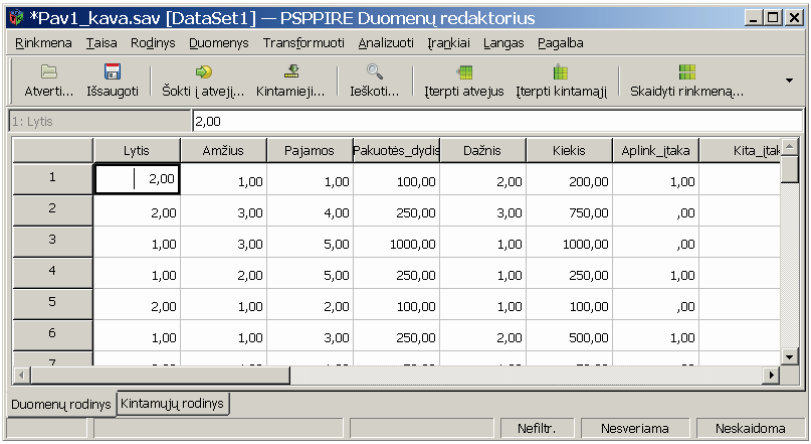
3.1. PSPP PROGRAMŲ PAKETO DUOMENŲ REDAKTORIUS

PSPP turi tris langus:

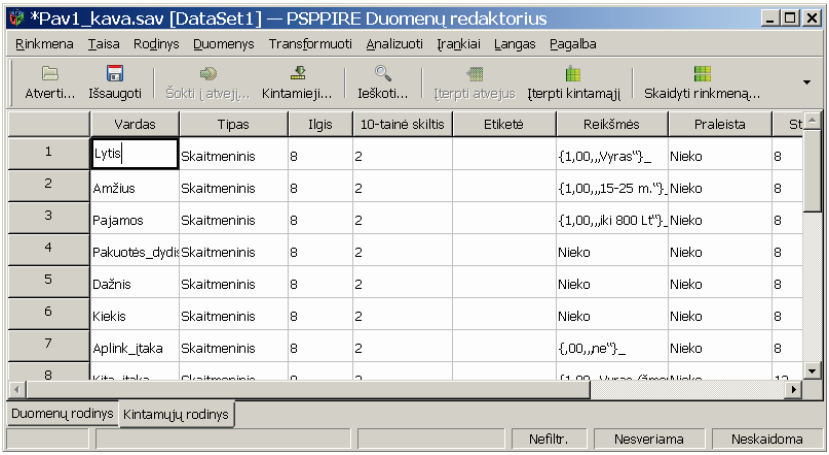
- Duomenų redaktorių (duomenų peržiūra ir redagavimas, duomenų analizė pasitelkus meniu);
- Sintaksės redaktorių (duomenų redagavimas ir analizė PSPP pasitelkus sintaksę);
- Rezultatų langą (pateikiami skaičiavimų rezultatai lentelių ir grafikų pavidalu).

Duomenų redaktorius

Duomenų redaktorius sukurtas elektroninės lentelės principu: eilutėse turime stebėjimų duomenis (atvejus), stulpeliuose – kintamuosius. Pasinaudosime turimu naudojant SPSS paketą sukurtu failu *Pav1_kava.sav*. PSPP aplinkoje atverkime šį failą taikydami komandas *Rinkmena* → *Atverti...* arba spragtelėję mygtuką *Atverti...* Spragtelėję atitinkamą auselę redaktoriaus lango kairiajame apatiniame kampe, galėsite matyti duomenų ir kintamųjų rodinius (žr. 3.1 pav., 3.2 pav.):



3.1 pav. PSPP duomenų redaktorius. *Duomenų rodinys*



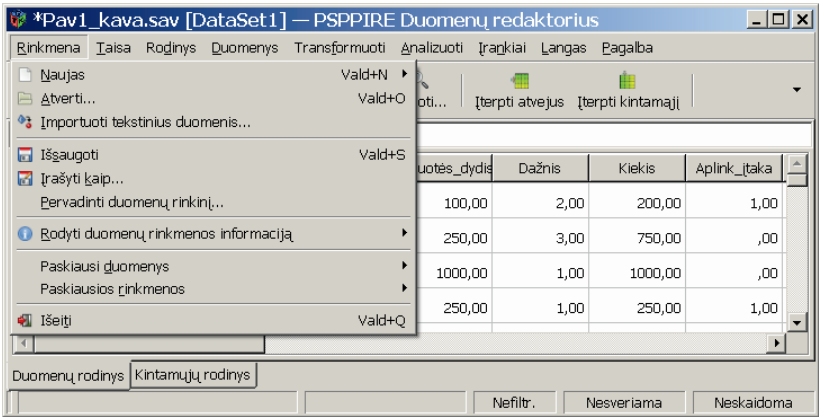
3.2 pav. PSPP duomenų redaktorius. *Kintamųjų rodinys*

Pridėti ar panaikinti (ištrinti) stebėjimo duomenis (atvejus, lentelės eilutes, įrašus) arba kintamuosius galima spragtelėjus dešiniuoju pelės mygtuku atitinkamai eilutės numerį ar stulpelio vardą. Kintamieji redaguojami kintamųjų rodinyje *Kintamųjų rodinys*, spragtelint kintamąjį ar pasirinktą jo savybių langelį. Panašiai galima redaguoti pasirinktų stebėjimo duomenų reikšmes duomenų rodinyje *Duomenų rodinys*.

Svarbu. Duomenys iš skaičiuoklės *OpenOffice Calc* į duomenų rodinį ir atvirkščiai gali būti **kopijuojami** tiesiai.

Meniu *Rinkmena*

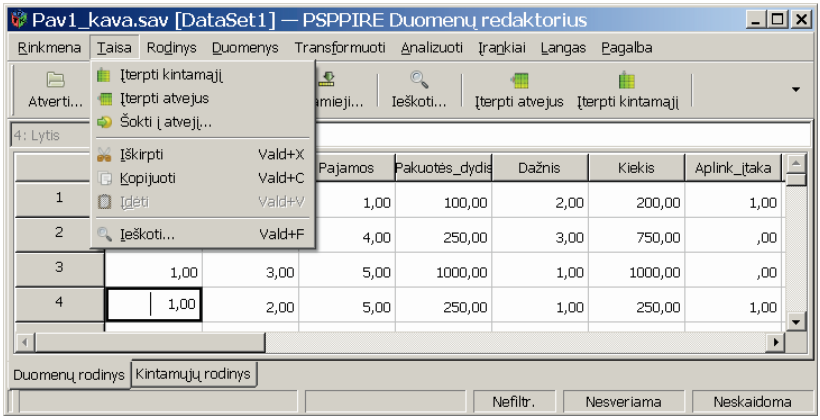
Meniu *Rinkmena* (žr. 3.3 pav.) komandomis galima atverti naują ar esamą duomenų ar sintaksės failą, importuoti tekstinius duomenis, išsaugoti failą, pervardyti failą, peržiūrėti failo informaciją. Taip pat galima atverti paskiausiai naudotus failus.



3.3 pav. Meniu *Rinkmena*

Meniu *Taisa*

Pasitelkęs meniu *Taisa* komandas (žr. 3.4 pav.) vartotojas gali įterpti kintamąjį ar stebėjimo duomenis (atvejus), eiti į norimus stebėjimo duomenis, iškirpti, kopijuoti ir įterpti duomenis. Ieškos operacija leidžia ieškoti pasirinktos kintamojo reikšmės.



3.4 pav. Meniu *Taisa*

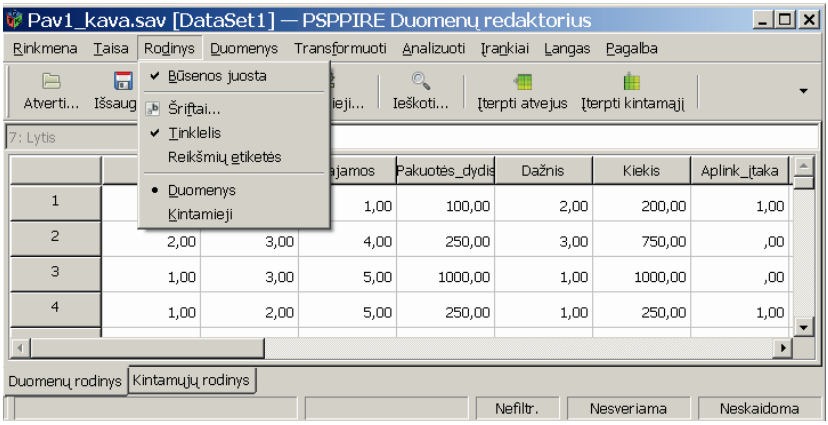
Meniu *Rodinys*

Pasitelkęs šio meniu komandas vartotojas gali rinktis iš duomenų *(Duomenų rodinys)* ir kintamųjų (*Kintamųjų rodinys)*, rodyti ar paslėpti

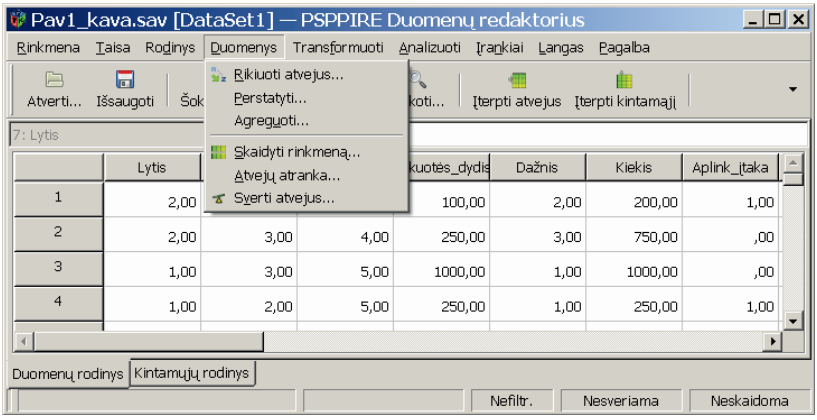
būsenos juostą (yra lango apačioje su informacija apie filtravimą, svėrimo būseną bei failo skaidymą), rodyti/slėpti tinklelį, keisti rodomų duomenų šriftus, rodyti kintamųjų reikšmes arba kintamųjų etiketes (žr. 3.5 pav.).

Meniu *Duomenys*

Šiame meniu (žr. 3.6 pav.) vartotojas gali rūšiuoti duomenis (pagal vieną ar kelis kintamuosius didėjančia ar mažėjančia tvarka), perkelti duomenis, sujungti duomenis, skaidyti failo duomenis, išrinkti duomenis pagal kriterijus ar pavyzdinius duomenis (nepasirinkti duomenys g.b. neaktyvūs ar pašalinti), sverti atvejus pagal pasirinktą dažnių kintamąjį.



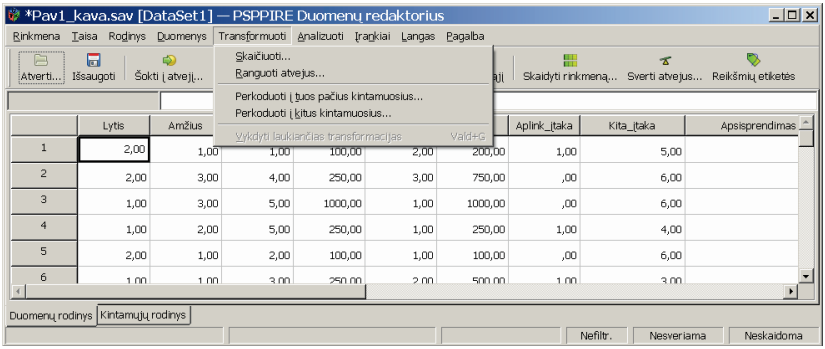
3.5 pav. Meniu *Rodinys*



3.6 pav. Meniu *Duomenys*

Meniu *Transformuoti*

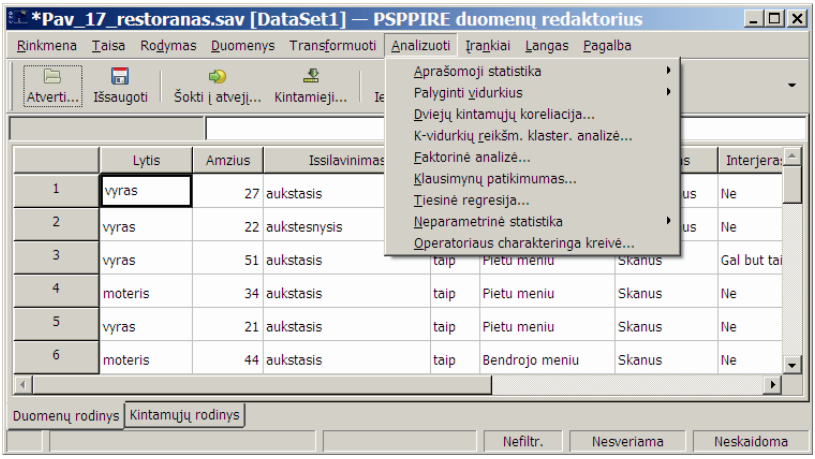
Meniu *Transformuoti* (žr. 3.7 pav.) vartotojas gali skaičiuoti naujus kintamuosius, ranguoti atvejus, perkoduoti kintamuosius.



3.7 pav. Meniu *Transformuoti*

Meniu *Analizuoti*

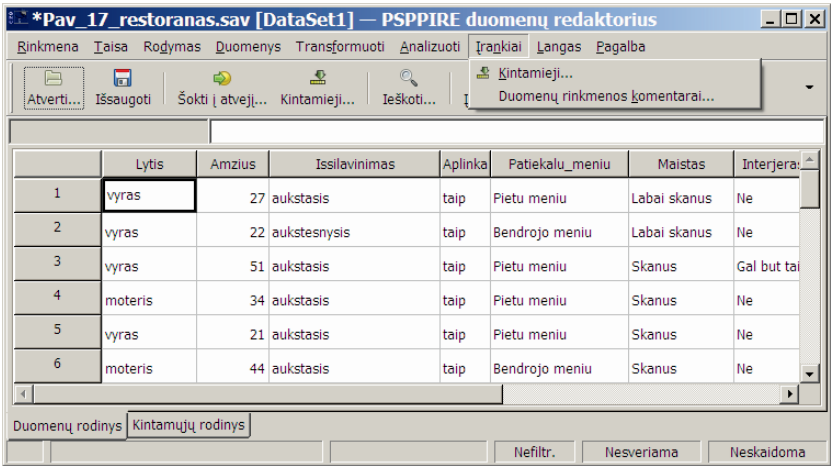
Tai bene svarbiausias PSPP meniu (žr. 3.8 pav.), kuriame vartotojas gali atlikti aprašomąją statistiką, palyginti vidurkius, skaičiuoti dviejų kintamųjų koreliaciją, faktorinę analizę, tikrinti klausimynų patikimumą, skaičiuoti tiesinę regresiją, atlikti neparametrinę statistiką, rasti operatoriaus charakteringą kreivę (angl. *ROC curve*).



3.8 pav. Meniu *Analizuoti*

Meniu *Įrankiai*

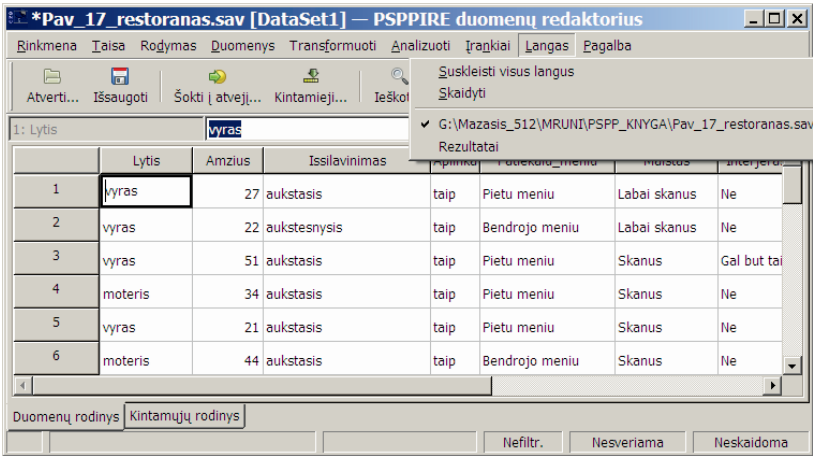
Vartotojas gali rasti informaciją apie duomenų failo kintamuosius, taip pat peržiūrėti ir tvarkyti duomenų failo komentarus (žr. 3.9 pav.).



3.9 pav. Meniu *Įrankiai*

Meniu *Langas*

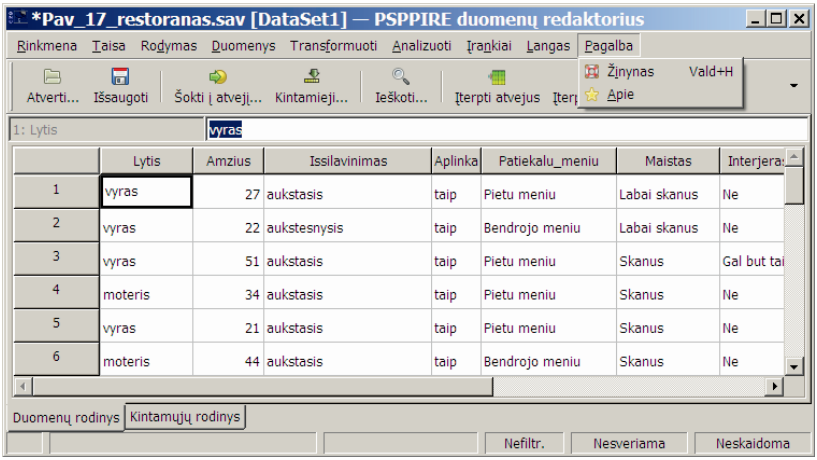
Šiame meniu (žr. 3.10 pav.) vartotojas gali rinktis duomenų redaktorių, sintaksės redaktorių ar rezultatų langą. Taip pat galima suskleisti visus langus ir skaidyti duomenų redaktoriaus langą.



3.10 pav. Meniu *Langas*

Meniu *Pagalba*

Meniu *Pagalba* (žr. 3.11 pav.) vartotojui prieinamas pagalbos žinynas ir informacija apie autorius bei licencijas.



3.11 pav. Meniu *Pagalba*

Duomenų suvedimas į kompiuterį – tai labai svarbus paruošiamasis statistinio duomenų apdorojimo etapas. Jis susideda iš nesugadintų (užpildytų, teisingai užpildytų) anketų atrankos, sugadintų ir neatsakytų anketų dalies nustatymo, anketos kodavimo schemos sudarymo, duomenų kodavimo bei duomenų bazės (duomenų failo) kūrimo.

Duomenys į failą gali būti įvedami tiesiai arba perkeliami iš įvairių šaltinių. Aptarsime

- duomenų tiesioginį įvedimą į PSPP duomenų bazę;
- duomenų įvedimą (skaitymą) iš PSPP/SPSS duomenų failų;
- duomenų įvedimą (importavimą) iš tekstinių failų.

Pastaba. Duomenys iš skaičiuoklės *OpenOffice Calc* į PSPP ir atvirkščiai gali būti kopijuojami / perkeliami tiesiai.

3.2. PSPP DUOMENŲ BAZĖS PILDYMAS

PSPP duomenų redaktorius

Duomenų redaktorius sukurtas elektroninės lentelės principu: eilutėse turime stebėjimų duomenis, stulpeliuose – kintamuosius (žr. 3.13 pav.).

Spragtelėję duomenų redaktoriaus lango apačioje atitinkamą auselę galime pasirinkti rodinius.

- *Duomenų rodinys*. Matome duomenų reikšmes arba duomenų apibūdinimo žymes.
- *Kintamųjų rodinys*. Matome informaciją apie kintamuosius: kintamojo ir kintamojo žymenos pavadinimą, duomenų tipą, matavimo skalę, praleistas reikšmes ir kt.

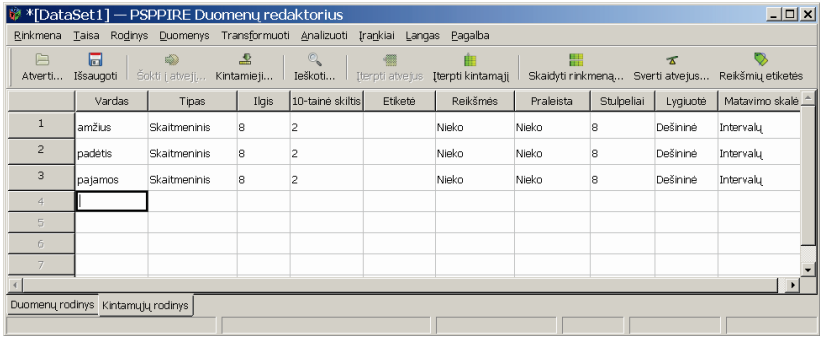
Abiejų rodinių languose galima faile esančią informaciją papildyti, keisti arba pašalinti.

Skaitmeninių duomenų įvedimas

Nedidelius duomenų failus galima tiesiogiai įvesti į PSPP duomenų redaktorių. Taip pat PSPP duomenų redaktorius gali būti naudojamas ir didelių failų duomenims redaguoti.

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje (žr. 3.12 pav.).

Apibrėšime kintamuosius, kuriuos naudosime mūsų pavyzdyje: *amžius*, *šeiminė padėtis* ir *metinės pajamos*.



3.12 pav. Kintamųjų vardai *Kintamųjų rodinys* lange

- Pirmame stulpelyje pirmoje eilutėje įrašykite *amžius*, antroje eilutėje – *padėtis*, trečioje eilutėje – *pajamos*.

Kintamojo vardui keliami šie reikalavimai:

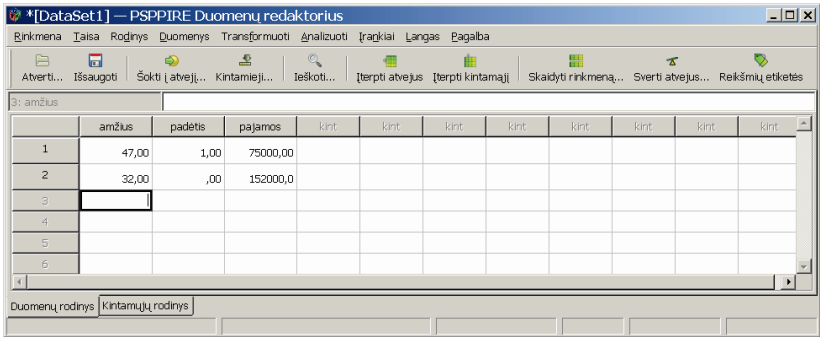
- Kintamojo vardas turi prasidėti raide. Kiti simboliai gali būti lotyniškos raidės, skaitmenys, taškas, simboliai @, #, _, $. Neleidžiami tarpai ir specialūs simboliai, pvz., !, ?, ‘, *.
- Paskutinis kintamojo vardo simbolis negali būti taškas arba pabraukimas.
- Didžiosios ir mažosios raidės suprantamos vienodai.
- Kintamojo vardo ilgis iki 64 simbolių.

Jei kintamiesiems vardas neįvedamas, PSPP automatiškai sukuria unikalius vardus, pvz., *Kint00000*, *Kint00001* ir t. t. Patariama tokių kintamųjų vengti, nes jie neatspindi jokios kintamųjų prasmės.

Naujiems kintamiesiems automatiškai suteikiamas *Skaitmeninis* tipas. Šis tipas gali būti keičiamas vedant, pvz., tekstinius kintamuosius, tačiau mes vedame koduotus duomenis, todėl jų kodai turi skaitmeninį tipą.

Pradedant kurti naują duomenų bazę, matavimo skalė *Matavimo skalė* automatiškai pasirenkama *Intervalų*. Ranginiams kintamiesiems matavimo skalę reikia pakeisti į *Rangų*, o vardiniams kintamiesiems – į *Pavadinimų*:

- Spragtelėkite stulpelio *Matavimo skalė* atitinkamame langelyje. Langelio dešinėje pamatysite juodą trikampėlį.
- Spragtelėkite juodą trikampėlį.
- Atsivėrusiame sąraše pasirinkite reikiamą matavimo skalę.
- Spragtelėkite duomenų peržiūros *Duomenų rodinys* auselę – įvesime kintamųjų reikšmes:

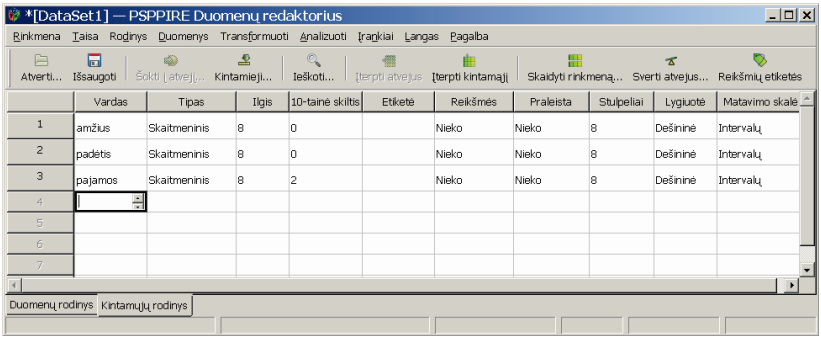


3.13 pav. Kintamųjų reikšmės *Duomenų rodinys* lange

- Stulpelyje *amžius* įrašykite 47.
- Stulpelyje *padėtis* įrašykite 1.
- Stulpelyje *pajamos* įrašykite 75000.
- Perkelkite teksto žymeklį į antrą eilutę ir atitinkamuose langeliuose įrašykite 32, 0 ir 152000.

Kaip matome, stulpeliuose *amžius* ir *padėtis* duomenys vaizduojami dešimtainiu formatu. Jei reikia, galime trumpeninę šių kintamųjų dalį paslėpti (žr. 3.14):

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Eilučių *amžius* ir *padėtis* stulpelyje *10-tainė skiltis* įrašykite 0.

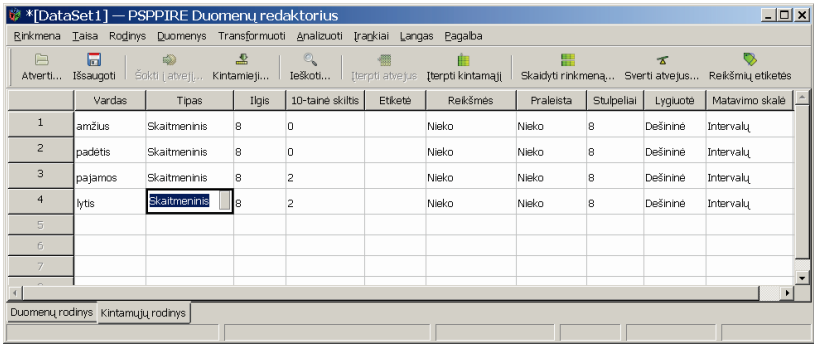


3.14 pav. Koduoti kintamieji *amžius* ir *padėtis* – rodomi be trumpeninės dalies

Tekstinių duomenų įvedimas

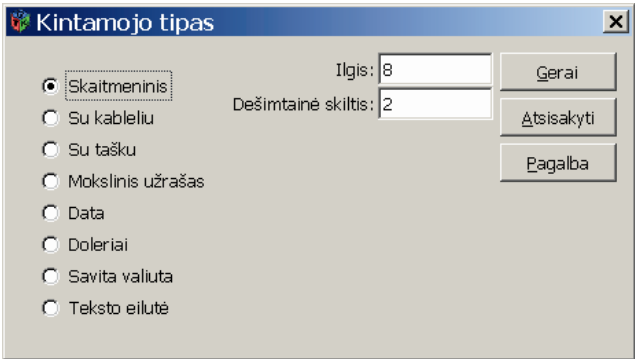
Šalia skaitmeninių duomenų į PSPP duomenų redaktorių įveskime ir tekstinius – teksto eilutes.

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Pirmos tuščios eilutės pirmame langelyje įrašykite naujo kintamojo vardą *lytis*.
- Spragtelėkite naujojo kintamojo langelyje *Tipas*. Langelio *Tipas* dešinėje atsiranda pilkas mygtukas (žr. 3.15 pav.).



3.15 pav. Mygtukas kintamojo *lytis* tipui pasirinkti

- Spragtelėkite pilką mygtuką langelio *Tipas* dešinėje. Pamatysite kintamojo tipo pasirinkimo dialogo langą (žr. 3.16 pav.):



3.16 pav. Kintamojo tipo pasirinkimo langas

- Pasirinkite kintamojo tipą *Teksto eilutė* ir spragtelėkite *Gerai* mygtuką.

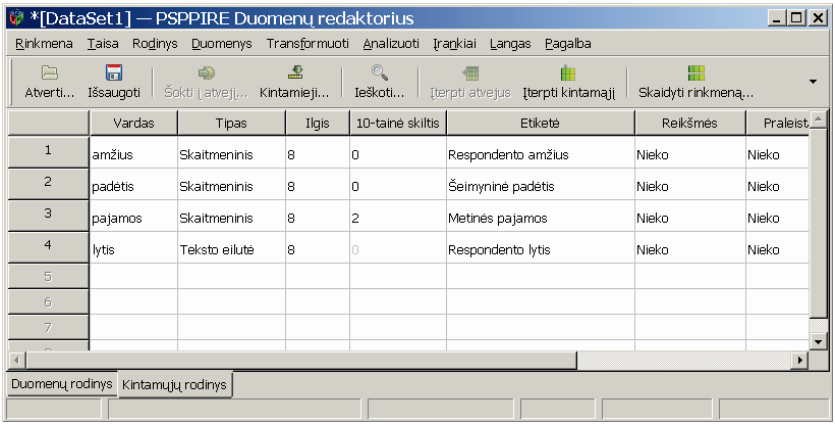
Duomenų aprašymas

Kartu su kintamųjų tipų aprašymu galima papildomai aprašyti kintamųjų vardus ir duomenų reikšmes. Šie aprašymai yra naudojami statistikos ataskaitose ir diagramose.

Kintamųjų žymenos *Etiketė*

Kintamųjų aprašymui naudojamos žymenos *Etiketė*. Paprastai tai yra platesnis kintamojo vardo variantas. Žymenos gali būti iki 255 simbolių ilgio.

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus apačioje (žr. 3.17 pav.).
- Eilutės *amžius* stulpelyje *Etiketė* įrašykite „Respondento amžius“.
- Eilutės *padėtis* stulpelyje *Etiketė* įrašykite „Šeimyninė padėtis“.
- Eilutės *pajamos* stulpelyje *Etiketė* įrašykite „Metinės pajamos“.
- Eilutės *lytis* stulpelyje *Etiketė* įrašykite „Respondento lytis“.

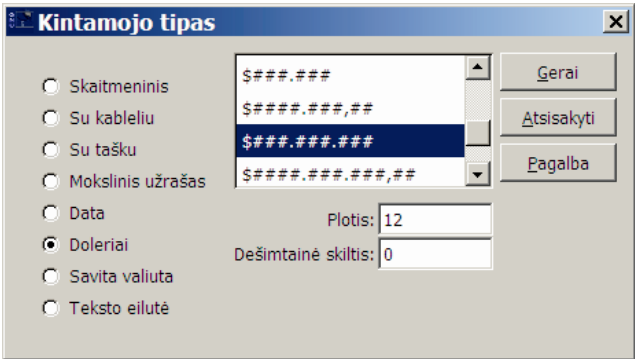


3.17 pav. Kintamųjų žymenos *Etiketė* stulpelyje

Kintamųjų tipo ir formato keitimas

Stulpelyje *Tipas* matome kiekvieno kintamojo tipą einamuoju momentu. Dažniausiai tai skaitmeniniai ar tekstiniai kintamieji. PSPP palaiko ir kitokius kintamųjų formatus. Mūsų pavyzdyje kintamasis *pajamos* yra skaitmeninio tipo.

- Spragtelėkite eilutės *pajamos* langelį *Tipas*. Spragtelėkite šiame langelyje atsiradusį pilką mygtuką. Bus atvertas kintamojo tipo *Kintamojo tipas* dialogo langas (žr. 3.18 pav.):



3.18 pav. Kintamųjų tipo pasirinkimo langas

- Spragtelėkite *Doleriai*
- Pasirinkite valiutos formatą $###.###.###
- Spragtelėkite *Gerai*.

Skaitmeninių kintamųjų žymenos

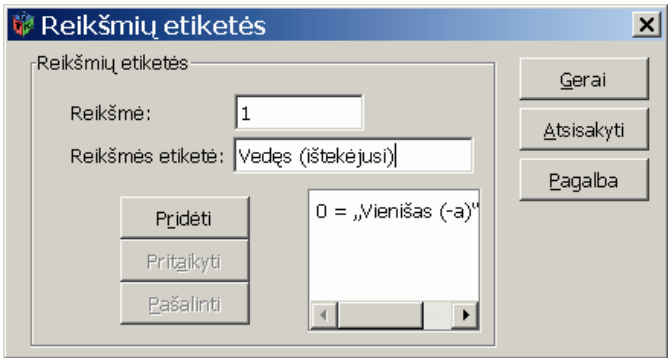
Reikšmių žymenos leidžia aprašyti skaitmeninių kintamųjų tam tikras reikšmes teksto eilutėmis. Mūsų pavyzdyje kintamasis *padėtis* turi dvi reikšmes: 0, tai reiškia, kad subjektas vienišas (nevedęs, netekėjusi) ir 1, tai reiškia, kad subjektas vedęs (ištekėjusi).

- Spragtelėkite eilutės *padėtis* langelį *Reikšmės*. Tada spragtelėkite šio langelio dešinėje esantį mygtuką. Atsivers reikšmių žymenų *Reikšmių etiketės* dialogo langas (žr. 3.18 pav.).

Reikšmė – skaitmeninio kintamojo reikšmė.

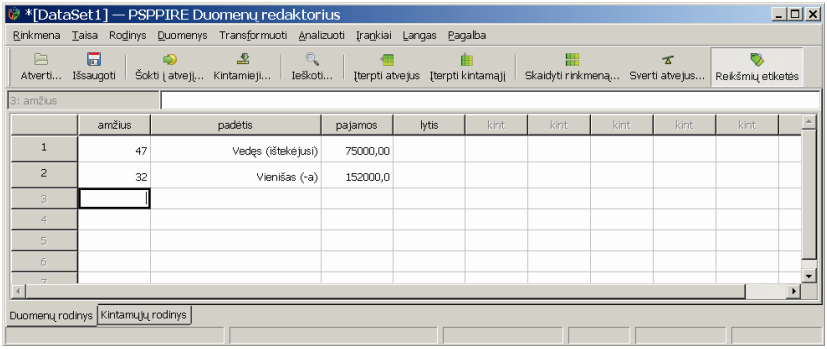
Reikšmės etiketė – tekstinė žymena, priskiriama pažymėtai skaitmeninio kintamojo reikšmei.

- Įrašykite „0“ *Reikšmė:* laukelyje.
- Įrašykite „Vienišas (-a)“ *Reikšmės etiketė:* laukelyje.
- Spragtelėkite *Pridėti,* norėdami šią žymeną priskirti sąrašui.
- Įrašykite „1“ *Reikšmė:* laukelyje.
- Įrašykite „Vedęs (ištekėjusi)“ *Reikšmės etiketė:* laukelyje.
- Spragtelėkite *Pridėti,* norėdami šią žymeną priskirti sąrašui.
- Spragtelėkite *Gerai*.



3.19 pav. Reikšmių žymenų pasirinkimo langas

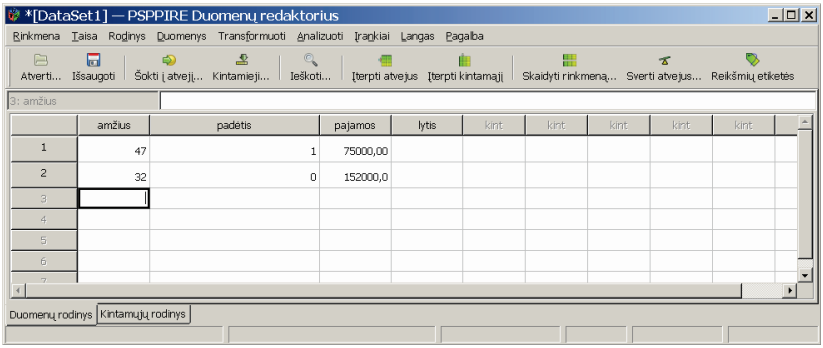
PSPP duomenų redaktoriuje mūsų pavyzdys atrodo (žr. 3.20 pav.):



3.20 pav. Duomenų reikšmių peržiūros *Duomenų rodinys* langas su kintamojo *padėtis* reikšmių žymenomis

Mūsų atveju reikšmių žymenos *Reikšmių etiketės* yra aktyvios.

Per meniu *Rodinys* → *Reikšmių etiketės* arba mygtuku *Reikšmių etiketės*  galima reikšmių žymenas *Reikšmių etiketės* padaryti neaktyvias (žr. 3.21 pav.):

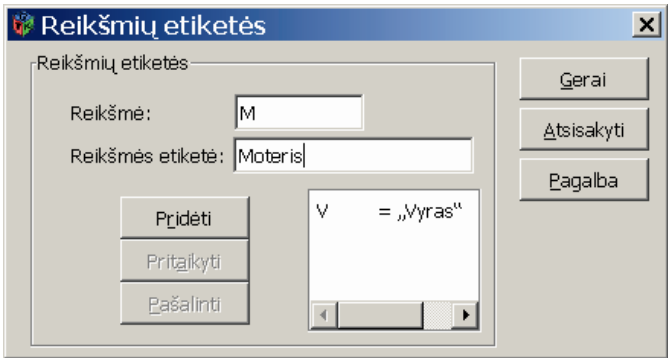


3.21 pav. Duomenų reikšmių peržiūros *Duomenų rodinys* langas be kintamojo *padėtis* reikšmių žymenų

Tekstinių kintamųjų žymenos

Kartais patogu tekstinius kintamuosius papildomai aprašyti. Pavyzdžiui, mūsų duomenyse *lytis* gali būti pažymėta raidėmis V ir M. Galėtume pasirinkti žymenas *Vyras* ir *Moteris*, atitinkamai žyminčias kintamuosius V ir M (žr. 3.22 pav.).

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Spragtelėkite eilutės *lytis* langelį *Reikšmės*. Spragtelėkite langelio dešinėje atsiradusį mygtuką. Atsivers reikšmių žymenų *Reikšmių etiketės* dialogo langas.
- Įrašykite „V“ į laukelį *Reikšmė:* ir „Vyras“ į laukelį *Reikšmės etiketė:*.
- Spragtelėkite *Pridėti*.
- Įrašykite „M“ į laukelį *Reikšmė:* ir „Moteris“ į laukelį *Reikšmės etiketė:*.
- Spragtelėkite *Pridėti*.



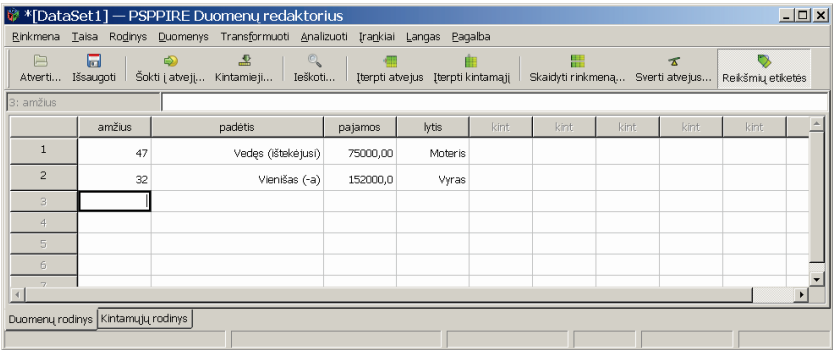
3.22 pav. Reikšmių žymenų pasirinkimo langas

- Spragtelėkite *Gerai*. Išsaugosite pakeitimus ir grįšite į duomenų redaktorių.

Dėmesio. Tekstiniuose kintamuosiuose didžiosios ir mažosios raidės suprantamos kaip skirtingos, todėl m yra ne tas pats kaip M.

Reikšmių žymenų naudojimas duomenims įvesti

- Spragtelėkite duomenų peržiūros *Duomenų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Pirmoje eilutėje pažymėkite langelį *lytis*.
- Įrašykite V.
- Analogiškai kitoje eilutėje įrašykite M.



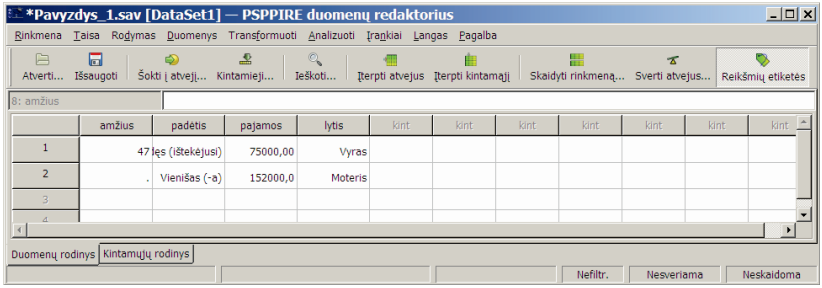
3.23 pav. Reikšmių žymenų naudojimas reikšmėms įvesti

Pagal tai, ar *Reikšmių etiketės* yra aktyvios, ar ne, stulpelyje *lytis* matome reikšmes *Moteris* ir *Vyras* ar M ir V. 3.23 pav. *Reikšmių etiketės* yra aktyvios.

Praleistų (Praleista) duomenų tvarkymas

Praleisti ar neteisingai suvesti duomenys pasitaiko pernelyg dažnai, kad juos galėtume ignoruoti. Kaip atsiranda tokie duomenys? Apklausos respondentai gali tam tikrus klausimus ignoruoti, gali nežinoti atsakymų, gali atsakyti netinkama forma. Jei tokių atvejų neidentifikuosime ar neatmesime, mūsų apklausos rezultatai gali nebūti teisingi.

Skaitmeniniams duomenims tušti ar neteisingo formato langeliai yra paverčiami sisteminiais praleistais, pažymint juos vienu tašku (žr. 3.24 pav.).



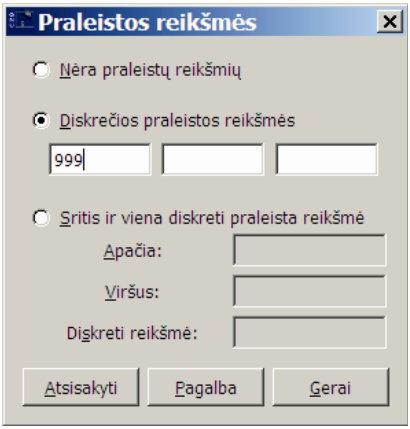
3.24 pav. Praleistos reikšmės pažymėtos taškais

Analizuojant duomenis nustatyti, kodėl reikšmė yra praleista (trūkstama), gali būti labai svarbu. Pavyzdžiui, gali būti svarbu nustatyti, kurie respondentai atmetė klausimus sąmoningai, o kurie tiesiog neatsakė į netinkamus klausimus.

Praleistos skaitmeninių kintamųjų reikšmės

- Spragtelėkite kintamųjų peržiūros *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Spragtelėkite eilutės *amžius* langelį *Praleista*. Spragtelėkite langelio dešinėje esantį mygtuką. Atsivers praleistų reikšmių *Praleistos reikšmės* dialogo langas (žr. 3.25 pav.).

Čia galite pažymėti iki trijų skirtingų praleistų reikšmių arba pažymėti reikšmių intervalą plius vieną papildomą atskirą reikšmę.

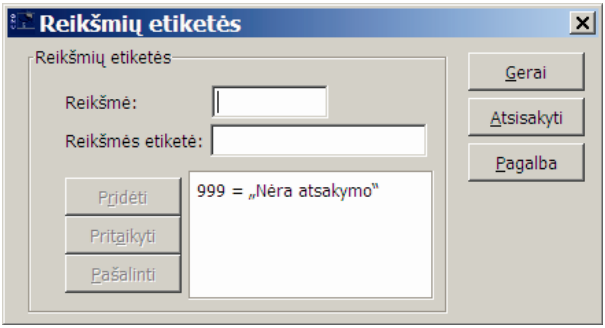


3.25 pav. Praleistų reikšmių *Praleistos reikšmės* dialogo langas

- Pasirinkite *Diskrečios praleistos reikšmės.*
- Įrašykite „999“ į pirmąjį langelį. Kitus palikite tuščius.
- Paspauskite *Gerai* ir grįžkite į PSPP duomenų redaktorių.

Dabar, kai pasižymėjome praleistų duomenų reikšmes, galime jas pritaikyti praleistų duomenų žymenose.

- Spragtelėkite eilutės *amžius* langelį *Reikšmės*. Spragtelėkite langelio dešinėje esantį mygtuką. Atsivers reikšmių žymenų *Reikšmių etiketės* dialogo langas (žr. 3.26 pav.).
- Įrašykite „999“ į laukelį *Reikšmė:*.
- Įrašykite „Nėra atsakymo“ į laukelį *Reikšmės etiketė:*.



3.26 pav. Reikšmių žymenų *Reikšmių etiketės* dialogo langas

- Spragtelėkite *Pridėti*
- Spragtelėkite *Gerai*, išsaugokite pakeitimus ir grįžkite į PSPP duomenų redaktorių.

Praleistos tekstinių kintamųjų reikšmės

Praleistos reikšmės tekstiniams kintamiesiems tvarkomos panašiai kaip ir skaitmeniniams kintamiesiems, tik tušti tekstinių kintamųjų laukai nelaikomi sisteminiais – praleistais, dažniau jie interpretuojami kaip tuščia eilutė.

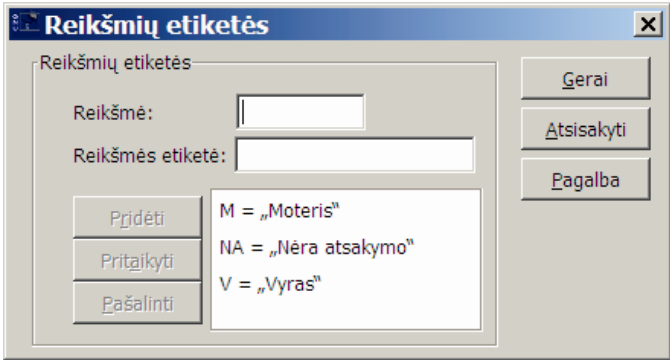
- Spragtelėkite *Kintamųjų rodinys* auselę duomenų redaktoriaus lango apačioje.
- Spragtelėkite eilutės *lytis* langelį *Praleista*. Spragtelėkite langelio dešinėje atsiradusį mygtuką. Atsivers praleistų reikšmių *Praleistos reikšmės* dialogo langas.
- Pasirinkite *Diskrečios praleistos reikšmės*
- Įrašykite *NA* į pirmąjį teksto langelį.

Dėmesio. Tekstiniuose kintamuosiuose didžiosios ir mažosios raidės traktuojamos kaip skirtingos, todėl *na* yra ne tas pats, kaip *NA*.

- Spragtelėkite *Gerai*, pakeitimų išsaugojimui ir grįžkite į PSPP duomenų redaktorių.

Dabar galime pridėti praleistoms reikšmėms žymenas.

- Spragtelėkite eilutės *lytis* langelį *Reikšmės*. Spragtelėkite langelio dešinėje esantį mygtuką. Atsivers reikšmių žymenų *Reikšmių etiketės* dialogo langas (žr. 3.27 pav.).
- Įrašykite „NA“ į laukelį *Reikšmė:*.
- Įrašykite „Nėra atsakymo“ į laukelį *Reikšmės etiketė:*.
- Spragtelėkite *Pridėti*.

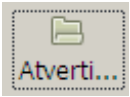


3.27 pav. Reikšmių žymenų *Reikšmių etiketės* dialogo langas

- Spragtelėkite *Gerai*, išsaugokite pakeitimus ir grįžkite į PSPP duomenų redaktorių.

Duomenų skaitymas iš PSPP/SPSS duomenų failų

Paketų PSPP ir SPSS sintaksės ir duomenų failai yra tarpusavyje suderinami. Atverti PSPP/SPSS duomenų failą galima meniu komandomis *Rinkmena* → *Atverti...* arba failų atvėrimo *Atverti...* mygtuko pagalba (žr. 3.28 pav.):

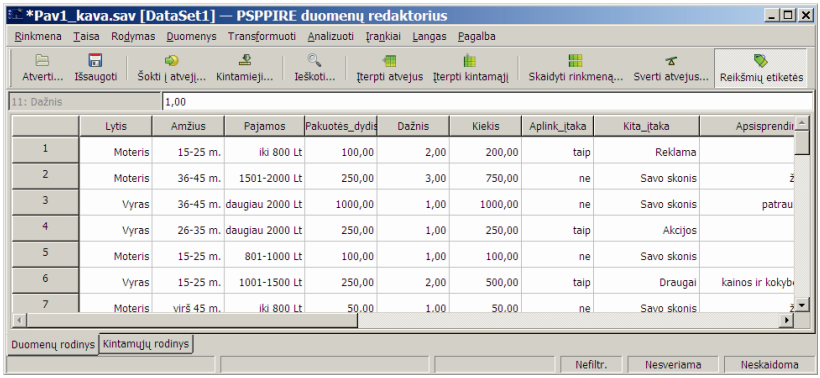


3.28 pav. Failo atvėrimo mygtukas *Atverti...*

Abiem atvejais pateikiamas failo atvėrimo dialogas.

Numatytieji PSPP duomenų failai turi plėtinį **.sav*, o sintaksės failai **.sps*.

3.29 pav. turime atvertą failą *Pav1_kava.sav*



3.29 pav. Failas *Pav1_kava.sav* PSPP duomenų redaktoriuje

PSPP duomenų failai sudaryti iš kintamųjų vardų (stulpeliai) ir kintamųjų reikšmių (eilutės). Mūsų pavyzdyje eilutėse turime respondentų atsakymus į klausimus, surašytus stulpeliuose.

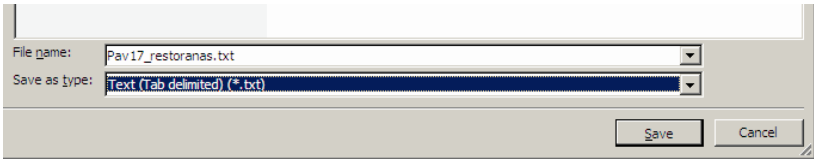
3.3. DUOMENŲ ĮKĖLIMAS IŠ KITŲ ŠALTINIŲ

Be tiesioginio duomenų suvedimo į PSPP duomenų redaktorių ir naujų ar esančių failų **.sav* atvėrimo, PSPP taip pat siūlo duomenų įkėlimą (importavimą) iš kitų šaltinių apriboto (angl. *delimited*) teksto formatu.

Duomenų įkėlimas iš skaičiuoklės failo

Svarbu. Iš *Open Office Calc* skaičiuoklės duomenis į PSPP galima kopijuoti tiesiogiai (ir atvirkščiai – iš PSPP į *Open Office Calc*).

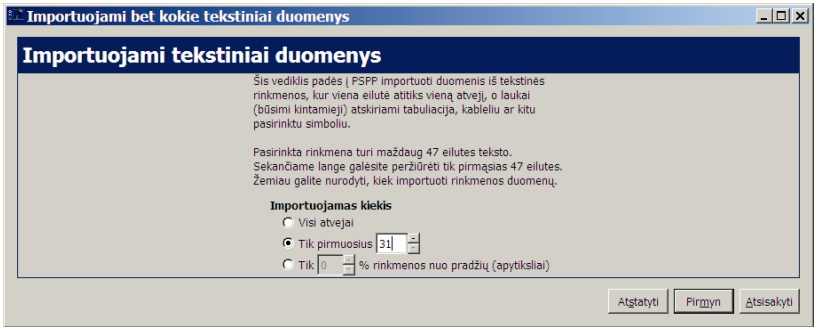
Jei turime *MS Excel* failą, pirma jį reikia išsaugoti tekstiniu formatu (žr. 3.30 pav.). Rekomenduojama tekstiniam failui duoti kitą vardą, negu turėjo *MS Excel* failas:



3.30 pav. Skaičiuoklės failo išsaugojimas **.txt* formatu

Norėdami įkelti duomenis iš tekstinio failo *Pav17_restoranas.txt* į PSPP, atlikite komandas:

- *Rinkmena* → *Importuoti tekstinius duomenis...*
- Dialogo lange *Importuoti bet kokius tekstinius duomenis* pasirinkite failo tipą *Visos rinkmenos*
- Pasirinkite falą *Pav17_restoranas.txt*
- Paspauskite mygtuką *Atverti…* Pateikiamas dialoginis langas (žr. 3.31 pav.)



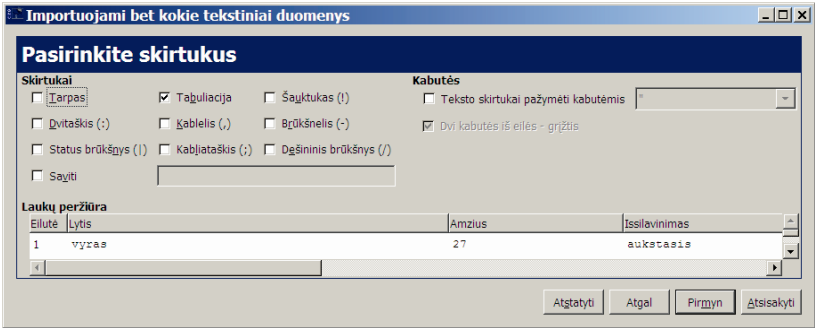
3.31 pav. Importuojamų duomenų kiekio pasirinkimas

- Jei *Excel* stulpelių vardai netenkina PSPP kintamųjų vardų sudarymo taisyklių, jie pertvarkomi pagal reikalavimus ir išsaugomi kaip PSPP kintamųjų vardai. Jei norime importuoti dalį skaičiuoklės duomenų, langelyje *Tik pirmuosius* pažymėkime reikiamų *Excel* eilučių kiekį. Mūsų atveju reikia įkelti 31 eilutę.
- Spragtelėkite mygtuką *Pirmyn.*



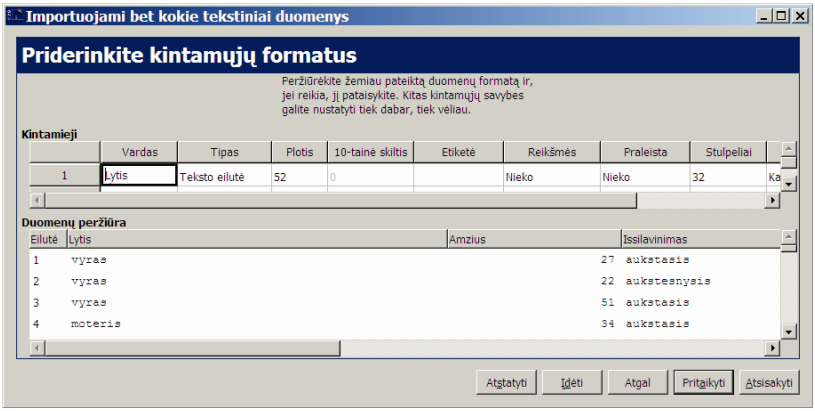
3.32 pav. Importuotinų duomenų pasirinkimas

- Pasirinkite pirmąją importuotinų duomenų eilutę: kadangi pirmojoje failo eilutėje yra kintamųjų vardai, reikia rinktis antrąją failo eilutę (žr. 3.32 pav.).
- Pažymėkite parinktį ☑ *Virš pasirinktosios eilutės yra kintamųjų vardai.*
- Spragtelėkite mygtuką *Pirmyn.*



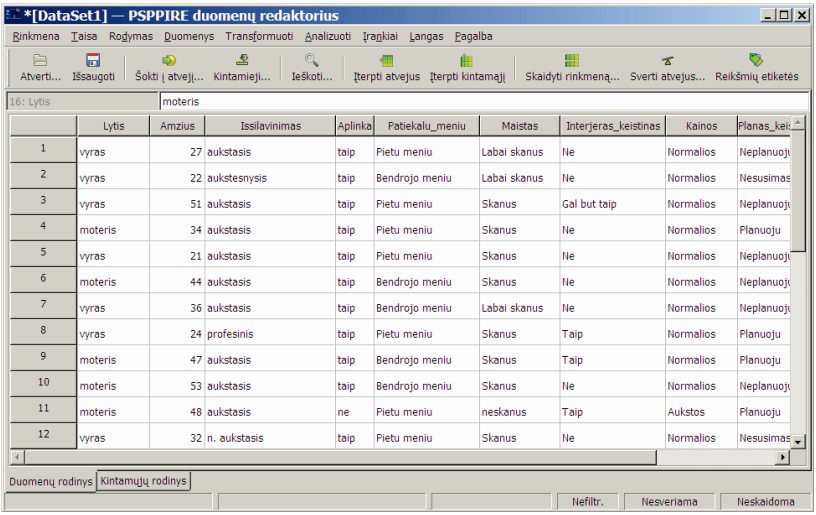
3.33 pav. Skirtukų tarp duomenų pasirinkimas

- Mums tinka skirtukas ☑ *Tabuliacija*. Spragtelėkite mygtuką *Pirmyn* (žr. 3.33 pav.).



3.34 pav. Kintamųjų formatų priderinimas

- Jei duomenų formatas netinka, galima spragtelėti mygtuką *Atgal* ir pasirinkus kitus duomenų skirtukus, priderinti kintamųjų formatą (žr. 3.34 pav.).
- Mūsų atveju kintamųjų formatai parinkti tinkamai. Spragtelėkite mygtuką *Pritaikyti*.



3.35 pav. Importuoti duomenys iš *Pav17_restoranas.txt* failo

Duomenis dabar matome PSPP duomenų redaktoriuje. Stulpelių vardai tapo kintamųjų vardais (žr. 3.35 pav.).

KONTROLINIAI KLAUSIMAI

3.1. Kokiais būdais į PSPP duomenų bazę galima suvesti tyrimo duomenis?

3.2. Kokie reikalavimai keliami kintamojo vardui PSPP programoje?

3.3. Ar importuojant duomenis iš skaičiuoklės į PSPP kintamųjų vardai pakinta?

3.4. Ar galima įkelti į PSPP duomenų bazę skaičiuoklės duomenis komandomis *Kopijuoti*/*Įdėti*?

4. PIRMINĖ STATISTINĖ DUOMENŲ ANALIZĖ

Statistinio tyrimo metu surinkti duomenys apie kiekvieną objekto vienetą – didžiulis kiekis medžiagos, kuri dažnai neteikia išsamesnės informacijos apie patį objektą. Medžiagą būtina susisteminti, susumuoti ir apdoroti. Šio proceso pradžioje kiekvieną kintamąjį (požymį) reikia nagrinėti atskirai – *vienmatė* statistinė analizė. Tik tuomet bus galima išsiaiškinti būdingus statistinės visumos bruožus, nustatyti dėsningumus. Šiai analizei atlikti naudojami *aprašomosios statistikos* metodai. Aprašomosios statistikos metodus sudaro:

- duomenų grupavimas;
- dažnių lentelės;
- statistinių charakteristikų (duomenų padėties ir sklaidos charakteristikų) skaičiavimas;
- grafinis stebėjimų vaizdavimas.

Sudaromos kintamųjų dažnių lentelės, kuriose pateikiami nominalinėje arba (esant nedideliam kategorijų skaičiui) rangų skalėje išmatuotų kintamųjų dažniai.

Be vienmačių dažnių lentelių, sudaromos ir porinės (gali būti ir su daugiau kintamųjų) dažnių lentelės. Jos parodo tarp kintamųjų esančius ryšius. Porinės dažnių lentelės sudaromos nominaliniams ir ranginiams kintamiesiems. Kiekybiniai kintamieji porinėms lentelėms sudaryti nenaudojami.

Aprašomosios statistikos metodų taikymo priklausomybė nuo kintamųjų tipo pateikta 4.4 skyriuje (žr. 4.1 lentelę).

4.1. DUOMENŲ KEITIMAS IR ATRANKA

Pradiniai duomenys ne visada yra tokie, kokius norime matyti ataskaitose ar toliau analizuoti. Mums gali prireikti:

- sukurti kokybinius (kategorinius) kintamuosius iš kiekybinių (intervalinių) kintamųjų;

- sujungti keletą atsakymų į vieną;
- sukurti naują kintamąjį, kuris yra gaunamas iš jau esančių kintamųjų naudojant įvairias matematines funkcijas.

Naudosime duomenų failą *Apklausa_men_pajamos.sav*.

Kokybinių (kategorinių) kintamųjų sukūrimas iš kiekybinių (intervalinių) kintamųjų

Faile *Apklausa_men_pajamos.sav* kintamasis *Men_pajamos* yra kiekybinis (intervalų). Esant dideliam atsakymų kiekiui patogu pajamų dydžius sugrupuoti į laisvai apibrėžtus intervalus – kategorijas. Taip iš kiekybinio kintamojo *Men_pajamos* gausime kokybinį (kategorinį) kintamąjį *Men_pajamos_kat*. Šis kintamasis sveikais skaičiais nuo 1 iki 11 atvaizduos pajamų kategorijas: mažiau už 300, 300–599; 600–899; 900–1199; 1200–1499; 1500–1799; 1800–2099; 2100–2399; 2400–2699; 2700–2999; 3000 ir daugiau.

Sukursime kokybinį kintamąjį *Men_pajamos_kat*:

- PSPP duomenų redaktoriuje pasirinkite: *Transformuoti* → *Perkoduoti į kitus kintamuosius...* Atveriamas *Perkoduoti į kitus kintamuosius* dialogo langas (žr. 4.1 pav.).

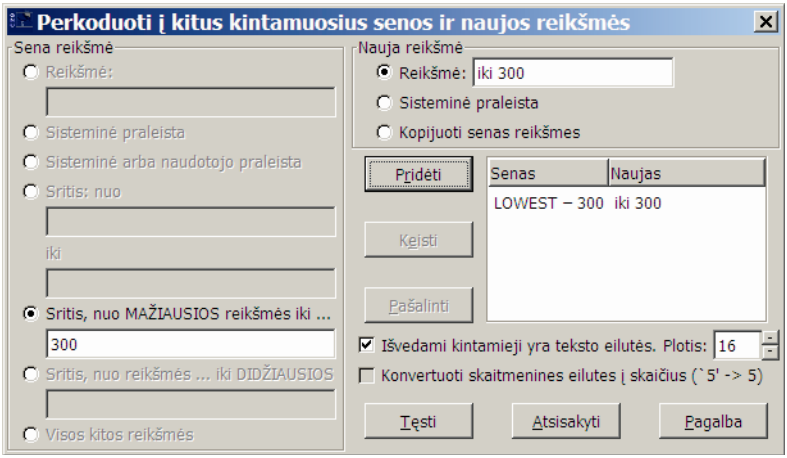


4.1 pav. Pradinis *Perkoduoti į kitus kintamuosius* dialogo langas

Pradiniame *Perkoduoti į kitus kintamuosius* dialogo lange pasirenkame kiekybinius (*intervalų*) kintamuosius, kuriems sukursime naujus,

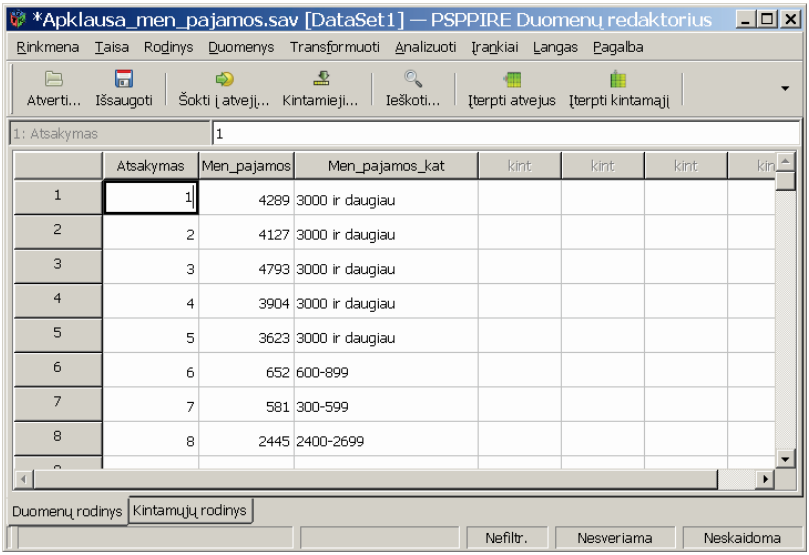
sugrupuotus kokybinius (*rangų* arba *kategorinius*) kintamuosius. Procedūra leidžia dviem ar daugiau pasirinktų gretimų reikšmių sukurti atskiras kategorijas – reikšmių intervalus. Kitais žodžiais, kintamojo reikšmės yra skaitomos (skenuojamos), analizuojamos ir priskiriamos tam tikram intervalui – kategorijai.

- Pažymėkite kintamąjį *Men_pajamos* ir spragtelėję mygtuką perkelkite į langelį *Kintamieji:*. **Pastaba**. Jei kintamieji turi etiketes, kintamųjų sąraše matomi ne kintamųjų vardai, o kintamųjų etiketes. Jei norime matyti kintamųjų vardus, reikia ant kintamojo vardo spragtelėti dešiniu pelės mygtuku ir nuimti varnelę nuo parinkties *„Jei įmanoma, naudoti kintamųjų etiketes“*.
- Langelyje *Kintamieji*: pažymėkite perkeltą kintamąji *Men_pajamos*. Srityje *Išvedamas kintamasis* įveskite naujo kintamojo vardą *Men_pajamos_kat*. Spragtelėkite mygtuką *Pakeisti*.
- Langelyje *Kintamieji:* atsirado seno ir naujo kintamųjų vardai. Spragtelėkite mygtuką *Senos ir naujos reikšmės*. Atsiveria langas *Perkoduoti į kitus kintamuosius senos ir naujos reikšmės*. Lango dalyje *Sena reikšmė* pažymėkite *Sritis, nuo MAŽIAUSIOS reikšmės iki...* Įveskite *300*. Lango dalyje *Nauja reikšmė* langelyje *Reikšmė:* įveskite *iki 300*.
- Pažymėkite varnele *Išvedami kintamieji yra teksto eilutės. Plotis:* nustatykite *16*. Spragtelėkite mygtuką *Pridėti* (žr. 4.2 pav.).



4.2 pav. *Perkoduoti į kitus kintamuosius senos ir naujos reikšmės* dialogo langas

- Atitinkamai kitoms senoms reikšmėms priskirkite naujas. Spragtelėkite mygtuką *Tęsti*.
- Spragtelėkite mygtuką *Gerai*.
- Bus sukurtas naujas sugrupuotas kintamasis *Men_pajamos_kat* (žr. 4.3 pav.).



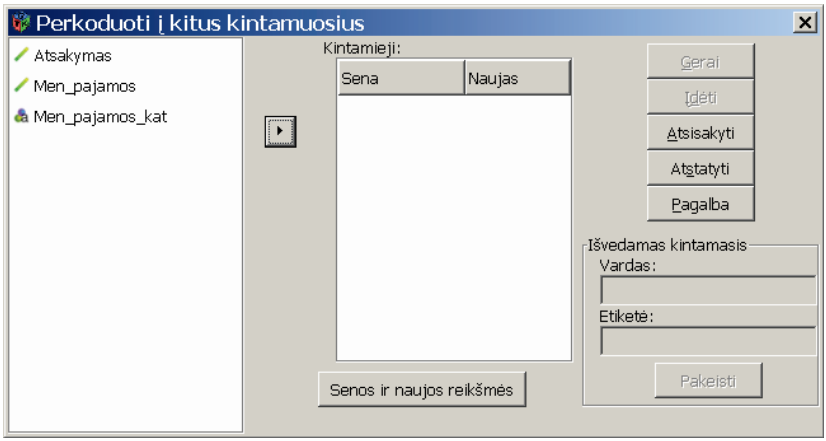
4.3 pav. PSPP duomenų redaktoriaus langas. Naujas kintamasis *Men_pajamos_kat*

Kiekybinių (intervalinių) kintamųjų sukūrimas iš kokybinių (kategorinių) kintamųjų

Apklausos metu mėnesio pajamos gali būti grupuojamos atskirose kategorijose. Faile *Apklausa_men_pajamos.sav* kintamasis *Men_pajamos_kat* yra kokybinis (kategorinis). Jis turi 11 pajamų kategorijų: mažiau už 300, 300–599; 600–899; 900–1199; 1200–1499; 1500–1799; 1800–2099; 2100–2399; 2400–2699; 2700–2999; 3000 ir daugiau. Kartais dėl įvairių priežasčių (pvz., ieškant padėties ir sklaidos charakteristikų) patogu vietoje kategorinių kintamųjų naudoti kiekybinius (intervalinius). Iš kokybinio (kategorinio) kintamojo *Men_pajamos_kat*

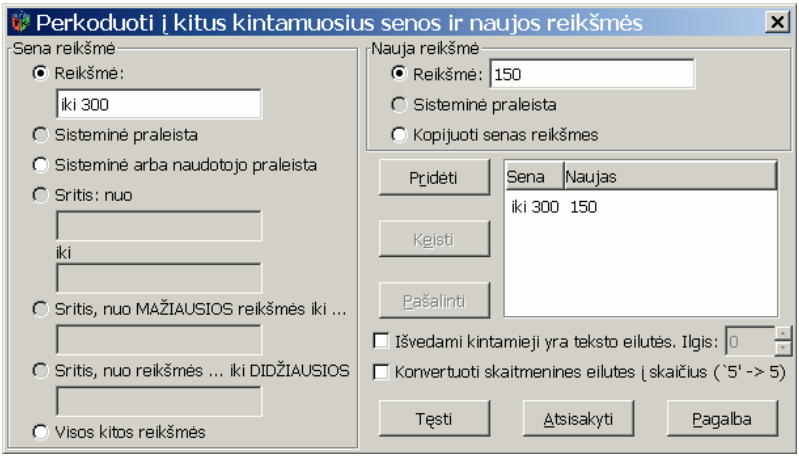
sukursime naują kiekybinį (intervalinį) kintamąjį *Men_pajamos_int*. Jam priskirsime vidurinę kategorijos intervalo reikšmę: intervalui 300–599 priskirsime 450, intervalui 600–899 priskirsime 750 ir t. t. Norėdami išlaikyti vienodus intervalus, mažiausių pajamų intervalui priskirsime reikšmę 150, o didžiausių – 3150.

- PSPP duomenų redaktoriuje pasirinkite: *Transformuoti* → *Perkoduoti į kitus kintamuosius...* Atveriamas *Perkoduoti į kitus kintamuosius* dialogo langas (žr. 4.4 pav.).



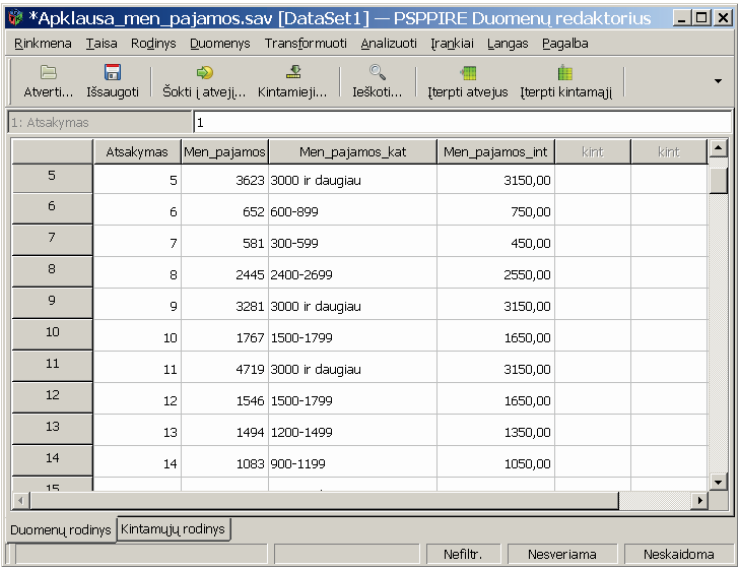
4.4 pav. Pradinis *Perkoduoti į kitus kintamuosius* dialogo langas

- Pažymėkite kintamąjį *Men_pajamos_kat* ir spragtelėję mygtuką ▸ perkelkite į langelį *Kintamieji:*.
- Langelyje *Kintamieji*: pažymėkite perkeltą kintamąjį *Men_pajamos_kat*. Srityje *Išvedamas kintamasis* įveskite naujo kintamojo vardą *Men_pajamos_int*.. Spragtelėkite mygtuką *Pakeisti*.
- Langelyje *Kintamieji:* atsirado seno ir naujo kintamųjų vardai. Spragtelėkite mygtuką *Senos ir naujos reikšmės*. Atsiveria langas *Perkoduoti į kitus kintamuosius senos ir naujos reikšmės*. Lango dalyje *Sena reikšmė* langelyje *Reikšmė:* įveskite *iki 300*. Lango dalyje *Nauja reikšmė* langelyje *Reikšmė:* įveskite *150*.
- Spragtelėkite mygtuką *Pridėti* (žr. 4.5 pav.).



4.5 pav. *Perkoduoti į kitus kintamuosius senos ir naujos reikšmės* dialogo langas

- Atitinkamai kitoms senoms reikšmėms priskirkite naujas. Spragtelėkite mygtuką *Tęsti*.
- Spragtelėkite mygtuką *Gerai*.
- Bus sukurtas naujas sugrupuotas kintamasis *Men_pajamos_int* (žr. 4.6 pav.).

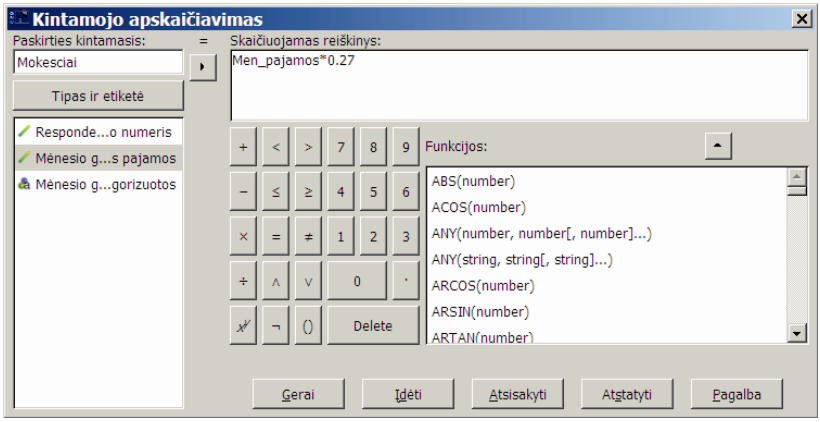


4.6 pav. PSPP duomenų redaktoriaus langas. Naujas kintamasis *Men_pajamos_kat*

Naujo kintamojo apskaičiavimas naudojant aritmetinę išraišką

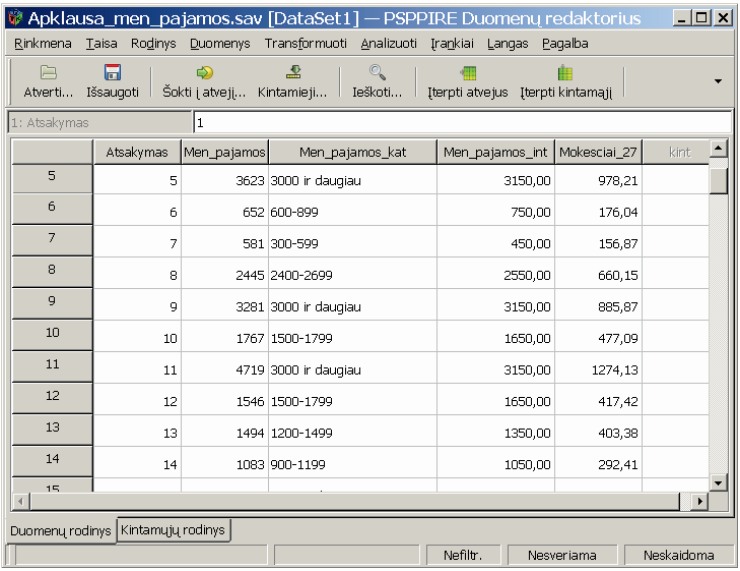
Sukursime naują kintamąjį *Mokesciai_27*, kuris nuo gaunamų mėnesinių pajamų apskaičiuos 27 proc. dydžio mokestį.

- Duomenų redaktoriuje pasirinkite *Transformuoti* → *Skaičiuoti…*
- *Paskirties kintamasis:* langelyje įveskite „Mokesciai_27“.
- Pasirinkite kintamųjų sąraše *Men_pajamos* ir spragtelėję rodyklę perkelkite kintamąjį į *Skaičiuojamas reiškinys:* langelį.
- *Skaičiuojamas reiškinys:* langelyje esančiu kalkuliatoriumi surinkite x 0.27 (žr. 4.7 pav.).



4.7 pav. Langas *Kintamojo apskaičiavimas*

- Spragtelėkite *Gerai*. Duomenų redaktoriuje matysite naują kintamąjį *Mokesciai_27* (žr. 4.8 pav.).



4.8 pav. Naujas kintamasis *Mokesciai_27* duomenų redaktoriuje

Naujo kintamojo apskaičiavimas naudojant loginę formulę

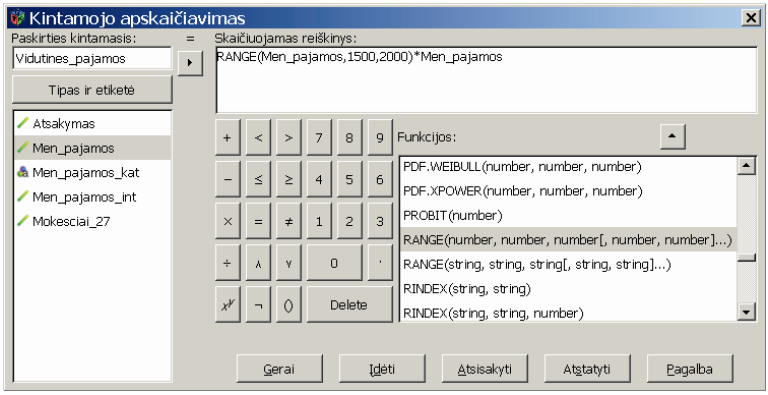
PSPP leidžia apskaičiuoti naujus kintamuosius iš esančių stebėjimo duomenų (atvejų) naudojant logines sąlygas. Jei loginės sąlygos yra tenkinamos, kiekvienai kintamojo reikšmei yra suteikiama loginė reikšmė *true*, jei sąlygos netenkinamos – *false*, jei reikšmė praleista – *missing*.

Jei loginės formulės rezultatas yra *true*, naujas kintamasis sukuriamas, jei loginės formulės rezultatas yra *false* arba *missing*, naujas kintamasis nesukuriamas.

Loginei (sąlygos) formulei sudaryti galima naudoti kintamųjų vardus, konstantas, aritmetinius operatorius, skaitmenines ir kt. funkcijas, dialogo langelyje esančius santykio operatorius: < (mažiau), > (daugiau), <= (mažiau arba lygu), >= (daugiau arba lygu), ≠ (nelygu) ir loginius operatorius: & (loginis IR), | (loginis ARBA), ¯ (loginis NE).

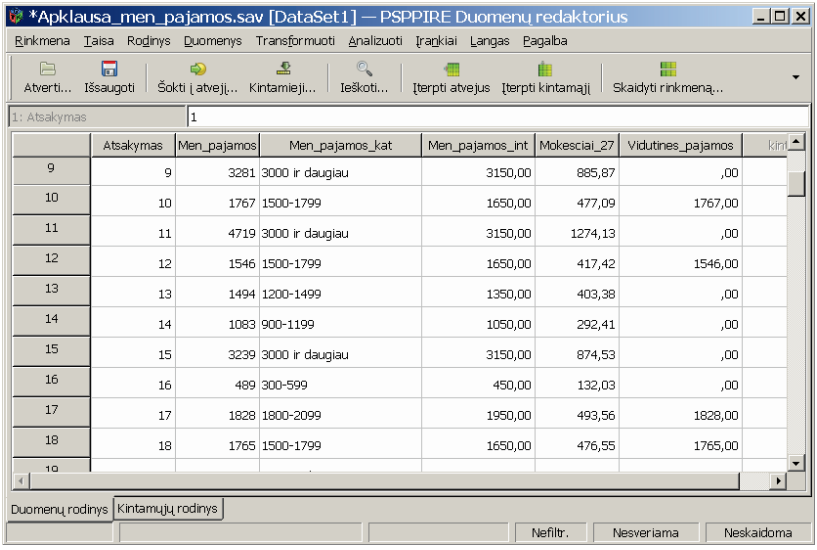
Sukursime naują kintamąjį *Vidutines_pajamos*. Šiam kintamajam logine funkcija priskirsime kintamojo *Men_pajamos* reikšmes nuo 1500 iki 2000 Lt imtinai.

- Atverkite failą *Apklausa_men_pajamos.sav*
- Duomenų redaktoriuje pasirinkite *Transformuoti* → *Skaičiuoti…*Bus atvertas dialoginis langas *Kintamojo apskaičiavimas*. Langelyje *Paskirties kintamasis:* įveskite naujo kintamojo vardą *Vidutines_pajamos*. Langelyje *Funkcijos*: pasirinkite *RANGE (number, number, number)*. Pirmas argumentas bus mūsų kintamasis *Men_pajamos*, antras argumentas – mažiausia vidutinių pajamų reikšmė 1500, o trečias argumentas – didžiausia vidutinių pajamų reikšmė – 2000. Jei *Men_pajamos* reikšmė bus iš diapazono nuo 1500 iki 2000 imtinai, funkcijos rezultatas bus *true* arba 1, jei ne iš šio diapazono – *false* arba 0. Dauginsime naująjį kintamąjį iš kintamojo *Men_pajamos*: jei funkcijos rezultatas bus *true* – naujojo kintamojo reikšmė bus lygi *Men_pajamos*, jei funkcijos reikšmė bus *false* – naujojo kintamojo reikšmė – 0.
- Langelyje *Skaičiuojamas reiškinys:* pažymėkite pasirinktos funkcijos RANGE(?,?) pirmąjį argumentą, kintamųjų sąraše pasirinkite kintamąjį *Men_pajamos* ir spragtelėję trikampėlį perkelkite jį į funkcijos pirmąjį argumentą. Įveskite antrąjį ir trečiąjį argumentus, atitinkamai 1500 ir 2000. Funkciją padauginkite iš kintamojo *Men_pajamos* (žr. 4.9 pav.):



4.9 pav. Langas *Kintamojo apskaičiavimas*

- Spragtelėkite *Gerai*. Duomenų redaktoriuje matysite naują kintamąjį *Vidutines_pajamos* (žr. 4.10 pav.).

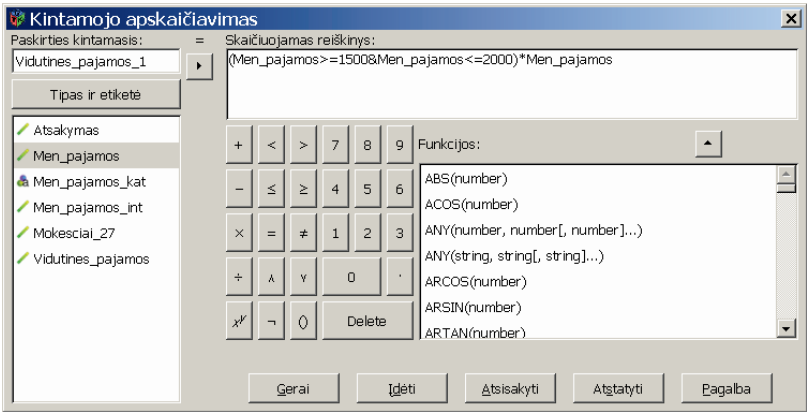


4.10 pav. Naujas kintamasis *Vidutines_pajamos*

Analogišką rezultatą pasiektume naudodami ne loginę funkciją, o loginius operatorius. Sukursime naują kintamąjį *Vidutines_pajamos_1*. Šiam kintamajam loginiais operatoriais priskirsime kintamojo *Men_pajamos* reikšmes nuo 1500 iki 2000 Lt imtinai. Tam naudosime loginį operatorių IR (&).

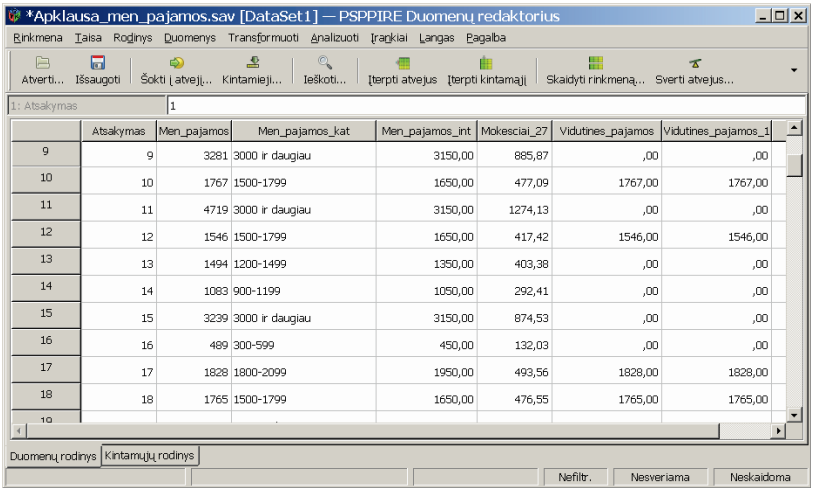
- Duomenų redaktoriuje pasirinkite *Transformuoti* → *Skaičiuoti…* Bus atvertas dialoginis langas *Kintamojo apskaičiavimas*. Langelyje *Paskirties kintamasis:* įveskite naujo kintamojo vardą *Vidutines_pajamos_1*. Naujam kintamajam loginiais operatoriais turime priskirti kintamojo *Men_pajamos* reikšmes iš diapazono nuo 1500 iki 2000 imtinai. Tada loginės operacijos rezultatas bus *true* arba 1, priešingu atveju – *false* arba 0. Dauginsime naująjį kintamąjį iš kintamojo *Men_pajamos*: jei funkcijos rezultatas bus *true* – naujojo kintamojo reikšmė bus lygi *Men_pajamos*, jei funkcijos reikšmė bus *false* – naujojo kintamojo reikšmė – 0.
- Kintamųjų sąraše pasirinkite kintamąjį *Men_pajamos* ir spragtelėję trikampėlį perkelkite jį į langą *Skaičiuojamas reiškinys:*. Įveskite loginę sąlygą >=1500. Loginiu IR (&) sujunkite su ant-

ra logine sąlyga <=2000. Viską apskliaudę skliausteliais padauginkite iš kintamojo *Men_pajamos* (žr. 4.11 pav.):



4.11 pav. Langas *Kintamojo apskaičiavimas*

- Spragtelėkite *Gerai*. Duomenų redaktoriuje matysite naują kintamąjį *Vidutines_pajamos1* (žr. 4.12 pav.).



	Atsakymas	Men_pajamos	Men_pajamos_kat	Men_pajamos_int	Mokesciai_27	Vidutines_pajamos	Vidutines_pajamos_1
9	9	3281	3000 ir daugiau	3150,00	885,87	,00	,00
10	10	1767	1500-1799	1650,00	477,09	1767,00	1767,00
11	11	4719	3000 ir daugiau	3150,00	1274,13	,00	,00
12	12	1546	1500-1799	1650,00	417,42	1546,00	1546,00
13	13	1494	1200-1499	1350,00	403,38	,00	,00
14	14	1083	900-1199	1050,00	292,41	,00	,00
15	15	3239	3000 ir daugiau	3150,00	874,53	,00	,00
16	16	489	300-599	450,00	132,03	,00	,00
17	17	1828	1800-2099	1950,00	493,56	1828,00	1828,00
18	18	1765	1500-1799	1650,00	476,55	1765,00	1765,00

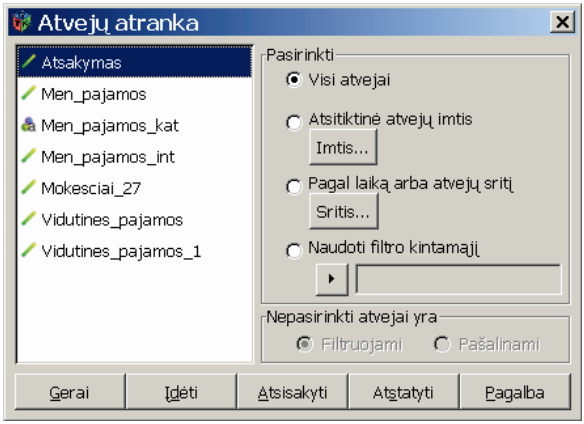
4.12 pav. Naujas kintamasis *Vidutines_pajamos_1*

4.2. DUOMENŲ ATRANKA (FILTRAVIMAS)

Dažnai analizei reikalingi tik tie stebėjimų duomenys (atvejai), kurie tenkina nustatytus kriterijus. Atvejai, kurie netenkina atrankos kriterijų, gali būti neaktyvūs (lieka faile, bet neanalizuojami) arba pašalinti.

Jei norite atrinkti stebėjimų duomenis:

- Atverkite failą su analizuojamais duomenimis (mes naudojame failą *Apklausa_men_pajamos.sav*).
- Nurodykite komandas *Duomenys* → *Atvejų atranka…* (žr. 4.13 pav.).



4.13 pav. Dialogo langas *Atvejų atranka*

- Dialogo lange *Atvejų atranka* pasirinkite reikiamą duomenų atrankos metodą:

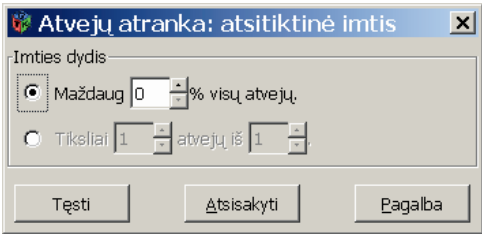
1. *Visi atvejai* – tai variantas pagal nustatymą.
2. *Atsitiktinė atvejų imtis*. Galima rinktis atsitiktinius atvejus procentais iš visų arba tiksliai tam tikrą skaičių atsitiktinių atvejų.
3. *Pagal laiką arba atvejų sritį*. Galima pasirinkti atvejus iš diapazono, nuo kurio iki kurio tiriame. Tam tikros rūšies duomenims gali būti nurodytas laiko diapazonas.
4. *Naudoti filtro kintamąjį*. Ši atranka naudojama tada, kai norime turėti duomenų bazę be praleistų duomenų arba be nulinių filtruojančio kintamojo reikšmių.

- Visais atvejais galima nurodyti, ką daryti su neparinktais atvejais – filtruoti (praleisti) ar pašalinti.

Atsitiktinė duomenų atranka

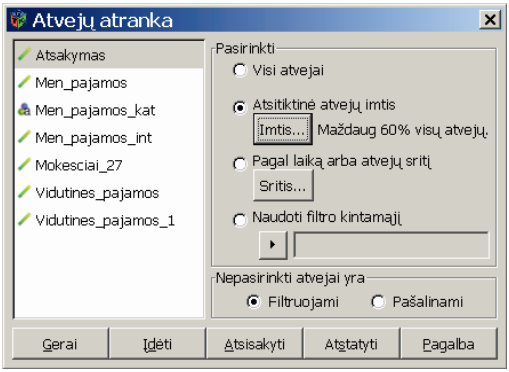
Kai stebėjimų skaičius yra didelis, pirminei hipotezei patikrinti gali būti tikslinga taikyti nedidelę atsitiktinę imtį.

- Atsitiktinei duomenų atrankai iš stebėjimų visumos dialogo lange *Atvejų atranka* pažymėkite variantą *Atsitiktinė atvejų imtis* ir spragtelėkite mygtuką *Imtis…* Atsivers dialogo langelis *Atvejų atranka: atsitiktinė imtis* (žr. 4.14 pav.), kuriame galima pasirinkti šiuos imties dydžio nurodymo būdus:
  - *Maždaug*. Įrašykite į laukelį procentinę atsitiktinės imties dydžio reikšmę – PSPP programa sukurs atsitiktinę imtį, savo dydžiu apytiksliai atitinkančią nurodytą visų stebėjimų dalį.
  - *Tiksliai*. Šiuo atveju nurodykite tikslų stebėjimo atvejų skaičių atsitiktinėje imtyje. Taip pat nurodykite stebėjimų, iš kurių bus paimta atsitiktinė imtis, skaičių. Šis skaičius neturi viršyti visų stebėjimų (atvejų) skaičiaus duomenų faile.



4.14 pav. Dialogo langas *Atvejų atranka: atsitiktinė imtis*

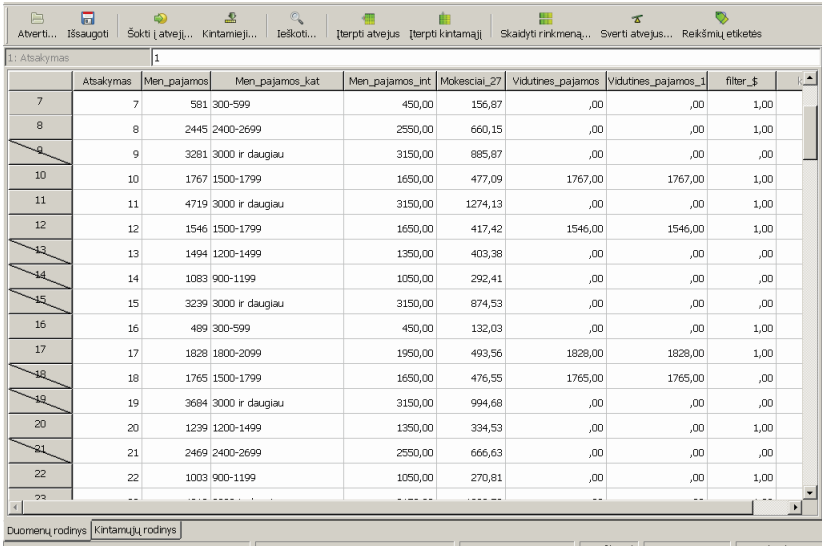
- Dialogo lange *Atvejų atranka: atsitiktinė imtis* pasirinkite *Maždaug* ir įrašykite procentinę atsitiktinės imties reikšmę 60. Spragtelėkite mygtuką *Tęsti* (žr. 4.15 pav.).



4.15 pav. Dialogo langas *Atsitiktinė atranka*. Atrankos būdas *Atsitiktinė atvejų imtis*

- Paspauskite mygtuką *Gerai*.

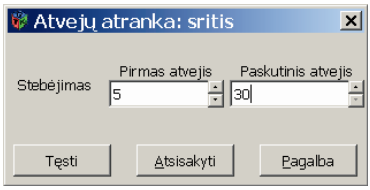
PSPP programa palieka apytikriai 60 procentų visų stebėjimo duomenų dydžio imtį (žr. 4.16 pav.).



4.16 pav. Atsitiktinės atrankos (filtravimo) rezultatas

Diapazoninė duomenų atranka

- Dialogo lange *Atvejų atranka* pažymėkite variantą *Pagal laiką arba atvejų sritį* ir spragtelėkite mygtuką *Sritis…*
- Atsivėrusiame dialogo langelyje *Atvejų atranka: sritis* laukelyje *Pirmas atvejis* nurodykite diapazono pradžią, o laukelyje *Paskutinis atvejis* – diapazono pabaigą. Datos ir laiko diapazonai gali būti nurodyti tik tam tikro formato duomenims (žr. 4.17 pav.).

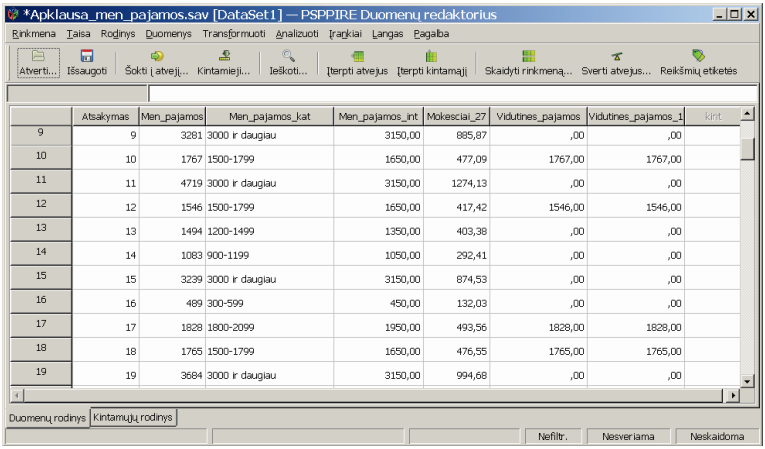


4.17 pav. Diapazoninė duomenų atranka. Diapazono pasirinkimas

- Spragtelėkite mygtuką *Tęsti*, paskui dialogo langelyje *Atvejų atranka* paspauskite mygtuką *Gerai*.

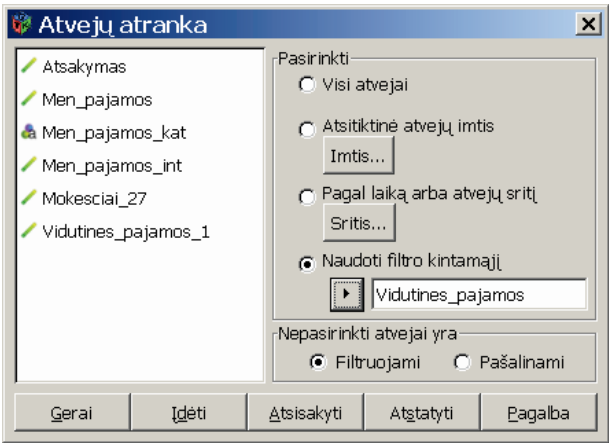
Duomenų atranka pagal filtruojantį kintamąjį

Mūsų duomenų bazėje kintamasis *Vidutines_pajamos* nenulines reikšmes turi tik tada, kai kintamasis *Men._pajamos* yra tarp 1500 ir 2000 imtinai (žr. 4.18 pav.). Atrinksime atvejus, kai kintamasis *Vidutines_pajamos* turi nelygias nuliui reikšmes.



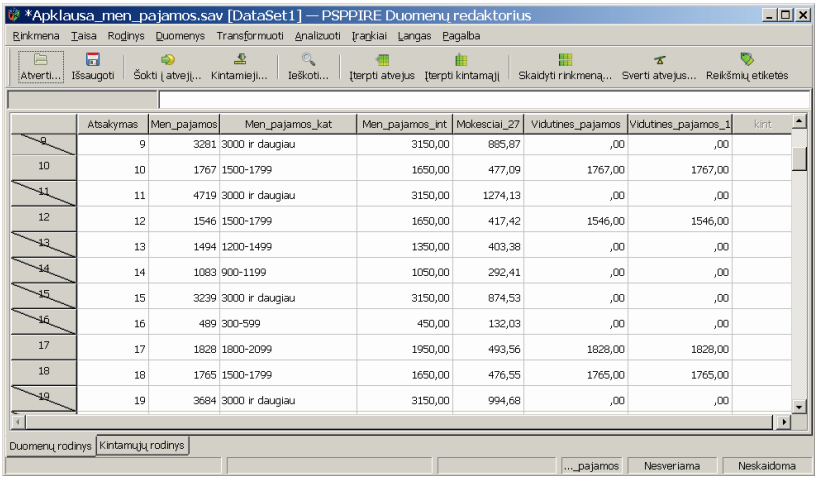
4.18 pav. Kintamojo *Vidutines_pajamos* reikšmės

- Dialogo langelyje *Atvejų atranka* pažymėkite variantą *Naudoti filtro kintamąjį* ir į kintamojo langelį įkelkite kintamąjį *Vidutines_pajamos* (žr. 4.19 pav.).



4.19 pav. Atvejų atranka pagal filtruojantį kintamąjį

- Paspauskite mygtuką *Gerai*. Atvejai su nulinėmis kintamojo *Vidutines_pajamos* reikšmėmis bus filtruoti (žr. 4.20 pav.).



4.20 pav. Atrankos pagal filtruojantį kintamąjį rezultatas

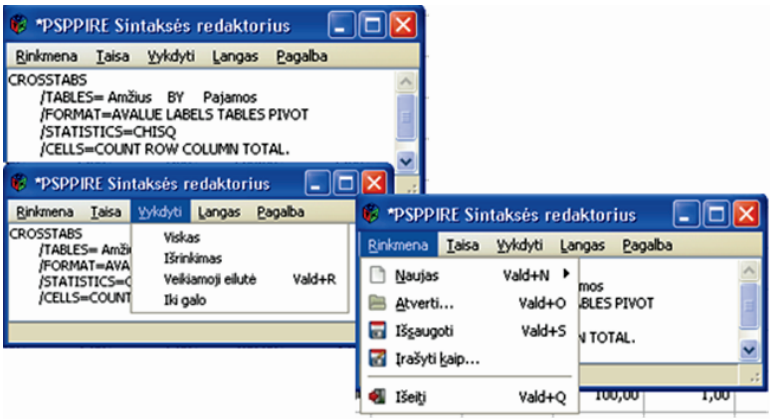
4.3. PSPP PAKETO SINTAKSĖS REDAKTORIUS

PSPP paketas (kaip ir SPSS paketas) turi sintaksės redaktorių. Sintaksės redaktorius leidžia atlikti skaičiavimus nenaudojant valdymo skydelio mygtukų (meniu komandų) ir dialogo langų, bet taikant specialią veiksmų komandų kalbą. Komandos užrašomos vadinamuosiuose *Sintax* failuose, kurie iš esmės yra programos, užrašytos PSPP arba SPSS programavimo kalba. Abi kalbos yra panašios.

Naudojant *Sintax* programas, galima automatizuoti dažnai kartojamus veiksmus. Gerai išmanant programavimo kalbą, galima sukurti papildomų statistinės analizės metodų, kurių neįmanoma atlikti per meniu komandas, programas.

PSPP terpė

Sintaksės redaktoriaus dialogo langas (žr. 4.21 pav.) *PSPPIRE Sintaksės redaktorius* iškviečiamas komanda *Rinkmena* → *Naujas* → *Sintaksė*. Šiame lange galima rinkti ir vykdyti PSPP komandas.

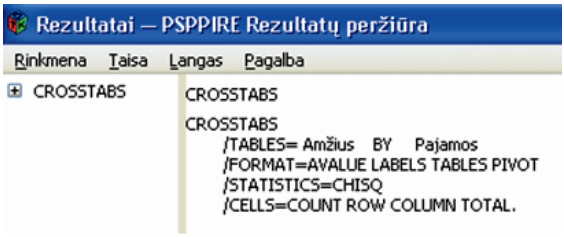


4.21 pav. PSPP sintaksės redaktoriaus langai

Programinis failas vykdomas komanda *Vykdyti*. Komandos meniu leidžia pasirinkti viso failo vykdymą (*Viskas*) arba jo dalies. Komanda *Rinkmena* leidžia išsaugoti komandų rinkinį atskirame faile.

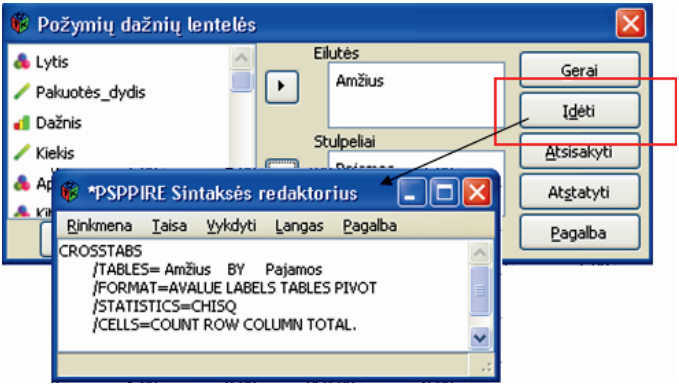
Paprasčiausias būdas sukurti sintaksės programą – nukopijuoti ją iš *Rezultatai – PSPPIRE Rezultatų peržiūra* lango (žr. 4.22 pav.), kur

sintaksės komandos automatiškai įterpiamos atlikus skaičiavimus pasitelkus meniu. Šią programą vėliau galima redaguoti keičiant komandų parametrus.



4.22 pav. PSPP sintaksės komandos *Rezultatai – PSPPIRE Rezultatų peržiūra* lange

Bet kurios PSPP komandos dialogo langas turi mygtuką *Įdėti* (žr. 4.23 pav.). Mygtuko *Įdėti* paspaudimas atveria *PSPPIRE Sintaksės redaktorius* langą. Šiame lange pateikiama atliekamos procedūros sintaksė. Šiuo atveju atlikti skaičiavimus galima pasitelkus meniu *Vykdyti*.



4.23 pav. PSPP sintaksės redaktoriaus iškvietimas dialogo lange

SPSS terpė

Sintaksės redaktoriaus dialogo langas *SPSS Sintax Editor* iškviečiamas komanda *File* → *New* → *Sintax*. Šiame lange galima rinkti ir vykdyti SPSS komandas.

Paprasčiausias būdas sukurti *Sintax* programą – nukopijuoti ją iš *Output-SPSS Viewer* lango, kur jos automatiškai įterpiamos atlikus skaičiavimus pasitelkus meniu. Šią programą vėliau galima redaguoti

keičiant komandų parametrus. Programinis failas vykdomas komanda *Run*. Komandos meniu leidžia pasirinkti viso failo (*All*) arba jo dalies vykdymą. Jei SPSS duomenų failas yra aktyvus, SPSS sintaksės lange galima iš karto nurodyti vykdomas komandas. Priešingu atveju programą reikia pradėti nurodant duomenų failo iškvietimo komandą. SPSS sintaksės programos failą galima išsaugoti.

SPSS komandų sintaksės aprašymas pateikiamas surinkus komandą *Help* → *Command Syntax Reference*. Laikomasi šių SPSS komandų sintaksės taisyklių:

- Kiekviena komanda pradedama iš naujos eilutės ir baigiama tašku.
- Komanda gali užimti keletą eilučių. Kiekviena vienos komandos eilutė pradedama nuo tarpo.
- Pagalbinės komandos atskiriamos ženklu „/".
- Eilutės ilgis turi būti ne daugiau kaip 80 ženklų.
- Dešimtainė trupmena atskiriama tašku.

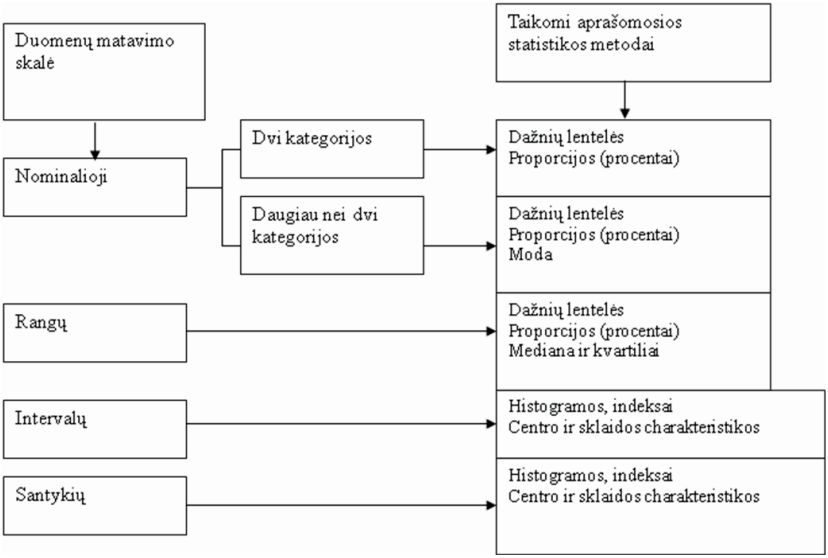
Sintaksės komandų kalba yra sudėtinga, todėl šioje knygoje išsamiai nenagrinėjama, tačiau kiekvienam nagrinėjamam analizės pavyzdžiui pateikiami PSPP ir SPSS sintaksės programų tekstai.

4.4. APRAŠOMOJI STATISTIKA

Statistinio stebėjimo metu surinkti duomenys apie kiekvieną objekto vienetą – didelis kiekis medžiagos, kuri dažnai neteikia išsamesnės informacijos apie patį objektą. Šią medžiagą būtina susisteminti, susumuoti ir apdoroti. Tik tuomet bus galima išsiaiškinti būdingus statistinės visumos bruožus, nustatyti dėsningumus.

Prieš atliekant nuodugnesnę surinktų duomenų statistinę analizę, pirmiausia reikia panagrinėti kiekvieną kintamąjį (požymį) atskirai. Statistinė analizė, kai kiekvienas kintamasis nagrinėjamas atskirai, vadinama *vienmate*. Šiai analizei atlikti naudojami *aprašomosios statistikos* metodai.

4.1 lentelė. Aprašomosios statistikos metodų taikymo priklausomybė nuo kintamųjų tipo



Aprašomosios statistikos metodus sudaro

- duomenų grupavimas ir dažnių (pasikartojimų) lentelės;
- statistinių charakteristikų (duomenų padėties ir sklaidos charakteristikų) skaičiavimas;
- grafinis stebėjimų vaizdavimas.

Su kintamųjų tipu siejasi matavimo skalės. Kintamųjų reikšmės gali būti išmatuotos nominalinėje, ranginėje ar metrinėje (intervalų ar santykių) skalėje. Matavimų skalė lemia duomenų analizės metodo parinkimą.

Dažnių lentelės sudaromos kintamiesiems, kurie išmatuoti nominalinėje arba, esant nedideliam kategorijų skaičiui, rangų skalėje.

Pradinei duomenų analizei, be grupuotų dažnių lentelių (vienmačių), dažnai dar sudaromos porinės (ir daugiau kintamųjų) dažnių lentelės. Porinės dažnių lentelės skirtos tarp kintamųjų esantiems ryšiams pavaizduoti. Tokios dažnių lentelės paprastai sudaromos nominaliniams ir ranginiams kintamiesiems su santykinai mažu stebimų reikšmių (kategorijų) skaičiumi. Porinėms lentelėms sudaryti nenaudojami kiekybiniai kintamieji. Jeigu į tokią lentelę būtina įtraukti kiekybinį kin-

tamąjį (pavyzdžiui, pajamas), tai pirmiausia šį kintamąjį reikia perkoduoti taip, kad jo reikšmės būtų sudėtos į keletą intervalų (pavyzdžiui, mažos pajamos, vidutinės, didelės).

Aprašomoji statistika, be dažnių lentelių ir diagramų, nominaliniams kintamiesiems jokių kitokių metodų nesiūlo. Ranginiams kintamiesiems ir kiekybiniams kintamiesiems, išmatuotiems intervalų ar santykių skalėse, bet neturintiems normaliojo skirstinio, šalia dažnių lentelių ir diagramų dažniausiai dar skaičiuojama vidurinė (centrinė) atsitiktinio dydžio reikšmė (mediana) ir abu kvartiliai. Kvartiliais vadinami skaičiai, dalijantys variacinę eilutę į keturias lygias dalis (variacinė eilutė – duomenų eilutė, išdėstyta nemažėjimo tvarka. Jei duomenys išdėstyti didėjimo ar mažėjimo tvarka, turime ranžiruotą eilutę). Kintamiesiems, išmatuotiems intervalų ar santykių skalėse ir turintiems normalųjį skirstinį, skaičiuojamos dažnių lentelės, piešiamos diagramos (histogramos, linijinės diagramos ir pan.) ir skaičiuojamos centro bei sklaidos charakteristikos.

Kadangi dauguma metodų, naudojamų intervalų ir santykių skalės kintamųjų analizei, remiasi prielaida, kad šie kintamieji turi normalųjį skirstinį, pradinei duomenų analizei priskiriami paprasčiausi normališkumo tikrinimo kriterijai.

Esant normaliajam skirstiniui dauguma reikšmių grupuojasi apie vidurkį, o tolstant nuo jo į abi puses, stebėjimų tankis tolygiai mažėja.

Tam, kad būtų galima įsitikinti, ar imtis turi normalųjį skirstinį, naudojami du paprasti grafiniai būdai: histograma ir normalumo tiesė.

Histograma braižoma pasirinkus šį diagramų tipą ir papildomai brėžiant normaliojo skirstinio kreivę.

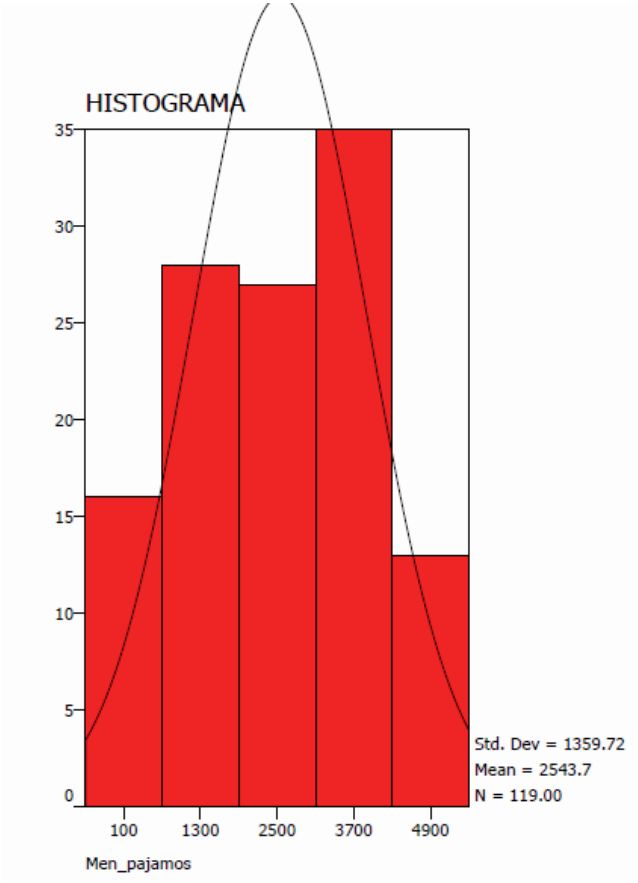
Pavyzdys. Failas *Apklausa_men_pajamos.sav*. Apklausoje gauti respondentų atsakymai apie jų gaunamas pajamas – intervalinis kintamasis *Men_pajamos*.

PSPP terpė

Pasitelkę PSPP nubrėšime kintamojo *Men_pajamos* histogramą ir normalumo kreivę. Normalumo tiesės šis paketas nebrėžia.

- Atidarykite failą su analizuojamais duomenimis.
- Nurodykite komandas *Analizuoti* → *Aprašomoji statistika* → *Dažniai*…

- Įkelkite kintamąjį *Men_pajamos* į lauką *Kintamieji*:.
- Spragtelėkite mygtuką *Diagramos…*
- Pažymėkite *Braižyti histogramas* ir *Uždėti normaliąją kreivę*. Paspauskite mygtuką *Tęsti*.
- Sugrįšite į dialogo langą *Dažniai*. Paspauskite mygtuką *Gerai*.

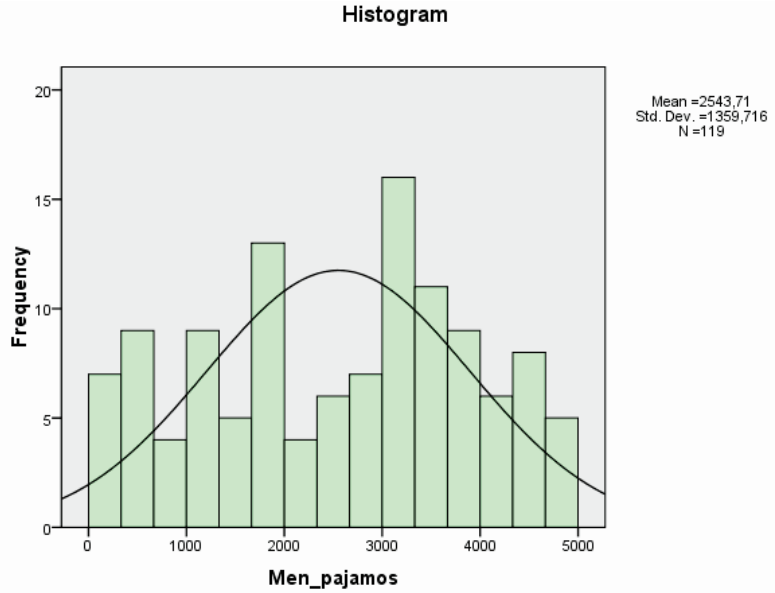


4.24 pav. Kintamojo *Men_pajamos* histograma ir normalumo kreivė

SPSS terpė

Pasitelkę SPSS nubrėšime kintamojo *Men_pajamos* histogramą ir normalumo kreivę.

- Atidarykite failą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Frequencies*…
- Įkelkite kintamąjį *Men_pajamos* į lauką *Variable(s)*:.
- Spragtelėkite mygtuką *Charts*…
- Lauke *Chart type* pasirinkite *Histograms:* ir pažymėkite *With normal curve.*
- Spragtelėkite mygtuką *Continue*.
- Spragtelėkite mygtuką *OK*.



4.25 pav. Dažnių histograma

Brėžiant histogramą, svarbu parinkti tinkamą grupių skaičių. Kompiuterinės programos dažniausiai automatiškai parenka grupių skaičių, kurį galima keisti.

Grupių skaičius nustatomas pagal bendras taisykles. Beveik į kiekvieną grupę turi patekti vidutiniškai ne mažiau kaip penki duomenys, o grupių turėtų būti ne mažiau kaip 5 ir ne daugiau kaip 15.

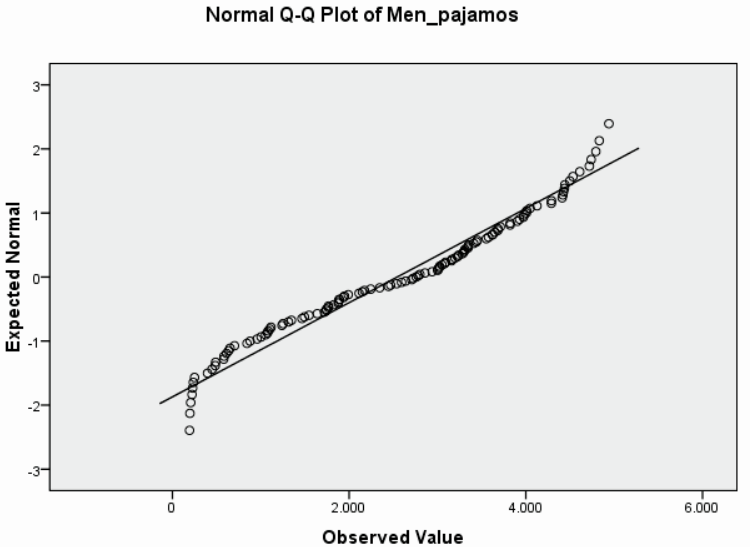
Grupės turi būti vienodo ilgio, kad jas būtų galima lyginti tarpusavyje. Kai duomenų daug (iki 100), grupių skaičių n nustatyti galima pagal Sterdžeso formulę

$$n = 1 + 3{,}322 * \log N; \tag{4.1}$$

čia: log – dešimtainis logaritmas, N – visų duomenų skaičius.

Brėžiant normalumo tiesę (statistinėse programose vadinamą *Normal probability plot* arba *Normal Q-Q plot* (SPSS)) duomenys transformuojami tokiu būdu, kad tuo atveju, kai duomenys turi normalųjį skirstinį, jie sudarytų tiesę. SPSS programoje:

- Atidarykite failą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Q-Q Plots*…
- Įkelkite kintamąjį *Men_ pajamos* ir paspauskite *OK*.



4.26 pav. Normalumo tiesė

Jei duomenų skirstinys nėra normalusis, taškai skirsis nuo tiesės. Iš vaizdo, pateikto diagramoje, galima spręsti, ar duomenys turi normalųjį skirstinį. Netgi nedidelė darbo su realiais duomenimis patirtis leidžia

žmogui gana užtikrintai atskirti duomenų nukrypimą nuo normaliojo skirstinio. Mūsų atveju taškai nėra tiesėje – skirstinys nėra normalusis.

Jei netenkina vizuali kontrolė, patikrinimui galime naudoti Kolmogorovo–Smirnovo ir Šapiro–Vilko (kai duomenų yra mažiau nei 50) kriterijus.

- Atidarykite failą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Explore*…
- Įkelkite kintamąjį *Men_ pajamos*.
- Paspauskite mygtuką *Plots*…
- Įsitikinkite, kad pasirinkta *Normality plots with tests*. Paspauskite mygtukus *Continue* ir *OK*.

4.2 lentelė. Skirstinio normalumo tikrinimo rezultatai

Tests of Normality						
	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Statistic	df	Sig.	Statistic	df	Sig.
Men_pajamos	,094	119	,011	,955	119	,001
a. Lilliefors Significance Correction						

Kadangi stebimasis reikšmingumo lygmuo *Sig.* (iš angl. *Significance* – reikšmingumas) yra mažesnis už reikšmingumo lygmenį 0,05, galime daryti išvadą, kad skirstinys nėra normalusis.

4.4.1. Vienmatė dažnių analizė

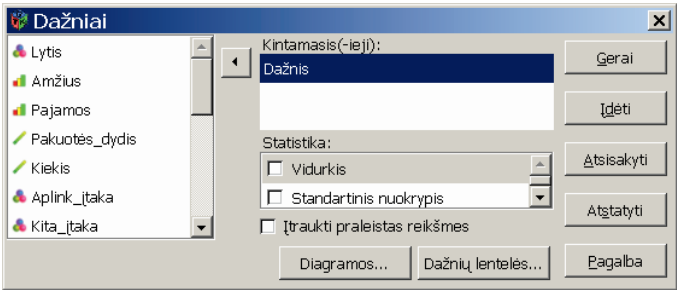
Dažnių analizės rezultatai ar grafinė medžiaga yra pakankami apibūdinti daugeliui kintamųjų ir yra pirmas duomenų statistinės analizės etapas.

Vienmatė dažnių lentelė sudaroma analizuojant pavadinimų (nominalius) ir ranginius kintamuosius. Šia lentele naudojamasi tiriant, kaip atskiro kintamojo reikšmės pasiskirsčiusios imtyje.

Pavyzdys. Apklausoje buvo tirta, kaip dažnai per mėnesį respondentai perka kavą. Kintamasis *Dažnis*. Pasitelkę PSPP/SPSS gausime šio kintamojo dažnių lentelę.

PSPP terpė

- Atidarykite failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodykite komandas *Analizuoti* → *Aprašomoji statistika* → *Dažniai…*
- Dialogo langelyje *Dažniai* pažymėkite analizuojamą kintamąjį *Dažnis* ir spragtelėję mygtuką su rodyklės ženklu įkelkite jį į analizuojamų kintamųjų sąrašą *Kintamasis(-ieji):*.
- Kadangi kintamasis *Dažnis* yra ranginis, langelyje *Statistika*: pašaliname visas varneles – padėties ir sklaidos rodikliai neskaičiuojami (žr. 4.27 pav.).



4.27 pav. Kintamasis *Dažnis* dialogo lange *Dažniai*

- Spragtelėkite mygtuką *Gerai*. Rezultatus matote rodinio *Rezultatai-PSPPIRE Rezultatų peržiūra* lange (žr. 4.28 pav.).

Dažnis

Reikšmės etiketė	*Reikšmė*	*Dažnis*	*Procentai*	*Galiojanti procentinė dalis*	*Sukauptoji procentinė dalis*
	1,00	14	31,11	31,11	31,11
	2,00	18	40,00	40,00	71,11
	3,00	12	26,67	26,67	97,78
	4,00	1	2,22	2,22	100,00
	Iš viso	45	100,0	100,0	

4.28 pav. PSPP dažnių lentelė

Kintamasis *Dažnis* vaizduoja respondentų atsakymus į klausimą, kiek kartų per mėnesį jie perka kavą. Kintamojo reikšmės 1, 2, 3, 4.

Kiekviena dažnių lentelės eilutė aprašo vieną kintamojo reikšmę. Pirmame stulpelyje *Reikšmės etiketė* pateikiamos kintamojo reikšmių etiketės (mūsų atveju jų neturime), antrame stulpelyje *Reikšmė* – visos

galiojančios kintamojo reikšmės. Trečiame stulpelyje *Dažnis* pateiktas kiekvieno atsakymo į anketos klausimą varianto dažnis, t. y. kiek respondentų pirko kavą per mėnesį vieną kartą (1), kiek du (2) ir t. t. Ketvirtame stulpelyje *Procentai* pateiktas kiekvieno atsakymo procentinis dažnis, kuris gaunamas kiekvieno atsakymo variantų skaičių dalijant iš stebėjimų skaičiaus, t. y. bendro respondentų skaičiaus. Penktame stulpelyje *Galiojanti procentinė dalis* pateikiamas atsakiusiųjų procentinis dažnis. Kai nėra praleistų duomenų, ketvirto ir penkto stulpelių reikšmės sutampa. Šeštame stulpelyje *Sukauptoji procentinė dalis* pateikiamas sukauptasis dažnis, kuris gaunamas sumuojant prieš tai buvusias reikšmes.

Pavyzdžiui, bendras respondentų, atsakiusių, kad per mėnesį perka kavą vieną ar du kartus, procentas yra 71,11.

SPSS terpė

- Atidarykite failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Frequencies...*
- Dialogo langelyje *Frequencies* pažymėkite analizuojamą kintamąjį *Dažnis* ir spragtelėję mygtuką su rodyklės ženklu įkelkite jį į analizuojamų kintamųjų sąrašą *Variable(s)*.
- Įsitikinę, kad pažymėtas laukelis *Display frequency tables* (rodyti dažnių lenteles) spragtelėkite mygtuką *OK*. Rezultatus matome rodinio (*Viewer*) lange.

Statistics		
Dažnis		
N	Valid	45
	Missing	0

Dažnis					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	14	31,1	31,1	31,1
	2,00	18	40,0	40,0	71,1
	3,00	12	26,7	26,7	97,8
	4,00	1	2,2	2,2	100,0
	Total	45	100,0	100,0	

4.29 pav. SPSS dažnių lentelė

4.3 lentelė. Vienmatės dažnių analizės sintaksė

PSPP	SPSS
GET GETFILE="C:\…\Pav1_kava.sav". FREQUENCIES FREQUENCIES /VARIABLES=Dažnis /FORMAT=AVALUE TABLE /STATISTICS=NONE.	GET FILE='C:\...\Pav1_kava.sav'. FREQUENCIES VARIABLES=Dažnis /ORDER=ANALYSIS.

4.4.2. Porinės dažnių lentelės

Porinės dažnių lentelės atskleidžia tarp kintamųjų esančius ryšius. Tokios dažnių lentelės paprastai sudaromos nominaliniams ir ranginiams kintamiesiems su santykinai mažu stebimų reikšmių (kategorijų) skaičiumi. Porinėms lentelėms sudaryti nenaudojami kiekybiniai kintamieji. PSPP tuo tikslu naudojamos vadinamosios požymių priklausomumo lentelės ir didelis testų pasirinkimas priklausomybės laipsniui tarp nagrinėjamų kintamųjų įvertinti.

Jeigu į tokią lentelę būtina įtraukti kiekybinį kintamąjį (pajamas ir pan.), tai pirmiausia šio kintamojo reikšmes reikia suskirstyti į keletą intervalų (mažos pajamos, vidutinės, didelės, labai didelės).

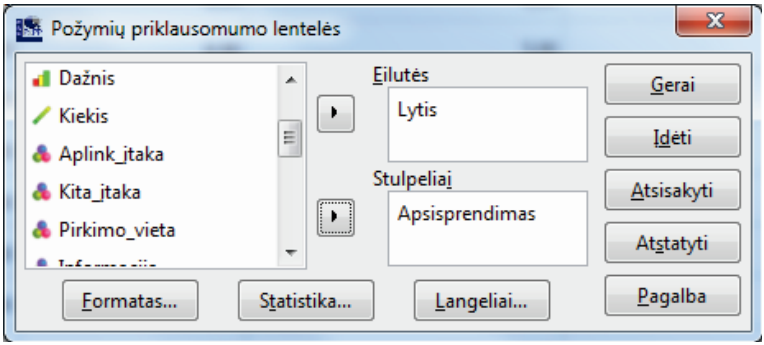
Paprasčiausia porinės dažnių lentelės forma yra 2×2 lentelė, gaunama tuo atveju, kai tiriami du kintamieji, kurių kiekvienas gali įgyti tik dvi reikšmes (dichotominiai kintamieji).

Pavyzdys. Apklausoje buvo tirta, pagal kokius kriterijus vyrai ir moterys perka kavą. Buvo galima rinktis iš penkių kriterijų: „rūšis“, „žema kaina“, „patraukli pakuotė“, „kokybė“, „kainos ir kokybės santykis“. Pirmas kintamasis (respondentų lytis) pavadintas *Lytis* ir jam suteiktas skaitmeninis formatas (kintamojo reikšmės 1 ir 2), antras kintamasis pavadintas *Apsisprendimas*. Jam taip pat suteiktas skaitmeninis formatas. Šio kintamojo reikšmės 1, 2, 3, 4, 5.

Sudarysime porines dažnių lenteles.

PSPP terpė

- Atidarykite failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodykite komandas *Analizuoti* → *Aprašomoji statistika* → *Požymių priklausomumo lentelės...*
- Dialogo langelyje *Požymių priklausomumo lentelės* pažymime kintamąjį *Lytis* ir spragtelėję mygtuką su rodyklės ženklu įkeliame jį į laukelį *Eilutės*. Šis kintamasis bus dažnių lentelės eilutėse. Atitinkamai kintamąjį *Apsisprendimas* įkeliame į laukelį *Stulpeliai*. Šis kintamasis bus dažnių lentelės stulpeliuose.



4.30 pav. Dialogo langelis *Požymių priklausomumo lentelės*

- Spragtelėkite *Gerai*.

4.31 pav. rodomos absoliučios stebėjimų skaičiaus reikšmės ir jų dažnių procentinė išraiška.

Taip pat pateikiamos tam tikros statistikinės charakteristikos, kuriomis remiantis galima daryti išvadas apie nagrinėjamų kintamųjų ryšį. Tai bus nagrinėjama vėlesniuose skyriuose.

4.31 pav. rezultatai gaunami pagal PSPP standartinius nustatymus. Dialogo langelio *Požymių priklausomumo lentelės* mygtukais *Formatas…*, *Statistika…*, *Langeliai…* galima atsisakyti stebėjimų skaičiaus reikšmių dažnių procentinės išraiškos, statistinių charakteristikų.

Santrauka.

	Atvejų					
	Galiojančių		Praleista		Iš viso	
	N	Procentai	N	Procentai	N	Procentai
Lytis * Apsisprendimas	45	100.0%	0	0.0%	45	100.0%

Lytis * Apsisprendimas [kiekis, % eilutė, % stulpelis, iš viso %].

	Apsisprendimas					
Lytis	rūšis	žema kaina	patraukli pakuotė	kokybė	kainos ir kokybės santykis	Iš viso
Vyras	3,0	2,0	3,0	3,0	4,0	15,0
	20,0%	13,3%	20,0%	20,0%	26,7%	100,0%
	27,3%	22,2%	37,5%	33,3%	50,0%	33,3%
	6,7%	4,4%	6,7%	6,7%	8,9%	33,3%
Moteris	8,0	7,0	5,0	6,0	4,0	30,0
	26,7%	23,3%	16,7%	20,0%	13,3%	100,0%
	72,7%	77,8%	62,5%	66,7%	50,0%	66,7%
	17,8%	15,6%	11,1%	13,3%	8,9%	66,7%
Iš viso	11,0	9,0	8,0	9,0	8,0	45,0
	24,4%	20,0%	17,8%	20,0%	17,8%	100,0%
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
	24,4%	20,0%	17,8%	20,0%	17,8%	100,0%

Chi-kvadrato kriterij.

Statistika	*Reikšmė*	*skirt.*	*Asimp. p-reikšmė (2-pusė)*
Pirsono Chi-kvadratas	1,74	4	,78
Tikėtinumo santykis	1,73	4	,79
Dvitiesinė priklausomybė	1,18	1	,28
N galiojančių atvejų	45		

4.31 pav. PSPP paketas. Požymių dažnių analizės rezultatai

SPSS terpė

- Atidarykite failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Crosstabs...*
- Dialogo langelyje *Crosstabs* pažymime kintamąjį *Lytis* ir spragtelėję mygtuką su rodyklės ženklu įkeliame jį į laukelį *Row(s):*. Šis kintamasis bus dažnių lentelės eilutėse. Atitinkamai kintamąjį *Apsisprendimas* įkeliame į laukelį *Column(s):*. Šis kintamasis bus dažnių lentelės stulpeliuose.
- Spragtelėkite *OK*.

Row(s) sąraše yra 2 kintamieji, *Column(s)* – 5 kintamieji, gausime 2*5 = 10 dažnių lentelę.

Lytis * Apsisprendimas Crosstabulation							
Count							
		Apsisprendimas					Total
		1,00	2,00	3,00	4,00	5,00	
Lytis	1,00	3	2	3	3	4	15
	2,00	8	7	5	6	4	30
Total		11	9	8	9	8	45

4.32 pav. SPSS paketas. Standartinio formato požymių dažnių lentelė

Ši lentelė gaunama pagal SPSS standartinius nustatymus: rodomos absoliučios stebėjimų skaičiaus reikšmės. *Crosstabs* dialogo lange esančių mygtukais *Exact…*, *Statistics…*, *Cells…*, *Format…* galima lentelėje parodyti daugiau informacijos: procentinę dažnių išraišką, tikėtinas dažnių reikšmes ir kt.

Lyginant tikėtinas dažnių reikšmes su stebėtomis (eksperimentinėmis), galima padaryti tam tikras išvadas apie vieno kintamojo reikšmių priklausomybę nuo kito kintamojo. Tai bus nagrinėjama vėlesniuose skyriuose.

4.4 lentelė. Porinės dažnių analizės sintaksė

PSPP	SPSS
GET GET FILE="C:\...\Pav1_kava.sav". CROSSTABS CROSSTABS /TABLES=Lytis BY Apsisprendimas /FORMAT=AVALUE LABELS TABLES PIVOT /STATISTICS=CHISQ /CELLS=COUNT ROW COLUMN TOTAL.	GET FILE='C:\...\Pav1_kava.sav'. CROSSTABS /TABLES=Lytis BY Apsisprendimas /FORMAT=AVALUE TABLES /CELLS=COUNT /COUNT ROUND CELL.

4.4.3. Pagrindiniai grafikai: histogramos ir skritulinės diagramos

Statistinės programos siūlo įvairiausių tipų diagramas. Standartiniai grafikai sudaromi naudojantis vartotojo aplinkos meniu procedūromis. Toks diagramų sudarymo būdas yra labai lengvas. Pasirinkus norimą diagramos tipą, nurodomas kintamojo vardas ir pagal iš anksto nustatytą schemą iš kintamojo reikšmių yra sudaroma diagrama. Sudarytą diagramą lengva redaguoti. Redagavimo principas daugelyje šiuolaikinių statistinių programų yra vienodas. Norint pakeisti kurį nors užrašą, liniją ar spalvą, tereikia du kartus spustelėti kairįjį pelės klavišą ant norimo redaguoti elemento.

Dažnių lentelių rezultatus galima pavaizduoti grafiškai. Dažniai vaizduojami stulpeliais, kurių viršuje galima užrašyti apskaičiuotas dažnių reikšmes. Tokia dažnių diagrama vadinama stulpeline diagrama arba histograma.

Histogramomis vaizduojamas intervalų skalės kintamųjų pasiskirstymas. Kintamojo reikšmės yra suskirstomos į intervalus, nustatomas į šiuos intervalus patenkančių reikšmių skaičius (dažnis) ir rodikliai pavaizduojami stačiakampių stulpelių, kurių kiekvieno aukštis yra proporcingas atitinkamo intervalo duomenų dažniui, seka. Pagal pradinius nustatymus intervalų skaičių ir plotį programa parenka automatiškai, tačiau šiuos parametrus gali nustatyti ir vartotojas.

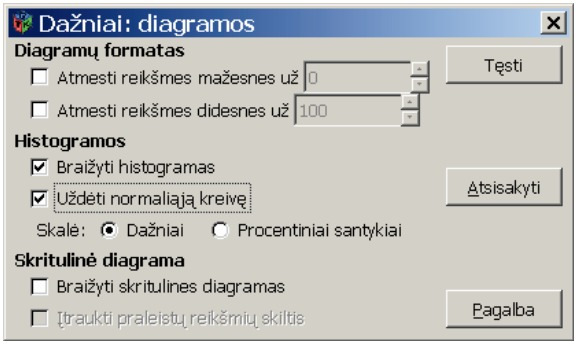
Duomenys, išmatuoti nominalioje skalėje, dažniausiai vaizduojami stulpelinėmis (angl. *bar charts*) ir skritulinėmis (angl. *pie charts*) diagramomis. Pastarosios ypač tinka, kai požymio reikšmių skaičius yra nedidelis.

Pavyzdys. Apklausoje buvo tirta, kaip dažnai per mėnesį respondentai perka kavą. Kintamasis *Dažnis*. Pasitelkę PSPP/SPSS pavaizduosime šį kintamąjį grafiškai.

PSPP terpė

Šiuo metu PSPP pakete galima brėžti tik histogramas ir skritulines diagramas. Grafinė vartotojo sąsaja jų redaguoti neleidžia. Ateityje ketinama grafinę vartotojo sąsają tobulinti.

- Atidarome failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodome komandas *Analizuoti* → *Aprašomoji statistika* → *Dažniai...*
- Dialogo langelyje *Dažniai* pažymime analizuojamą kintamąjį *Dažnis* ir spragtelėję mygtuką su rodyklės ženklu įkeliame jį į analizuojamų kintamųjų sąrašą *Kintamasis(-ieji):*.
- Spragtelėkite mygtuką *Diagramos...* Atsidarys dialogo langelis *Dažniai: diagramos*
- Pasirinkite *Braižyti histogramos* ir *Uždėti normaliąją kreivę*. Skalė *Dažniai* (žr. 4.33 pav.).



4.33 pav. Dialogo langelis *Dažniai: diagramos*

- Spragtelėkite mygtuką *Tęsti*.
- Spragtelėkite mygtuką *Gerai*.

Rodinio *Rezultatai – PSPPIRE Rezultatų peržiūra* lange pasirodys kintamojo *Dažnis* dažnių lentelė ir histograma (žr. 4.34 pav.).

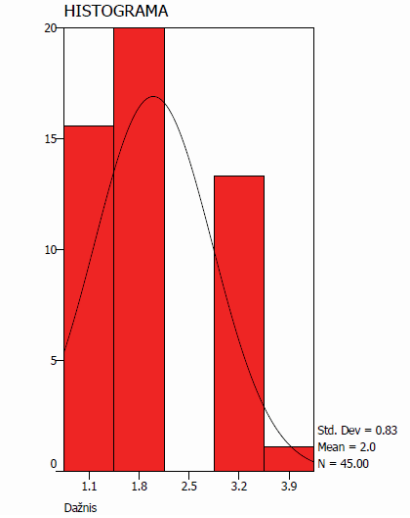
Langelyje *Dažniai: diagramos* pasirinkę *Braižyti skritulines diagramas*, gauname kintamojo *Dažnis* skritulinę diagramą (žr. 4.35 pav.).

Dažnis

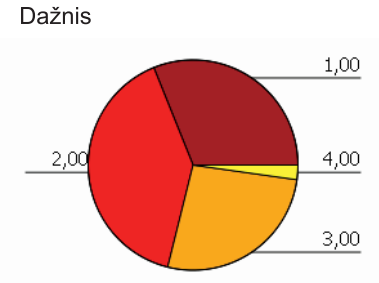
Reikšmės etiketė	Reikšmė	Dažnis	Procentai	Galiojanti procentinė dalis	Sukauptoji procentinė dalis
	1,00	14	31,11	31,11	31,11
	2,00	18	40,00	40,00	71,11
	3,00	12	26,67	26,67	97,78
	4,00	1	2,22	2,22	100,00
	Iš viso	45	100,0	100,0	

Dažnis

N	Galiojančių	45
	Praleista	0
Vidurkis		2,00
Std. nuokrypis		,83
Mažiausia		1,00
Didžiausia		4,00



4.34 pav. PSPP paketas. Kintamojo *Dažnis* dažnių lentelė ir histograma



4.35 pav. PSPP paketas. Kintamojo *Dažnis* skritulinė diagrama

SPSS terpė

- Atidarome failą su analizuojamais duomenimis *Pav1_kava.sav*.
- Nurodome komandas *Analyze* → *Descriptive Statistics* → *Frequencies...*
- Dialogo langelyje *Frequencies* pažymime analizuojamą kintamąjį *Dažnis* ir spragtelėję mygtuką su rodyklės ženklu įkeliame jį į analizuojamų kintamųjų sąrašą *Variable(s)*.
- Spragtelėkite mygtuką *Charts...*(diagramos). Atsidarys dialogo langelis *Frequencies: Charts*.
- Pasirinkite diagramos tipą *Histograms:*, pažymėkite laukelį *With normal curve*, spragtelėkite dialogo langelio *Frequencies: Charts* mygtuką *Continue*, tada — dialogo langelio *Frequencies* mygtuką *OK*.

Rodinio (*Viewer*) lange pasirodys Jūsų pasirinkta diagrama.

Dialogo langelyje *Frequencies: Charts* grupėje *Chart Type* pasirinkus *Pie charts,* kintamasis *Dažnis* bus atvaizduotas skritulinėje diagramoje (žr. 4.37 pav.).

4.5 lentelė. Histogramos brėžimo sintaksė

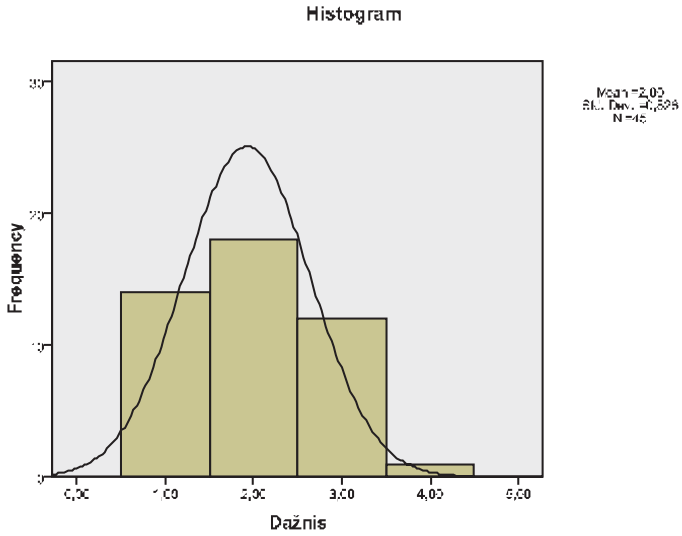
PSPP	SPSS
FREQUENCIES FREQUENCIES /VARIABLES=Dažnis /FORMAT=AVALUE TABLE /HISTOGRAM=NORMAL.	FREQUENCIES VARIABLES=Dažnis /HISTOGRAM NORMAL /ORDER=ANALYSIS.

4.6 lentelė. Skritulinės diagramos brėžimo sintaksė

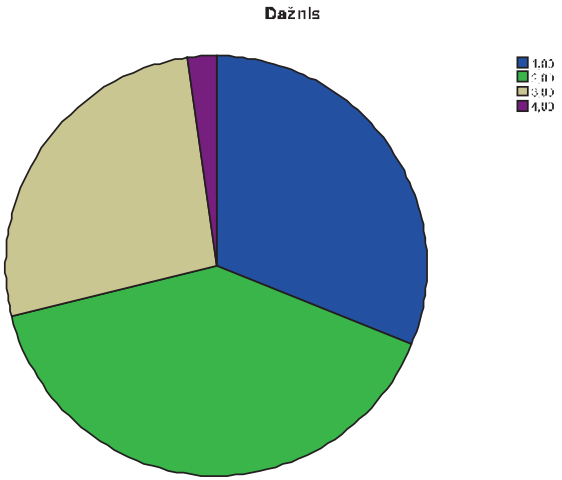
PSPP	SPSS
FREQUENCIES FREQUENCIES /VARIABLES=Dažnis /FORMAT=AVALUE TABLE /PIECHART=.	FREQUENCIES VARIABLES=Dažnis /PIECHART FREQ /ORDER=ANALYSIS.

Statistics		
Dažnis		
N	Valid	45
	Missing	0

Dažnis					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	14	31,1	31,1	31,1
	2,00	18	40,0	40,0	71,1
	3,00	12	26,7	26,7	97,8
	4,00	1	2,2	2,2	100,0
	Total	45	100,0	100,0	



4.36 pav. SPSS paketas. Dažnių lentelė ir histograma *Viewer* lange



4.37 pav. SPSS paketas. Skritulinė diagrama

4.4.4. Padėties ir sklaidos charakteristikos

Kintamiesiems, išmatuotiems intervalų ar santykių skalėmis ir turintiems normalųjį skirstinį, dažniausiai skaičiuojamas vidurkis ir standartinis nuokrypis. Kartais skaičiuojamas vidurkio standartinis nuokrypis (standartinė klaida). Statistinės programos pateikia daug galimų duomenų padėties ir sklaidos charakteristikos skaičiavimo pasirinkimo galimybių. Tačiau pateikiant skaičiavimo rezultatus patartina apsiriboti viena, geriausiai duomenims tinkančia, centro charakteristika (vidurkiu arba mediana) ir viena sklaidos charakteristika (dažniausiai pateikiamas standartinis nuokrypis arba standartinė klaida).

Statistinės duomenų analizės programos centro ir sklaidos charakteristikas dažniausiai skirsto į kelias grupes.

Pirmąją grupę sudaro šios duomenų padėties (centro) charakteristikos

Vidurkis arba **aritmetinis vidurkis** (angl. *Mean*) – tai populiariausias duomenų vidurio (centro) matas. Jis apskaičiuojamas susumuojant visas reikšmes ir padalijant iš reikšmių skaičius. Vidurkis,

apskaičiuotas iš imties, yra empirinis ir žymimas $\bar{x}$. Skirtingai nuo empirinio vidurkio, populiacijos vidurkis žymimas raide μ. Vidurkis yra kiekybinis rodiklis, kuris apibendrina ir išreiškia vienarūšių reiškinių tam tikro varijuojančio požymio tipišką lygį konkrečiomis vietos ir laiko sąlygomis.

Tegul tam tikrą objektų visumą tiriame pagal mus dominantį požymį X, o šio požymio galimos reikšmės (variantai) visumos objektams yra x_1, x_2, x_3, …, x_n. Aritmetinis vidurkis skaičiuojamas pagal formulę

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum x_i}{n}. \qquad (4.2)$$

Tuo atveju, jei variantų reikšmės kartojasi, skaičiuojamas svertinis aritmetinis vidurkis

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum x_i f_i}{\sum f_i}. \qquad (4.3)$$

Čia: f_i – yra reikšmės x_i pasikartojimų skaičius – dažniai.

Pavyzdys.

Tirtas studentų kelionės į universitetą laikas.

Laikas, min.	*0–5*	*5–10*	*10–15*	*15–20*	*20–25*	*25–30*
Studentų skaičius	12	4	3	15	10	6

Apskaičiuosime, kiek vidutiniškai užtrunka studento kelionė į universitetą.

Tuo atveju, kai požymio reikšmės yra duotos intervalais, variantų reikšmėmis imamos intervalų vidurio reikšmės

$$x_i = \frac{x_{i1} + x_{i2}}{2}, \qquad (4.4)$$

x_{i1} – apatinė intervalo reikšmė, x_{i2} – viršutinė intervalo reikšmė.

Pertvarkysime duotų intervalų pasiskirstymo eilutę į diskrečiąją, variantų reikšmėmis imdami intervalo vidurio taškus.

x_i	2,5	7,5	12,5	17,5	22,5	27,5
f_i	12	4	3	15	10	6

Skaičiuojame svertinį vidurkį:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{2,5*12 + 7,5*4 + 12,5*3 + 17,5*15 + 22,5*10 + 27,6*6}{12 + 4 + 3 + 15 + 10 + 6} = 15.$$

Mediana (angl. *Median*) – tai matavimų skalės taškas, kuris stebimas reikšmes padalija į dvi lygias dalis. Pavyzdžiui, jei išmatavę gavome reikšmes

7 1 8 8 9 1 6 8 5,

tai pirmiausia jas reikia išrikiuoti didėjimo tvarka, t. y. sudaryti ranžiruotą variacinę eilutę:

1 1 5 6 7 8 8 8 9

ir surasti vidurinę reikšmę (šiuo atveju 7). Jei stebėjimų skaičius yra lyginis, tai surandamas dviejų vidurinių reikšmių aritmetinis vidurkis. Kurių narių vidurkį imti randame pagal formulę (N+1)/2.

Pavyzdžiui, jei

N = 10. Pagal formulę 11/2 = 5,5. Mediana bus 5 ir 6 ranžiruotos eilutės narių aritmetinis vidurkis.

N = 9. Pagal formulę 10/2 = 5. Mediana bus 5 ranžiruotos eilutės narys.

Moda (angl. *Mode*) – tai dažniausiai pasikartojanti reikšmė. Jeigu kelios reikšmės kartojasi vienodą skaičių kartų, tai pasirenkama mažiausia reikšmė.

Antrąją grupę sudaro duomenų sklaidos charakteristikos

Požymio atskirų reikšmių skirtumai tam tikros visumos viduje statistikoje vadinami požymio sklaida. Visumos elementai apibūdinami kiekybinėmis ir kokybinėmis požymių reikšmėmis. Sklaida statistikoje suprantama kaip nagrinėjamą reiškinį apibūdinančio požymio dydžio kiekybiniai pokyčiai dėl skirtingų veiksnių neprognozuojamos įtakos. Kaip jau minėta, būdingiausius visumos bruožus apibūdina vidurkis. Sklaidos rodikliai įvertina požymio reikšmių išsisklaidymo apie vidurkį dydį ir amplitudę.

Sklaidos plotis – tai didžiausios ir mažiausios požymio reikšmės skirtumas.

$$R = x_{\max} - x_{\min}, \tag{4.5}$$

čia x_{max} – didžiausia požymio reikšmė, x_{min} – mažiausia požymio reikšmė.

Standartinis nuokrypis (angl. *Standard Deviation*) – tai matuojamų dydžių sklaidos apie vidurkį matas. Jis apskaičiuojamas ištraukiant kvadratinę šaknį iš dispersijos. Imties (empirinis) standartinis nuokrypis žymimas raide s, o populiacijos standartinis nuokrypis – raide σ. Jei duomenų skirstinys yra normalusis, tai į intervalą, kuris gaunamas atidėjus į abi puses nuo vidurkio po standartinį nuokrypį s $(\bar{x} - s, \bar{x} + s)$, patenka apytiksliai 68 proc. visų duomenų. Į intervalą $(\bar{x} - 2s, \bar{x} + 2s)$ patenka apytikriai 95,4 proc. visų duomenų, o į intervalą $(\bar{x} - 3s, \bar{x} + 3s)$ patenka beveik visi duomenys (tiksliau, 99,7 proc.).

$$s = \sqrt{s^2}. \tag{4.6}$$

Dispersija (angl. *Variance*) apskaičiuojama susumuojant kiekvienos reikšmės atstumo (skirtumo) nuo vidurkio kvadratus ir gautą sumą padalijant iš reikšmių skaičiaus minus 1. Dispersija, kaip ir standartinis nuokrypis, parodo duomenų sklaidą apie vidurkį. Kadangi dispersija apskaičiuojama keliant reikšmes kvadratu, tai ir dispersijos matavimo vienetai yra pakelti kvadratu. Standartinis nuokrypis, kuris yra šaknis iš dispersijos, matuojamas tokiais pačiais matavimo vienetais kaip ir patys duomenys, todėl standartinį nuokrypį lengviau suprasti ir interpretuoti.

Dispersija

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}. \tag{4.7}$$

Sugrupuotiems duomenims

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}. \tag{4.8}$$

Vidurkio standartinis nuokrypis arba **standartinė klaida** (angl. *Standard Error of Mean*) yra apskaičiuojamas imties standartinį nuokrypį s padalijant iš kvadratinės šaknies, ištrauktos iš visų duomenų

skaičiaus (imties tūrio n). Vidurkio standartinis nuokrypis, kaip ir imties standartinis nuokrypis, padeda nustatyti atskirų reikšmių išsibarstymo intervalą. Su 99,7 proc. tikimybe galime teigti, kad populiacijos vidurkis yra intervale, kuris gaunamas atidėjus į abi puses nuo vidurkio po tris vidurkio standartinius nuokrypius $(\overline{x}-3s,\overline{x}+3s)$; su 95,4 proc. tikimybe galime teigti, kad populiacijos vidurkis yra intervale, gautame atidėjus į abi puses nuo vidurkio po du vidurkio standartinius nuokrypius $(\overline{x}-2s,\overline{x}+2s)$; o jei nuo vidurkio atidėsime po vieną vidurkio standartinį nuokrypį $(\overline{x}-s,\overline{x}+s)$, tai tikimybė, kad į šį intervalą pateks populiacijos vidurkis, yra apytikriai 68 proc.

Vidurkio patikimumo intervalas (angl. *ConfidenceLimits of Mean*). Vidurkio standartinis nuokrypis arba standartinė klaida leidžia apskaičiuoti vidurkio patikimumo (pasikliautinąjį) intervalą. Su 95,4 proc. tikimybe galime teigti, kad intervale, kurį sudaro du vidurkio standartiniai nuokrypiai, atidėti į abi puses nuo imties vidurkio, bus populiacijos vidurkis. Jei atidėsime ne po 2, o po 1,96 vidurkio standartinius nuokrypius, gausime tikslią 95 proc. tikimybę. Šis intervalas vadinamas 95 proc. vidurkio patikimumo (pasikliautinuoju) intervalu.

Variacijos (kitimo) koeficientas. Šis rodiklis nustato sklaidos laipsnį:

$$CV = \frac{s}{x} * 100\%. \tag{4.9}$$

Vertinama taip: iki 10 % – sklaida maža; 10–20 % – sklaida vidutinė; 20–30 % – sklaida didelė; 30 % ir daugiau – sklaida labai didelė.

Į atskirą grupę (trečiąją) išskirti procentiliai

Skaičiai, kurie suskirsto variacinę eilutę į 100 apytikriai lygių dalių, yra vadinami *procentiliais*, o skaičiai, dalijantys variacinę eilutę į keturias maždaug lygias dalis, vadinami *kvartiliais*. Dažniausiai kvartiliai žymimi raidėmis Q1, Q2, Q3. Pirmasis kvartilis Q1 – tai toks taškas išmatuotų reikšmių skalėje, nuo kurio kairėje pusėje lieka 25 proc. reikšmių. Antrasis kvartilis Q2 – tai taškas, kurio kairėje pusėje yra 50 proc. reikšmių. Antrasis kvartilis sutampa su mediana. Trečiojo kvartilio Q3 kairėje pusėje yra 75 proc. reikšmių.

Kaip viena iš galimų duomenų centro charakteristikų naudojamas pirmojo ir trečiojo kvartilių vidurkis $(Q_1 + Q_3)/2$.

Skaičiai, kurie duomenų eilutę sudalija į 10 lygių dalių, vadinami *deciliais*.

Ketvirtąją grupę sudaro du asimetriškumo matai

Asimetrijos koeficientas (angl. *Skewness*) matuoja, ar duomenų išsidėstymas skiriasi nuo simetrinio išsidėstymo, t. y. nuo tokio duomenų išsidėstymo, kai reikšmės išsidėsto vienodais atstumais į abi puses nuo vidurkio. Jei duomenys turi normalųjį skirstinį, tai asimetrijos koeficientas lygus nuliui. Asimetrijos koeficientą galima naudoti kaip vieną iš duomenų normališkumo tikrinimo būdų. Jei asimetrijos koeficientas reikšmingai skiriasi nuo nulio, tai duomenų skirstinys nėra normalusis. Kai asimetrijos koeficientas didesnis už nulį, asimetrija teigiama, kai mažesnis – neigiama.

Asimetrijos koeficientas skaičiuojamas pagal formulę

$$As = \frac{1}{\frac{n-1}{s^3}} \sum (x_i - \bar{x})^3. \tag{4.10}$$

Eksceso koeficientas (*Kurtosis*) parodo ar histograma yra aukšta su trumpomis „uodegomis“, ar, atvirkščiai, – plokščia su ilgomis uodegomis. Eksceso koeficientas parinktas taip, kad tuo atveju, kai duomenų skirstinys yra normalusis, eksceso koeficientas lygus nuliui.

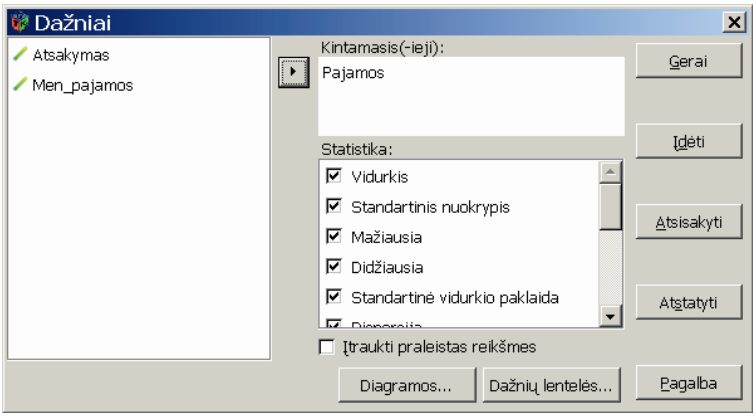
$$Ex = \frac{1}{\frac{n-1}{s^4}} \sum (x_i - \bar{x})^4 - 3. \tag{4.11}$$

Pasitelkę PSPP panagrinėsime kintamųjų padėties ir sklaidos charakteristikas. Drauge pateiksime skaičiavimų rezultatus, gautus pasitelkus SPSS paketą ir *MS Excel* skaičiuoklę.

Kaip pavyzdį imkime apklausą, kurioje sukaupti respondentų atsakymai apie jų gaunamas pajamas – intervalinis kintamasis *Pajamos*.

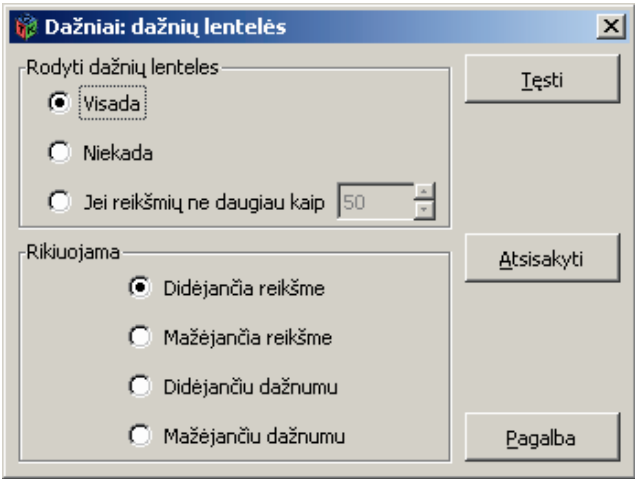
PSPP terpė

- Atidarykite failą *Apklausa_pajamos.sav* su analizuojamais duomenimis.
- Nurodykite komandas *Analizuoti* → *Aprašomoji statistika* → *Dažniai...*
- Įkelkite kintamąjį *Pajamos* į kintamųjų sąrašą *Kintamasis (-ieji)* (4.38 pav.).



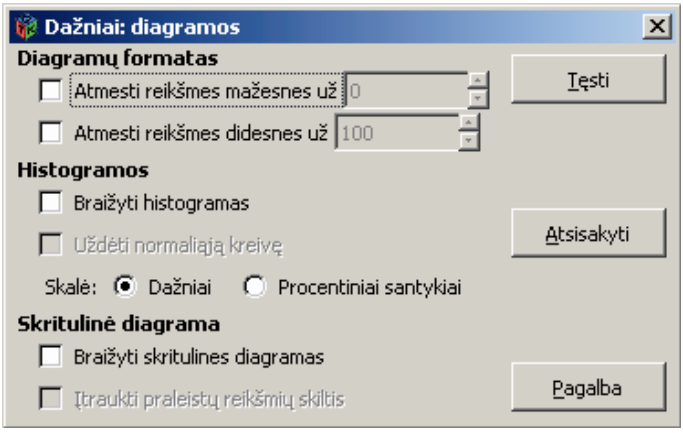
4.38 pav. Kintamasis *Pajamos* dialogo lange *Dažniai*

- Grupėje *Statistika*: pasirinkite (pažymėkite varnele) visas sąraše esančias statistines charakteristikas: *Vidurkis*, *Standartinis nuokrypis*, *Mažiausia*, *Didžiausia*, *Standartinė vidurkio paklaida*, *Dispersija*, *Asimetrija*, *Standartinė asimetrijos paklaida*, *Sritis*, *Veiksena*, *Ekscesas*, *Standartinė ekscesо paklaida*, *Mediana*, *Suma*.
- Spragtelėkite mygtuką *Dažnių lentelės...* Bus atvertas dialogo langelis *Dažniai: dažnių lentelės* (4.39 pav.). Čia galima pasirinkti, kaip ir kada rodyti dažnių lenteles ir duomenų rikiavimo variantus.



4.39 pav. Dialogo langelis *Dažniai: dažnių lentelės*

- Spragtelėkite mygtuką *Tęsti*. Grįšite į dialogo langelį *Dažniai*.
- Spragtelėkite mygtuką *Charts*... Bus atvertas dialogo langelis *Dažniai: diagramos* (4.40 pav.).



4.40 pav. Dialogo langelis *Dažniai: diagramos*

- Čia galima pasirinkti diagramų formatą ir diagramos tipą: histograma ar skritulinė diagrama. Kitų diagramos tipų PSPP pakete šiuo metu pasirinkti negalima. Pasirinkite *Braižyti histo-*

gramas ir spragtelėkite mygtuką *Tęsti*. Grįšite į dialogo langelį *Dažniai.*

- Spragtelėkite mygtuką *Gerai*.
- Rezultatų faile galime išskirti atliktos procedūros sintaksės dalį (žr. 4.7 lentelė.), kintamojo *Pajamos* dažnių lentelę (4.41 pav.), kintamojo *Pajamos* statistikinių charakteristikų lentelę (4.42 pav.) ir kintamojo *Pajamos* histogramą (4.43 pav.):

Pajamos

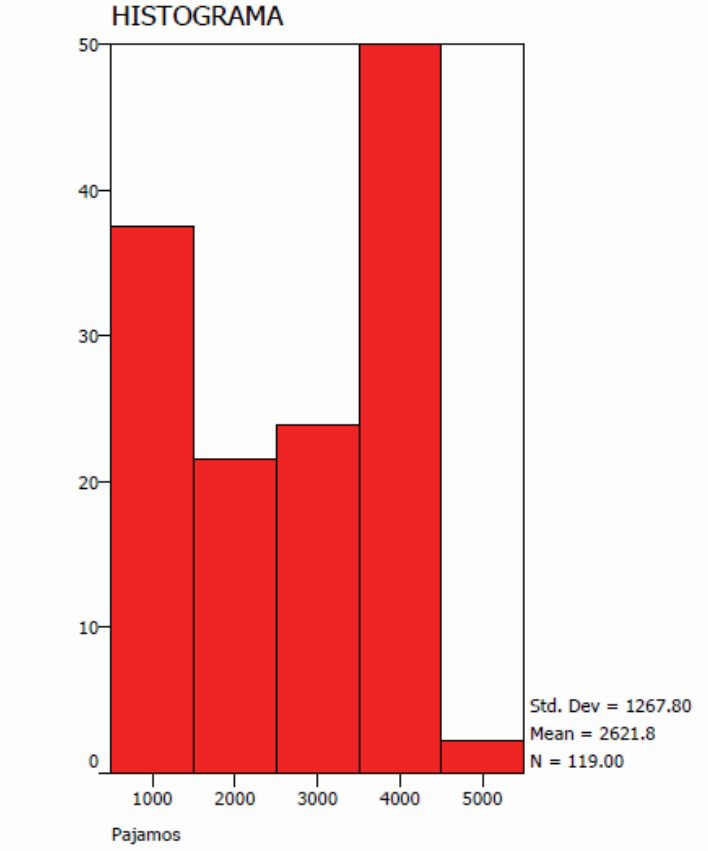
Reikšmės etiketė	Reikšmė	Dažnis	Procentai	Galiojanti procentinė dalis	Sukauptoji procentinė dalis
	800	17	14,29	14,29	14,29
	1200	16	13,45	13,45	27,73
	2000	19	15,97	15,97	43,70
	2800	21	17,65	17,65	61,34
	3600	27	22,69	22,69	84,03
	4400	17	14,29	14,29	98,32
	5200	2	1,68	1,68	100,00
	Iš viso	119	100,0	100,0	

4.41 pav. PSPP rezultatų failas. Kintamojo *Pajamos* dažnių lentelė

Pajamos

N	Galiojančių	119
	Praleista	0
Vidurkis		2621,85
Vidur. s.pakl.		116,22
Veiksena		3600,00
Std. nuokrypis		1267,80
Dispersija		1607315
Ekscesas		-1,22
Eksc. s.pakl.		,44
Asimetrija		,01
Asim. s.pakl.		,22
Sritis		4400,00
Mažiausia		800,00
Didžiausia		5200,00
Suma		312000,0
	50 (mediana)	2800

4.42 pav. PSPP rezultatų failas. Kintamojo *Pajamos* statistikinės charakteristikos



4.43 pav. PSPP rezultatų failas. Kintamojo *Pajamos* histograma

SPSS terpė

- Atverkite failą *Apklausa_pajamos.sav* su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Frequencies...*
- Įkelkite kintamąjį *Pajamos* į kintamųjų sąrašą *Variable(s).*
- Paspauskite dialogo langelio *Frequencies* mygtuką *Statistics…* Pasirodžiusiame dialogo langelyje *Frequencies: Statistics* pažymėkite visus *Central Tendency*, *Dispersion* ir *Distribution* žymimuosius rodiklių langelius ir spragtelėkite mygtuką *Continue.*

- Paspauskite dialogo langelio *Frequencies* mygtuką *Charts…* Pasirodžiusiame dialogo langelyje *Frequencies: Charts* pažymėkite *Histograms* ir spragtelėkite mygtuką *Continue*
- Lange *Frequencies* paspauskite mygtuką *OK*.

Rezultatai matomi rodinio *Viewer* lange (žr. 4.44 pav., 4.45 pav., 4.46 pav.).

Pajamos

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	800	17	14,3	14,3	14,3
	1200	16	13,4	13,4	27,7
	2000	19	16,0	16,0	43,7
	2800	21	17,6	17,6	61,3
	3600	27	22,7	22,7	84,0
	4400	17	14,3	14,3	98,3
	5200	2	1,7	1,7	100,0
	Total	119	100,0	100,0	

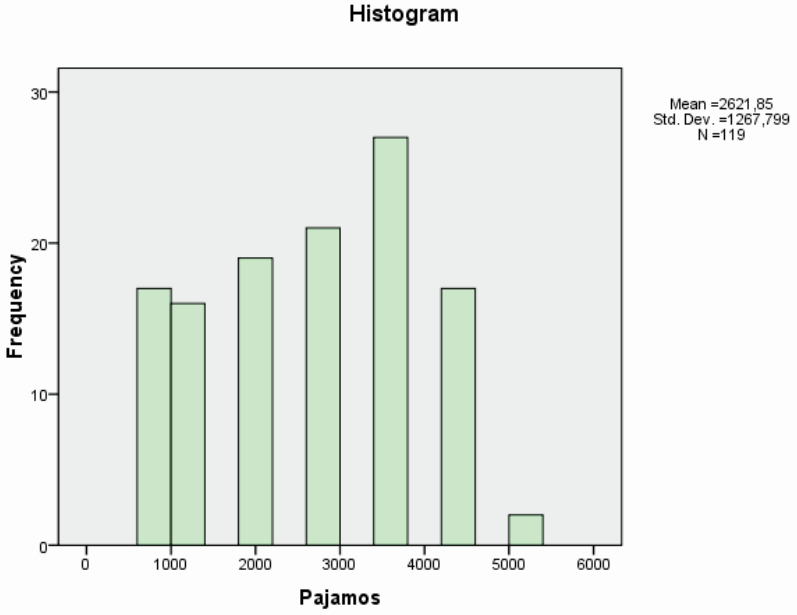
4.44 pav. SPSS rezultatų failas. Kintamojo *Pajamos* dažnių lentelė

Statistics

Pajamos

N	Valid	119
	Missing	0
Mean		2621,85
Std. Error of Mean		116,219
Median		2800,00
Mode		3600
Std. Deviation		1267,799
Variance		1607315,197
Skewness		,008
Std. Error of Skewness		,222
Kurtosis		-1,219
Std. Error of Kurtosis		,440
Range		4400
Minimum		800
Maximum		5200
Sum		312000

4.45 pav. SPSS rezultatų failas. Kintamojo *Pajamos* statistikinės charakteristikos



4.46 pav. SPSS rezultatų failas. Kintamojo *Pajamos* histograma

4.7 lentelė. Aprašomosios statistikos sintaksė

PSPP	SPSS
FREQUENCIES FREQUENCIES /VARIABLES=Pajamos /FORMAT=AVALUE TABLE /STATISTICS=ALL /HISTOGRAM=NONORMAL.	FREQUENCIES VARIABLES=Pajamos /STATISTICS=STDEV VARIANCE RANGE MINIMUM MAXIMUM STMEAN MEAN MEDIAN MODE SUM SKEWNESS SESKEW KURTOSIS SEKURT /HISTOGRAM /ORDER=ANALYSIS.

Skaičiuoklės MS Excel terpė

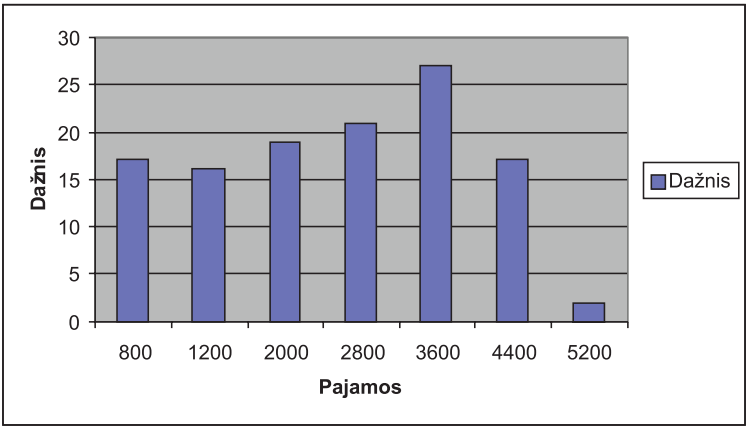
- Atverkite failą *Men_pajamos.xls* su analizuojamais duomenimis.
- Pasirinkite įrankį *Data Analysis* įrankius (jų buvimo vieta priklauso nuo Excel versijos).
- Įrankių sąraše pažymėkite *Descriptive Statistics* ir spragtelėkite mygtuką *OK*.
- Atsivėrusiame dialogo lange *Descriptive Statistics* nurodykite kintamojo *Pajamos* koordinates, rezultatų lentelės koordina-

tes, pažymėkite, kad kintamojo vardas yra pirmoje eilutėje, ir spragtelėkite *OK*.

Pajamos	
Mean	2621,848739
Standard Error	116,2189761
Median	2800
Mode	3600
Standard Deviation	1267,799352
Sample Variance	1607315,197
Kurtosis	-1,218973429
Skewness	0,007948162
Range	4400
Minimum	800
Maximum	5200
Sum	312000
Count	119

4.47 pav. MS Excel. Kintamojo *Pajamos* statistikinės charakteristikos

Kintamojo *Pajamos* histograma brėžiama įprastai, naudojantis *Excel* komandomis *Insert* → *Chart*…



4.48 pav. *MS Excel.* Kintamojo *Pajamos* histograma

Kaip matyti iš skaičiavimų PSPP, SPSS bei *Excel*, rezultatai sutampa.

KONTROLINIAI KLAUSIMAI

4.1. Kokie yra pagrindiniai aprašomosios statistikos tikslai (uždaviniai)?

4.2. Kokie aprašomosios statistikos metodai naudojami kintamiesiems, išmatuotiems nominalioje matavimo skalėje?

4.3. Kokie aprašomosios statistikos metodai naudojami kintamiesiems, išmatuotiems intervalinėje matavimo skalėje?

4.4. Koks skirstinys vadinamas normaliuoju?

4.5. SPSS programa tikrindami tyrimo rezultatų stebimąjį reikšmingumą, gavome Sig. = 0,01. Esame pasirinkę reikšmingumo lygmenį 0,05. Kokią išvadą galime padaryti apie tyrimo rezultatų skirstinio normalumą?

4.6. Kada vidurkis lygus medianai?

4.7. Tarkime, kad sklaidos plotis yra 105,4, o standartinis nukrypimas yra 260,6. Kokią galite padaryti išvadą?

4.8. Gavote dispersiją, lygią –45,5. Kokią galite padaryti išvadą?

UŽDUOTYS

4.1. Pridedame po 100 prie kiekvieno sekos nario. Kas atsitiks standartiniam nuokrypiui, sklaidos pločiui, vidurkiui, medianai?

4.2. Lentelėje duotos planetų charakteristikos. Raskite medianą kiekvienam dydžiui. Jei kas sakytų, kad gavo diametro medianą 88838, kokią klaidą jis padarė?

Planeta	**Atstumas nuo Saulės (mln. mylių)**	**Skersmuo (myliomis)**	**Metai (dienomis)**
Merkurijus	36	3030	88
Venera	67	7520	225
Žemė	93	7926	365
Marsas	142	4217	687
Jupiteris	484	88 838	4332
Saturnas	887	74 896	10 760
Uranas	1765	31 762	30 684
Neptūnas	2791	30 774	30 188
Plutonas	3654	1428	90 467

5. STATISTINĖS HIPOTEZĖS

5.1. IMTIES SKIRSTINIAI. ĮVERČIAI

Pagrindinė statistikos problema – remiantis imtimi gauti išvadas apie visą populiaciją. Bendrasis atsitiktinės imties modelis aprašomas taip. Tarkime, kad n kartų matuojame atsitiktinį dydį X. Tuomet atsitiktinę imtį sudaro atsitiktinis vektorius (C), kurio visi atsitiktiniai dydžiai $X_1, X_2, ..., X_n$ yra nepriklausomi, vienodai pasiskirstę ir turintys tą patį kaip ir išmatuotas dydis X skirstinį. Atsitiktinė imtis $X_1, X_2, ..., X_n$ ir bet kuri jos funkcija yra atsitiktiniai dydžiai. Statistinėms išvadoms naudojama kokia nors imties duomenų funkcija. Atsitiktinės imties funkcija $f(X_1, X_2, ..., X_n)$ vadinama *statistika*. Statistikos realizacija (konkretus skaičius) skiriasi nuo pačios statistikos (atsitiktinio dydžio). Statistiką žymėsime didžiosiomis raidėmis, o jos realizaciją – mažosiomis. Pavyzdžiui, $\overline{X}$ yra statistika, o $\bar{x}$ – realizacija.

Dažniausiai mus domina ne visa informacija apie stebimojo atsitiktinio dydžio skirstinį, o tik kai kurios skaitinės skirstinio charakteristikos. Kitaip sakant, remdamiesi imties duomenimis, norime įvertinti populiacijos parametrą. Matematiškai tokią situaciją aprašo parametrinis modelis. Parametrinis modelis $X \sim P_\theta$: stebimasis atsitiktinis dydis X, kurio skirstinys P_θ priklauso nuo nežinomo parametro θ iš aibės Θ. Statistika, kuri naudojama nežinomam parametrui θ įvertinti, vadinama θ taškiniu įverčiu (toliu – įverčiu) ir žymima $\hat{\theta}$. Nežinomas parametras θ yra skaičius, parametro įvertis $\hat{\theta}$ yra atsitiktinis dydis, įverčio realizacija yra skaičius, randamas konkrečiai imties realizacijai.

Parametro įverčio realizacija – tai apytikslė parametro reikšmė. Dažnai prireikia populiacijos parametrą palyginti su konkrečiu skaičiumi arba kitos populiacijos analogišku parametru. Taikant matematinės statistikos metodus galima remiantis imtimi priimti sprendimą apie nežinomą populiacijos parametro reikšmę.

5.2. STATISTINIŲ HIPOTEZIŲ TIKRINIMAS

Hipotezė – tai teiginys, aprašantis ir aiškinantis reiškinį, kurio teisingumas iš anksto nežinomas. Statistinė hipotezė – tai prielaida apie empirinį atsitiktinio dydžio pasiskirstymo dėsnį ir (arba) apie jo empirines skaitines charakteristikas. Statistinės hipotezės – tai teiginiai apie masinius, pasikartojančius reiškinius. Statistinė hipotezė gali būti patikrinta naudojant statistinius kriterijus. Statistiniai kriterijai pagrindžiami matematinės statistikos metodais. Formuluojant hipotezes tyrinėtojo nuomonė sistematizuojama ir pateikiama konkrečiu pavidalu.

Hipotezių tikrinimo procedūrą sudaro šie pagrindiniai žingsniai:

1. Imties skirstinio tipo nustatymas.
2. Hipotezių formulavimas.
3. Statistinio kriterijaus skaičiavimas.
4. Statistinio reikšmingumo lygmens parinkimas.
5. Sprendimo priėmimas.

1. Imties skirstinio tipo nustatymas. Atliekant statistinį tyrimą gaunami kintamojo X stebėjimo rezultatai $x_1, x_2, x_3, \ldots, x_i, \ldots x_n$ (n – imties dydis). Sąryšis, siejantis atsitiktinio dydžio reikšmes ir jų dažnius, vadinamas atsitiktinio dydžio skirstiniu. Prieš parenkant statistinį kriterijų svarbu žinoti, ar išmatuotų duomenų skirstinys yra normalusis. Normalusis arba Gauso skirstinys grafiškai turi varpo formą. Gauso skirstinio tikimybės tankis aprašomas formule:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (5.1)$$

čia σ – dispersija, μ – vidurkis, $e = 2{,}71$ – Neperio skaičius.

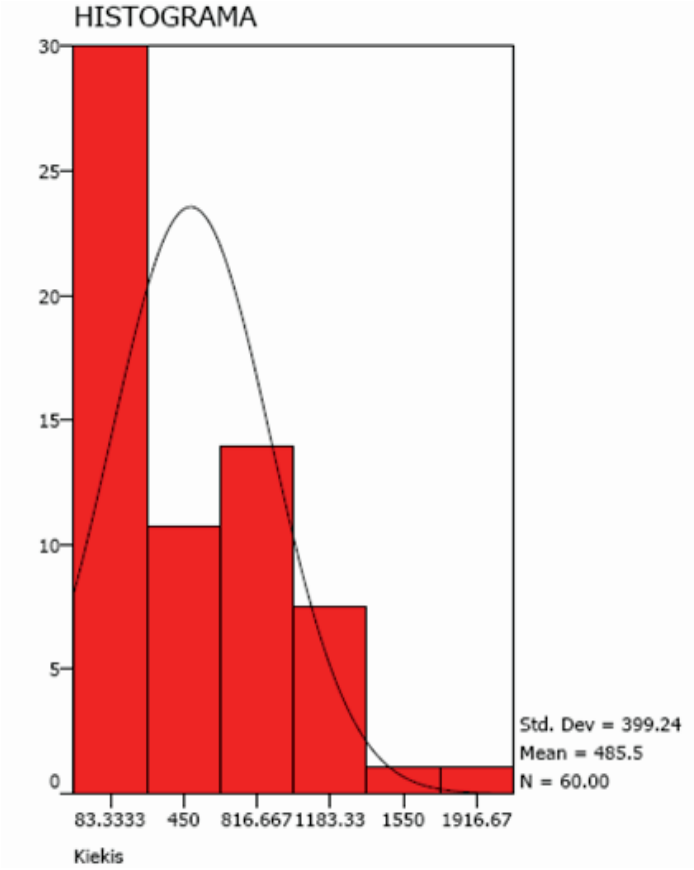
Skirstinio normalumas ypač svarbus mažų imčių atveju. Nuo imties skirstinio tipo priklauso statistinio kriterijaus parinkimas.

KINTAMOJO SKIRSTINIO NORMALUMO TIKRINIMAS

PSPP terpė

PSPP terpėje nenumatyta speciali imties normalumo tikrinimo procedūra. Galima pritaikyti tik nagrinėjamo kintamojo histogramos ir normaliosios kreivės palyginimą.

Komandos *Analizuoti* → *Aprašomoji statistika* → *Dažniai* atveria dialogo langą *Dažniai.* Mygtukas *Diagramos* atveria langą *Dažniai: diagramos.* Šiame lange reikia pažymėti *Histogramos* → *Braižyti histogramos* ir *Uždėti normaliąją kreivę.* Gautas rezultatas pateiktas 5.1 pav.



5.1 pav. Normalumo tikrinimas PSPP terpėje

SPSS terpė

Kiekybinio kintamojo skirstinio normalumą galima patikrinti taikant aprašomosios statistikos procedūrą *Explore.*

Patikrinsime, ar analizuojamų įmonių darbuotojų skaičius turi normalų skirstinį. Dialogo langas *Explore* atveriamas komanda *Analyze → Descriptive Statistics → Explore*). Kintamasis *Darbuotoju_sk* perkeliamas į langelį *Dependent List.* Mygtuko *Plots* paspaudimas atveria dialogo langą *Explore: Plots.* Šiame lange reikia pažymėti langelį *Normality plots with tests.* Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 5.2 pav.).

Tests of Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Darbuotojų_sk	,117	30	,200(*)	,974	30	,644

* This is a lower bound of the true significance.
a Lilliefors Significance Correction

5.2 pav. Imties skirstinio normalumo tikrinimo rezultatas

Lentelėje *Tests of Normality* pateikti Kolmogorovo–Smirnovo *(Kolmogorov-Smirnov(a))*ir Šapiro–Vilko *(Shapiro-Wilk)* kriterijų skaičiavimo rezultatai. Kriterijų reikšmingumo lygmuo *p-level (Sig.)* > α = *0,05*, kintamojo skirstinio skirtumas nuo normaliojo statistiškai nereikšmingas.

2. Hipotezių formulavimas. Suformuluotos hipotezės išreiškia spėjimus apie tiriamus duomenis. Statistinių hipotezių klasifikacija pateikta 5.1 lentelėje.

Statistinės hipotezės skirstomos į nulines ir alternatyvias:

Nulinė hipotezė H_0 – tai spėjimas apie skirtumų nebuvimą. Nulinė hipotezė teigia, kad lyginamųjų populiacijų parametrų, skirstinių arba populiacijų statistikų skirtumas yra atsitiktinis ir nereikšmingas.

5.1 lentelė. Statistinių hipotezių klasifikacija

Statistinės hipotezės					
Parametrinės hipotezės			**Neparametrinės hipotezės**		
Nulinė	Altenatyvioji		Nulinė	Altenatyvioji	
$\theta \in \Theta_0$	$\theta \in \Theta_1$		*X* ir *Y* skirstiniai nesiskiria	*X* ir *Y* skirstiniai skiriasi	
	Dvipusė	Vienpusė		Dvipusė	Vienpusė
$\theta_1 = \theta_0$	$\theta_1 \neq \theta_0$	$\theta_1 > \theta_0$	$P(X>t) = P(Y>t)$ kiekvienam t	$P(X>t) > P(Y>t)$ arba $P(X>t) < P(Y>t)$ kiekvienam t	$P(X>t) > P(Y>t)$ kiekvienam t
		$\theta_1 < \theta_0$			$P(X>t) < P(Y>t)$ kiekvienam t

Alternatyvioji hipotezė H_1 – tai spėjimas apie egzistuojančius skirtumus, kurių negalima paaiškinti atsitiktiniais svyravimais. Alternatyvioji hipotezė dažniausiai yra tai, ką mes norime įrodyti.

Nulinė ir alternatyvioji hipotezės gali būti *parametrinės* ir *neparametrinės*. Tuo atveju, kai duomenų skirstinys artimas normaliajam ir duomenys išmatuoti kiekybinėje skalėje, galima tikrinti teiginius apie tiriamos populiacijos parametrus (dažniausiai vidurkį ir dispersiją). Bet koks teiginys apie populiacijos parametro reikšmę vadinamas parametrine hipoteze. Statistinę parametrinę hipotezę sudaro du alternatyvūs teiginiai apie galimas parametro reikšmes. Problema formuluojama kaip spėjimas apie galimas parametro reikšmes, priklausančias aibei Θ_0, pateikiant alternatyvą, kad θ priklauso aibei Θ_1:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1, \end{cases} \tag{4.2}$$

čia H_0 – parametrinė (nulinė) hipotezė, o H_1 – alternatyva (alternatyvioji hipotezė). Alternatyvos skirstomos į dvipusės $\theta_1 \neq \theta_0$ ir vienpuses $\theta_1 > \theta_0$; $\theta_1 < \theta_0$.

Tais atvejais, kai apie duomenų skirstinį nieko nežinome arba kai tiriamos mažos imtys, formuluojami teiginiai, nesusiję su parametrų įvertinimu. Tokios hipotezės vadinamos neparametrinėmis. Jei duomenys išmatuoti pavadinimų ir rangų skalėse, dažniausiai formuluojamos neparametrinės hipotezės. Neparametrinių hipotezių atveju dažniausiai lyginami skirstiniai. Neparametrinės hipotezės taip pat būna vienpusės ir dvipusės. Tarkime, norime palyginti kintamųjų *X* ir *Y* skirstinius. Neparametrinės hipotezės formuluojamos:

$$\begin{cases} H_0\text{: } X \text{ ir } Y \text{ skirstiniai nesiskiria,} \\ H_1\text{: } X \text{ ir } Y \text{ skirstiniai skiriasi.} \end{cases} \tag{4.3}$$

Matematiškai dėl dvipusių hipotezių tai galima užrašyti kaip:

$$\begin{cases} H_0\text{: } P(X>t) = P(Y>t) \text{ kiekvienam } t, \\ H_1\text{: } P(X>t) > P(Y>t) \text{ arba } P(X>t) < P(Y>t) \text{ kiekvienam } t. \end{cases} \tag{4.4}$$

Vienpusių alternatyvų statistinės hipotezės matematiškai užrašomos kaip:

$$\begin{cases} H_0\text{: } P(X>t) = P(Y>t) \text{ kiekvienam } t, \\ H_1\text{: } P(X>t) > P(Y>t) \text{ kiekvienam } t, \end{cases} \tag{4.5}$$

arba

$$\begin{cases} H_0\text{: } P(X>t) = P(Y>t) \text{ kiekvienam } t, \\ H_1\text{: } P(X>t) < P(Y>t) \text{ kiekvienam } t. \end{cases} \tag{4.6}$$

3. Statistinio kriterijaus skaičiavimas. Taisyklė, pagal kurią iš imties rezultatų darome išvadą apie hipotezės teisingumą ar klaidingumą, vadinama statistiniu kriterijumi. Statistiniai kriterijai skirstomi į parametrinius ir neparametrinius. Dauguma parametrinių hipotezių tikrinimo kriterijų gali būti taikomi tik tada, kai kintamojo skirstinys yra normalus (kintamieji pasiskirstę pagal Gauso dėsnį). Praktiškai ši sąlyga ne visada tenkinama. Be to, mažoms imtims šios sąlygos patikrinti praktiškai neįmanoma. Yra grupė kriterijų, kuriuos taikant nereikalaujama, kad stebimojo kintamojo skirstinys būtų normalus. Šie kriterijai vadinami nepriklausomais nuo skirstinio arba neparametriniais kriterijais. Jie nėra skirti hipotezėms apie populiacijų reikšmes tikrinti. Neparametriniai kriterijai gali būti taikomi mažoms imtims. Parametriniai ir neparametriniai kriterijai papildo vieni kitus.

Parametriniai kriterijai yra galingesni, suteikia didesnę tikimybę nustatyti lyginamų populiacijų parametrų skirtumus, tačiau jie ne tokie universalūs. Neparametriniai kriterijai yra universalesni, mažiau jautrūs išskirtims, tačiau jie silpnesni.

4. Statistinio reikšmingumo lygmens parinkimas. Tikrinant hipotezę dažniausiai taikomas toks principas: hipotezę atmetame, jei atsitinka tai, ko, esant teisingai hipotezei, atsitikti praktiškai negalėjo, arba tokio įvykio tikimybė yra labai maža. Tam įvedama reikšmingumo lygmens sąvoka. Matematikoje reikšmingumo lygmuo paprastai žymimas α raide, $\alpha > 0$. Sutariame įvykį laikyti praktiškai negalimu, jei jo

tikimybė mažesnė už α. Kuo mažesnė α reikšmė, tuo didesnė tikimybė, kad bus priimtas teisingas sprendimas. Priimdami arba atmesdami hipotezę $H_{0,}$ galima padaryti dviejų rūšių klaidas: pirmosios rūšies klaida: H_0 atmetame, kai jį teisinga, antrosios rūšies klaida: H_0 priimame, kai jį klaidinga. Kalbėdami formalizuotai, galime pasakyti, kad reikšmingumo lygmuo – tai tikimybė atmesti H_0 hipotezę, kai ji yra teisinga (pirmosios rūšies klaida). Taisyklė, pagal kurią iš imties rezultatų darome išvadą apie hipotezės teisingumą ar klaidingumą, vadinama statistiniu kriterijumi. Galimos statistinio kriterijaus taikymo baigtys pateiktos 5.2 lentelėje.

5.2 lentelė. Pirmosios ir antrosios rūšies klaidos

	H_0 **teisinga**	H_0 **neteisinga**
atmetame H_0	I rūšies klaida	neteisingas sprendimas
neatmetame H_0	teisingas sprendimas	II rūšies klaida

Reikšmingumo lygmuo, kuriuo remiantis hipotezė atmetama, yra sutartinis. Daugelyje mokslo sričių naudojamas 5 proc. reikšmingumo lygmuo. Tačiau reikia atminti, kad 5 proc. reikšmingumo lygmuo susijęs su gana didele klaidos tikimybe (5 proc.). Pagal bendrai priimtą terminologiją išvados, padarytos remiantis klaidos tikimybe $\alpha \leq 0,05$ vadinamos reikšmingomis, remiantis klaidos tikimybe $\alpha \leq 0,01$ labai reikšmingomis, o esant klaidos tikimybei $\alpha \leq 0,001$ reikšmingiausiomis. Kartais literatūroje tokios išvados žymimos atitinkamai viena, dviem arba trimis žvaigždutėmis (žr. 5.3 lentelę).

Kompiuterinėse duomenų analizės programose skaičiuojamas *mažiausias* reikšmingumo lygmuo, su kuriuo teisinga nulinė hipotezė gali būti atmesta turimiems duomenims. Šis reikšmingumo lygmuo vadinamas *stebimuoju reikšmingumo lygmeniu p-level*.

5.3 lentelė. Klaidos tikimybės interpretacija

Klaidos tikimybė	**Reikšmingumas**	**Žymėjimas**
$\alpha > 0,05$	Nereikšminga	ns (non-significant)
$\alpha \leq 0,05$	Reikšminga	*
$\alpha \leq 0,01$	Labai reikšminga	**
$\alpha \leq 0,001$	Reikšmingiausia	***

5. Sprendimo priėmimas. Visos galimos statistinio kriterijaus reikšmės dalijamos į dvi sritis: pirmoji sritis, į kurią patekus apskaičiuoto kriterijaus reikšmei nulinė hipotezė atmetama, ir antroji, į kurią patekus apskaičiuoto kriterijaus reikšmei nulinės hipotezės atmesti negalime, t. y. jį priimama. Skaičiai C_1, C_2, …, kurie atskiria hipotezės atmetimo ir neatmetimo sritis, vadinami *kritinėmis reikšmėmis*. Jei apskaičiuota (empirinė) statistinio kriterijaus reikšmė patenka į kritinę sritį, nulinė hipotezė atmetama. Jei apskaičiuota (empirinė) statistinio kriterijaus reikšmė nepatenka į kritinę sritį, nulinė hipotezė neatmetama (priimama).

Skaičiuojant kompiuteriu rezultatuose pateikiama apskaičiuota statistinio kriterijaus reikšmė ir stebimasis reikšmingumo lygmuo *p-level*. Tada sprendimo priėmimo taisyklė formuluojama taip:

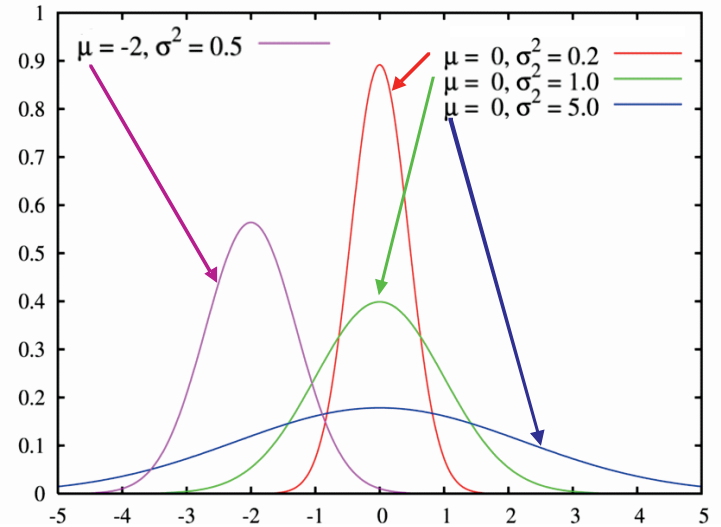
- Jeigu apskaičiuotas stebimasis reikšmingumo lygmuo *p-level* yra mažesnis už pasirinktą reikšmingumo lygmenį α (*p-level* < α), nulinė hipotezė atmetama.
- Jeigu apskaičiuotas stebimasis reikšmingumo lygmuo *p-level* yra didesnis arba lygus α (*p-level* ≥ α), nulinė hipotezė neatmetama (priimama).

Standartiniai skirstiniai. Kritinės kriterijų reikšmės nustatomos nagrinėjant standartinius teorinius skirstinius. Dažniausiai naudojami Gauso (normalusis) skirstinys, Pirsono χ^2 skirstinys, Stjudento *t* skirstinys, Fišerio *F* skirstinys.

Gauso skirstinio tikimybės tankis apibūdintas 5.1 formulę, grafikai pateikti 5.3 pav. Normalieji skirstiniai pasižymi šiomis savybėmis:

- vidurkio (μ), modos ir medianos reikšmės sutampa,
- skirstinio tikimybių tankio kreivė yra dvipusiai simetriška, o simetrijos ašis yra ties vidurkiu,
- skirstinio kreivės padėtis priklauso nuo vidurkio vietos skaičių ašyje,
- skirstinio tikimybių tankio kreivės plotis ir amplitudė priklauso nuo dispersijos (σ^2).

Standartinis normalusis skirstinys turi vidurkį $\mu = 0$ ir standartinį nuokrypį $\sigma = 1$. Šis skirstinys taikomas, pavyzdžiui, tikrinant hipotezes apie vidurkių lygybę.



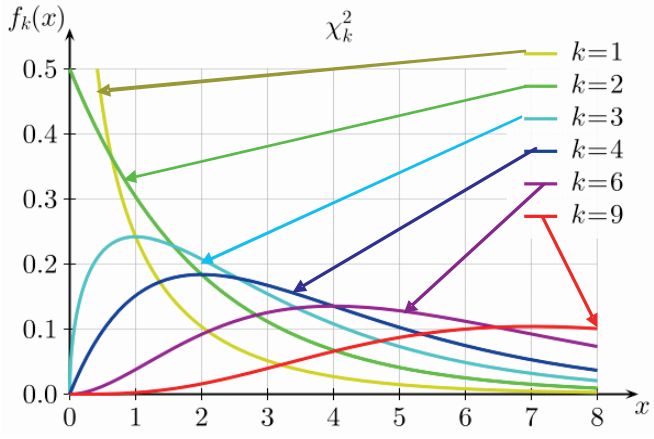
5.3 pav. Gauso skirstinys su skirtingais vidurkiais ir dispersijomis

Pirsono χ^2 skirstinio tikimybės tankis:

$$f(x) = \frac{\left(\frac{1}{2}\right)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \tag{5.7}$$

čia k – laisvės laipsnių skaičius.

Skirstinio tikimybės tankio funkcijos grafikas pateiktas 5.4 pav. Grafiko forma priklauso nuo laisvės laipsnių skaičiaus. Šis skirstinys taikomas, pavyzdžiui, neparametrinės statistikos hipotezėms tikrinti, labiausiai žinomas χ^2 kriterijus, χ^2 skirstinio α lygmens kritinės reikšmės pateiktos 1 priedo lentelėje.

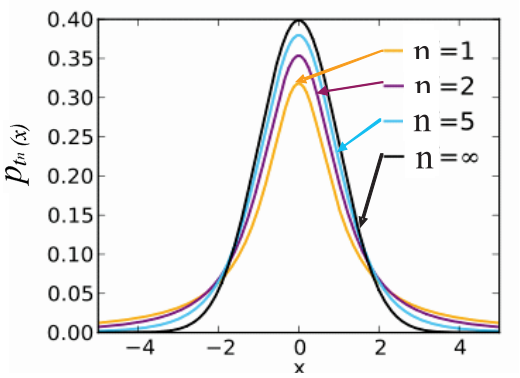


5.4 pav. Pirsono χ^2 skirstinys dėl skirtingo laisvės laipsnio skaičiaus

Stjudento skirstinio tikimybės tankis:

$$p_{tn}(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \qquad (5.8)$$

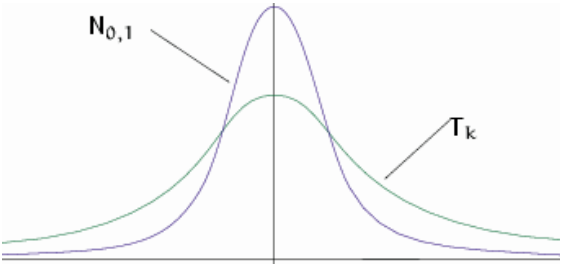
čia n – laisvės laipsnių skaičius.



5.5 pav. Stjudento skirstinys dėl skirtingo laisvės laipsnio skaičiaus

Skirstinio tikimybės tankio funkcijos grafikas pateiktas 5.5 pav. Grafiko forma priklauso nuo laisvės laipsnių skaičiaus. T kriterijus

dažniausiai taikomas dviejų priklausomų arba dviejų nepriklausomų imčių vidurkių lygybės analizei. Stjudento skirstinio α lygmens kritinės reikšmės $t_{\alpha}(n)$ pateiktos 2 priedo lentelėje. Stjudento skirstinio ir standartinio normalaus skirstinio tikimybės tankio grafikų palyginimas pateiktas 5.6 pav.



Stjudento skirstinio (Tk) ir standartinio normalaus skirstinio (N0,1) tikimybės tankio grafikų palyginimas

5.6 pav. Stjudento skirstinio ir standartinio normalaus skirstinio grafikas

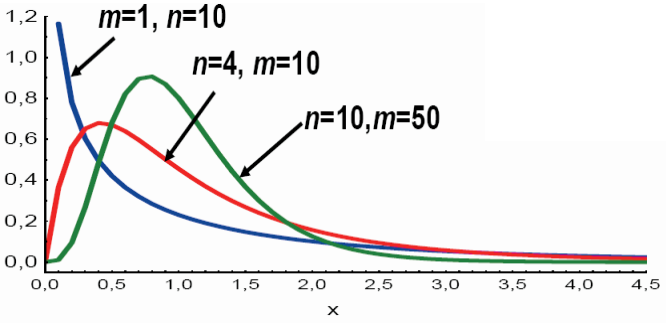
Fišerio skirstinys. Jeigu yra du nepriklausomi kintamieji X_1 ir X_2, kiekvienas kintamasis turi χ^2 skirstinį su n ir m laisvės laipsniais:

$$X_1 \sim \chi^2(n) \text{ ir } X_2 \sim \chi^2(m), \tag{5.9}$$

tai atsitiktinis dydis:

$$F(n,m) = \frac{mX_1}{nX_2} \tag{5.10}$$

turi Fišerio F skirstinį su n ir m laisvės laipsniais. Fišerio skirstinys dažniausiai taikomas dispersinei analizei.



5.7 pav. Fišerio skirstinys dėl skirtingo laisvės laipsnio skaičiaus

KONTROLINIAI KLAUSIMAI

5.1. Nuo ko priklauso parametrinių arba neparametrinių hipotezių pasirinkimas?

5.2. Kaip priimamas sprendimas priimti arba atmesti hipotezę H_0?

5.3. Ką reiškia pirmos ir antros rūšies klaidos?

5.4. Ką reiškia reikšmingumo lygmuo?

5.5. Pateikite statistinio kriterijaus sampratą.

5.6. Kada taikomi parametriniai ir neparametriniai kriterijai?

5.7. Kokie dažniausiai taikomi standartiniai teoriniai skirstiniai?

UŽDUOTYS

5.8. Tikrinant respondentų amžiaus skirstinio normalumą gautas rezultatas (žr. 5.8 pav.). Paaiškinkite gautą rezultatą. Ar galima šiam kintamajam taikyti parametrinius kriterijus?

Tests of Normality

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Amžius	,198	60	,000	,874	60	,000

a Lilliefors Significance Correction

5.8 pav. Imties skirstinio normalumo tikrinimo rezultatas

6. RYŠIŲ ANALIZĖ

Ryšių analizė leidžia ne tik analizuoti kintamuosius atskirai, bet ir nustatyti jų tarpusavio ryšį. Galima skirti parametrinius ir neparametrinius ryšius.

6.1. KINTAMŲJŲ, IŠMATUOTŲ NOMINALIOJE SKALĖJE ARBA RANGŲ SKALĖJE SU NEDIDELIU KATEGORIJŲ SKAIČIUMI, STATISTINIAI RYŠIAI

Praktiškai atliekant bet kokį tyrimą, duomenų analizė pradedama nuo dažnių lentelių, ypač kai analizuojami nominalieji ir ranginiai kintamieji. Paprasčiausia porinės dažnių lentelės forma yra 2×2 lentelė, gaunama tuo atveju, kai tiriami du kintamieji, kurių kiekvienas gali įgyti tik dvi reikšmes (dichotominiai kintamieji). Sudarius dviejų ar daugiau kintamųjų dažnių lenteles, naudojami suderinamumo kriterijai (Pirsono χ^2, Fišerio ir pan.), kuriais tikrinama, ar tarp stebimų ir tikėtinų dažnių yra reikšmingas skirtumas. Nors Pirsono χ^2 suderinamumo kriterijų galima taikyti bet kokio tipo kintamiesiems, geriausią rezultatą ši analizė duoda analizuojant nominalius ir ranginius kintamuosius. Pirsono χ^2 kriterijų galima taikyti ir 2×2, ir didesnėms lentelėms.

6.1.1. Pirsono χ^2 suderinamumo kriterijus

Pirsono χ^2 (*Pearson chi-square*) kriterijus remiasi tuo, kad apskaičiuojami *tikėtini* dažniai, t. y. dažniai, kurių reikėtų *tikėtis* tuo atveju, kai tarp kintamųjų nėra jokio ryšio. Pavyzdžiui, apklausta 30 vyrų ir 30 moterų, ar pirkdami kavą jie pasiduoda aplinkinių įtakai. Jei nėra jokio ryšio tarp lyties ir aplinkinių poveikio, reikėtų *tikėtis,* kad tiek pasiduodančių ir nepasiduodančių aplinkinių įtakai vyrų ir moterų skaičius bus maždaug vienodas.

Pavyzdžiui, reikia nustatyti, ar aplinkinių įtaka perkant kavą priklauso nuo lyties (žr. 6.1 lentelę).

6.1 lentelė. Porinė dažnių lentelė

		Lytis		**Iš viso**
		Vyras	Moteris	
Aplinkinių įtaka	ne	$f_{11} = 10$	$f_{12} = 12$	$f_{1*} = 22$
	taip	$f_{21} = 20$	$f_{22} = 18$	$f_{2*} = 38$
Iš viso		$f_{*1} = 30$	$f_{*2} = 30$	60

Pirsono χ^2 kriterijaus taikymas susijęs su kai kuriais apribojimais, taikomais stebėjimų skaičiui *n*. Šis kriterijus taikomas, kai stebėjimų skaičius *n* didesnis nei 30 ir bent 75 proc. dažnių lentelės narvelių dažniai ne mažesni kaip 5. Jei šios sąlygos netenkinamos, rezultatai gali būti nepatikimi.

Suformuluokime hipotezes:

Hipotezė H_0: aplinkinių įtaka perkant kavą nepriklauso nuo lyties;

Hipotezė H_1: aplinkinių įtaka perkant kavą priklauso nuo lyties.

Statistinio reikšmingumo lygmens parinkimas

Reikšmingumo lygmuo, kuriuo remiantis hipotezė atmetama, yra sutartinis. Daugelyje mokslo sričių naudojamas 5 proc. reikšmingumo lygmuo. Pasirenkame $\alpha = 0{,}05$.

Pirsono χ^2 kriterijaus skaičiavimas

Analizuojant statistinius ryšius pagal χ^2 kriterijų, pirmiausia apskaičiuojami tikėtini dažniai, t. y. kokios turėtų būti lentelės narvelių reikšmės, jeigu tarp kintamųjų nebūtų statistinio ryšio. Du kintamieji yra tarpusavyje nepriklausomi, jei stebimi dažniai sutampa su tikėtinais dažniais. Tikėtini dažniai apskaičiuojami pagal formulę:

$$q_{ij} = \frac{f_{i*} f_{*j}}{n}; \qquad (6.1)$$

čia q_{ij} – tikėtinas dažnumas narvelyje, esančiame *i*-tosios eilutės ir *j*-tojo stulpelio sankirtoje,

f_{i*} – *i*-tosios eilutės dažnių suma, $f_{i*} = f_{i1} + f_{i2}$,

f_{*j} – *j*-tojo stulpelio dažnių suma, $f_{*j} = f_{1j} + f_{2j}$,

n – nagrinėjamų elementų skaičius, šiuo atveju $n = 60$.

Pirsono χ^2 kriterijus apskaičiuojamas pagal formulę:

$$\chi^2 = \sum_{i,j=1}^{n} \frac{\left(f_{ij} - q_{ij}\right)^2}{q_{ij}}, \qquad (6.2)$$

čia: f_{ij} – gardelės, esančios *i*-tosios eilutės ir *j*-tojo stulpelio sankirtoje, tikrasis (stebėtas) dažnis. Iš 6.1 lentelės duomenų gauta reikšmė $\chi^2 = 0{,}287$.

Paskui apskaičiuojami lentelės laisvės laipsniai:

$$df = (s - 1)(t - 1), \qquad (6.3)$$

čia: *df* – lentelės laisvės laipsnių skaičius, *s* – eilučių skaičius, *t* – stulpelių skaičius. Nagrinėjamą lentelę sudaro dvi eilutės ir du stulpeliai. Taigi $df = (2 - 1)(2 - 1) = 1$.

Statistinių išvadų formulavimas, skaičiavimui nenaudojant statistinės programos

Reikia nustatyti, ar gauta χ^2 kriterijaus reikšmė yra pakankama, kad būtų galima teigti, jog tiriami požymiai yra priklausomi. Hipotezė apie kintamųjų nepriklausomumą yra atmetama, kai apskaičiuotos statistikos χ^2 reikšmė yra didesnė už kritinę reikšmę. Pasirinkome reikšmingumo lygmenį a = 0,05. Kritinę χ^2 reikšmę galima rasti χ^2 skirstinio lentelėje (žr. 1 priedą). Šioje lentelėje rasime, kad reikšmingumo lygmens a = 0,05 kritinė reikšmė yra: $\chi^2_{0,05}(1) = 3{,}84$ ir ji yra didesnė už iš duomenų, pateiktų 6.1 lentelėje, apskaičiuotą reikšmę $\chi^2 = 0{,}287$, todėl hipotezės H_0 atmesti negalima.

Dažnių lentelių ryšio stiprumo matai

Φ koeficientas (Phi coefficient)

Apskaičiavus χ^2 kriterijų nustatoma, ar tarp požymių yra statistinis ryšys. Norint tiksliau nustatyti ryšio stiprumą, naudojami nominaliųjų kintamųjų koreliacijos matai. Vienas iš tokių matų – *Φ* koeficientas. Šiuo matu įvertinamas koreliacinio ryšio stiprumas 2×2 lentelėse. *Φ* koeficiento reikšmės kinta nuo 0 (tarp požymių nėra jokio ryšio) iki 1 (tarp požymių yra labai stiprus ryšys):

$$\Phi^2 = \frac{\chi^2}{n} \text{ arba } \Phi = \sqrt{\frac{\chi^2}{n}}, \quad (6.4)$$

čia n – imties dydis.

Šis koeficientas gali būti naudojamas ir didesnėms nei 2×2 lentelėms, tačiau jose kartais Φ koeficiento reikšmės būna didesnė už 1. Norint išvengti šio trūkumo, skaičiavimams naudojami papildomi koeficientai.

Kramerio V koeficientas (*Cramer's V*) apskaičiuojamas pagal formulę:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}}, \quad (6.5)$$

čia: r – dažnių lentelės eilučių skaičius, c – stulpelių skaičius, *min(r – 1,c – 1)* parenkama mažiausia reikšmė. Šiuo kriterijumi įvertinama, kad eilučių ir stulpelių skaičius gali būti skirtingas. *V koeficientas* gali būti naudojamas didesnėms nei 2×2 lentelėms.

6.1.2. Pirsono χ^2 suderinamumo kriterijaus skaičiavimas (nominalūs kintamieji)

Turint omenyje, kad PSPP paketas suderinamas su SPSS paketu, pateiksime skaičiavimų atlikimo ir rezultatų pateikimo PSPP ir SPSS terpėje palyginimą.

Patikrinsime, ar priklauso nuo lyties aplinkinių įtaka kavos pirkimui (nominalūs kintamieji). Pirmiausia suformuluosime hipotezes:

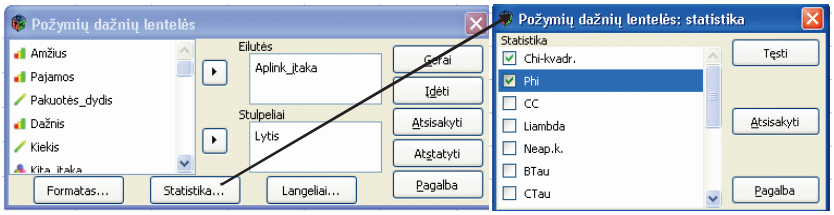
H_0 – aplinkinių įtaka nepriklauso nuo lyties;

H_1 – aplinkinių įtaka priklauso nuo lyties.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*...(žr. 6.1 pav.). Kintamasis *Aplink_įtaka* perkeliamas į langelį *Eilutės*, o kintamasis *Lytis* per-

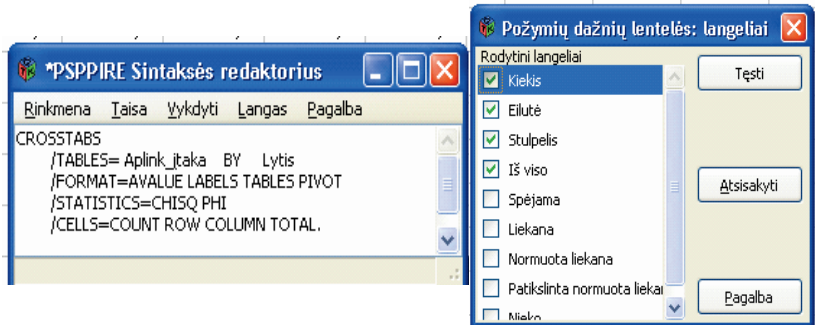
keliamas į langelį *Stulpeliai.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelį *Chi-kvadr.*, galima pažymėti ir langelį *Phi,* jei reikia nustatyti ne tik ryšį, bet ir jo stiprumą. Paspaudę *Tęsti* mygtuką, grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu.



6.1 pav. Dialogo langai *Požymių dažnių lentelės* ir *Požymių dažnių lentelės: statistika*

Mygtuko *Įdėti* paspaudimas atveria *PSPPIRE Sintaksės redaktorius* langelį (žr. 6.2 pav.). Šiame langelyje pateikiama atliekamos procedūros sintaksė. Šiuo atveju atlikti skaičiavimus galima pasitelkus meniu *Vykdyti.*

Porinės dažnių lentelės duomenų pateikimą galima pakeisti langelyje *Požymių dažnių lentelės:langeliai*, kuris iškviečiamas mygtuko *Langeliai...* paspaudimu (žr. 6.2 pav.). Jei pažymėtas tik langelis *Kiekis* – porinių dažnių lentelėje rodomi tik kintamųjų dažnių skaičiai (žr. 6.3 pav.(A)), langelio *Eilutė* žymėjimas pateikia kintamųjų procentinį dažnį eilutėje (žr. 6.3 pav. (B)), langelio *Stulpelis* – kintamųjų procentinį dažnį stulpelyje (žr. 6.3 pav. (C)). Langelio *Iš viso* žymėjimas pateikia kiekvieno narvelio procentinį dažnį visoje imtyje (žr. 6.3 pav. (D)).



6.2 pav. Dialogo langai *PSPPIRE Sintaksės redaktorius* ir *Požymių dažnių lentelės: langeliai*

Aplink_įtaka * Lytis [kiekis].

	Lytis		A
Aplink_įtaka	1	2	Iš viso
1	10,0	12,0	22,0
2	20,0	18,0	38,0
Iš viso	30,0	30,0	60,0

Aplink_įtaka * Lytis [% eilutė].

	Lytis		B
Aplink_įtaka	1	2	Iš viso
1	45,5%	54,5%	100,0%
2	52,6%	47,4%	100,0%
Iš viso	50,0%	50,0%	100,0%

Aplink_įtaka * Lytis [% stulpelis].

	Lytis		C
Aplink_įtaka	1	2	Iš viso
1	33,3%	40,0%	36,7%
2	66,7%	60,0%	63,3%
Iš viso	100,0%	100,0%	100,0%

Aplink_įtaka * Lytis [iš viso %].

	Lytis		D
Aplink_įtaka	1	2	Iš viso
1	16,7%	20,0%	36,7%
2	33,3%	30,0%	63,3%
Iš viso	50,0%	50,0%	100,0%

6.3 pav. Porinių dažnių lentelių pateikimo būdai

Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai. Pirmiausia pateikiama analizuojamų kintamųjų dažnių lentelė: *Aplink_įtaka*Lytis [kiekis]* (žr 6.4 pav.). PSPP paketas neinformuoja, ar duomenys tinka analizei. Pirsono χ^2 kriterijus taikomas, kai bendras imties dydis $m \geq 30$ elementų bei 75 proc. porinių dažnių lentelės narvelių dažniai ≥ 5. Jei šios sąlygos netenkinamos, rezultatai gali būti nepatikimi. Nagrinėjami kintamieji tinkami analizei.

Lentelėje *Chi-kvadrato kriterij.* pateikti analizės rezultatai (žr. 6.4 pav.). *Pirsono Chi-kvadratas* eilutės stulpelyje *Reikšmė* rodoma χ^2 kriterijaus reikšmė, stulpelyje *skirt.* – lentelės laisvės laipsnių skaičius, stulpelyje *Asimp. p-reikšmė (2-pusė)* – stebimojo reikšmingumo lygmens *p-level* reikšmė. Skaičiavimui naudojant statistinę programą, dėl hipotezės H_0 atmetimo sprendžiama pagal stebimąjį reikšmingumo lygmenį *p-level*. *P-level = 0,59 > α = 0,05,* negalima atmesti hipotezės H_0, kad aplinkinių įtaka nepriklauso nuo lyties. Eilutė *Dvitiesinė priklausomybė* leidžia daryti išvadą dėl *nenominalių* kintamųjų tiesinio ryšio – nagrinėjamu atveju šis parametras reikšmės neturi.

Ryšio stiprumui nustatyti dėl nominalių kintamųjų taikomi *Phi and Cramer's V* testai (žr. 6.4 pav.). Lentelėje *Simetriniai matavimai* pateikti šių testų reikšmės. Nagrinėjamu atveju šie koeficientai nereikšmingi.

Aplink_įtaka * Lytis [kiekis].

	Lytis		
Aplink_įtaka	1	2	Iš viso
1	10,0	12,0	22,0
2	20,0	18,0	38,0
Iš viso	30,0	30,0	60,0

Chi-kvadrato kriterij.

Statistika	Reikšmė	skirt.	Asimp. p-reikšmė (2-pusė)
Pirsono Chi-kvadratas	,29	1	,59
Tikėtinumo santykis	,29	1	,59
Tolydumo pataisa	,07	1	,79
Dvitiesinė priklausomybė	,28	1	,60
N galiojančių atvejų	60		

Simetriniai matavimai.

Kategorija	Statistika	Reikšmė	Asimpt. std. paklaida	Apytiksl. T	Apytiksl. p-reikšmė
Pavadinimų x pavadinimų	Phi	,07			
	Kramerio V	,07			
N galiojančių atvejų		60			

6.4 pav. Pirsono χ^2 kriterijaus skaičiavimo rezultatas PSPP terpėje

Galima palyginti PSPP terpėje atliktą analizę su analize SPSS terpėje.

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze* → *Descriptive Statistics* → *Crosstabs*.... Kintamasis *Aplink_įtaka* perkeliamas į langelį *Row(s)*, o kintamasis *Lytis* perkeliamas į langelį *Column(s)*. Tam, kad būtų iškart nubraižytas porinių dažnių grafikas, reikia pažymėti langelį *Display clustered bar charts*. Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics*. Šiame lange reikia pažymėti langelį *Chi-square*, galima pažymėti ir langelį *Phi and Cramer's V*, jei reikia nustatyti ne tik ryšį, bet ir jo stiprumą. (Ryšio stiprumui nustatyti dėl nominalių kintamųjų taikomi *Phi and Cramer's V* koeficientai – dialogo lango *Crosstabs: Statistics* sritis *Nominal*). Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.5 pav.).

Aplink_įtaka * Lytis Crosstabulation

Count

		Lytis		Total
		Vyras	Moteris	1
Aplink_įtaka	ne	10	12	22
	taip	20	18	38
Total		30	30	60

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,287(b)	1	,592		
Continuity Correction(a)	,072	1	,789		
Likelihood Ratio	,287	1	,592		
Fisher's Exact Test				,789	,395
Linear-by-Linear Association	,282	1	,595		
N of Valid Cases	60				

a Computed only for a 2x2 table

b 0 cells (,0%) have expected count less than 5. The minimum expected count is 11,00.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-,069	,592
	Cramer's V	,069	,592
N of Valid Cases		60	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

6.5 pav. Pirsono χ^2 kriterijaus, *Phi ir Cramer's V* koeficientų skaičiavimo rezultatas SPSS terpėje

Pirmiausia pateikiama analizuojamų kintamųjų dažnių lentelė: *Aplink_įtaka*Lytis Crosstabulation*. Lentelėje *Chi-Square Tests* pateikti analizės rezultatai. Užrašas po lentele informuoja, ar kintamieji tinka analizei. Nagrinėjami kintamieji tinkami Pirsono χ^2 kriterijaus analizei. Nurodytas skaičius narvelių su dažniais mažiau kaip 5 (*0 cells have exspected count less than 5)* yra mažesnis negu leidžiamas tokių narvelių skaičius *(The minimum expected count is 11,00). Pearson Chi-Square* eilutės stulpelyje *Value* rodoma χ^2 kriterijaus reikšmė, stulpelyje *df* – lentelės laisvės laipsnių skaičius, stulpelyje *Asimp. Sig. (2-sided)* – stebi-

mojo reikšmingumo lygmens *p-level* reikšmė. Skaičiavimui naudojant statistinę programą, dėl hipotezės H_0 atmetimo sprendžiama pagal stebimąjį reikšmingumo lygmenį *p-level*. *P-level = 0,592 > α = 0,05,* negalima atmesti hipotezės H_0, kad aplinkinių įtaka nepriklauso nuo lyties. Eilutė *Linear-by-Linear Association* leidžia daryti išvadą dėl *nenominalių* kintamųjų tiesinio ryšio – nagrinėjamu atveju šis parametras reikšmės neturi. Ryšio stiprumui nustatyti dėl nominalių kintamųjų taikomi *Phi and Cramer's V* testai. Lentelėje *Symmetric Measures* pateikti šių testų reikšmės ir jų stebimas reikšmingumo lygmuo *p-level* (stulpelis *Approx. Sig.) P-level = 0,592 > α = 0,05,* koeficientai nereikšmingi. 6.5 pav. pateikta porinių dažnių stulpelinė diagrama.

Atliktos procedūros sintaksė pateikta 6.2 lentelėje. Galima teigti, kad sintaksė panaši, tačiau negalima tiesiogiai perkopijuoti komandų iš SPSS į PSPP redaktorių, nes PSPP terpėje ne visos komandos yra realizuotos.

6.2 lentelė. Pirsono χ^2 kriterijaus skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES= Aplink_įtaka BY Lytis /FORMAT=AVALUE LABELS TABLES PIVOT /STATISTICS=CHISQ PHI /CELLS=COUNT.	CROSSTABS /TABLES=Aplink_įtaka BY Lytis /FORMAT= AVALUE TABLES /STATISTIC=CHISQ PHI /CELLS= COUNT /COUNT ROUND CELL.

6.1.3. Pirsono χ^2 suderinamumo kriterijaus skaičiavimas (ranginiai kintamieji)

χ^2 kriterijumi **patikrinsime, kaip pirkimo dažnis priklauso nuo amžiaus** (ranginiai kintamieji). Suformuluosime hipotezes:

H_0 – pirkimo dažnis nepriklauso nuo amžiaus;

H_1 – pirkimo dažnis priklauso nuo amžiaus.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės...* (žr. 6.6 pav.). Kintamasis *Aplink_įtaka* perkeliamas į langelį *Eilutės*, o kintamasis *Lytis* perkeliamas į langelį *Stulpeliai.* Tam, kad būtų iškart nubraižytas porinių dažnių grafikas, reikia pažymėti langelį *Display clustered bar charts.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelį *Chi-square*, galima pažymėti ir langelį *Gamma,* jei reikia nustatyti ne tik ryšį, bet ir jo stiprumą. Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.7 pav.).

Dažnių lentelė rodo, kad nagrinėjami kintamieji tinkami analizei. *Pirsono Chi-kvadratas* eilutės stulpelyje *Reikšmė* rodoma χ^2 kriterijaus reikšmė, stulpelyje *skirt.* – lentelės laisvės laipsnių skaičius, stulpelyje *Asimp. p-reikšmė (2-pusė)* – stebimojo reikšmingumo lygmens *p-level* reikšmė. Skaičiavimui naudojant statistinę programą, dėl hipotezės H_0 atmetimo sprendžiama pagal stebimąjį reikšmingumo lygmenį *p-level. P-level = 0,18 > α = 0,05,* negalima atmesti hipotezės H_0, kad pirkimo dažnis nepriklauso nuo amžiaus.



6.6 pav. Dialogo langai *Požymių dažnių lentelės* ir *Požymių dažnių lentelės: statistika*

Ryšio stiprumui nustatyti dėl *ranginių* kintamųjų taikomas *Gamma* testas. Lentelėje *Simetriniai matavimai* pateikta šio testo reikšmė (*Gamma = –0,07)* Nagrinėjamu atveju šis koeficientas nereikšmingas.

Dažnis * Amžius1 [kiekis].

	Amžius1			
Dažnis	15-25 m.	26-35 m.	virš 36 m.	Iš viso
1	8,0	6,0	6,0	20,0
2	6,0	7,0	8,0	21,0
3	11,0	1,0	7,0	19,0
Iš viso	25,0	14,0	21,0	60,0

Chi-kvadrato kriterij.

Statistika	Reikšmė	skirt.	Asimp. p-reikšmė (2-pusė)
Pirsono Chi-kvadratas	6,24	4	,18
Tikėtinumo santykis	7,31	4	,12
Dvitiesinė priklausomybė	,14	1	,71
N galiojančių atvejų	60		

Simetriniai matavimai.

Kategorija	Statistika	Reikšmė	Asimpt. std. paklaida	Apytiksl. T	Apytiksl. p-reikšmė
Rangų x rangų	Gamma	-,07	,18	-,37	
N galiojančių atvejų		60			

6.7 pav. Pirsono χ^2 kriterijaus skaičiavimo rezultatas PSPP terpėje

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze →Descriptive Statistics → Crosstabs*. Kintamasis *Dažnis* perkeliamas į langelį *Row(s)*, o kintamasis *Amžius* perkeliamas į langelį *Column(s).* Tam, kad būtų iškart nubraižytas porinių dažnių grafikas, reikia pažymėti langelį *Display clustered bar charts.* Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics.* Šiame lange reikia pažymėti langelį *Chi-square*, galima pažymėti ir langelį *Gamma* (dialogo lango *Crosstabs: Statistics* sritis *Ordinal*), jei reikia nustatyti ne tik ryšį, bet ir jo stiprumą. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.8 pav.).

Lentelėje *Chi-Square Tests* pateikti analizės rezultatai. Užrašas po lentele informuoja, kad kintamieji tinka analizei. *P-level = 0,182 > α = 0,05,* negalima atmesti hipotezės H_0, kad pirkimo dažnis nepriklauso nuo amžiaus. Eilutė *Linear-by-Linear Association* leidžia padaryti išvadą dėl nenominalių kintamųjų tiesinio ryšio, *p-level = 0,707 > α = 0,05,* kintamųjų tiesinis ryšys nereikšmingas. Ryšio stiprumui nustatyti dėl *ranginių* kintamųjų taikomas *Gamma* testas. Lentelėje *Symmetric*

Measures pateikta šio testo reikšmė (*Gamma = –0,068)* ir jo stebimas reikšmingumo lygmuo *p-level* (stulpelis *Approx. Sig.). P-level = 0,710 > α = 0,05, Gamma* koeficientas nereikšmingas.

Dažnis * Amžius Crosstabulation

Count

		Amžius			Total
		15–25 m.	26–35 m.	per 36 m.	
Dažnis	1,00	8	6	6	20
	2,00	6	7	8	21
	3,00	11	1	7	19
Total		25	14	21	60

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6,237(a)	4	,182
Likelihood Ratio	7,311	4	,120
Linear-by-Linear Association	,141	1	,707
N of Valid Cases	60		

a 1 cells (11,1%) have expected count less than 5. The minimum expected count is 4,43.

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Gamma	-,068	,182	-,372	,710
N of Valid Cases		60			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

6.8 pav. Pirsono χ^2 kriterijaus skaičiavimo rezultatas SPSS terpėje

Atliktos procedūros sintaksė pateikta 6.3 lentelėje.

6.3 lentelė. Pirsono χ^2 kriterijaus skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES= Dažnis BY Amžius /FORMAT=AVALUE LABELS TABLES PIVOT /STATISTICS=CHISQ GAMMA /CELLS=COUNT.	CROSSTABS /TABLES=Amžius BY Dažnis /FORMAT= AVALUE TABLES /STATISTIC=CHISQ GAMMA /CELLS= COUNT /COUNT ROUND CELL.

6.1.4. Kiti χ^2 kriterijaus skaičiavimo būdai

PSPP ir SPSS programose χ^2 kriterijus skaičiuojamas trejopai: pagal Pirsono formulę (6.2), pagal tikėtinumo santykio formulę (6.6) bei pagal Mantelio–Haenzelio formulę (6.7).

Tikėtinumo santykio formulę:

$$\chi^2 = -2 \sum_{i,j=1}^{n} f_{ij} \ln \frac{q_{ij}}{f_{ij}} \qquad (6.6),$$

čia q_{ij} – tikėtinas dažnumas narvelyje, esančiame *i*-tosios eilutės ir *j*-tojo stulpelio sankirtoje,

f_{ij} – gardelės, esančios *i*-tosios eilutės ir *j*-tojo stulpelio sankirtoje, tikrasis (stebėtas) dažnis,

n – nagrinėjamų elementų skaičius.

Pagal Pirsono ir tikėtinumo santykio formules gaunami artimi rezultatai, o esant didelėms imtims jie praktiškai sutampa.

Mantelio–Haenzelio formulė χ^2 kriterijaus reikšmei apskaičiuoti:

$$\chi^2 = r^2(n-1) \qquad (6.7),$$

čia *r* – Pirsono koreliacijos koeficientas, *n* – imties dydis.

Mantelio–Haenzelio formulė χ^2 kriterijui skaičiuoti taikoma hipotezėms apie tiesinį ranginių kintamųjų nepriklausomumą patikrinti, netaikoma nominaliems kintamiesiems.

Nagrinėjant χ^2 kriterijaus skaičiavimo rezultatą **PSPP** terpėje (žr. 6.7 pav.), reikia žiūrėti *Chi-kvadrato kriterij.* lentelę. Šioje lentelėje pateikta Pirsono χ^2 kriterijaus reikšmė (*Pirsono Chi-kvadratas = 6,24),* χ^2 kriterijaus reikšmė pagal tikėtinumo formulę (*Tikėtinumo santykis = 7,31)* bei pagal Mantelio–Haenzelio formulę (*Dvitiesinė priklausomybė = 0,14).* Imtis nedidelė, *n = 60,* todėl *Pirsono Chi-kvadratas = 6,24* ir *Tikėtinumo santykis = 7,31* nedaug skiriasi. *Dvitiesinės priklausomybės* reikšmė rodo, kad nėra nagrinėjamų kintamųjų tiesinio ryšio.

SPSS terpėje (žr. 6.8 pav.) rezultatai pateikti *Chi-Square Tests* lentelę. Pirsono χ^2 kriterijaus reikšmė (*Pearson Chi-Square = 6,237),* χ^2 kriterijaus reikšmė pagal tikėtinumo formulę (*Likelihood Ratio = 7,311)* ir pagal Mantelio–Haenzelio formulę (*Linear-by-Linear = 0,141).*

6.1.5. Nominaliųjų kintamųjų ryšio matai

Nominaliųjų kintamųjų, turinčių daugiau kaip dvi kategorijas, tarpusavio ryšiui nustatyti paprastai taikomas χ^2 testas. Tačiau PSPP ir SPSS programoje taip pat yra pateikiami ryšio tarp dviejų nominaliųjų kintamųjų įvertinimo kiekybiniai kriterijai. Šie kriterijai parodo dviejų priklausančių nominalinei skalei kintamųjų priklausomumo ar nepriklausomumo laipsnį. Kriterijaus reikšmė, lygi nuliui, atitinka visišką kintamųjų nepriklausomumą, o reikšmė, lygi vienetui, – didžiausią priklausomumą. Šie tarpusavio ryšio matai negali turėti neigiamų reikšmių, nes negalint išrikiuoti kintamųjų, negalima nustatyti priklausomybės krypties.

6.1.1 skyrelyje jau pateikti *Φ koeficientas* (6.4) ir *Kramerio V koeficientas* (6.5). Išvardytieji kriterijai yra skaičiuojami remiantis χ^2 kriterijumi. Kiti matuojamų pagal nominalinę skalę kintamųjų tarpusavio ryšio matai: *Lambda, Gudmano–Kruskalio tau* (*lambda* modifikacija), *Somers'D* koeficientai skaičiuojami remiantis vadinamąja proporcingo klaidos mažinimo koncepcija. Jie įvertina priklausomo kintamojo nuspėjamumo santykinį klaidos sumažėjimą, kai žinomas nepriklausomas kintamasis. Išnagrinėsime kintamųjų *Lytis* ir *Aplink_įtaka* ryšį.

PSPP terpėje šie koeficientai skaičiuojami vykdant komandas *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*.... Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelius *Lambda, D.* Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.9 pav.).

Lentelėje *Kriptingi matavimai* pateikti *Lambda, Gudmano-Kruskalio tau, Somers'D* koeficientai. Sprendimas dėl šių koeficientų reikšmingumo priimamas *Stjudento t* statistikos pagrindu, lyginant apskaičiuotą *t* reikšmę (*Apytiksl. T*) su kritine. Parenkamas reikšmingumo lygmuo α. Hipotezė H_0 atmetama, jei absoliuti *t* reikšmė viršija Stjudento skirstinio su *(n–2)* laisvės laipsniais *α/2* lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

Kryptingi matavimai.

Kategorija	*Statistika*	*Tipas*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Pavadinimų x pavadinimų	Liambda	Simetrinis	,04	,09	,43	
		Aplink_įtaka priklausomas	,00	NaN	NaN	
		Lytis priklausomas	,07	,15	,43	
	Gudmano-Kruskaliotau	Aplink_įtaka priklausomas	,00			
		Lytis priklausomas	,00			
Rangų x rangų	Somers d	Simetrinis	-,07	,00	-,54	
		Aplink_įtaka priklausomas	-,07	,12	-,54	
		Lytis priklausomas	-,07	,13	-,54	

6.9 pav. Nominaliųjų kintamųjų ryšio skaičiavimas PSPP terpėje

SPSS terpėje šie koeficientai skaičiuojami vykdant komandas *Analyze →Descriptive Statistics → Crosstabs*. Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics.* Šiame lange reikia pažymėti langelius *Lambda, Somers'd*. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.10 pav.).

Lentelėje *Direktional Measures* pateikti apskaičiuoti koeficientai. Sprendimas dėl šių koeficientų reikšmingumo priimamas pagal *p-level* (*Approx. Sig.*) reikšmę. Šiuo atveju po lentele pateikti gautų rezultatų aiškinimai.

Directional Measures

			Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,038	,088	,427	,669
		Aplink_įtaka Dependent	,000	,000	.(c)	.(c)
		Lytis Dependent	,067	,151	,427	,669
	Goodman and Kruskal tau	Aplink_įtaka Dependent	,005	,018		,595(d)
		Lytis Dependent	,005	,018		,595(d)
Ordinal by Ordinal	Somers' d	Symmetric	-,069	,129	-,537	,591
		Aplink_įtaka Dependent	-,067	,124	-,537	,591
		Lytis Dependent	-,072	,134	-,537	,591

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.
c Cannot be computed because the asymptotic standard error equals zero.
d Based on chi-square approximation

6.10 pav. Nominaliųjų kintamųjų ryšio skaičiavimas SPSS terpėje

6.2. KORELIACINĖ ANALIZĖ

Duomenys, apibūdinantys tiriamą objektą, dažnai būna susiję tarpusavyje arba priklauso vieni nuo kitų. Tokius ryšius galima nustatyti tik su tam tikra tikimybe. Statistinio ryšio (koreliacijos) stiprumą galima įvertinti koreliacijos koeficientais. Kalbant apie koreliaciją, vartojami du terminai *koreliacinis ryšys* ir *koreliacinė priklausomybė*. Priklausomybė reiškia vieno kintamojo įtaką kitam, o ryšys – bendrą ar darnų dviejų kintamųjų kitimą. Koreliacijos koeficientas neparodo, kuris kintamasis yra priežastis, o kuris pasekmė. Jis tik nustato, ar kintamųjų kitimas yra susietas. Jei galima stebėti, kad vieno kintamojo kitimas veikia kitus kintamuosius, šis kintamasis vadinamas nepriklausomu. Kintamieji, kurie, mūsų nuomone, kinta veikiant nepriklausomiems kintamiesiems, vadinami priklausomais. Dažniausiai tyrime sunku nustatyti, kuris kintamasis yra priklausomas, o kuris nepriklausomas. Tai tyrėjas gali numatyti analizuodamas logiškai nagrinėjamo proceso veiksnius.

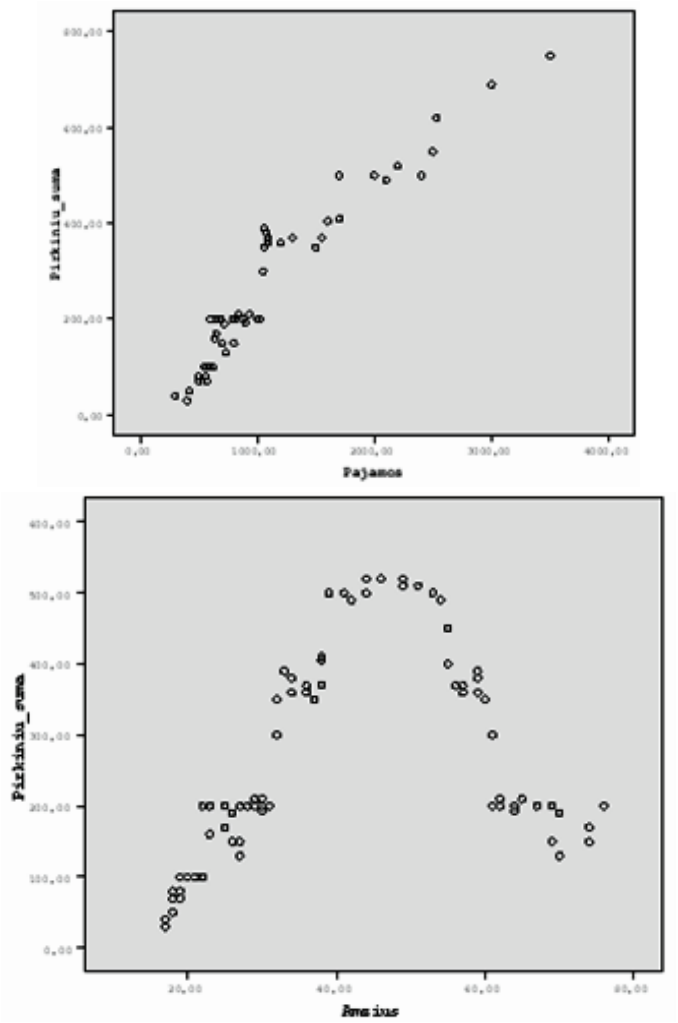
Koreliacinė analizė skirta *ranginių* arba *kiekybinių* kintamųjų ryšiui, jo stiprumui ir krypčiai nustatyti. Kiekybinių kintamųjų tiesinio ryšio įvertinimui skaičiuojamas Pirsono (*Pearson*) koreliacijos koeficientas. Ranginių kintamųjų (tik skaitmeninių reikšmių) ryšio įvertinimui yra skaičiuojamas Spirmeno (*Spearman*) koreliacijos koeficientas. Turinčius dvi kategorijas (binarinius) nominaliuosius kintamuosius koreliacinėje analizėje galima laikyti ranginiais kintamaisiais.

Didelėms imtims (kai n daugiau už 30) yra nusistovėjusios koreliacijos stiprumo nustatymo taisyklės (žr. 6.4 lentelę). Tiksliau įvertinti koreliacinį ryšį galima patikrinus hipotezę dėl koreliacijos koeficiento lygybės nuliui.

6.4 lentelė. Koreliacijos stiprumo interpretacija

Koreliacijos koeficiento reikšmė	Interpretacija
Nuo 0,9 iki 1,0 arba nuo –0,9 iki –1,0	Labai stipri koreliacija
Nuo 0,7 iki 0,9 arba nuo –0,7 iki –0,9	Stipri koreliacija
Nuo 0,5 iki 0,7 arba nuo –0,5 iki –0,7	Vidutinė koreliacija
Nuo 0,3 iki 0,5 arba nuo –0,3 iki –0,5	Silpna koreliacija
Nuo 0,3 iki –0,3	Koreliacija nereikšminga

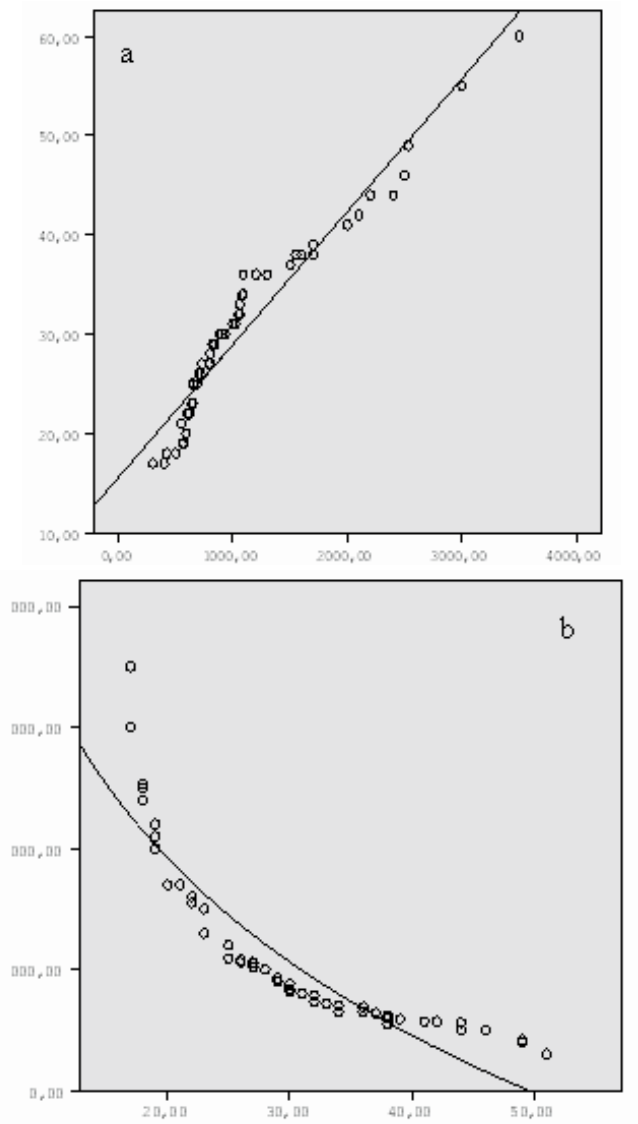
Koreliaciniai ryšiai skiriasi pagal formą, kryptį ir stiprumą. Pagal *formą* koreliaciniai ryšiai skirstomi į tiesinius ir kreivinius. Ryšio formą galima pamatyti analizuojant taškinius (*Scatter*) diagramas (žr. 6.11 pav.).



6.11 pav. Tiesinis ir netiesinis koreliacinis ryšys

Pagal *kryptį* koreliaciniai ryšiai yra teigiami (*tiesioginiai*) ir neigiami (*atvirkštiniai*). Esant teigiamam ryšiui, vieno kintamojo reikšmėms

didėjant, didėja ir kito kintamojo reikšmės (6.12 a pav.). Esant neigiamam ryšiui, vieno kintamojo reikšmėms didėjant, kito kintamojo reikšmės mažėja (6.12 b pav.).



6.12 pav. Teigiamas (tiesinis) ir neigimas (netiesinis) koreliacinis ryšys

6.2.1. Pirsono koreliacijos koeficientas

Kiekybinių kintamųjų ryšio *stiprumą* galima išmatuoti koreliacijos koeficientu. Didelės šio koeficiento reikšmės, nepaisant, ar jos teigiamos, ar neigiamos, atitinka tai, ką vadiname stipria koreliacija, o mažos reikšmės – silpna koreliacija. Jei koreliacija yra nereikšminga, tai nereiškia, kad koreliacijos koeficientas tiksliai lygus nuliui, tačiau jo reikšmė yra arti nulio. Kiekybiniams kintamiesiems tirti naudojami keli skirtingi koreliacijos koeficientai. Dažniausiai naudojamas tiesinio ryšio Pirsono koreliacijos koeficientas *r*. Jei Pirsono koreliacijos koeficientas rodo kintamųjų statistinio ryšio nebuvimą, tai nereiškia, kad šio ryšio visiškai nėra, tarp kintamųjų nėra *tiesinio* ryšio, bet gali būti sudėtingesnės formos ryšys.

Empirinis Pirsono koreliacijos koeficientas yra tiesinio ryšio stiprumo tarp kintamųjų *X* ir *Y* matas. Jis apskaičiuojamas iš turimų stebėjimų pagal formulę:

$$r = \frac{\frac{1}{n-1}\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{S_x S_y}, \tag{6.8}$$

čia $\overline{x}$ ir $\overline{y}$ atitinkamai yra stebėjimų *x* ir *y* vidurkiai, o S_x ir S_y – standartiniai nuokrypiai.

Koreliacijos koeficiento reikšmingumas tikrinamas *t* kriterijumi. Suformuluokime statistines hipotezes:

H_0: koreliacijos koeficientas lygus nuliui.

H_1: koreliacijos koeficientas nelygus nuliui.

Hipotezės H_0 teisingumui patikrinti naudosime *t* kriterijų:

$$t = r\sqrt{\frac{n-2}{1-r^2}}, \tag{6.9}$$

čia: *r* – apskaičiuota koreliacijos koeficiento reikšmė, *n* – imties dydis.

Parenkamas reikšmingumo lygmuo α. Hipotezė H_0 atmetama, jei absoliuti *t* reikšmė viršija Stjudento skirstinio su *(n–2)* laisvės laipsniais *α/2* lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

Determinacijos koeficientas. Vertinant pasirinkto kintamųjų ryšio modelio tinkamumą analizuojamos liekanos ir skaičiuojamas *determinacijos* koeficientas. Determinacijos koeficientas parodo, kuri vieno požymio bendro kitimo dalis gali būti paaiškinta kito požymio reikšmių kitimu. Šis koeficientas interpretuojamas taip:

$$R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2} = \frac{\text{Variacijos dalis, kurią paaiškina modelis}}{\text{Visa variacija}}, \quad (6.10)$$

čia: S_ε^2 – liekanų (ε) dispersija, S_y^2 – stebimų kintamojo (*Y*) reikšmių dispersija.

Determinacijos koeficientas R^2 tiesinės priklausomybės atveju yra lygus koreliacijos tarp kintamųjų (požymių) *X* ir *Y* koeficiento kvadratui.

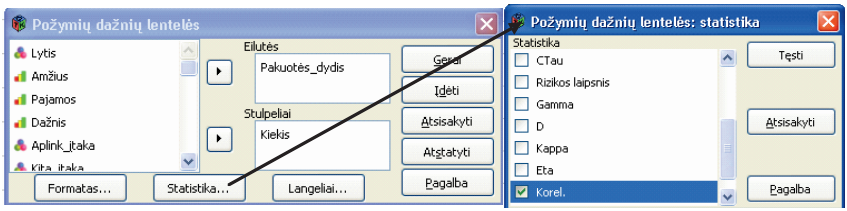
$$R^2 = r_{xy}^2. \quad (6.11)$$

6.2.2. Pirsono koreliacijos koeficiento skaičiavimas

Patikrinsime, ar yra **ryšys tarp pakuotės dydžio ir perkamos kavos kiekio**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*...(žr. 6.13 pav.). Kintamasis *Pakuotės_dydis* perkeliamas į langelį *Eilutės*, o kintamasis *Kiekis* perkeliamas į langelį *Stulpeliai.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelį *Korel.* Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai* – *PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.14 pav.).



6.13 pav. Dialogo langai *Požymių dažnių lentelės* ir *Požymių dažnių lentelės: statistika*

Lentelėje *Simetriniai matavimai* pateikti du koreliacijos koeficientai: Pirsono ir Spirmeno. Tyrėjas turi žinoti, koks koeficientas tinkamas. Nagrinėjamas kiekybinių kintamųjų ryšys, Pirsono koreliacijos koeficientas $R = 0{,}760$, koreliacija stipri ir teigiama. Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Spirmano koreliacija	,89	,02	8,58	
Intervalų x intervalų	Pirsono R	,76	,06	3,70	
N galiojančių atvejų		60			

6.14 pav. Pirsono koreliacijos koeficiento skaičiavimo rezultatas PSPP terpėje

Hipotezės H_0 teisingumui patikrinti naudosime *t* kriterijų. Lentelėje pateikta t kriterijaus reikšmė (*Apytiksl. T*), $t = 3{,}7$.

Parenkamas reikšmingumo lygmuo $\alpha = 0{,}05$. Hipotezė H_0 atmetama, jei absoliuti *t* reikšmė viršija Stjudento skirstinio su *(n–2)* laisvės laipsniais *α/2* lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

$t_{\alpha/2(n-2)} = t_{0{,}025/2(60-2)} = 2 < t = 3{,}7$. Patvirtinta hipotezė H_1, koreliacija reikšminga.

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze* →*Descriptive Statistics* → *Crosstabs*.... Kintamasis *Pakuotės_dydis* perkeliamas į langelį *Row(s)*, o kintamasis *Kiekis* perkeliamas į langelį *Column(s)*. Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics*. Šiame lange reikia pažymėti langelį *Correlations*. Paspaudę *Continue* mygtuką

grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.15 pav.).

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Interval by Interval	Pearson's R	,760	,058	8,899	,000(c)
Ordinal by Ordinal	Spearman Correlation	,894	,026	15,186	,000(c)
N of Valid Cases		60			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.
c Based on normal approximation.

6.15 pav. Pirsono koreliacijos koeficiento skaičiavimo rezultatas SPSS terpėje

Lentelėje *Symmetric Measures* pateikti du koreliacijos koeficientai: Pirsono ir Spirmeno. Nagrinėjamas kiekybinių kintamųjų ryšys, Pirsono koreliacijos koeficientas (*Pearson's R*) $r = 0{,}760$, koreliacija stipri ir teigiama. Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Stebimasis reikšmingumo lygmuo (*Approx. Sig.*) $p\text{-level} = 0{,}000$, tai mažiau už pasirinktą reikšmingumo lygmenį $\alpha = 0{,}05$. Parvirtinta hipotezę H_1. Lentelėje *Symmetric Measures* pateikta ir *t* kriterijaus reikšmė (*Approx. T(b)*). Norint koreliacijos koeficiento reikšmingumą galima patikrinti ir pasitelkus *T* statistiką. Galima pastebėti, kad rezultatai nesiskiria, pagal tai, koks taikomas skaičiavimo būdas.

Atliktos procedūros sintaksė pateikta 6.5 lentelėje.

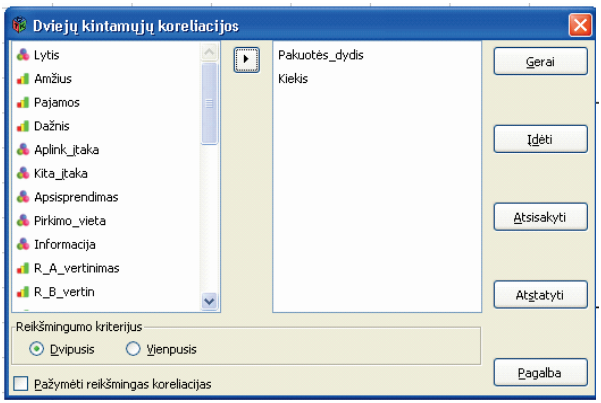
6.5 lentelė. Pirsono koreliacijos koeficiento skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES=Pakuotės_dydis BY Kiekis /FORMAT=AVALUELABELS TABLES PIVOT /STATISTICS=CORR /CELLS=COUNT ROW COLUMN TOTAL.	CROSSTABS /TABLES=Pakuotės_dydis BY Kiekis /FORMAT= AVALUE TABLES /STATISTIC=CORR /CELLS= COUNT /COUNT ROUND CELL.

Galima kitaip apskaičiuoti pakuotės dydžio ir perkamos kavos kiekio Pirsono koreliacijos koeficientą.

PSPP terpė

Analizės dialogo langas *Dviejų kintamųjų koreliacijos* atveriamas komanda *Analizuoti* → *Dviejų kintamųjų koreliacija...* Kintamieji *Pakuotės_dydis* ir *Kiekis* perkeliami į analizuojamų kintamųjų langelį. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.16 pav.).



Koreliacijos

		Pakuotės_dydis	*Kiekis*
Pakuotės_dydis	*Pirsono koreliacija*	1,00	,76
	p-reikšmė (2-pusė)		,00
	N	60	60
Kiekis	*Pirsono koreliacija*	,76	1,00
	p-reikšmė (2-pusė)	,00	
	N	60	60

6.16 pav. Dialogo langas *Požymių dažnių lentelės: statistika* ir Pirsono koreliacijos koeficiento skaičiavimo rezultatas PSPP terpėje

Pirsono koreliacijos koeficientas $r = 0,76$, koreliacija stipri ir teigiama. Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Stebimasis reikšmingumo lygmuo *p-level (p-reišmė)* = *0,00*, tai mažiau už pasirinktą reikšmingumo lygmenį $\alpha = 0,05$. Parvirtinta hipotezė H_1.

SPSS terpė

Analizės dialogo langas *Bivariate Correlations* atveriamas komanda *Analyze* → *Correlate* → *Bivariate...* Kintamieji *Pakuotės_dydis* ir *Kiekis* perkeliami į langelį *Variables.* Paskui reikia pažymėti langelį *Pearson*. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.17 pav.). Pirsono koreliacijos koeficientas $r = 0{,}760$, koreliacija stipri ir teigiama. Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Stebimasis reikšmingumo lygmuo *p-level (Sig. (2-tailed))= 0,000*, tai mažiau už pasirinktą reikšmingumo lygmenį $\alpha = 0{,}05$. Parvirtinta hipotezė H_1.

Correlations

		Pakuotės_dydis	Kiekis
Pakuotės_dydis	Pearson Correlation	1	,760**
	Sig. (2-tailed)		,000
	N	60	60
Kiekis	Pearson Correlation	,760**	1
	Sig. (2-tailed)	,000	
	N	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

6.17 pav. Pirsono koreliacijos koeficiento skaičiavimo rezultatas SPSS terpėje

Atliktos procedūros sintaksė pateikta 6.6 lentelėje.

6.6 lentelė. Pirsono koreliacijos koeficiento skaičiavimo sintaksė

PSPP	SPSS
CORRELATION /VARIABLES=Pakuotės_dydis Kiekis /PRINT=TWOTAIL SIG.	CORRELATIONS /VARIABLES=Pakuotės_dydis Kiekis /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.

PSPP terpėje vienu metu galima nustatyti kelių kiekybinių kintamųjų koreliacijos koeficientus. Analizės dialogo langas *Dviejų kintamųjų koreliacijos* atveriamas komanda *Analizuoti* → *Dviejų kintamųjų*

koreliacija... Analizuojami kintamieji perkeliami į analizuojamų kintamųjų langelį. Procedūra baigiama mygtuko *Gerai* paspaudimu.

Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.18 pav.). Lentelėje *Koreliacijos* pateikti kintamųjų poriniai koreliacijos koeficientai ir jų reikšmingumo įverčiai.

Atliktos procedūros sintaksė:

```
CORRELATION
  /VARIABLES = Pakuotės_dydis Kiekis A_kaina
B_kaina Amžius2 Pajamos2
  /PRINT = TWOTAIL SIG.
```

Koreliacijos

		Pakuotės_dydis	Kiekis	A_kaina	B_kaina	Amžius2	Pajamos2
Pakuotės_dydis	Pirsono koreliacija	1,00	,76	,25	-,21	,24	,39
	p-reikšmė (2-pusė)		,00	,06	,10	,07	,00
	N	60	60	60	60	60	60
Kiekis	Pirsono koreliacija	,76	1,00	,10	-,05	,17	,37
	p-reikšmė (2-pusė)	,00		,43	,71	,20	,00
	N	60	60	60	60	60	60
A_kaina	Pirsono koreliacija	,25	,10	1,00	-,30	,35	,07
	p-reikšmė (2-pusė)	,06	,43		,02	,01	,57
	N	60	60	60	60	60	60
B_kaina	Pirsono koreliacija	-,21	-,05	-,30	1,00	-,26	-,11
	p-reikšmė (2-pusė)	,10	,71	,02		,04	,42
	N	60	60	60	60	60	60
Amžius2	Pirsono koreliacija	,24	,17	,35	-,26	1,00	,39
	p-reikšmė (2-pusė)	,07	,20	,01	,04		,00
	N	60	60	60	60	60	60
Pajamos2	Pirsono koreliacija	,39	,37	,07	-,11	,39	1,00
	p-reikšmė (2-pusė)	,00	,00	,57	,42	,00	
	N	60	60	60	60	60	60

6.18 pav. Kelių kiekybinių kintamųjų Pirsono koreliacijos koeficiento skaičiavimo rezultatas

6.2.3. Spirmeno (*Spearman*) koreliacijos koeficientas

Ryšio stiprumui tarp dviejų ranginių kintamųjų, arba kiekybinių kintamųjų, kurių skirstinys nėra normalusis, taikomas Spirmeno koreliacijos koeficientas, kuris ryšio stiprumui įvertinti naudoja ne pačias kintamųjų reikšmes, o jų rangus.

Skaičiuojant šį koreliacijos koeficientą palyginami dviejų sekų rangai:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n^3 - n}, \tag{6.12}$$

čia r_s – Spirmeno koreliacijos koeficientas, n – imties dydis, D_i – skirtumas tarp vienos poros rangų.

Šis paprastas Spirmeno koeficiento skaičiavimo būdas naudojamas tada, kai visi rangai skirtingi, t. y. kai tarp stebimų kintamųjų nebuvo pasikartojančių reikšmių. Jeigu yra sutampančių rangų, Spirmeno koreliacijos koeficientas skaičiuojamas pagal apibendrintą formulę:

$$r_S = \frac{\sum_{i=1}^{n}\left(R_{xi} - \frac{n+1}{2}\right)\left(R_{yi} - \frac{n+1}{2}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{xi} - \frac{n+1}{2}\right)^2}\sqrt{\sum_{i=1}^{n}\left(R_{yi} - \frac{n+1}{2}\right)^2}}, \tag{6.13}$$

čia $R_{x1}, R_{x2}, \cdots, R_{xn}$ – pirmosios sekos rangai, o $R_{y1}, R_{y2}, \cdots, R_{yn}$ – antrosios sekos rangai.

Jeigu imties dydis yra $n \geq 10$, tai Spirmeno koreliacijos koeficiento skirstinys yra apytiksliai Stjudento t skirstinys su $n - 2$ laisvės laipsniais. Hipotezę dėl šio koreliacijos koeficiento reikšmingumo galima tikrinti t kriterijumi:

$$t = r_S\sqrt{\frac{n-2}{1-r_S^2}}, \tag{6.14}$$

čia: r_S – apskaičiuotas Spirmeno koreliacijos koeficientas, n – imties dydis.

Sprendimo priėmimo taisyklė pasirenkama pagal skaičiavimo rezultatų pateikimo būdą (ar skaičiavimams naudota statistinė programa, ar kitokia skaičiavimo priemonė).

Kai statistinio t kriterijaus reikšmė apskaičiuota ir stebimasis reikšmingumo lygmuo *p-level* nežinomas, sprendimą galima priimti lyginant apskaičiuotą ir kritinę t reikšmes. Parenkamas reikšmingumo

lygmuo α, statistinėje lentelėje (žr. 2 priedą) randama Stjudento skirstinio su (*n – 2*) laisvės laipsniais *α/2* lygmens kritinė reikšmė $t_{\alpha/2}(n-2)$, čia: *n* – imties dydis. Hipotezė H_0 atmetama, jei apskaičiuota *t* kriterijaus absoliuti reikšmė didesnė nei $t_{\alpha/2}(n-2)$.

Kai skaičiavimui naudojama statistinė programa, ar atmesti hipotezę H_0, sprendžiama pagal stebimąjį reikšmingumo lygmenį *p-level*. Jei stebimojo reikšmingumo lygmens *p-level* reikšmė mažesnė už pasirinktą reikšmingumo lygmenį α (*p-level*<α), tai hipotezė H_0 atmetama ir daroma išvada, kad koreliacija yra reikšminga. Jei *p-level* reikšmė didesnė ar tokia pati, kaip reikšmingumo lygmuo α (p-level ≥ α), tai koreliacija nereikšminga, o kintamieji nepriklausomi.

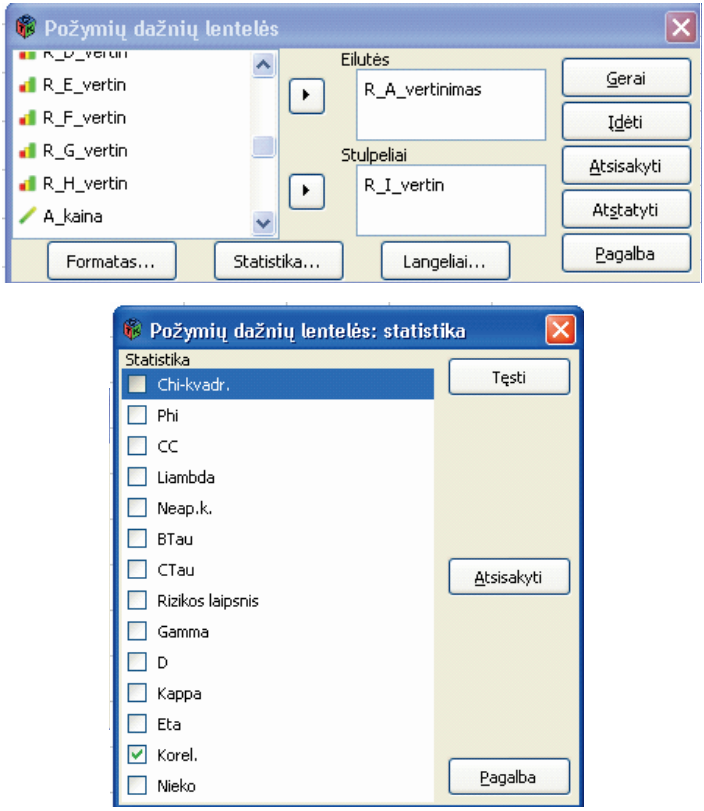
6.2.4. Spirmeno koreliacijos koeficiento skaičiavimas

Patikrinsime, ar yra **ryšys tarp skirtingos kavos rūšies įvertinimo**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės...* (žr. 6.19 pav.). Tam, kad būtų galima nustatyti ryšį tarp A rūšies ir I rūšies įvertinimo, kintamasis *R_A_Vertinimas* perkeliamas į langelį *Eilutės,* o kintamasis *R_I_Vertin* perkeliamas į langelį *Stulpeliai.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelį *Korel.* Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.20 pav.).

Lentelėje *Simetriniai matavimai* pateikti du koreliacijos koeficientai: Pirsono ir Spirmeno. Nagrinėjamas ranginių kintamųjų ryšys, Spirmeno koreliacijos koeficientas $r_s = -0{,}56$, koreliacija vidutinė, neigiama: kuo geriau vertinama A rūšis, tuo blogiau vertinama I rūšis.



6.19 pav. Dialogo langai *Požymių dažnių lentelės* ir *Požymių dažnių lentelės: statistika*

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Spirmano koreliacija	-,56	,09	-9,52	
Intervalų x intervalų	Pirsono R	-,54	,09	-9,78	
N galiojančių atvejų		60			

6.20 pav. Spirmeno koreliacijos koeficiento skaičiavimo rezultatas PSPP terpėje

Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Hipotezės H_0 teisingumui patikrinti naudosime *t* kriterijų. Lentelėje pateikta t kriterijaus reikšmė (*Apytiksl. T*), $| t | = 9{,}52$.

Parenkamas reikšmingumo lygmuo α = 0,05. Hipotezė H_0 atmetama, jei absoliuti *t* reikšmė viršija Stjudento skirstinio su *(n–2)* laisvės laipsniais *α/2* lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

$t_{\alpha/2(n-2)} = t_{0{,}025/2(60-2)} = 2 < | t | = 9{,}52$. Parvirtinta hipotezė H_1, koreliacija reikšminga.

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze →Descriptive Statistics → Crosstabs....* Kintamasis *R_A_Vertinimas* perkeliamas į langelį *Row(s)*, o kintamasis *R_I_Vertin* perkeliamas į langelį *Column(s).* Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics.* Šiame lange reikia pažymėti langelį *Correlations.* Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.21 pav.).

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Interval by Interval	Pearson's R	-,544	,091	-4,932	,000(c)
Ordinal by Ordinal	Spearman Correlation	-,556	,088	-5,093	,000(c)
N of Valid Cases		60			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.
c Based on normal approximation.

6.21 pav. Spirmeno koreliacijos koeficiento skaičiavimo rezultatas SPSS terpėje

Šiuo atveju lentelėje *Symmetric Measures* pateikti du koreliacijos koeficientai: Pirsono ir Spirmeno. Nagrinėjamas ranginių kintamųjų ryšys, Spirmeno koreliacijos koeficientas (*Spearman Correlation*) $r_s = -0{,}556$, koreliacija vidutinė, neigiama: kuo geriau vertinama A rūšis, tuo blogiau vertinama I rūšis. Kad būtų galima patikrinti koreliacijos koeficiento reikšmingumą, suformuluojame hipotezes:

H_0: koreliacijos koeficientas lygus nuliui,

H_1: koreliacijos koeficientas nelygus nuliui.

Stebimasis reikšmingumo lygmuo (*Approx. Sig.*) *p-level = 0,000*, tai mažiau už pasirinktą reikšmingumo lygmenį $\alpha = 0,05$. Parvirtinta hipotezė H_1. Galima pastebėti, kad rezultatai nesiskiria pagal tai, koks taikomas skaičiavimo būdas. Lentelėje *Symmetric Measures* pateikta ir *t* kriterijaus reikšmė (*Approx. T(b)*). Norint koreliacijos koeficiento reikšmingumą galima patikrinti ir pasitelkus *T* statistiką (žr. 6.2.1 poskyrį).

Atliktos procedūros sintaksė pateikta 6.7 lentelėje.

6.7 lentelė. Spirmeno koreliacijos koeficiento skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES= R_A_vertinimas BY R_I_vertin /FORMAT=AVALUE LABELS TABLES PIVOT /STATISTICS=CORR /CELLS=COUNT ROW COLUMN TOTAL.	CROSSTABS /TABLES=R_A_vertinimas BY R_I_vertin /FORMAT= AVALUE TABLES /STATISTIC=CORR /CELLS= COUNT /COUNT ROUND CELL.

6.2.5. Kiti kokybinių kintamųjų ryšio matai

Be dažniausiai taikomo Spirmeno ranginės koreliacijos koeficiento, yra keli koeficientai, kurių skaičiavimai remiasi nukrypimų (vadinamųjų inversijų) nuo nustatytos tvarkos skaičiumi. Inversijų skaičius nustatomas surikiavus didėjimo tvarka vieno iš dviejų kintamųjų, tarp kurių norima nustatyti tarpusavio ryšio laipsnį, reikšmes ir šalia jų užrašant atitinkamas kito kintamojo reikšmes. Antro kintamojo nukrypimų nuo nustatytos tvarkos skaičius ir bus inversijų skaičius. Inversijų ir sutapimų skaičiaus pagrindu (arba kitaip – suderintų ir nesuderintų porų skaičiaus skirtumo pagrindu) skaičiuojamas kintamųjų tarpusavio ryšio koeficientas, kurio reikšmės yra nuo –1 iki +1. Tai yra ***Gamma***, ***Somers'd***, ***Kendall's tau-b*** (taikomas kvadratinių dažnių lentelių atveju, t. y. kai eilučių ir stulpelių skaičius vienodas), ***Kendall's tau-c*** (taikomas bet kokio tipo lentelėms).

Patikrinsime, ar yra **ryšys tarp skirtingos kavos rūšies įvertinimo**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*.... Tam, kad būtų galima nustatyti ryšį tarp A rūšies ir I rūšies įvertinimo, kintamasis *R_A_Vertinimas* perkeliamas į langelį *Eilutės,* o kintamasis *R_I_Vertin* perkeliamas į langelį *Stulpeliai.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelius *BTau, CTau, Gamma, D.* Paspaudę *Tęsti* mygtuką, grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.22 pav.).

Gautos koeficientų reikšmės pateiktos lentelėse *Simetriniai matavimai* ir *Kryptingi matavimai.*

Kad būtų galima patikrinti koeficientų reikšmingumą, suformuluojame hipotezes:

H_0: koeficientas lygus nuliui,

H_1: koeficientas nelygus nuliui.

Hipotezės H_0 teisingumui patikrinti naudosime *t* kriterijų (*Apytiksl. T*).

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Kendall tau-b	-,41	,08	-5,44	
	Kendall tau-c	-,40	,07	-5,44	
	Gamma	-,45	,08	-5,44	
N galiojančių atvejų		60			

Kryptingi matavimai.

Kategorija	*Statistika*	*Tipas*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Somers d	Simetrinis	-,41	,01	-5,44	
		R_A_vertinimas priklausomas	-,41	,08	-5,44	
		R_I_vertin priklausomas	-,41	,08	-5,44	

6.22 pav. Kokybinių kintamųjų ryšio matų skaičiavimas PSPP terpėje

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze* →*Descriptive Statistics* → *Crosstabs*.... Kintamasis *R_A_Vertinimas* perkeliamas į langelį *Row(s)*, o kintamasis *R_I_Vertin* perkeliamas į langelį *Column(s).* Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statis-*

tics. Šiame lange reikia pažymėti langelius *Gamma, Sommers'd, Kendall's tau b, Kendall's tau c.* Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.23 pav.).

Kad būtų galima patikrinti koeficientų reikšmingumą, suformuluojame hipotezes:

H_0: koeficientas lygus nuliui,

H_1: koeficientas nelygus nuliui.

Stebimasis reikšmingumo lygmuo (*Approx. Sig.*) *p-level = 0,000*, tai mažiau už pasirinktą reikšmingumo lygmenį $\alpha = 0,05$. Parvirtinta hipotezė H_1. Galima pastebėti, kad rezultatai nesiskiria pagal tai, koks taikomas skaičiavimo būdas. Lentelėje *Symmetric Measures* pateikta ir *t* kriterijaus reikšmė (*Approx. T(b)*). Norint koreliacijos koeficiento reikšmingumą galima patikrinti ir *T* statistika (žr. 6.2.1 poskyrį).

Directional Measures

			Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	-,411	,076	-5,440	,000
		R_A_vertinimas Dependent	-,409	,076	-5,440	,000
		R_I_vertin Dependent	-,412	,077	-5,440	,000

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	-,411	,076	-5,440	,000
	Kendall's tau-c	-,396	,073	-5,440	,000
	Gamma	-,451	,085	-5,440	,000
N of Valid Cases		60			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

6.23 pav. Kokybinių kintamųjų ryšio matų skaičiavimas SPSS terpėje

Kappa* koeficientas** taikomas dviejų ekspertų, vertinančių tą patį objektą ar reiškinį, išvadų suderinamumui nustatyti. Koeficiento reikšmė 1 rodo visišką ekspertų vertinimų sutapimą. ***Kappa koeficientas taikomas tik tada, kai abu kintamieji turi tas pačias kategorijų reikšmes ir vienodą kategorijų skaičių.

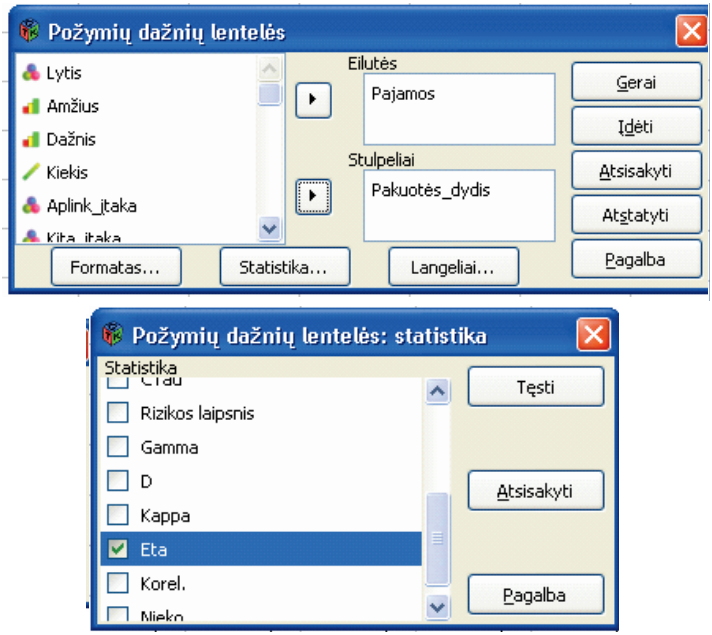
6.2.6. Kiti tarpusavio ryšio matai

***Eta (Nominal by Interval)* koeficientas** yra taikomas, kai nepriklausomas kintamasis yra nominalusis arba ranginis, o priklausomas kintamasis – kiekybinis. Nominalus arba ranginis kintamasis turi turėti skaitmeninį kodavimą (*Numeric* tipo).

Patikrinsime, ar yra **pajamų ir pakuotės dydžio ryšys**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*...(žr. 6.24 pav.). Kintamasis *Pajamos* perkeliamas į langelį *Eilutės*, o kintamasis *Pakuotės_dydis* perkeliamas į langelį *Stulpeliai*. Mygtuko *Statistika*... paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika*. Šiame lange reikia pažymėti langelį *Korel*. Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.25 pav.).

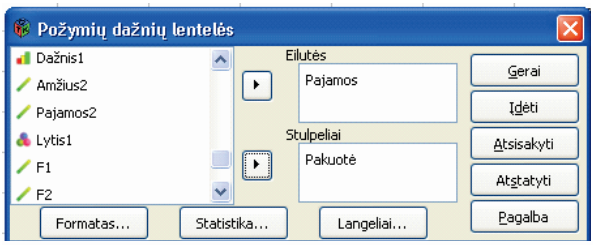


6.24 pav. *Eta* koreliacijos koeficiento skaičiavimas PSPP terpėje

Kryptingi matavimai.

Kategorija	Statistika	Tipas	Reikšmė	Asimpt. std. paklaida	Apytiksl. T	Apytiksl. p-reikšmė
Pavadinimų x intervalų	Eta	Pajamos priklausomas	,57			
		Pakuotės_dydis priklausomas	,62			

6.25 pav. *Eta* koreliacijos koeficiento skaičiavimo rezultatas PSPP terpėje



Simetriniai matavimai.

Kategorija	Statistika	Reikšmė	Asimpt. std. paklaida	Apytiksl. T	Apytiksl. p-reikšmė
Rangų x rangų	Spirmano koreliacija	,53	,09	5,03	
Intervalų x intervalų	Pirsono R	,46	,10	5,79	
N galiojančių atvejų		54			

6.26 pav. Spirmeno koreliacijos koeficiento skaičiavimo rezultatas PSPP terpėje

Nagrinėjamu atveju *Pajamos* – ranginis kintamasis (nepriklausomas), *Pakuotės_dydis* – kiekybinys kintamasis (priklausomas). Pasirenkame koeficiento reikšmę iš eilutės *Pakuotės_dydis priklausomas.* Eta koeficiento reikšmė *0,62*. Yra vidutinė teigiama pakuotės dydžio priklausomybė nuo pajamų.

Jeigu suranguotume pakuotės dydžius (naujas kintamasis *Pakuotė*) ir apskaičiuotume kintamųjų *Pajamos* ir *Pakuotė* Spirmeno koreliacijos koeficientą, gautume $r_s = 0,53$ (žr. 6.26 pav.). Šis koeficientas rodo silpnesnį kintamųjų ryšį nei *Eta* koeficientas.

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze* →*Descriptive Statistics* → *Crosstabs*.... Kintamasis *Pajamos* perkeliamas į langelį *Row(s)*, o kintamasis *Pakuotės_dydis* perkeliamas į langelį *Column(s).*

Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics.* Šiame lange reikia pažymėti langelį *Eta* srityje *Nominal by Interval.* Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.27 pav.).

Nagrinėjamu atveju *Pajamos* – ranginis kintamasis (nepriklausomas), *Pakuotės_dydis* – kiekybinis kintamasis (priklausomas). Pasirenkame koeficiento reikšmę iš eilutės *Pakuotės_dydis Dependent.* Eta koeficiento reikšmė *0,616*. Yra vidutinė teigiama pakuotės dydžio priklausomybė nuo pajamų.

			Value
Nominal by Interval	Eta	Pajamos Dependent	,569
		Pakuotės_dydis Dependent	,616

6.27 pav. *Eta (Nominal by Interval)* koreliacijos koeficiento skaičiavimo rezultatas SPSS terpėje

Atliktos procedūros sintaksė pateikta 6.8 lentelėje.

6.8 lentelė. *Eta* koreliacijos koeficiento skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES=Pajamos BY Pakuotės_dydis /FORMAT=AVALUELABELS TABLES PIVOT /STATISTICS=CHISQ ETA /CELLS=COUNT ROW COLUMN TOTAL.	CROSSTABS /TABLES=Pajamos BY Pakuotės_dydis /FORMAT= AVALUE TABLES /STATISTIC=ETA /CELLS= COUNT /COUNT ROUND CELL.

Risk (rizikos laipsnis) yra skaičiuojamas 2×2 dažnių lentelėms, sudarytoms laikantis tam tikrų toliau išdėstytų reikalavimų. Skaičiuojant rizikos matą, analizuojamas vadinamasis rizikos kintamasis, kuris turi dvi kategorijas ir nurodo, įvyko tam tikras įvykis ar ne. Rizikos kintamasis analizuojamas priežastinio (nepriklausomo) kintamojo, kuris irgi turi būti binarinis (turintis dvi kategorijas), atžvilgiu. Pavyzdžiui,

atlikta 60 respondentų apklausa apie aplinkinių įtaką kavos pirkimui. Apklausos rezultatai pateikti 6.9 lentelėje.

6.9 lentelė. Porinė dažnių lentelė

Aplinkinių įtaka	Taip	Ne
Vyrai	a = 20	b = 10
Moterys	c = 18	d = 12

Aplinkinių įtaka yra rizikos kintamasis, o *lytis* – nepriklausomas (priežastinis) kintamasis. SPSS rizikos matas yra apskaičiuojamas pagal šias formules:

$$R_0 = \frac{a \cdot d}{b \cdot c}, \text{ galimybių santykis (}Odds\ Ratio\text{)}, \tag{6.15}$$

$$R_1 = \frac{a \cdot (c + d)}{c \cdot (a + b)}, \text{ santykinės rizikos koeficientas,} \tag{6.16}$$

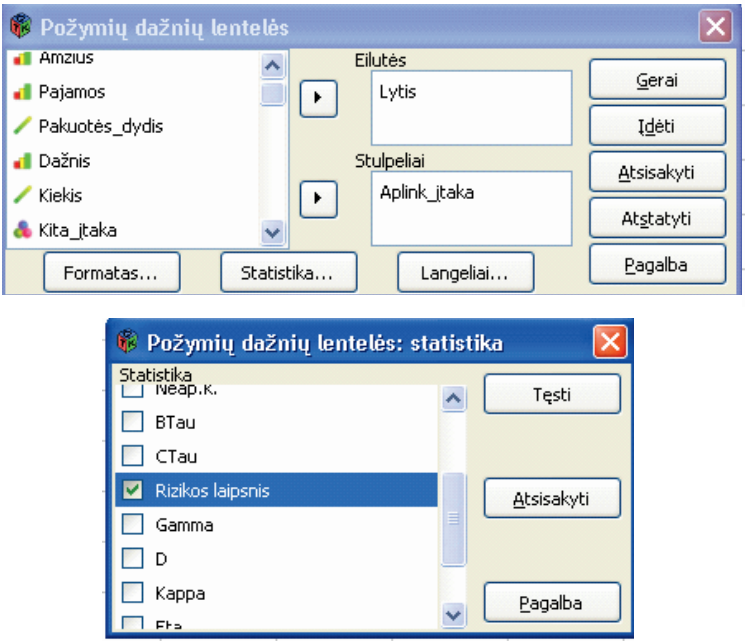
$$R_2 = \frac{b \cdot (c + d)}{d \cdot (a + b)}, \text{ santykinės rizikos koeficientas,} \tag{6.17}$$

čia *a, b, c, d* – lentelėje pateikti dažniai.

Apskaičiuosime **riziką pasiduoti aplinkinių įtakai**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Aprašomoji statistika* → *Požymių dažnių lentelės*...(žr. 6.28 pav.). Kintamasis *Lytis* perkeliamas į langelį *Eilutės*, o kintamasis *Aplink_įtaka* perkeliamas į langelį *Stulpeliai.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Požymių dažnių lentelės: statistika.* Šiame lange reikia pažymėti langelį *Rizikos laipsnis.* Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 6.29 pav.).



6.28 pav. Rizikos laipsnio skaičiavimas PSPP terpėje

Rizikos įvertis lentelėje pateiktas dydis – *Galimybių santykis: Aplink_įtaka (0/1)* – rodo, kad rizika pasiduoti aplinkinių įtakai moterims (kodas **1**) vidutiniškai yra 1,333 karto didesnė negu vyrams (kodas **2**). Reikšmės *Apačia = 0,46* ir *Viršus = 3,82* rodo 95 proc. rizikos pasikliautinojo intervalo $[\bar{x} - 2s; \bar{x} + 2s]$ ribas, čia $\bar{x}$ – rizikos santykio reikšmės vidurkis, s – šios reikšmės standartinis nuokrypis. Galima teigti, kad su patikimumu 95 proc. moterų rizika pasiduoti aplinkinių įtakai sudaro nuo 0,465 iki 3,823 vyrų rizikos. Kiti du koeficientai vadinami santykinės rizikos koeficientais.

Rizikos įvertis.

Statistika	*Reikšmė*	*95% pasikliautinasis intervalas*	
		Apačia	*Viršus*
Galimybių santykis: Aplink_įtaka (0 / 1)	1,33	,46	3,82
Kohortai Lytis = 1	1,20	,61	2,34
Kohortai Lytis = 2	,90	,61	1,32
N galiojančių atvejų	60		

6.29 pav. Rizikos laipsnio skaičiavimo rezultatas PSPP terpėje

SPSS terpė

Analizės dialogo langas atveriamas komanda *Analyze →Descriptive Statistics → Crosstabs....* Nepriklausomas kintamasis *Lytis* perkeliamas į langelį *Row(s)*, o rizikos kintamasis *Aplink_įtaka* perkeliamas į langelį *Column(s).* Mygtuko *Statistics* paspaudimas atveria dialogo langą *Crosstabs: Statistics.* Šiame lange reikia pažymėti langelį *Risk.* Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 6.30 pav.).

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Lytis (Moteris / Vyras)	1,333	,465	3,823
For cohort Aplink_įtaka = ne	1,200	,614	2,344
For cohort Aplink_įtaka = taip	,900	,611	1,325
N of Valid Cases	60		

6.30 pav. Rizikos laipsnio įvertinimo rezultatas

Risk Estimate (rizikos įvertinimas) lentelėje pateiktas dydis – galimybių santykis (*Odds Ratio for Lytis (Moteris / Vyras)*) – rodo, kad rizika pasiduoti aplinkinių įtakai moterims vidutiniškai yra 1,333 karto didesnė negu vyrams. Reikšmės *Lower = 0,465* ir *Upper = 3,823* rodo 95 proc. rizikos pasikliautinojo intervalo $[\bar{x} - 2s; \bar{x} + 2s]$ ribas, čia $\bar{x}$ – rizikos santykio reikšmės vidurkis, *s* – šios reikšmės standartinis nuokrypis. Galima teigti, kad su patikimumu 95 proc. moterų rizika pasiduoti aplinkinių įtakai sudaro nuo 0,465 iki 3,823 vyrų rizikos. Kiti du koeficientai vadinami santykinės rizikos koeficientais.

Atliktos procedūros sintaksė pateikta 6.10 lentelėje.

6.10 lentelė. Rizikos laipsnio skaičiavimo sintaksė

PSPP	SPSS
CROSSTABS /TABLES=Lytis BY Aplink_įtaka /FORMAT=AVALUELABELS TABLES PIVOT /STATISTICS=CHISQ RISK /CELLS=COUNT ROW COLUMN TOTAL.	CROSSTABS /TABLES=Lytis1 BY Aplink_įtaka /FORMAT= AVALUE TABLES /STATISTIC=CORR RISK /CELLS= COUNT /COUNT ROUND CELL.

KONTROLINIAI KLAUSIMAI

6.1. Kokiais atvejais kintamųjų ryšiui nustatyti taikomas Pirsono χ^2 suderinamumo kriterijus?

6.2. Ar galima Pirsono χ^2 suderinamumo kriterijų taikyti kiekybiniams kintamiesiems?

6.3. Paaiškinkite *tikėtinų dažnių* sampratą.

6.4. Kaip nustatomas statistinio ryšio stiprumas, taikant Pirsono χ^2 suderinamumo kriterijų?

6.5. Kokie Pirsono χ^2 suderinamumo kriterijaus taikymo apribojimai?

6.6. Kokiais atvejais naudojamas Spirmeno koreliacijos koeficientas?

6.7. Kokiais atvejais naudojamas Pirsono koreliacijos koeficientas?

6.8. Kaip priimamas sprendimas, ar atmesti hipotezę H_0: koreliacijos koeficientas lygus nuliui?

6.9. Kaip koreliaciniai ryšiai skirstomi pagal formą?

6.10. Kaip koreliaciniai ryšiai skirstomi pagal kryptį?

6.11. Ką rodo determinacijos koeficientas?

UŽDUOTYS

6.1. Tikrinant pirkimo dažnio nuo lyties priklausomybę, gautas rezultatas pateiktas 6.31 pav. Suformuluokite hipotezes ir paaiškinkite gautą rezultatą.

Santrauka.

	Atvejų					
	Galiojančių		Praleista		Iš viso	
	N	Procentai	N	Procentai	N	Procentai
Lytis * Dažnis1	60	100.0%	0	0.0%	60	100.0%

Lytis * Dažnis1 [kiekis].

Lytis	*Dažnis1*			Iš viso
	1,00	2,00	3,00	
1	13,0	7,0	10,0	30,0
2	7,0	14,0	9,0	30,0
Iš viso	20,0	21,0	19,0	60,0

Chi-kvadrato kriterij.

Statistika	*Reikšmė*	*skirt.*	*Asimp. p-reikšmė (2-pusė)*
Pirsono Chi-kvadratas	4,19	2	,12
Tikėtinumo santykis	4,26	2	,12
Dvitiesinė priklausomybė	,63	1	,43
N galiojančių atvejų	60		

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Pavadinimų x pavadinimų	Phi	,26			
	Kramerio V	,26			
N galiojančių atvejų		60			

6.31 pav. Pirsono χ^2 suderinamumo kriterijaus skaičiavimo rezultatas

6.2. Tikrinant priimtinos jogurto kainos (litais) priklausomybę nuo amžiaus (metais), gautas rezultatas pateiktas 6.32 pav. Suformuluokite hipotezes ir paaiškinkite gautą rezultatą.

6.3. 6.33 pav. pateikta lyties – pajamų porinių dažnių lentelė. Paaiškinkite, ar nagrinėjamu atveju kintamųjų ryšiui nustatyti galima taikyti Pirsono χ^2 suderinamumo kriterijų? Kodėl?

6.4. Tyrimo metu gauti kintamieji: respondentų amžius (kiekybinis), respondentų vidutinės pajamos (ranginis), vidutinis pinigų kiekis, išleidžiamas per mėnesį prekei A (kiekybinis), priimtina prekės A kaina (kiekybinis). Apdorojant rezultatus nustatyta, kad amžiaus ir vidutinio pinigų kiekio, išleidžiamo per mėnesį prekei A, duomenų skirstinys yra normalusis, priimtinos prekės A kainos duomenų skirstinys neatitinka normalumo reikalavimų. Numatykite, kokius koreliacijos koeficientus galima apskaičiuoti duomenų ryšiams nustatyti.

Santrauka.

	Atvejų					
	Galiojančių		Praleista		Iš viso	
	N	Procentai	N	Procentai	N	Procentai
Amžius2 * A_kaina	480	100.0%	0	0.0%	480	100.0%

Amžius2 * A_kaina [kiekis].

	A_kaina					
Amžius2	1	2	3	4	5	Iš viso
20,00	80,0	48,0	16,0	40,0	16,0	200,0
30,00	8,0	16,0	24,0	32,0	32,0	112,0
40,00	8,0	16,0	56,0	72,0	16,0	168,0
Iš viso	96,0	80,0	96,0	144,0	64,0	480,0

Chi-kvadrato kriterij.

Statistika	Reikšmė	skirt.	Asimp. p-reikšmė (2-pusė)
Pirsono Chi-kvadratas	151,79	8	,00
Tikėtinumo santykis	152,30	8	,00
Dvitiesinė priklausomybė	65,56	1	,00
N galiojančių atvejų	480		

Simetriniai matavimai.

Kategorija	Statistika	Reikšmė	Asimpt. std. paklaida	Apytiksl. T	Apytiksl. p-reikšmė
Pavadinimų x pavadinimų	Phi	,56			
	Kramerio V	,40			
Rangų x rangų	Spirmano koreliacija	,36	,04	11,54	
Intervalų x intervalų	Pirsono R	,37	,04	15,46	
N galiojančių atvejų		480			

6.32 pav. Pirsono χ^2 suderinamumo kriterijaus skaičiavimo rezultatas

Lytis * Pajamos [kiekis].

	Pajamos					
Lytis	1	2	3	4	5	Iš viso
1	,0	2,0	5,0	13,0	10,0	30,0
2	5,0	10,0	2,0	7,0	6,0	30,0
Iš viso	5,0	12,0	7,0	20,0	16,0	60,0

6.33 pav. Porinių dažnių lentelė

6.5. Analizuojant pajamų ir pirkimo dažnio ryšį, buvo apskaičiuotas Spirmeno koreliacijos koeficientas (žr. 6.34 pav.). Suformuluokite statistines hipotezes ir paaiškinkite gautą rezultatą.

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Spirmano koreliacija	-,01	,13	-,14	
Intervalų x intervalų	Pirsono R	,03	,13	,41	
N galiojančių atvejų		60			

6.34 pav. Koreliacinės analizės rezultatas

6.6. Analizuojant amžiaus ir pinigų, išleistų pirkiniams ryšį, buvo apskaičiuotas Pirsono koreliacijos koeficientas (žr. 6.35 pav.). Suformuluokite statistines hipotezes ir paaiškinkite gautą rezultatą.

Simetriniai matavimai.

Kategorija	*Statistika*	*Reikšmė*	*Asimpt. std. paklaida*	*Apytiksl. T*	*Apytiksl. p-reikšmė*
Rangų x rangų	Spirmano koreliacija	,64	,10	5,41	
Intervalų x intervalų	Pirsono R	,66	,10	3,60	
N galiojančių atvejų		51			

6.35 pav. Koreliacinės analizės rezultatas

7. REGRESINĖ ANALIZĖ

Apskaičiavus koreliacijos koeficientą nustatomas ryšių egzistavimas, tačiau negalima nustatyti ryšio dėsnio. Kintamųjų ryšio apibūdinimui ir prognozavimui taikomi regresijos modeliai. Regresinės analizės prognozės yra kiekybinės – regresijos funkcija aprašoma tiriamo (priklausomo – *pasekmės*) kintamojo priklausomybė nuo kito (nepriklausomo – *priežasties*) kintamojo reikšmių kitimo. Regresinė analizė leidžia parinkti kintamuosius siejančią funkciją (sudaromas regresijos modelis). Modeliui sudaryti taikomas aproksimacijos arba *kreivės pritaikymo* metodas, kai turimiems duomenims parenkama geriausiai tinkama funkcijos kreivė. Preliminariai nustatyti priklausomybės formą (funkcijos tipą) galima nagrinėjant analizuojamų duomenų grafiką.

7.1. TIESINIS (PAPRASTAS) REGRESIJOS MODELIS

Tiesinės regresijos modelis nustato kiekybinių kintamųjų tiesinę priklausomybę. Šis modelis dažnai taikomas socialiniuose moksluose, nes paprastai procesai visuomenėje kinta lėtai ir taikant tiesinį modelį jie prognozuojami ganėtinai gerai. Tiesinė kintamojo Y nuo kintamojo X priklausomybė užrašoma lygtimi:

$$y = \alpha + \beta x + \varepsilon, \tag{7.1}$$

čia α ir β yra nežinomi koeficientai (konstantos), o ε – atsitiktinė paklaida. Atsitiktinę paklaidą gali sudaryti matavimo paklaidos arba neįvardytos ir netiriamos priežastys.

Parinkę įvairias koeficientų α ir β reikšmes, gausime tiesių šeimą. Geriausiai tiesę atitinkantys koeficientai apskaičiuojami taikant matematinį *mažiausių kvadratų metodą*. Šiuo metodu gautos koeficientų α ir β reikšmės – įverčiai (jas žymėsime raidėmis a ir b), randami minimizuojant reiškinį:

$$SSE = \sum_{i=1}^{n} \left(y_i - \alpha - \beta x_i \right)^2. \tag{7.2}$$

Sumą SSE minimizuoja šie koeficientų α ir β įverčiai:

$$b = \frac{n\left(\sum_{i=1}^{n} x_i y_i \right) - \left(\sum_{i=1}^{n} x_i \right)\left(\sum_{i=1}^{n} y_i \right)}{n\left(\sum_{i=1}^{n} x_i^2 \right) - \left(\sum_{i=1}^{n} x_i \right)^2}, \tag{7.3}$$

$$a = \frac{\sum_{i=1}^{n} y_i}{n} - b\frac{\sum_{i=1}^{n} x_i}{n}. \tag{7.4}$$

Lygtis $\hat{y}(x) = a + bx$, kurios koeficientai apibrėžti (7.3) ir (7.4) lygybėmis, yra regresijos tiesės lygtis, o $\varepsilon = y(x) - \hat{y}(x)$ yra liekamoji paklaida, arba liekana. Koeficientai *a* ir *b*, apskaičiuoti iš imties, skirsis nuo koeficientų α ir β, gaunamų iš populiacijos. Tiesinės regresijos atveju koeficientų α ir β skirstiniai yra normalieji su vidurkiais *a* ir *b*.

Regresijos modelio reikšmingumas. Dažniausiai tikrinama hipotezė, kad tarp kintamųjų *X* ir *Y* nėra jokio tiesinio ryšio, t. y. kad koeficientas β lygus nuliui.

H_0: $\beta_1 = 0$,

H_1: $\beta_1 \neq 0$.

Šios hipotezės siejamos su regresijos lygties reikšmingumu. Jeigu hipotezė H_0 neatmetama, tai reiškia, kad tarp kintamųjų *X* ir *Y* tiesinio ryšio nėra. Norint patikrinti šią hipotezę, galima taikyti *t* statistiką:

$$t = \frac{\beta}{s_\beta}, \tag{7.5}$$

čia s_β – β koeficiento standartinis nuokrypis.

Jei nulinė hipotezė, kad nėra jokio tiesinio ryšio, yra teisinga, ši statistika (7.5) turi Stjudento *t* skirstinį su (*n – 2*) laisvės laipsniais. Hipotezę, kad tiesė eina per koordinačių pradžią, t. y. kad koeficientas α lygus nuliui, galima patikrinti naudojantis kita *t* statistika:

$$t = \frac{\alpha}{s_\alpha}, \tag{7.6}$$

čia s_α – α koeficiento standartinis nuokrypis.

Statistika (7.6) taip pat turi Stjudento *t* skirstinį su (*n* – *2*) laisvės laipsniais.

Parinkto modelio tikimas turimiems duomenims

Sudarius modelį, būtina ištirti, ar šis modelis gerai tinka turimiems duomenims. Tiesinio modelio tinkamumui tikrinti naudojami: koreliacijos koeficientas, determinacijos koeficientas ir pataisytasis determinacijos koeficientas.

Determinacijos koeficientu R^2 įvertinamas modelio tikimas visai populiacijai dažnai būna netikslus. Pataisytasis determinacijos koeficientas R_a^2 tiksliau įvertina modelio tikimą duomenims:

$$R_a^2 = R^2 - \frac{p(1-R^2)}{N-p-1}, \tag{7.7}$$

čia: *p* yra nepriklausomų kintamųjų skaičius (nagrinėjamu atveju $p = 1$).

Didelė determinacijos koeficiento reikšmė reiškia, kad duomenys yra mažai nutolę nuo mažiausių kvadratų metodu gautos tiesės. Jei duomenys idealiai atitinka regresijos tiesės lygtį, t. y. visos y_i reikšmės patenka ant regresijos tiesės, determinacijos koeficientas yra 1. Jei determinacijos koeficientas artimas nuliui, tai regresijos tiesės lygtis visiškai netinka prognozei. Praktiškai taikant tiesinę regresinę analizę dažniausiai reikalaujama, kad $R^2 \geq 0,25$. Jei $R^2 < 0,25$, tiesinės regresijos modelio taikymas labai abejotinas.

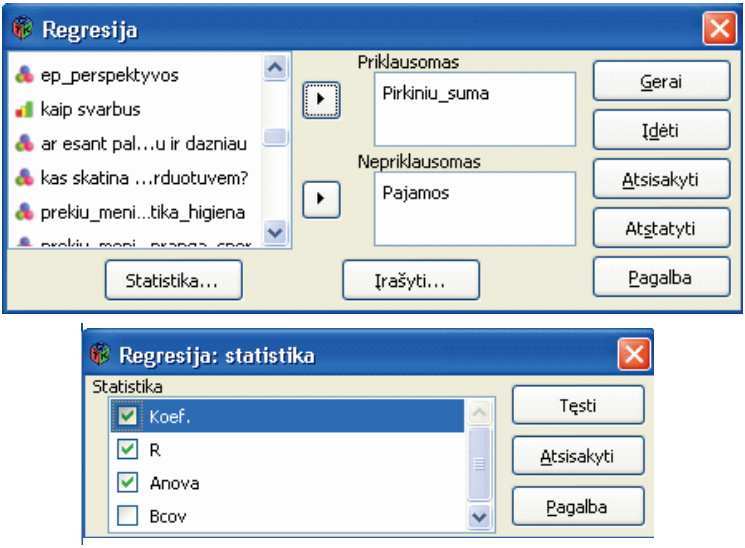
Statistinės programos koeficiento β (kartu ir regresijos lygties) reikšmingumui tikrinti taikomi dispersinės analizės metodai (žr. skyrių „Vienfaktorinė dispersinė analizė“). Šiuo metodu taip pat tikrinama hipotezė apie koeficiento β lygybę nuliui. Jeigu nulinė hipotezė H_0: $\beta = 0$ yra teisinga, tai *F* statistika yra pasiskirsčiusi pagal $F_{1,n-2}$ skirstinį ir nulinė hipotezė atmetama, jeigu apskaičiuota reikšmė $F_0 > F_{\alpha,1,n-2}$. Dispersinės analizės metodu gaunamas analogiškas rezultatas, kaip ir taikant *t* kriterijų. Abi procedūros duoda vienodas išvadas.

7.2. TIESINĖ REGRESINĖ ANALIZĖ PSPP IR SPSS TERPĖJE

Apskaičiuosime **pinigų, išleidžiamų pirkiniams, priklausomybę nuo pajamų**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Tiesinė regresija*...(žr. 7.1 pav.). Kintamasis *Pirkiniu_suma* perkeliamas į langelį *Priklausomas*, o kintamasis *Pajamos* perkeliamas į langelį *Nepriklausomas*. Mygtuko *Statistika*... paspaudimas atveria dialogo langą *Regresija: statistika*. Šiame lange reikia pažymėti langelius *Koef., R, Anova*. Paspaudę *Tęsti* mygtuką, grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 7.2 pav.).



7.1 pav. Regresijos skaičiavimas PSPP terpėje

Modelio tinkamumas nustatomas pagal determinacijos koeficiento R^2 *(R kvadratas)* reikšmę. Modelis laikomas tinkamu, jei determinacijos koeficiento reikšmė $R^2 > 0{,}25$. Analizės rezultatai bei regresijos lygties

koeficientai rodomi lentelėse *Modelio santrauka* ir *Koeficientai*. Lentelėje *Modelio santrauka* pateiktas kintamųjų Pirsono koreliacijos koeficientas $R = 0,66$, determinacijos koeficientas (*R kvadratas = 0,43),* koreguotas determinacijos koeficientas (*Koreguotas R kvadratas = 0,43*). Nagrinėjamu atveju determinacijos koeficientai $R^2 > 0,25$, tiesinis modelis yra tinkamas. Determinacijos koeficientas rodo, kokią tikro proceso dalį aprašo modelis. Nagrinėjamu atveju – tai tik apie 43 procentus. Regresijos lygties koeficientai pateikti lentelėje *Koeficientai.* Apskaičiuotų regresijos lygties koeficientų (konstantų) patikimumą galima patikrinti taikant stebimojo reikšmingumo lygmens *p-level* (*Reišmingumo lygmuo*) ir pasirinkto reikšmingumo lygmens α palyginimą. Tikrinamos hipotezės H_0: $a = 0$ ir H_0: $b = 0$. Hipotezė atmetama, jei *p-level* < α. Nagrinėjamu atveju lygties konstanta *a* (*Konstanta*) yra nepatikima, p-*level* = 0,36 > α = 0,05, kintamojo koeficientas *b* yra patikimas p-*level* = 0,000 < α = 0,05, koeficientas reikšmingai skiriasi nuo nulio. Pasitvirtina hipotezė H_0: $a = 0$ (*p-level = 0,36 > α = 0,05)* ir atmetama hipotezė H_0: $b = 0$ *(p-level = 0,000 < α = 0,05).* Tiesinės regresijos lygtis: *Pirkinių_suma = 29,04 + 0,15 · Pajamos.*

Stulpelyje *Beta* pateikiamas standartizuotų duomenų regresijos lygties koeficientas. Paprastosios tiesinės regresijos atveju šis koeficientas sutampa su pradinių duomenų koreliacijos koeficientu. Lentelėje *ANOVA* pateikti dispersinės analizės rezultatai. Stulpelyje *Kvadratų suma* pateikti: regresijos kvadratų suma (*Regresija*), liekamųjų paklaidų kvadratų suma (Liekana) ir visa kvadratų suma (Iš viso).

Modelio santrauka

R	R kvadratas	Koreguotas R kvadratas	Įverčio std. paklaida
,66	,43	,43	122,52

ANOVA

	Kvadratų suma	skirt.	Vidutinis kvadratas	F	Reikšmingumo lygmuo
Regresija	562015,2	1	562015,2	37,44	,00
Liekana	735553,4	49	15011,29		
Iš viso	1297569	50			

Koeficientai

	B	Std. paklaida	Beta	t	Reikšmingumo lygmuo
(Konstanta)	29,04	31,71	,00	,92	,36
Pajamos	,15	,02	,66	6,12	,00

7.2 pav. Regresinės analizės rezultatas PSPP terpėje

SPSS terpė

Regresijos modelio parametrų vertinimas gaunamas taikant *Regression* procedūrą. Analizės dialogo langas atveriamas komanda *Analyze → Regression → Curve Estimation...* Priklausomą kintamąjį *Pirkinių_suma* perkeliame į langelį *Dependent(s),* nepriklausomą kintamąjį *Pajamos* perkeliame į langelį *Independent.* Srityje *Models* galima pasirinkti norimą regresijos modelį. Norint nubraižyti regresijos grafiką, reikia pažymėti langelį *Plot Models.* Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 7.3 pav.).

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,658	,433	,422	122,521

The independent variable is Pajamos.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	562015,233	1	562015,233	37,439	,000
Residual	735553,395	49	15011,294		
Total	1297568,627	50			

The independent variable is Pajamos.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Pajamos	,150	,025	,658	6,119	,000
(Constant)	29,044	31,709		,916	,364

7.3 pav. Regresinės analizės rezultatas SPSS terpėje

Modelio tinkamumas nustatomas pagal determinacijos koeficiento R^2 *(R Square)* reikšmę. Analizės rezultatai bei regresijos lygties koeficientai rodomi lentelėse *Model Summary* ir *Coefficients*. Lentelėje *Model Summary* pateiktas kintamųjų Pirsono koreliacijos koeficientas *R = 0,658,* determinacijos koeficientas (*R Square = 0,433),* koreguotas determinacijos koeficientas (*Adjusted R Square = 0,422*). Nagrinėjamu atveju determinacijos koeficientai $R^2 > 0,25$, tiesinis modelis yra tinkamas. Determinacijos koeficientas rodo, kokį dalį tikro proceso aprašo modelis. Nagrinėjamu atveju – tai tik apie 43 procentus. Regresijos lyg-

ties koeficientai pateikti lentelėje *Coefficients.* Apskaičiuotų regresijos lygties koeficientų (konstantų) patikimumą galima patikrinti taikant stebimojo reikšmingumo lygmens *p-level* (*Sig.*) ir pasirinkto reikšmingumo lygmens α palyginimą. Tikrinamos hipotezės H_0: $a = 0$ ir H_0: $b = 0$. Hipotezė atmetama, jei $p\text{-level} < \alpha$. Nagrinėjamu atveju lygties konstanta *a* (*Constant*) yra nepatikima, $p\text{-level} = 0{,}364 > \alpha = 0{,}05$, kintamojo koeficientas *b* yra patikimas $p\text{-level} = 0{,}000 < \alpha = 0{,}05$, koeficientas reikšmingai skiriasi nuo nulio. Pasitvirtina hipotezė H_0: $a = 0$ ($p\text{-level} = 0{,}364 > \alpha = 0{,}05$) ir atmetama hipotezė H_0: $b = 0$ ($p\text{-level} = 0{,}000 < \alpha = 0{,}05$).Tiesinės regresijos lygtis: $Pirkinių_suma = 29{,}04 + 0{,}15 \cdot Pajamos$.

Stulpelyje *Beta* pateikiamas standartizuotų duomenų regresijos lygties koeficientas. Paprastosios tiesinės regresijos atveju šis koeficientas sutampa su pradinių duomenų koreliacijos koeficientu. Paskutiniuose dviejuose lentelės *Coefficients* stulpeliuose pateikti koeficientų a ir b 95 proc. pasikliautinųjų intervalų rėžiai. Lentelėje *ANOVA* pateikti dispersinės analizės rezultatai. Stulpelyje kvadratų suma (*Sum of Squares*): regresijos kvadratų suma (*Regression*), liekamųjų paklaidų kvadratų suma (*Residual*) ir visa kvadratų suma (*Total*).

Atliktos procedūros sintaksė pateikta 7.1 lentelėje.

7.1 lentelė. Regresinės analizės sintaksė

PSPP	SPSS
REGRESSION /VARIABLES=Pajamos /DEPENDENT=Pirkinių_suma /STATISTICS=COEFF R ANOVA.	* Curve Estimation. TSET NEWVAR=NONE. CURVEFIT /VARIABLES=Pirkinių_suma WITH Pajamos /CONSTANT /MODEL=LINEAR /PLOT FIT.

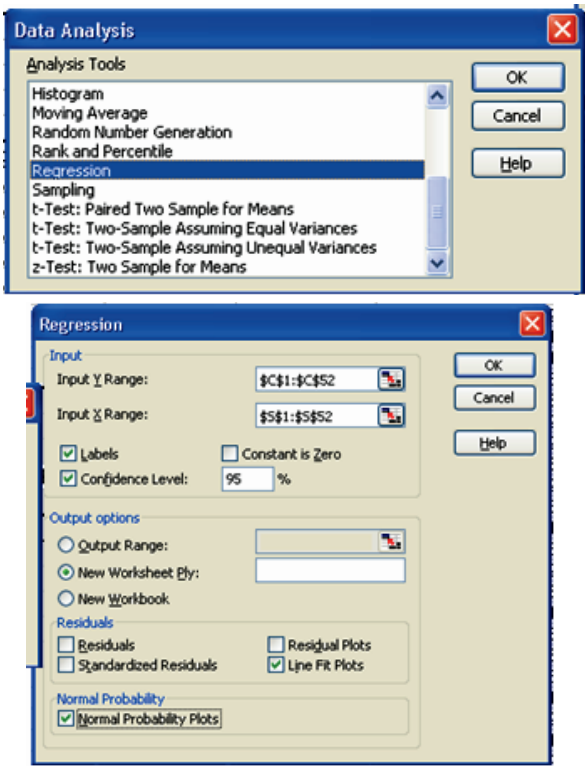
7.3. TIESINĖ REGRESINĖ ANALIZĖ MS EXCEL TERPĖJE

Microsoft Office Excel paketas turi visą statistinių funkcijų sąrašą. Tačiau galima panaudoti ir papildomas *MS Excel* galimybes. *MS Excel*

terpėje įkomponuotas modulis *Data Analysis.* Komanda *Data Analysis* atveria dialogo langą *Data Analysis* (žr. 7.4 pav.). Šiame lange pateiktas galimų statistinių operacijų sąrašas *Analysis Tools.*

Apskaičiuosime **pinigų, išleidžiamų pirkiniams, priklausomybę nuo pajamų**. Tiesinė regresinė analizė atliekama pažymėjus komanda *Regression*. Atveriamas dialogo langas *Regression.*

Nepriklausomo kintamojo *Pajamos* koordinates reikia nurodyti langelyje *Input X range,* priklausomo kintamojo *Pirkiniu_suma* koordinates – *Input Y range.* Jeigu duomenų masyvai pažymėti su pavadinimais, reikia pažymėti langelį *Labels.* Pažymėtas langelis *Confidence Level* nurodo pasikliautinojo intervalo reikšmę. Jeigu pažymėsime langelį *Line Fit Plots,* bus pateiktas regresijos grafikas, *Normal Probability Plots* – kaupiamųjų dažnių grafikas.



7.4 pav. Dialogo langai *Data Analysis* ir *Regression*

Regresinės analizės rezultatai pateikti 7.2 lentelėje ir 7.5 pav.

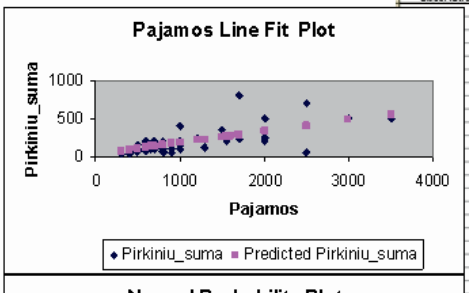
7.2 lentelė. Regresinės analizės rezultatas MS Excel terpėje

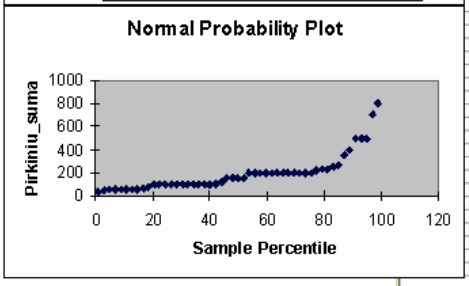
SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0,658125737
R Square	0,433129486
Adjusted R Square	0,4215607
Standard Error	122,5205851
Observations	51

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	562015,2325	562015,2	37,43949	1,53512E-07
Residual	49	735553,3949	15011,29		
Total	50	1297568,627			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	29,04402716	31,70940366	0,915944	0,364182	-34,67840403	92,76645834
Pajamos	0,149941525	0,024505126	6,118782	1,54E-07	0,100696631	0,199186418

Lentelės duomenys analogiški pateiktiems 7.3 pav. Modelio tinkamumas nustatomas pagal determinacijos koeficiento R^2 *(R Square)* reikšmę. Lentelėje pateiktas kintamųjų Pirsono koreliacijos koeficientas *Multiple R = 0,658125737,* determinacijos koeficientas (*R Square = 0,433129486),* koreguotas determinacijos koeficientas (*Adjusted R Square = 0,4215607*). Nagrinėjamu atveju determinacijos koeficientai $R^2 > 0,25$, tiesinis modelis yra tinkamas. Determinacijos koeficientas rodo, kokį dalį tikro proceso aprašo modelis. Nagrinėjamu atveju – tai tik apie 42 procentus. Stulpelyje *Coefficients* pateiktos konstantos (*Intercept)* ir kintamojo koeficiento (*Pajamos*) reikšmės. Apskaičiuotų regresijos lygties koeficientų (konstantų) patikimumą galima patikrinti taikant stebimojo reikšmingumo lygmens *p-level* (*P-value*) ir pasirinkto reikšmingumo lygmens α palyginimą.





RESIDUAL OUTPUT				PROBABILITY OUTPUT	
Observation	Predicted Pirkiniu_suma	Residuals		Percentile	Pirkiniu_suma
1	163,9913993	-60,9913993		0,980392	30
2	119,0089419	80,99105809		2,941176	40
3	74,02648453	-34,0264845		4,901961	50
4	134,0030944	15,99690563		6,862745	50
5	104,0147895	45,98521055		8,823529	50
6	119,0089419	-19,0089419		10,78431	50
7	119,0089419	-59,0089419		12,7451	60
8	163,9913993	-113,991399		14,70588	60
9	134,0030944	-34,0030944		16,66667	70
10	178,9855518	21,01444825		18,62745	80
11	104,0147895	45,98521055		20,58824	100
12	104,0147895	-24,0147895		22,54902	100
13	104,0147895	-34,0147895		24,5098	100
14	178,9855518	21,01444825		26,47059	100
15	253,956314	96,04368595		28,43137	100
16	119,0089419	80,99105809		30,39216	100
17	553,8393632	-53,8393632		32,35294	100
18	148,9972468	-48,9972468		34,31373	100
19	208,9738567	21,02614333		36,27451	100
20	328,9270763	-113,927076		38,23529	100
21	223,9680091	-98,9680091		40,19608	100
22	328,9270763	-78,9270763		42,15686	103
23	403,8978386	-353,897839		44,11765	125
24	478,8686009	21,13139906		46,07843	150
25	89,02063699	-59,020637		48,03922	150
26	178,9855518	221,0144482		50	150
27	328,9270763	-128,927076		51,96078	150
28	178,9855518	21,01444825		53,92157	200
29	403,8978386	296,1021614		55,88235	200
30	119,0089419	80,99105809		57,84314	200
31	148,9972468	51,00275317		59,80392	200
32	178,9855518	-78,9855518		61,76471	200
33	119,0089419	80,99105809		63,72549	200
34	328,9270763	171,0729237		65,68627	200
35	283,944619	-53,944619		67,64706	200
36	261,4533903	-61,4533903		69,60784	200
37	268,9504665	-8,95046651		71,56863	200
38	134,0030944	-34,0030944		73,52941	200
39	104,0147895	-4,01478945		75,4902	200
40	104,0147895	-4,01478945		77,45098	215
41	126,5060181	-26,5060181		79,41176	230
42	89,02063699	-39,020637		81,37255	230
43	148,9972468	51,00275317		83,33333	250
44	283,944619	516,055381		85,29412	260
45	178,9855518	-28,9855518		87,2549	350
46	148,9972468	-48,9972468		89,21569	400
47	134,0030944	65,99690563		91,17647	500
48	178,9855518	-78,9855518		93,13725	500
49	148,9972468	-88,9972468		95,09804	500
50	104,0147895	-4,01478945		97,05882	700
51	178,9855518	-78,9855518		99,01961	800

7.5 pav. Regresinės analizės rezultatas MS Excel terpėje

Tikrinamos hipotezės H_0: $a = 0$ ir H_0: $b = 0$. Hipotezė atmetama, jei *p-level* < α. Nagrinėjamu atveju lygties konstanta a (*Intercept*) yra nepatikima, *p-level* = 0,364 > α = 0,05, kintamojo koeficientas b yra patikimas *p-level* = 1,54E-07< α = 0,05, koeficientas reikšmingai skiriasi nuo nulio. Pasitvirtina hipotezė H_0: $a = 0$ (*p-level* = 0,364 > α = 0,05) ir atmetama hipotezė H_0: $b = 0$ (*p-level* = 1,54E-07 < α = 0,05).Tiesinės regresijos lygtis: *Pirkinių_suma = 29,04 + 0,15 · Pajamos.*

Paskutiniuose lentelės stulpeliuose pateikti koeficientų a ir b 95 proc. pasikliautinųjų intervalų rėžiai. Lentelėje *ANOVA* pateikti dispersinės analizės rezultatai. Stulpelyje kvadratų suma (*SS*): regresijos kvadratų suma (*Regression*), liekamųjų paklaidų kvadratų suma (*Residual*) ir visa kvadratų suma (*Total*).

Grafike *Line Fit Plots* (žr. 7.5 pav.) parodyta tiriamų kintamųjų priklausomybė (taškai *Pirkinių* suma) ir apskaičiuotas regresijos grafikas (*Predected Pirkiniu_suma*). Grafike *Normal Probability Plots* pateiktas priklausomo kintamojo kaupiamųjų dažnių grafikas. Tiksliai apskaičiuotos grafikų koordinatės pateikiamos lentelėse *RESIDUAL OUTPUT* ir *PROBABILITY OUTPUT*.

7.4. DAUGIALYPĖS TIESINĖS REGRESIJOS MODELIS

Paprasta tiesinė regresija ir porinis koreliacijos koeficientas tinka tada, kai tiriami du kintamieji – vienas priklausomas ir vienas nepriklausomas. Dažnai nepriklausomų kintamųjų būna keletas. Tada taikomas daugialypės tiesinės regresijos modelis ir daugialypė koreliacinė analizė. Daugialypės regresijos modelis yra tiesinio (paprasto) regresijos modelio apibendrinimas atvejui, kai nepriklausomų kintamųjų yra daugiau nei vienas:

$$y = \beta + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \varepsilon. \qquad (7.8)$$

Kaip ir paprastos tiesinės regresijos atveju, daugialypės tiesinės regresijos tyrimo tikslas – apskaičiuoti parametrų β_1, β_2, β_3,... reikšmes.

Apie regresijos lygties reikšmingumą sprendžiama analogiškai dviejų kintamųjų tiesinės regresijos atveju. Koeficientų reikšmingumas tikrinamas *t* kriterijumi, o nepriklausomų kintamųjų X įtaka priklausomam kintamajam *Y* vertinama apskaičiuojant determinacijos koeficientą arba taikant dispersinės analizės metodus. Šiuo atveju pasitelkiant dispersinę analizę tikrinamos hipotezės:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1: $\beta_j \neq 0$ bent vienam *j*.

Jeigu taikant *F* kriterijų nulinė hipotezė H_0 neatmetama, tai regresijos modelis netinka. Tuo tarpu atmesta nulinė hipotezė dar nereiškia, kad modelis tikrai tinka. Tai tik vienas iš daugelio rodiklių, pagal kurį sprendžiama apie modelio tinkamumą.

7.5. DAUGIALYPĖ TIESINĖ REGRESIJA PSPP IR SPSS TERPĖJE

Galima analizuoti vieno priklausomojo kintamojo ryšį su keletu nepriklausomų kintamųjų. Pavyzdžiui, nustatysime **perkamos kavos kiekio priklausomybę nuo amžiaus, pajamų ir pakuotės dydžio**.

PSPP terpė

Analizės dialogo langas atveriamas komanda *Analizuoti* → *Tiesinė regresija*.... Kintamasis *Kiekis* perkeliamas į langelį *Priklausomas*, o nepriklausomus kintamuosius *Amžius, Pajamos, Pakuotės_ dydis* perkeliame į langelį *Nepriklausomas.* Mygtuko *Statistika...* paspaudimas atveria dialogo langą *Regresija: statistika.* Šiame lange reikia pažymėti langelius *Koef., R, Anova.* Paspaudę *Tęsti* mygtuką, grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 7.6 pav.).

Modelio santrauka

R	*R kvadratas*	*Koreguotas R kvadratas*	*Įverčio std. paklaida*
,76	,59	,57	263,96

ANOVA

	Kvadratų suma	*skirt.*	*Vidutinis kvadratas*	*F*	*Reikšmingumo lygmuo*
Regresija	5502435	3	1834145	26,32	,00
Liekana	3901850	56	69675,89		
Iš viso	9404285	59			

Koeficientai

	B	*Std. paklaida*	*Beta*	*t*	*Reikšmingumo lygmuo*
(Konstanta)	97,51	140,25	,00	,70	,49
Amžius2	-1,91	4,13	-,04	-,46	,65
Pajamos2	,09	,09	,10	1,02	,31
Pakuotės_dydis	1,28	,17	,73	7,75	,00

7.6 pav. Daugialypės regresinės analizės rezultatas PSPP terpėje

Lentelėje *Modelio santrauka* pateikta determinacijos koeficiento reikšmė (*R kvadratas = 0,59)* ir patikslinta determinacijos koeficiento reikšmė (*Koreguotas R kvadratas = 0,57*). $R^2 = 0,59 > 0,25$, modelis tinkamas. Taip pat pateiktas daugialypės koreliacijos koeficientas $R = 0,76$.

Lentelėje *ANOVA* pateiktas kintamųjų ryšio modelio vertinimas. Gauname vieną iš dviejų išvadų: 1) H_0: visi koeficientai $b_j = 0$, t. y. regresijos modelis su duomenimis nesuderintas; 2) H_1: bent vienas koeficientas $b \neq 0$, t. y. regresijos modelis bent jau iš dalies prognozėms tinka. Hipotezė H_0: visi koeficientai lygūs nuliui atmetama, kai *p-level (*stulpelis *Reikšmingumo lygmuo)* < α, čia α — pasirinktas reikšmingumo lygmuo. Nagrinėjamu atveju *p-level =0,00* < α = *0,05,* Hipotezė H_0 atmetama. Lentelėje *Koeficientai* pateikiami daugialypės regresijos koeficientai. Regresijos lygtį galima užrašyti kaip:

Kiekis = 97,51 + 1,28 · Pakuotės_dydis – 1,91 · Amžius2 + 0,09 · Pajamos2.

Lentelėje *Coefficients(a)* pateikti ir standartizuotieji koeficientai (*Beta*). Didesnė absoliuti *Beta* koeficiento reikšmė rodo didesnę Y priklausomybę nuo X_j.

SPSS terpė

Linear Regression dialogo langas atveriamas komanda *Analyize* → *Regression* → *Linear*... Priklausomą kintamąjį *Kiekis* perkeliame į langelį *Dependent(s),* nepriklausomus kintamuosius *Amžius, Pajamos, Pakuotės_ dydis* perkeliame į langelį *Independent(s).* Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 7.7 pav.).

Lentelėje *Model Summary* pateikta determinacijos koeficiento reikšmė (*R Square = 0,585)* ir patikslinta determinacijos koeficiento reikšmė (*Adjusted R Square = 0,563*). $R^2 = 0,585 > 0,25$, modelis tinkamas. Taip pat pateiktas daugialypės koreliacijos koeficientas *R = 0,765.* Lentelėje *ANOVA(b)* pateiktas kintamųjų ryšio modelio vertinimas. Gauname vieną iš dviejų išvadų: 1) visi koeficientai $b_j = 0$, t. y. regresijos modelis su duomenimis nesuderintas; 2) bent vienas koeficientas $b \neq 0$, t. y. regresijos modelis bent jau iš dalies prognozėms tinka. Hipotezė H_0: visi koeficientai lygūs nuliui atmetama, kai *p-level (*stulpelis *Sig.)* < α, čia α – pasirinktas reikšmingumo lygmuo. Nagrinėjamu atveju *p-level =* < α = *0,05,* Hipotezė H_0 atmetama. Lentelėje *Coefficients(a)* pateikiami daugialypės regresijos koeficientai. Regresijos lygtį galima užrašyti kaip:

Kiekis = 97,514 + 1,282 · Pakuotės_dydis – 1,908 · Amžius2 + 0,088 · Pajamos2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,765(a)	,585	,563	263,96190

a Predictors: (Constant), Pajamos2, Amžius2, Pakuotės_dydis

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5502435,413	3	1834145,138	26,324	,000(a)
	Residual	3901849,587	56	69675,885		
	Total	9404285,000	59			

a Predictors: (Constant), Pajamos2, Amžius2, Pakuotės_dydis
b Dependent Variable: Kiekis

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	97,514	140,251		,695	,490
	Pakuotės_dydis	1,282	,165	,730	7,753	,000
	Amžius2	-1,908	4,134	-,043	-,462	,646
	Pajamos2	,088	,086	,101	1,022	,311

a Dependent Variable: Kiekis

7.7 pav. Daugialypės regresinės analizės rezultatas SPSS terpėje

Lentelėje *Coefficients(a)* pateikti ir standartizuotieji koeficientai (*Standardized Coefficients*, *Beta*). Didesnė absoliuti *Beta* koeficiento reikšmė rodo didesnę Y priklausomybę nuo X_j.

Atliktos procedūros sintaksė pateikta 7.3 lentelėje.

7.3 lentelė. Daugialypės regresinės analizės sintaksė

PSPP	SPSS
REGRESSION /VARIABLES= Amžius2 Pajamos2 Pakuotės_dydis /DEPENDENT=Kiekis /STATISTICS=COEFF R ANOVA.	REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT Kiekis /METHOD=ENTER Pakuotės_dydis Amžius2 Pajamos2.

7.6. DAUGIALYPĖ TIESINĖ REGRESIJA MS EXCEL TERPĖJE

Galima analizuoti vieno priklausomojo kintamojo ryšį su keletu nepriklausomų kintamųjų. Nustatysime **perkamos kavos kiekio priklausomybę nuo amžiaus, pajamų ir pakuotės dydžio**. Nepriklausomųjų kintamųjų *Amžius, Pajamos, Pakuotės_ dydis* koordinates reikia nurodyti langelyje *Input X range,* priklausomo kintamojo *Kiekis* koordinates – *Input Y range.* Jeigu duomenų masyvai pažymėti su pavadinimais, reikia pažymėti langelį *Labels.* Pažymėtas langelis *Confidence Level* nurodo pasikliautinojo intervalo reikšmę. Jeigu pažymėsime langelį *Normal Probability Plots,* bus pateiktas kaupiamųjų dažnių grafikas. Regresinės analizės rezultatai pateikti 7.4 lentelėje.

7.4 lentelė. Daugialypės regresinės analizės rezultatas *MS Excel* terpėje

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0,764917
R Square	0,585099
Adjusted R Square	0,562872
Standard Error	263,9619
Observations	60

ANOVA					
	df	SS	MS	F	Significance F
Regression	3	5502435	1834145	26,32396	9,4E-11
Residual	56	3901850	69675,89		
Total	59	9404285			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	97,5138	140,2507	0,695282	0,489754	-183,442	378,4698
Amžius2	-1,90845	4,134045	-0,46164	0,646126	-10,1899	6,373037
Pajamos2	0,08801	0,086154	1,021549	0,311387	-0,08458	0,260597
Pakuotės_ dydis	1,281742	0,165312	7,753452	1,97E-10	0,950582	1,612903

Lentelės duomenys analogiški pateiktiems 7.7 pav. Modelio tinkamumas nustatomas pagal determinacijos koeficiento R^2 *(R Square)* reikšmę. Lentelėje pateiktas daugialypės koreliacijos koeficientas *Multiple R = 0,764917,* determinacijos koeficientas (*R Square = 0,585099),* koreguotas determinacijos koeficientas (*Adjusted R Square = 0,562872*). Nagrinėjamu atveju determinacijos koeficientai $R^2 > 0,25$, tiesinis modelis yra tinkamas.

Lentelėje *ANOVA* pateiktas kintamųjų ryšio modelio vertinimas. Gauname vieną iš dviejų išvadų: 1) H_0*:* visi koeficientai $b_j = 0$, t. y. regresijos modelis su duomenimis nesuderintas; 2) H_1*:* bent vienas koeficientas $b \neq 0$, t. y. regresijos modelis bent jau iš dalies prognozėms tinka. Hipotezė H_0: visi koeficientai lygūs nuliui atmetama, kai *p-level (*stulpelis *Significance F)* < α, čia α — pasirinktas reikšmingumo lygmuo. Nagrinėjamu atveju *p-level = 9,4E-11* < α *= 0,05,* Hipotezė H_0 atmetama. Lentelėje *Coefficients* pateikiami daugialypės regresijos koeficientai. Regresijos lygtį galima užrašyti kaip:

Kiekis = 97,51 + 1,28 · Pakuotės_dydis – 1,91 · Amžius2 + 0,09 · Pajamos2.

KONTROLINIAI KLAUSIMAI

7.1. Kam reikalingas regresijos modelis?

7.2. Kaip nustatomas regresijos modelio tinkamumas?

7.3. Kokie būna regresijos modeliai?

7.4. Ką reiškia determinacijos koeficientas?

UŽDUOTYS

7.1. Analizuojant vidutiniškai per mėnesį išleidžiamų šampūnui pinigų priklausomybę nuo amžiaus, gautas rezultatas pateiktas 7.8 pav. Paaiškinkite gautą rezultatą.

Modelio santrauka

R	*R kvadratas*	*Koreguotas R kvadratas*	*Įverčio std. paklaida*
,12	,35	,35	3,07

ANOVA

	Kvadratų suma	*skirt.*	*Vidutinis kvadratas*	*F*	*Reikšmingumo lygmuo*
Regresija	56,26	1	56,26	5,98	,02
Liekana	319,74	34	9,40		
Iš viso	376,00	35			

Koeficientai

	B	*Std. paklaida*	*Beta*	*t*	*Reikšmingumo lygmuo*
(Konstanta)	12,41	1,49	,00	8,36	,00
Amžius	-,08	,03	-,39	-2,45	,02

7.8 pav. Regresinės analizės rezultatas

7.2. Užrašykite regresijos lygtį. Suformuluokite hipotezes ir patikrinkite lygties koeficientų patikimumą.

7.3. Analizuojant perkamos kavos kiekio priklausomybę nuo kainos gautas rezultatas, pateiktas 7.9 pav. Ar šiuo atveju tinkamas tiesinės regresijos modelis?

Modelio santrauka

R	*R kvadratas*	*Koreguotas R kvadratas*	*Įverčio std. paklaida*
,10	,01	,01	400,48

7.9 pav. Regresinės analizės rezultatas

8. SKIRTUMŲ ANALIZĖ

Rinkos tyrimų praktikoje, atliekant rinkos segmentavimą, dažnai reikia palyginti skirtingas respondentų grupes. Skirtumų analizės tikslas – nustatyti respondentų grupes, tarp kurių yra reikšmingas skirtumas. Statistinės procedūros, kurios leidžia nustatyti tokius skirtumus (*t* kriterijus ir dispersinė analizė), lygina kintamųjų vidurkius. *T* kriterijus taikomas dviejų respondentų grupių skirtumams nustatyti. Skirtumų analizei tinka kiekybiniai kintamieji. Ranginiai kintamieji gali būti apdorojami kaip kiekybiniai, jei rangų skaičius lygus 15 ir daugiau.

8.1. HIPOTEZĖS APIE VIDURKIŲ LYGYBĘ

Vidurkių lygybė dažniausiai tikrinama taikant *t* kriterijų arba dispersinę analizę. Dispersinė analizė – tai universalus metodas, *t* kriterijaus taikymas turi imčių skaičiaus apribojimus. Į dispersinę analizę galima žiūrėti kaip į *t* kriterijaus apibendrinimą: lyginant dvi imtis, vienfaktorinės dispersinės analizės ir *t* kriterijaus taikymo rezultatai sutampa. Teoriškai *t* kriterijus gali būti naudojamas net ir tada, kai imtis yra labai maža (pavyzdžiui, mažesnė nei 10), tačiau būtina, kad kiekvienos imties kintamųjų skirstinys būtų normalusis ir imčių dispersijų skirtumas neryškus. Normalumą galima patikrinti vizualiai, χ^2 arba *Kolmogorovo–Smirnovo* kriterijais.

8.1 lentelė. Kintamųjų tipai *t* kriterijui skaičiuoti

T kriterijus vienai imčiai			
Priklausomi kintamieji		Nepriklausomi kintamieji	
Kiekis	Tipas	Kiekis	Tipas
nėra	nėra	Neribojamas	Kiekybinis
T kriterijus dviem nepriklausomoms imtims			
Priklausomi kintamieji		Nepriklausomi kintamieji	
Kiekis	Tipas	Kiekis	Tipas
Vienas	Dichotominis Kiekybinis	Neribojamas	Kiekybinis

8.1 lentelės pabaiga

T kriterijus dviem priklausomoms imtims			
Priklausomi kintamieji		Nepriklausomi kintamieji	
Kiekis	Tipas	Kiekis	Tipas
nėra	nėra	Du	Kiekybinis

Nagrinėjami trys pagrindiniai *t* kriterijaus tipai:

- *t* kriterijus vienai imčiai.
- *t* kriterijus dviem nepriklausomoms imtims,
- *t* kriterijus dviem priklausomoms imtims,

Naudojamų *t* kriterijui skaičiuoti kintamųjų tipai pateikti 8.1 lentelėje.

Priklausomas kintamasis naudojamas tik *t* kriterijaus dviem nepriklausomoms imtims skaičiuoti. Kitais atvejais priklausomų kintamųjų nėra todėl, kad analizuojama praktiškai viena imtis. Nepriklausomas (analizuojamas) kintamasis turi būti tik kiekybinis.

8.2. *T* KRITERIJUS VIENAI IMČIAI

***T* kriterijus vienai imčiai** naudojamas, kai yra viena duomenų imtis ir reikia patikrinti, ar kintamojo vidurkis lygus kažkokiam fiksuotam skaičiui.

Dažnai reikia nustatyti, ar nežinomas populiacijos vidurkis skiriasi nuo tam tikro skaičiaus. Paprastai dispersijos reikšmė yra nežinoma. Nežinoma populiacijos dispersija keičiama jos įverčiu s^2. Iš imties apskaičiuota populiacijos dispersijos reikšmė nėra tiksli ir galimas tam tikras skirtumas nuo tikrosios populiacijos dispersijos. Tikrinama *hipotezė apie vidurkio lygybę skaičiui, kai dispersija nežinoma.* Pavyzdžiui, norima nustatyti, ar vartotojams priimtina A jogurto kaina statistiškai skiriasi nuo pardavėjų nustatytos 3,5 Lt kainos. Iš imties duomenų galima apskaičiuoti vidurkį $\bar{x}$ ir standartinį nuokrypį *s.* Formuluojamos statistinės hipotezės:

H_0: $\mu = a$ ($a = 3,5$, priimtina kaina nesiskiria nuo nustatytos);

H_1: $\mu \neq a$ ($a = 3,5$, priimtina kaina skiriasi nuo nustatytos).

Pasirenkamas reikšmingumo lygmuo α. Kadangi standartinis nuokrypis tiksliai nėra žinomas (jo reikšmė apytikriai apskaičiuota iš imties), hipotezei tikrinti naudojamas *t* kriterijus:

$$t = \frac{\bar{x} - a}{s / \sqrt{n}}, \qquad (8.1)$$

čia $\bar{x}$ – imties vidurkis, *s* – imties standartinis nuokrypis, *n* – imties dydis, *a* – konstanta, su kuria lyginamas vidurkis. Šio *t* kriterijaus reikšmės pasiskirsčiusios pagal Stjudento skirstinį su *(n-1)* laisvės laipsniu, kai populiacijos vidurkis $\mu=a$. Stjudento skirstinys simetriškas nulio atžvilgiu, todėl esant dvipusei alternatyvai H_1: $\mu \neq a$, kritinės reikšmės yra $t_{\alpha/2}(n-1)$. Sprendimo priėmimui reikia žinoti $t_{0,025}(n-1)$ reikšmę. Ją galime sužinoti iš statistinių lentelių. Jei apskaičiuota $t \leq t_{0,025}(n-1)$, H_0 atmesti negalima. Jei apskaičiuota $t > t_{0,025}(n-1)$, H_0 atmetama.

8.2.1. *T* kriterijaus vienai imčiai skaičiavimas

Norima nustatyti, ar **vartotojams priimtina A jogurto kaina statistiškai skiriasi nuo pardavėjų nustatytos** 3,5 Lt kainos

PSPP terpė

Meniu *Analizuoti* pasirenkame komandas *Palyginti vidurkius* → *T kriterijus vienai imčiai*…. Atsiveria dialogo langas *T kriterijus vienai imčiai* (žr. 8.1 pav.).



8.1 pav. *T kriterijus vienai imčiai* dialogo langas

Kintamasis *A_kaina* perkeliamas į langelį *Kriterijaus kintamasis (-ieji)*. Langelyje *Kriterijaus reikšmė* nurodoma standartinė reikšmė (3,5), su kuria bus lyginamas analizuojamo kintamojo vidurkis. Mygtukas *Parinktys…* leidžia nustatyti patikimumo lygmenį. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 8.2 pav.).

Vienos imties statistika

	N	Vidurkis	Std. nuokrypis	Vidur. s.pakl.
A_kaina	60	3,00	1,35	,17

Kriterijus vienai imčiai

	Kriterijaus reikšmė = 3.500000					
					Skirtumo 95% pasikliautinasis intervalas	
	t	skirt.	p-reikšmė (2-pusė)	Vidurkių skirtumas	Apačia	Viršus
A_kaina	-2,86	59	,01	-,50	-,85	-,15

8.2 pav. *T* kriterijaus vienai imčiai skaičiavimo rezultatas PSPP terpėje

Pirmoje lentelėje *Vienos imties statistika* pateikiamas apskaičiuotas kintamojo vidurkis (*Vidurkis)* ir standartinis nuokrypis (*Std. nuokrypis)*. Lentelė *Kriterijus vienai imčiai* leidžia įvertinti statistiškai reikšmingus skirtumus. Išvada daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*p-reikšmė (2-pusė)*) reikšmę. Jei $p\text{-level}>\alpha$, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju $p\text{-level} = 0{,}01 < \alpha = 0{,}05$, skirtumas 0,5 Lt (stulpelis *Vidurkių skirtumas*) yra statistiškai reikšmingas.

SPSS terpė

Meniu *Analyzeze* pasirenkame komandos *Compare Means* → *One - Sample T Test...* Atsiveria dialogo langas *One - Sample T Test.* Kintamasis *A_kaina* perkeliamas į langelį *Test Variable(s)*. Langelyje *Test Value* nurodoma standartinė reikšmė (3,5), su kuria bus lyginamas analizuojamo kintamojo vidurkis. Mygtukas *Options* leidžia nustatyti patikimumo lygmenį. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 8.3 pav.). Pirmoje lentelėje *One-Sample Statistics* pateikimas apskaičiuotas kintamojo vidurkis (*Mean)* ir standartinis nuokrypis (*Std. Deviation)*. Lentelė *One-Sample Test* leidžia įvertinti statistiškai reikšmingus skirtumus.

Išvada daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*Sig. (2 tailed)*) reikšmę. Jei *p-level*>α, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju *p-level = 0,006 < α = 0,05*, skirtumas 0,5 Lt (stulpelis *Mean Difference*) yra statistiškai reikšmingas.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
A_kaina	60	3,0000	1,35296	,17467

One-Sample Test

	Test Value = 3.5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
A_kaina	-2,863	59	,006	-,50000	-,8495	-,1505

8.3 pav. *T* kriterijaus skaičiavimo rezultatas SPSS terpėje

Atliktos procedūros sintaksė pateikta 8.2 lentelėje.

8.2 lentelė. *T* kriterijaus vienai imčiai skaičiavimo sintaksė

PSPP	SPSS

`T-TEST/TESTVAL=3.5`
`  /VARIABLES=A_kaina/MISSING=`
`ANALYSIS`
`  /CRITERIA=CIN(0.95).`
	`T-TEST`
	`  /TESTVAL = 3.5`
	`  /MISSING = ANALYSIS`
	`  /VARIABLES = A_kaina`
	`  /CRITERIA = CI(.95) .`

8.3. *T* KRITERIJUS DVIEM NEPRIKLAUSOMOMS IMTIMS

***T* kriterijus dviem nepriklausomoms imtims** yra naudojamas, kai tiriamos dvi atskiros imtys ir norima sužinoti, ar šių imčių vidurkiai reikšmingai skiriasi. Šiuo atveju nepriklausomos imtys suprantamos kaip imtys, atrinktos pagal binarinę (dichotominę kategoriją), pavyzdžiui, vyrai ir moterys, žmonės, kurie perka kokį nors produktą ir neperka jo, t. y. kai yra du variantai priklausomo (grupuojančio) kin-

tamojo ir keletas nepriklausomų kintamųjų, kurių analizė ir parodys skirtumą tarp šių grupių. Kai tiriamos dvi *atskiros* imtys ir norima sužinoti, ar šių imčių vidurkiai reikšmingai skiriasi, naudojamas *t* kriterijus, skirtas dviem nepriklausomoms imtims. Šios imtys gaunamos stebint du nepriklausomus atsitiktinius dydžius, turinčius normalųjį skirstinį. Šių dydžių vidurkiai yra nežinomi. Tikrinama *hipotezė apie dviejų imčių vidurkių lygybę, kai dispersija nežinoma*.

Pastaba: Atliekant tyrimą kompiuteriu, grupės gali būti atskiros arba joms atskirti galima naudoti grupavimo kintamąjį (pvz., *Lytis: vyrai/moterys*). Naudojant grupavimo kintamąjį tiriama, ar priklausomo kintamojo (pvz., *Kiekis*) vidurkiai lygūs.

Naudojant *t* kriterijų dviem imtims, būtina, kad abi imtys būtų atrinktos iš populiacijų, turinčių normaliuosius skirstinius ir turėtų vienodas dispersijas. Dažniausiai dispersijų lygybei tikrinti taikomas *F* kriterijus arba *Livyno* kriterijus. *F* kriterijus labai jautriai reaguoja į normalumo sąlygos netenkinimą. Jei populiacijos neturi normaliojo skirstinio, *F* kriterijus gali atmesti nulinę hipotezę, teigiančią dispersijų lygybę ir tais atvejais, kai populiacijų dispersijos iš tiesų yra lygios. Normalumo prielaidos pažeidimui mažiau jautrus yra *Livyno* kriterijus.

***F* kriterijus** nustatomas apskaičiavus imčių dispersijas s_1^2 ir s_2^2:

$$F = \frac{s_1^2}{s_2^2}. \tag{8.2}$$

Hipotezė H_0 atmetama (t. y. teigiama, kad dispersijos reikšmingai skiriasi), jei gauta reikšmė *F* yra didesnė už kritinę reikšmę $F > F_{\alpha/2}(n - 1, m - 1)$, čia α – pasirinktas reikšmingumo lygmuo, *n* ir *m* – pirmos ir antros imties dydžiai.

Suformuluokime statistines hipotezes:

H_0: $\mu_1 = \mu_2$;

H_1: $\mu_1 \neq \mu_2$.

Įvertinus dispersijų lygybę, renkamės vieną iš dviejų galimų *t* kriterijaus skaičiavimo būdų.

Stjudento *t* kriterijus, kai populiacijų dispersijos yra lygios

Kadangi populiacijos dispersijos nežinomos, jos keičiamos dispersijų svertiniu vidurkiu:

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}. \tag{8.3}$$

T kriterijus skaičiuojamas pagal formulę:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}}. \tag{8.4}$$

Kritinė sritis sudaroma remiantis tuo, kad statistika (1) turi Stjudento skirstinį su $(n + m - 2)$ laisvės laipsniais kai $\bar{x}_1 = \bar{x}_2$. H_0 hipotezė atmetama, jei apskaičiuota $| t | > t_{\alpha/2}(n + m - 2)$.

Stjudento *t* kriterijus, kai populiacijų dispersijos nelygios

Jei jungtinę dispersiją naudosime, kai dispersijos nelygios, apskaičiuota *t* kriterijaus reikšmė gali būti klaidinga. Klaidos dydis priklauso nuo imčių dydžių skirtumo ir dispersijų skirtumų. Didelėms imtims neatitikimai tarp rezultatų, gautų šiais dviem metodais, yra maži. Kai dispersijos nelygios, hipotezei tikrinti naudojamas *t* kriterijus:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}. \tag{8.5}$$

Kadangi alternatyva dvipusė, sprendimui priimti reikia naudoti $t_{\alpha/2}(k)$ reikšmę. Laisvės laipsnių skaičius *k* yra mažiausias sveikas skaičius, tenkinantis sąlygą:

$$k \le \frac{(s_1^2 / n + s_2^2 / m)^2}{s_1^4 / n^3 + s_2^4 / m^3}. \tag{8.6}$$

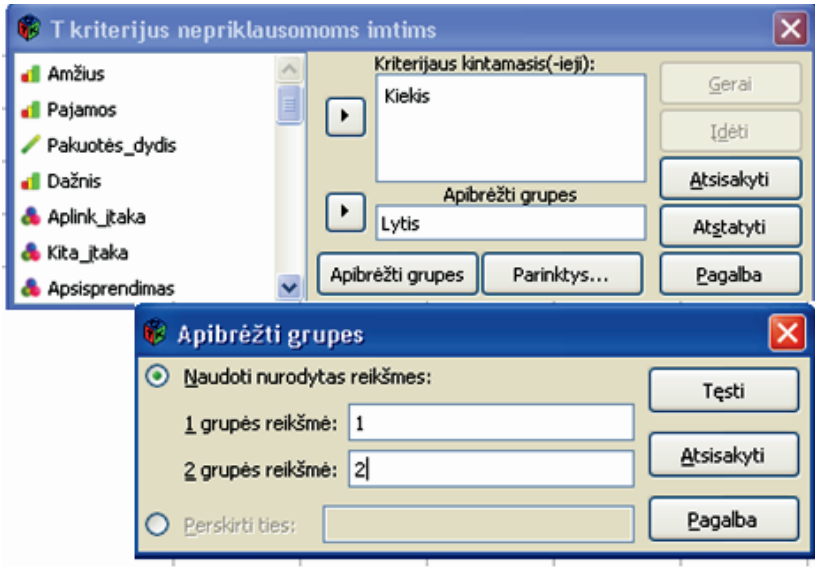
H_0 hipotezė atmetama, jei apskaičiuota $| t | > t_{\alpha/2}(k)$.

8.3.1. *T* kriterijaus dviem nepriklausomoms imtims skaičiavimas

Pavyzdžiui, **analizuojamas vyrų ir moterų suvartojamo per mėnesį kavos kiekio skirtumas**.

PSPP terpė

Meniu *Analizuoti* pasirenkame komandas *Palyginti vidurkius → T kriterijus nepriklausomoms imtims…*. Atsiveria dialogo langas *T kriterijus nepriklausomoms imtims* (žr. 8.4 pav.).



8.4 pav. Dialogo langai *T kriterijus nepriklausomoms imtims* ir *Apibrėžti grupes*

Nepriklausomą kintamąjį *Kiekis* perkeliame į *Kriterijaus kintamasis(-ieji)* lauką, priklausomąjį (grupuojantį) kintamąjį *Lytis* perkeliame į *Apibrėžti grupes* lauką. Šis kintamasis gali įgyti dvi užkoduotas reikšmes: 1 – vyras, 2 –moteris. Šias reikšmes reikia nurodyti dialogo lange, kuris atsiveria paspaudus mygtuką *Apibrėžti grupes* (žr. 8.4 pav.). Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 8.5 pav.).

Pirmoje lentelėje *Grupių statistika* pateikti kintamojo *Kiekis* vidurkiai (*Vidurkis*) respondentų grupėms vyrai (1) ir moteris (2). Antroji lentelė *Kriterijus nepriklausomoms imtims* leidžia nustatyti statistinius grupių skirtumus. Lentelės *Kriterijus nepriklausomoms imtims* analizę reikia pradėti nuo Livyno kriterijaus (*Levene kriterijus dispersijų lygybei*) analizės. Šis kriterijus taikomas hipotezei dėl analizuojamų

kintamųjų dispersijų lygybės tikrinti. Jeigu stebimojo reikšmingumo lygmens (*p-level*) reikšmė *p-reikšmė* stulpelyje *Levene kriterijus dispersijų lygybei* rodo, kad rezultatas statistiškai nereikšmingas (*p-level>α,* čia *α* – pasirenkamas reikšmingumo lygmuo, pvz., *α = 0,05*) (nagrinėjamu atveju *p-level = 0,92*), tai kintamųjų dispersijos statistiškai lygios ir vidurkių skirtumas nustatomas iš eilutės *Jei dispersijos būtų lygios* stulpelio *p-reikšmė (2-pusė).* Jei rezultatas statistiškai reikšmingas, tai kintamųjų dispersijos statistiškai skiriasi ir vidurkių skirtumas nustatomas iš eilutės *Jei dispersijos būtų nelygios* stulpelio *p-reikšmė (2-pusė).* Nagrinėjamu atveju Livyno kriterijus rodo statistinį reikšmingumą ir stulpelio *p-reikšmė (2-pusė)* reikšmė *0,81* (*p-level>α*) rodo, kad vyrų ir moterų suvartojamos kavos kiekio skirtumas statistiškai nereikšmingas.

Grupių statistika

Lytis	*N*	*Vidurkis*	*Std. nuokrypis*	*Vidur. s.pakl.*
Kiekis 1	30	498,00	356,34	65,06
2	30	473,00	443,83	81,03

Kriterijus nepriklausomoms imtims

	Levene kriterijus dispersijų lygybei		t kriterijus vidurkių lygybei						
								Skirtumo 95% pasikliautinasis intervalas	
	F	*p-reikšmė*	*t*	*skirt.*	*p-reikšmė (2-pusė)*	*Vidurkių skirtumas*	*Vidurkių skirtumo std. paklaida*	*Apačia*	*Viršus*
KiekisJei dispersijos būtų lygios	,01	,92	,24	58,00	,81	25,00	103,92	-183,01	233,01
Jei dispersijos būtų nelygios			,24	55,41	,81	25,00	103,92	-183,22	233,22

8.5 pav. *T* kriterijaus skaičiavimo rezultatas PSPP terpėje

SPSS terpė

Meniu *Analyze* pasirenkame komandas *Compare Means* → *Independent-Samples T Test*. Atsiveria dialogo langas *Independent-Samples T Test.*

Nepriklausomą kintamąjį *Kiekis* perkeliame į *Test Variable(s)* lauką, priklausomąjį (grupuojantį) kintamąjį *Lytis* perkeliame į *Grouping variable:* lauką. Šis kintamasis gali įgyti dvi užkoduotas reikšmes: 1 –

vyras, 2 – moteris. Šios reikšmes reikia nurodyti dialogo lange, kuris atsiveria paspaudus mygtuką *Define groups*. Mygtuko *Continue* paspaudimas grąžina *Independent-Samples T Test* langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 8.6 pav.).

Group Statistics

	Lytis	N	Mean	Std. Deviation	Std. Error Mean
Kiekis	Vyras	30	498,0000	356,33576	65,05771
	Moteris	30	473,0000	443,83245	81,03235

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Kiekis	Equal variances assumed	,011	,916	,241	58	,811	25,00000	103,91702	-183,01252	233,01252
	Equal variances not assumed			,241	55,412	,811	25,00000	103,91702	-183,21956	233,21956

8.6 pav. *T* kriterijaus skaičiavimo rezultatas SPSS terpėje

Pirmoje lentelėje *Group Statistics* pateikti kintamojo *Kiekis* vidurkiai (*Mean*) respondentų grupėms *Vyrai* ir *Moteris.* Antroji lentelė *Independent Samples Test* leidžia nustatyti statistinius grupių skirtumus. Lentelės *Independent Samples Test* analizę reikia pradėti nuo Livyno kriterijaus (*Levene's Test*) analizės. Šis kriterijus taikomas hipotezei dėl analizuojamų kintamųjų dispersijų lygybės tikrinti. Jeigu stebimojo reikšmingumo lygmens *p-level* reikšmė *Sig.* stulpelyje *Levene's Test for Equality of Variances* rodo, kad rezultatas statistiškai nereikšmingas (*p-level*>α, čia α – pasirenkamas reikšmingumo lygmuo, pvz., $\alpha = 0,05$) (nagrinėjamu atveju *p-level = 0,916*), tai kintamųjų dispersijos statistiškai lygios ir vidurkių skirtumas nustatomas iš eilutės *Equal variances assumed* stulpelio *Sig. (2-tailed).* Jei rezultatas statistiškai reikšmingas, tai kintamųjų dispersijos statistiškai skiriasi ir vidurkių skirtumas nustatomas iš eilutės *Equal variances not assumed* stulpelio *Sig. (2-tailed).* Nagrinėjamu atveju Livyno kriterijus rodo statistinį reikšmingumą ir stulpelio *Sig. (2-tailed)* reikšmė *0,811* (*p-level*>α) rodo, kad vyrų ir moterų suvartojamos kavos kiekio skirtumas statistiškai nereikšmingas.

Atliktos procedūros sintaksė pateikta 8.3 lentelėje.

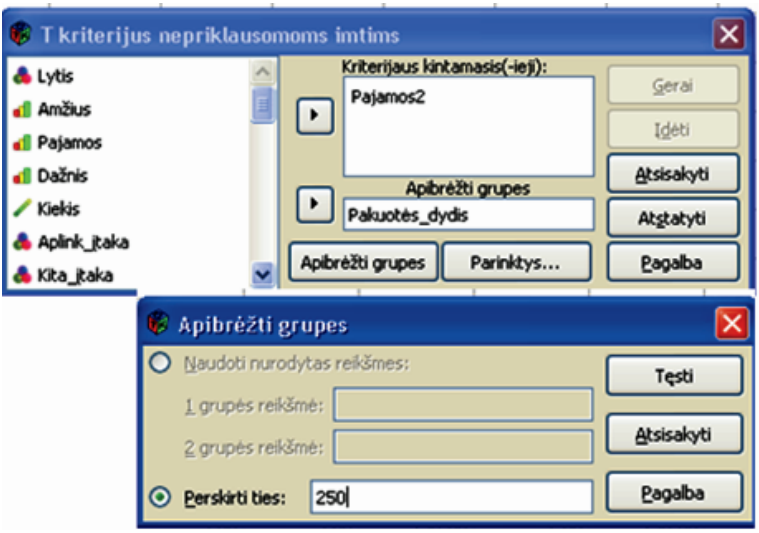
8.3 lentelė. *T* kriterijaus dviem nepriklausomoms imtims skaičiavimo sintaksė

PSPP	SPSS
`T-TEST/VARIABLES=Kiekis /GROUPS=Lytis(1,2)/MISSING=ANALYSIS /CRITERIA CIN (0.95).`	`T-TEST GROUPS = Lytis(1 2) /MISSING = ANALYSIS /VARIABLES = Kiekis /CRITERIA = CI(.95).`

Išnagrinėsime atvejį, kai grupuojantis kintamasis ne dichotominis, o kiekybinis. Pavyzdžiui, išnagrinėsime atvejį, kaip skiriasi respondentų **pajamos, priklausomai nuo perkamos kavos pakuotės dydžio**.

PSPP terpė

Šiuo atveju lange *Apibrėžti grupes* galima nustatyti atskirties tašką (*Perskirti ties*), kuris suskirstys visas galimas kintamojo *Pakuotės_dydis* reikšmes į dvi grupes (žr. 8.7 pav.). Daugiau analizės procedūra nesiskiria. Gautas rezultatas pateiktas 8.8 pav.



8.7 pav. Dialogo langai *T kriterijus nepriklausomoms imtims* ir *Apibrėžti grupes*

Pirmoje lentelėje *Grupių statistika* pateikti kintamojo *Pajamos* vidurkiai (*Vidurkis*) dėl skirtingų respondentų grupių (respondentai, kurie perka pakuotes *>=250 gr* ir *<250 gr*). Lentelė *Kriterijus nepriklausomoms imtims* leidžia nustatyti grupių statistinius skirtumus. Lentelės *Kriterijus nepriklausomoms imtims* analizę reikia pradėti nuo Livyno kriterijaus (*Levene kriterijus dispersijų lygybei*) analizės.

Grupių statistika

Pakuotės_dydis	*N*	*Vidurkis*	*Std. nuokrypis*	*Vidur. s.pakl.*
Pajamos2 >=250	37	1636,49	426,17	70,06
<250	23	1304,35	444,64	92,71

Kriterijus nepriklausomoms imtims

		Levene kriterijus dispersijų lygybei		t kriterijus vidurkių lygybei						
									Skirtumo 95% pasikliautinasis intervalas	
		F	*p-reikšmė*	*t*	*skirt.*	*p-reikšmė (2-pusė)*	*Vidurkių skirtumas*	*Vidurkių skirtumo std. paklaida*	*Apačia*	*Viršus*
Pajamos2	Jei dispersijos būtų lygios	4,87	,03	-2,89	58,00	,01	-332,14	116,21	-564,76	-99,52
	Jei dispersijos būtų nelygios			-2,86	45,28	,01	-332,14	116,21	-566,16	-98,12

8.8 pav. *T* kriterijaus skaičiavimo rezultatas PSPP terpėje

Stebimojo reikšmingumo lygmens (*p-level*) reikšmė *p-reikšmė* stulpelyje *Levene kriterijus dispersijų lygybei* rodo, kad *p-level = 0,3 > α=0,05,* čia *α* – pasirenkamas reikšmingumo lygmuo, kintamųjų dispersijos statistiškai lygios ir vidurkių skirtumas nustatomas iš eilutės *Jei dispersijos būtų lygios* stulpelio *p-reikšmė (2-pusė).).* Nagrinėjamu atveju stulpelio *p-reikšmė (2-pusė)* reikšmė *0,01* (*p-level<α*) rodo, kad skirtingo dydžio pakuotes perkančių respondentų pajamų skirtumas statistiškai reikšmingas.

SPSS terpė

Šiuo atveju lange *Define Groups* galima nustatyti atskirties tašką (*Cut point*), kuris suskirstys visas galimas kintamojo *Pakuotės_dydis* reikšmes į dvi grupes. Daugiau analizės procedūra nesiskiria. Gautas rezultatas pateiktas 8.9 pav. Sprendimų priėmimas analogiškas anksčiau pateiktam pavyzdžiui. Pirmoje lentelėje *Group Statistics* pateikti kintamojo *Pajamos* vidurkiai (*Mean*) dėl skirtingų respondentų grupių (respondentai, kurie perka pakuotes *>=250 gr* ir *<250 gr*).

Group Statistics

	Pakuotės_dydis	N	Mean	Std. Deviation	Std. Error Mean
Pajamos2	>= 250,00	37	1636,4865	426,16648	70,06134
	< 250,00	23	1304,3478	444,64313	92,71450

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pajamos2	Equal variances assumed	1,597	,211	2,887	58	,005	332,13866	115,04482	101,85142	562,42590
	Equal variances not assumed			2,858	45,277	,006	332,13866	116,20916	98,12089	566,15643

8.9 pav. *T* kriterijaus skaičiavimo rezultatas SPSS terpėje

Lentelės *Independent Samples Test* stebimojo reikšmingumo lygmens *p-level* reikšmė *Sig.* stulpelyje *Levene's Test for Equality of Variances* rodo, kad *p-level = 0,211 > α = 0,05,* čia *α* – pasirenkamas reikšmingumo lygmuo, kintamųjų dispersijos statistiškai lygios ir vidurkių skirtumas nustatomas iš eilutės *Equal variances assumed* stulpelio *Sig. (2-tailed).* Nagrinėjamu atveju stulpelio *Sig. (2-tailed)* reikšmė *0,005* (*p-level<α*) rodo, kad skirtingo dydžio pakuotes perkančių respondentų pajamų skirtumas statistiškai reikšmingas.

Atliktos procedūros sintaksė pateikta 8.4 lentelėje.

8.4 lentelė. *T* kriterijaus dviem nepriklausomoms imtims skaičiavimo sintaksė

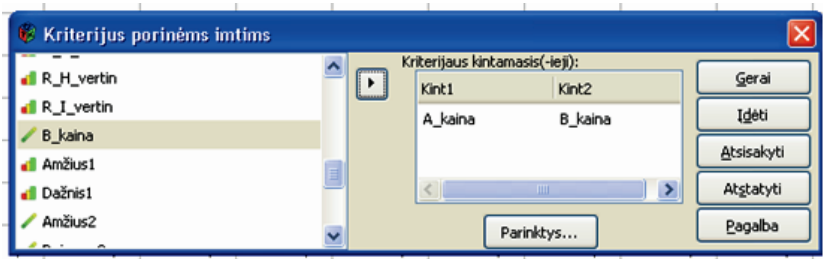
PSPP	SPSS
`T-TEST/VARIABLES=Pajamos2 /GROUPS=Pakuotės_dydis(250)/ MISSING=ANALYSIS /CRITERIA=CIN (0.95).`	`T-TEST GROUPS = Pakuotės_dydis(250) /MISSING = ANALYSIS /VARIABLES = Pajamos2 /CRITERIA = CI(.95).`

8.4. *T* KRITERIJAUS DVIEM PRIKLAUSOMOMS (PORINĖMS) IMTIMS SKAIČIAVIMAS PSPP IR SPSS TERPĖJE

***T* kriterijus dviem priklausomoms (porinėms) imtims** naudojamas tada, kai dvi duomenų grupės sudaro poras. Tai reiškia, kad kiekvieną vienos grupės reikšmę atitinka kitos grupės reikšmė. Šio tipo testai taikomi, kai į skirtingus klausimus atsako ta pati respondentų grupė. Pavyzdžiui, vartotojai nurodė A ir B jogurtų priimtiną 500 gr pakuotės kainą. Reikia **nustatyti, ar priimtinos kainos įvertinimas statistiškai reikšmingas.**

PSPP paketas

Meniu *Analizuoti* pasirenkame komandas *Palyginti vidurkius* → *T kriterijus porinėms imtims.* Atsiveria dialogo langas *Kriterijus porinėms imtims.* Kintamieji *A_kaina* ir *B_kaina* iš eilės perkeliami į *Kriterijaus kintamasis(-ieji):* langelį (žr. 8.10 pav.). Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 8.11 pav.).



8.10 pav. Dialogo langas *Kriterijus porinėms imtims*

Pirmoje lentelėje *Porinių imčių statistika* pateikti apskaičiuoti kintamųjų vidurkiai. Lentelėje *Porinių imčių koreliacija* pateiktas analizuojamų kintamųjų Pirsono koreliacijos koeficientas. Lentelė *Kriterijus porinėms imtims* leidžia įvertinti statistiškai reikšmingus kintamųjų skirtumus. Išvada daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*p-reikšmė (2-pusė)*) reikšmę. Jei *p-level*>α, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas.

Nagrinėjamų atveju *p-level = 0,1 >α = 0,05*, skirtumas 0,4 Lt (žr. *Porinių imčių statistika → Vidurkis*) yra statistiškai nereikšmingas.

Porinių imčių statistika

		Vidurkis	*N*	*Std. nuokrypis*	*Vidur. s.pakl.*
o pora	*A_kaina*	3,00	60	1,35	,17
	B_kaina	3,40	60	1,14	,15

Porinių imčių koreliacija

		N	*Koreliacija*	*p-reikšmė*
o pora	*A_kaina ir B_kaina*	60	-,30	,02

Kriterijus porinėms imtims

	Porinių imčių skirtumai							
				Skirtumo 95% pasikliautinasis intervalas				
	Vidurkis	*Std. nuokrypis*	*std. vidurkio paklaida*	*Apačia*	*Viršus*	*t*	*skirt.*	*p-reikšmė (2-pusė)*
o A_kaina - pora B_kaina	-,40	2,01	,26	-,92	,12	-1,54	59	,13

8.11 pav. *T* kriterijaus skaičiavimo rezultatas PSPP terpėje

SPSS paketas

Meniu *Analyze* pasirenkame komandas *Compare Means → Paired - Samples T Test*. Atsiveria dialogo langas *Paired - Samples T Test.* Kintamieji *A_kaina* ir *B_kaina* iš eilės perkeliami į *Paired Variables* langelį. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 8.12 pav.).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	A_kaina	3,0000	60	1,35296	,17467
	B_kaina	3,4000	60	1,13794	,14691

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	A_kaina & B_kaina	60	-,297	,021

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	A_kaina - B_kaina	-,40000	2,01014	,25951	-,91928	,11928	-1,541	59	,129

8.12 pav. *T* kriterijaus skaičiavimo rezultatas SPSS terpėje

Pirmoje lentelėje *Paired Samples Statistics* pateikti apskaičiuoti kintamųjų vidurkiai. Lentelėje *Paired Samples Correlation* pateiktas analizuojamų kintamųjų Pirsono koreliacijos koeficientas. Lentelė *Paired Samples Test* leidžia įvertinti statistiškai reikšmingus kintamųjų skirtumus. Išvada daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*Sig. (2 tailed)*) reikšmę. Jei $p\text{-level} > \alpha$, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju $p\text{-level} = 0{,}129 > \alpha = 0{,}05$, skirtumas 0,4 Lt (žr. *Paired Samples Statistics* → *Mean*) yra statistiškai nereikšmingas.

Atliktos procedūros sintaksė pateikta 8.5 lentelėje.

8.5 lentelė. *T* kriterijaus dviem nepriklausomoms imtims skaičiavimo sintaksė

PSPP	SPSS
T-TEST PAIRS=A_kaina WITH B_kaina (PAIRED) /MISSING=ANALY SIS /CRITERIA=CIN(0.95).	T-TEST PAIRS = A_kaina WITH B_kaina (PAIRED) /CRITERIA = CI(.95) /MISSING = ANALYSIS.

8.5. DISPERSINĖ ANALIZĖ

Atliekant rinkos tyrimus dažnai reikia palyginti daugiau kaip dvi respondentų grupes, pvz., ar kelių skirtingo amžiaus grupių atstovai išleidžia pirkiniams vienodą pinigų kiekį. Tuomet tenka lyginti kelių nepriklausomų imčių vidurkius ir, remiantis tyrimu, daryti išvadas apie populiacijas, iš kurių atrinktos imtys. Šiems uždaviniams spręsti naudojama dispersinė analizė. Į dispersinę analizę galima žiūrėti kaip į *t* kriterijaus apibendrinimą.

Dispersinėje analizėje vieni kintamieji nagrinėjami kaip priežastys, o kiti – kaip šių priežasčių veikimo pasekmės. Pirmojo tipo kintamieji yra *nepriklausomi* ir vadinami veiksniais, o jų veikimą atspindintys antrojo tipo kintamieji vadinami *priklausomais* kintamaisiais. Pavyzdžiui, tiriamų asmenų amžius arba pirkimo vieta yra veiksniai, o sumokėti pinigai – priklausomas kintamasis. Dispersinės analizės atveju priklausomasis kintamasis tas, kuris nustato skirtingų tiriamųjų gru-

pių vidurkių skirtumą. Kintamasis, kuris apibūdina grupių išskyrimą, vadinamas nepriklausomuoju.

Vienfaktorinė dispersinė analizė (*vienfaktorinė ANOVA*) taikoma tada, kai populiacijos viena nuo kitos skiriamos tik pagal vieną kintamąjį (veiksnį). Dispersinės analizės tikslas – nuspręsti, ar priklausomo kintamojo, išmatuoto skirtingose populiacijose, vidurkiai statistiškai reikšmingai skiriasi. Metodas leidžia atsakyti į klausimą, ar iš kelių nepriklausomų imčių bent dviejų vidurkiai statistiškai reikšmingai skiriasi. Veiksnio parinkimas dažnai yra subjektyvus ir specialistų nuomonės apie tai, kuris kintamasis yra veiksnys, o kuris priklausomas kintamasis, gali skirtis.

Dispersinės analizės procedūra ANOVA reikalauja, kad būtų patenkintos šios prielaidos:

- kiekviena tiriama imtis turi būti nepriklausoma nuo kitų,
- tiriamas kintamasis kiekvienoje populiacijoje pasiskirstęs pagal normalųjį dėsnį,
- imčių dispersijos lygios.

Ar patenkinami šie reikalavimai, galima patikrinti pradinės duomenų analizės procedūromis (apskaičiuojant dispersijas) bei naudojant statistinius kriterijus, skirtus duomenų normaliajam skirstiniui ir dispersijų lygybei tikrinti. Taikant vienfaktorinę dispersinę analizę labai svarbu, kad visos imtys būtų panašaus dydžio. Kuo daugiau stebėjimų kiekvienoje imtyje, tuo statistiškai reikšmingesni tampa vidurkių skirtumai. Kai imtys yra panašaus dydžio, ANOVA yra gana atspari nukrypimams nuo normaliojo skirstinio ir dispersijų lygybės.

Normalumo prielaida. Kadangi ANOVA gana atspari normalumo prielaidos pažeidimams, ji taikoma net ir tuo atveju, kai kintamieji matuojami pagal rangų skalę. Svarbu, kad rangų būtų pakankamai daug ir vidurkis turėtų aiškią prasmę. Nėra griežtų taisyklių, reglamentuojančių įgyjamų rangų skaičių. Dažniausiai manoma, kad turėtų būti bent 7 skirtingi rangai. Kartais ANOVA naudojama ir kai duomenys išmatuoti 5 skirtingų rangų skalėje.

Dispersijų lygybės prielaida. Dispersijų lygybę galima patikrinti įvairiais būdais. Pats paprasčiausias yra reikalavimas, kad didžiausioji iš imčių dispersijų būtų didesnė už mažiausiąją ne daugiau kaip tris kartus. Griežtesniam dispersijų lygybės tikrinimui naudojami *F* ir *Lyveno* kriterijai. Jei šie kriterijai atmeta dispersijų lygybę, daromos paklaidos

dydis labai priklauso nuo imčių dydžio. Jei didesnė imtis atrinkta iš populiacijos su didesne dispersija, apskaičiuotas stebimasis reikšmingumo lygmuo *p-level* gaunamas daug didesnis už tikrąją reikšmę. Todėl vidurkių skirtumo reikšmingumas pripažįstamas daug rečiau, nei turėtų būti. Jei iš populiacijos su didesne dispersija atrinkta maža imtis, apskaičiuota *p-level* reikšmė gaunama mažesnė už tikrąją ir skirtumo reikšmingumas pripažįstamas dažniau nei reikėtų. Kuo mažesni imčių dydžių skirtumai, tuo mažiau ANOVA jautri dispersijų skirtumams.

Tikrinamos hipotezės:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \ldots = \mu_k, \\ H_1: \text{bent du vidurkiai skiriasi,} \end{cases}$$

čia μ_j – nepriklausomų kintamųjų vidurkiai, k – imčių skaičius.

Dispersinė analizė remiasi faktu, kad dispersija gali būti padalyta. Dispersija yra funkcija nuokrypių nuo vidurkio kvadratų sumos. Visa kvadratų suma *SST (sum of squares total*), kuri apima visų duomenų ir bendrojo vidurkio skirtumus, yra suskaidoma į dviejų komponenčių – vidinės kvadratų sumos *SSW (sum of squares within*), įvertinančios kiekvienos imties duomenų sklaidą apie imties vidurkį ir grupių kvadratų sumos *SSB (sum of squares between*), įvertinančios imčių vidurkių ir bendrojo vidurkio skirtumus, sumą:

$$SST = SSW + SSB. \tag{8.7}$$

Nulinei hipotezei $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (visų populiacijų vidurkiai lygus) tikrinti naudojamas F kriterijus:

$$F = \frac{MSB}{MSW}, \tag{8.8}$$

čia $MSW = \frac{SSW}{N-k}$, $MSB = \frac{SSB}{k-1}$, N – bendras visų stebėjimų skaičius, k – imčių skaičius. Skirtumai

($N - k$) ir ($k - 1$) vadinami laisvės laipsniais.

Jei gauta F reikšmė daug didesnė už vienetą, tai tikėtina, kad vidurkiai skiriasi, jei artima vienetui – vidurkių skirtumai maži. Kai nulinė

hipotezė teisinga, tai kriterijus *F* turi Fišerio skirstinį su (*k –1*) ir (*N – k*) laisvės laipsniais. Apskaičiuotą *F* reikšmę atitinkantis stebimasis reikšmingumo lygmuo *p-level* pateikiamas rezultatų lentelėje. Jei stebimasis reikšmingumo lygmuo *p-level* mažesnis už pasirinktą reikšmingumo lygmenį α, hipotezė apie vidurkių lygybę atmetama. Nulinė hipotezė apie vidurkių lygybę atmetama, jei *bet kurie du* vidurkiai yra nelygūs.

ANOVA procedūros pagalba gauta *F* kriterijaus reikšmė rodo, ar imčių vidurkiai yra lygūs, tačiau negali pasakyti, kurios imties vidurkis skiriasi nuo kitų. Tam, kad būtų galima nustatyti, kurie vidurkiai reikšmingai skiriasi nuo kitų, naudojamos *daugkartinio lyginimo procedūros* ir skaičiuojami daugkartinio lyginimo kriterijai. Daugkartinio lyginimo procedūros apsaugo nuo klaidingų vidurkių skirtumų nustatymo. Daugkartinio lyginimo kriterijai yra skirstomi į planuotus iš anksto (*apriorinius*) kriterijus ir *Post Hoc* (*aposteriorinius*) kriterijus. Jeigu sprendžiame prieš vieno veiksnio (faktoriaus) analizę, turime apriorinius, iš anksto planuotus lyginimus, jeigu paskui, kai jau žinomi analizės rezultatai, turime aposteriorinius (*Post Hoc*) lyginimus. Tie patys kriterijai, tik su skirtingais reikšmingumo lygmenimis, gali būti taikomi ir kaip *aprioriniai*, ir kaip *aposterioriniai.* Daugkartinio lyginimo procedūrų yra gana daug. Skirtingos procedūros taikomos lygių ir nelygių dispersijų atveju. Pagrindiniai **aposterioriniai kriterijai**:

Fišerio LSD kriterijus (LSD test) yra liberaliausias *Post Hoc* kriterijus, t. y. jis dažniausiai randa statistiškai reikšmingus vidurkių skirtumus. Po to, kai ANOVA nustato reikšmingą vidurkių skirtumą, visos imčių poros lyginamos naudojant *t*-kriterijų.

Pagal *Bonferroni* kriterijų visos imčių poros lyginamos taikant Stjudento kriterijų, esant reikšmingumo lygmeniui $\alpha = \alpha_E/C$; čia α_E – eksperimento reikšmingumo lygmuo (t. y. tikimybė lyginant visas įmanomas poras nors kartą neteisingai nustatyti statistiškai reikšmingą dviejų imčių vidurkių skirtumą), α – reikšmingumo lygmuo lyginant vieną imčių porą, $C = k(k-1)/2$, *k* – imčių skaičius. *Bonferroni* kriterijus netaikytinas, kai imčių yra daug, nes labai sumažėja α ir nefiksuojamas statistiškai reikšmingas vidurkių skirtumas, nors tikrieji populiacijų vidurkiai ir skiriasi (t. y. labai padidėja antrosios rūšies klaidos tikimybė).

Tjukio HSD kriterijus (*Tukey HSD test*) siūlo lyginant *i*-tosios ir *j*-tosios imčių vidurkius skaičiuoti statistiką:

$$Q(i,j) = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSW / n}}. \tag{8.9}$$

Vidurkiai $\overline{x_i}$ ir $\overline{x_j}$ reikšmingai skiriasi, jei $|Q(i,j)| > Q_\alpha(nk-k,k)$; čia $Q_\alpha(nk-k,k)$ yra Q statistikos α lygmens kritinė reikšmė, k – imčių skaičius. Visų imčių dydžiai vienodi ir lygūs n. Jis taikomas, kai imčių daug. *Tjukio* kriterijus yra labai konservatyvus, t. y. labiausiai nelinkęs atmesti nulinės hipotezės.

Stjudento–Njumeno–Keulso SNK kriterijus (*Student–Newman–Keuls test*). Šis kriterijus skiriasi nuo *Tjukio HSD* kriterijaus tuo, kad atsižvelgia į skaičių imčių, kurių vidurkiai yra tarp $\overline{x_i}$ ir $\overline{x_j}$. Šis kriterijus, kaip ir *LSD* kriterijus, statistiškai reikšmingus vidurkių skirtumus linkęs rasti ir tada, kai populiacijų vidurkiai nesiskiria. Kai kurie statistikai siūlo jį naudoti tik tuo atveju, kai yra trys imtys.

Šefe kriterijus. Šis kriterijus yra konservatyvus ir statistiškai reikšmingam vidurkių skirtumo pripažinimui reikalauja didesnio vidurkių skirtumo nei kiti kriterijai.

Daugelis kitų *post-hoc* kriterijų yra panašūs į aptartuosius. Dažnai tai įvairios šių kriterijų modifikacijos.

Apriorinіai kriterijai

Prie apriorinių priskiriami kriterijai, kurie taikomi hipotezėms apie tiesines populiacijų vidurkių kombinacijas (kontrastus) tikrinti. Taigi yra tikrinamos ne visos imtys poromis, o pasirinktos kombinacijos. Tiesiniu populiacijų vidurkių kontrastu vadinama suma:

$$C_1\mu_1 + C_2\mu_2 + \ldots + C_k\mu_k, \tag{8.10}$$

čia k – imčių skaičius, *koeficientai* C_i tenkina lygybę:

$$C_1 + C_2 + + C_k = 0. \tag{8.11}$$

Kontrasto koeficientus, t. y. konkrečius skaičius, pasirenkame patys. Koeficientai gali būti teigiami, neigiami, sveikieji ar trupmeniniai skaičiai. Imdami tik du nenulinius koeficientus 1 ir –1, gauname hipotezę apie dviejų vidurkių lygybę. Norėdami atskirti vieną nuo ki-

tos kelias populiacijų grupes, pvz., palyginti pirmųjų dviejų populiacijų kombinaciją su penktąja, galime pasirinkti tokius koeficientus: –1 –1 0 0 2. Populiacijos, kurios atitinka nulinius koeficientus, iš analizės pašalinamos.

Tikrinamos hipotezės:

$$\begin{cases} H_0: C_1\mu_1 + C_2\mu_2 + \ldots + C_k\mu_k = 0, \\ H_1: C_1\mu_1 + C_2\mu_2 + \ldots + C_k\mu_k \neq 0, \end{cases}$$

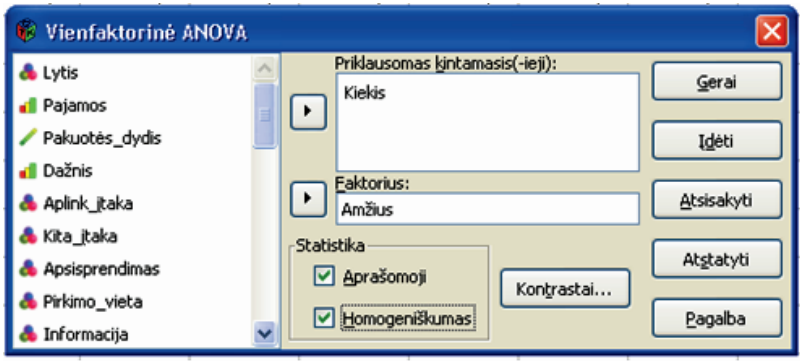
čia C_i – iš anksto pasirinkti skaičiai, tenkinantys (8.10) lygybę.

8.6. VIENFAKTORINĖ DISPERSINĖ ANALIZĖ PSPP IR SPSS TERPĖJE

Pavyzdžiui, **išnagrinėsime amžiaus poveikį perkamos kavos kiekiui**. Respondentai pagal amžių paskirstyti į keturias grupes: 15–25 m., 26–35 m., 36–45 m., daugiau kaip 45 m. Reikia nustatyti, ar jie perka vienodą kavos kiekį.

PSPP terpė

Vienfaktorinę dispersinę analizę galima atlikti taikant procedūrą *Analizuoti → Palyginti vidurkius → Vienfaktorinė ANOVA*. Atsiveria dispersinės analizės dialogo langas *Vienfaktorinė ANOVA* (žr. 8.13 pav.). Priklausomą kintamąjį *Kiekis* perkeliame į langelį (*Priklausomas kintamasis(-ieji)*), į nepriklausomų kintamųjų langelį (*Factorius*) perkeliame kintamąjį *Amžius.* Dispersijų lygybę nustato *Livyno* kriterijus. Kad būtų galima patikrinti *Livyno* kriterijų, reikia pažymėti langelį *Homogeniškumas*. Reikia pažymėti langelį *Deskriptive*, kad būtų galima parodyti aprašomosios statistikos duomenis. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 8.14 pav.).



8.13 pav. Dispersinės analizės dialogo langas *Vienfaktorinė ANOVA*

Lentelėje *Aprašomoji* (aprašomoji statistika) pateiktos kiekvienos grupės skaitinės charakteristikos: stebėjimų skaičius (*N*), vidurkis (*Vidurkis*), standartinis nuokrypis (*Std. nuokrypis*), standartinė paklaida (*Std. paklaida*), vidurkio 95 % pasikliautinasis intervalas (*95 % pasikliautinasis intervalas vidurkiams*), mažiausia reikšmė (*Mažiausia*), didžiausia reikšmė (*Didžiausia*).

Aprašomoji

		N	Vidurkis	Std. nuokrypis	Std. paklaida	95% pasikliautinasis intervalas vidurkiams		Mažiausia	Didžiausia
						Apačia	Viršus		
Kiekis	1	25	398,40	292,44	58,49	277,69	519,11	50,00	750,00
	2	14	509,29	499,05	133,38	221,14	797,43	80,00	2000,00
	3	19	628,42	418,27	95,96	426,82	830,02	100,00	1500,00
	4	2	50,00	,00	,00	50,00	50,00	50,00	50,00
	Iš viso	60	485,50	399,24	51,54	382,36	588,64	50,00	2000,00

Dispersijų homogeniškumo kriterijus

	Statistika (Levene)	skirt1	skirt2	Reikšmingumo lygmuo
Kiekis	2,00	3	56	,12

ANOVA

		Kvadratų suma	skirt.	Vidutinis kvadratas	F	Reikšmingumo lygmuo
Kiekis	Tarp grupių	965003,5	3	321667,8	2,13	,11
	Vidinė	8439281	56	150701,5		
	Iš viso	9404285	59			

8.14 pav. Vienfaktorinės dispersinės analizės rezultatas PSPP terpėje

Lentelėje *Dispersijų homogeniškumo kriterijus p-level* (stulpelis *Reikšmingumo lygmuo*) rodo Livyno kriterijaus (*Statistika (Levene)*) kriterijaus reikšmingumą. Jei $p\text{-}level > \alpha = 0{,}05$ (kriterijus nereikšmingas), dispersijos lygios. Nepriklausomų kintamųjų grupių skirtumų reikšmingumą rodo *F* kriterijaus *p-level* (žr. stulpelį *Reikšmingumo lygmuo*) lentelėje *ANOVA*. Jei *Livyno* kriterijus parodė dispersijų skirtumą, tai patikimumo lygmenį α reikia padidinti nuo *0,05* iki *0,01*. Nagrinėjamu atveju dispersijos lygios, $p\text{-}level = 0{,}106 > \alpha = 0{,}05$, skirtumas nereikšmingas. Negalima atmesti H_0 hipotezės dėl vidurkių lygybės.

Pastaba: PSPP paketas neskaičiuoja aposteriorinių kriterijų.

SPSS terpė

Vienfaktorinę dispersinę analizę galima atlikti taikant procedūrą *Analyze→Compare Means→One-way ANOVA*. Atsiveria dispersinės analizės dialogo langas *One-Way ANOVA*. Priklausomą kintamąjį *Kiekis* perkeliame į langelį (*Dependent list*), į nepriklausomų kintamųjų langelį (*Factor*) perkeliame kintamąjį *Amžius*. Dispersijų lygybę nustato *Livyno* kriterijus. Dispersijų tikrinimo parametrai nustatomi lange *One-Way ANOVA: Options*, kuris atsiveria paspaudus *Options* mygtuką. Kad būtų galima patikrinti *Livyno* kriterijų, reikia pažymėti langelį *Homogeneity of variance test*. Tame pačiame lange reikia pažymėti langelį *Deskriptive*, kad būtų galima parodyti aprašomosios statistikos duomenis. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 8.15 pav.).

Lentelėje *Descriptives* (aprašomoji statistika) pateiktos kiekvienos grupės skaitinės charakteristikos: stebėjimų skaičius (*N*), vidurkis (*Mean*), standartinis nuokrypis (*Std. Deviation*), standartinė paklaida (*Std. Error Mean*), vidurkio 95 % pasikliautinasis intervalas (95% *Confidence Interval for Mean*), mažiausia reikšmė (*minimum*), didžiausia reikšmė (*maximum*).

Lentelėje *Test of Homogeneity of Variances p-level* (stulpelis *Sig.*) rodo Livyno kriterijaus (*Levene Statistic*) reikšmingumą. Jei $p\text{-}level > \alpha = 0{,}05$ (kriterijus nereikšmingas), dispersijos lygios. Nepriklausomų kintamųjų grupių skirtumų reikšmingumą rodo *F* kriterijaus *p-level* (žr. stulpelį *Sig.*) lentelėje *ANOVA*. Jei *Livyno* kriterijus parodė dispersijų skirtumą, tai patikimumo lygmenį α reikia padidinti nuo *0,05* iki *0,01*.

Nagrinėjamu atveju dispersijos lygios, *p-level = 0,106 > α = 0,05*, skirtumas nereikšmingas. Negalima atmesti H_0 hipotezės dėl vidurkių lygybės.

Descriptives

Kiekis

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	25	398,4000	292,44202	58,48840	277,6859	519,1141	50,00	750,00
2	14	509,2857	499,05240	133,37736	221,1414	797,4300	80,00	2000,00
3	19	628,4211	418,26710	95,95705	426,8228	830,0193	100,00	1500,00
4	2	50,0000	,00000	,00000	50,0000	50,0000	50,00	50,00
Total	60	485,5000	399,24261	51,54200	382,3647	588,6353	50,00	2000,00

Test of Homogeneity of Variances

Kiekis

Levene Statistic	df1	df2	Sig.
1,997	3	56	,125

ANOVA

Kiekis

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	965003,511	3	321667,837	2,134	,106
Within Groups	8439281,489	56	150701,455		
Total	9404285,000	59			

8.15 pav. Vienfaktorinės dispersinės analizės rezultatas SPSS terpėje

Pastaba: SPSS paketas gali apskaičiuoti aposteriorinius kriterijus. Tuo tikslu paspaudus mygtuką *Post Hoc* atsiveria dialogo langas *One-Way ANOVA: Post Hoc Multiple Comparisons.* Šiame lange reikia pažymėti pasirinktus kriterijus.

Atliktos procedūros sintaksė pateikta 8.6 lentelėje.

8.6 lentelė. Vienfaktorinės dispersinės analizės ANOVA sintaksė

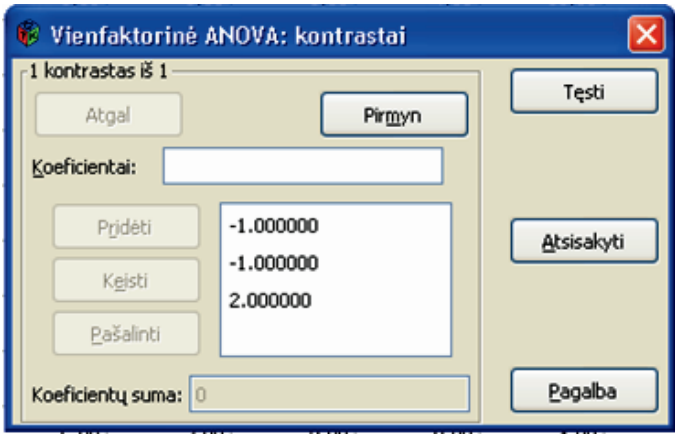
PSPP	SPSS
ONEWAY ONEWAY/VARIABLES=Kiekis BY Amžius /STATISTICS=DESCRIPTIVES HOMOGENEITY.	ONEWAY Kiekis BY Amžius /STATISTICS DESCRIPTIVES HOMOGENEITY /MISSING ANALYSIS.

8.6.1. Kontrastų tikrinimas

Panaudosime anksčiau išnagrinėtą pavyzdį.

PSPP terpė

Atliekant komandas *Analizuoti → Palyginti vidurkius →Vienfaktorinė ANOVA*, atsiveria dispersinės analizės dialogo langas *Vienfaktorinė ANOVA* (žr. 8.14 pav.). Priklausomą kintamąjį *Kiekis* perkeliame į langelį (*Priklausomas kintamasis(-ieji)*), į nepriklausomų kintamųjų langelį (*Factorius*) perkeliame kintamąjį *Amžius.* Paspaudus mygtuką *Kontrastai,* atsiveria langas *Vienfaktorinė ANOVA: kontrastai* (žr. 8.16 pav.). Šiame lange langelyje *Koeficientai* iš eilės įvedami kontrasto koeficientai, kiekvieno įvedimas patvirtinamas mygtuko *Pridėti* paspaudimu. Paspaudę *Tęsti* mygtuką, grįžtame į pradinį langą.



8.16 pav. Dialogo langas *Vienfaktorinė ANOVA: kontrastai*

Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 8.17 pav.).

Livyno kriterijus (*Dispersijų homogeniškumo kriterijaus lentelė)* rodo, kad nagrinėjamu atveju dispersijos lygios, $p\text{-level} = 0{,}23 > \alpha = 0{,}05$. Lentelėje *Kontrastų koeficientai* pateikti pasirinkti koeficientai. Lentelėje *Kontrastų kriterijus p-level* (*p-reikšmė (2-pusė))(Manant, kad dispersijos yra lygios)*, yra didesnė už pasirinktą reikšmingumo lygmenį

p-level = 0,28 > α = 0,05, negalima atmesti nulinės hipotezės apie kontrasto reikšmę, t. y. dviguba trečios grupės vidurkio reikšmė statistiškai reikšmingai nesiskiria nuo pirmos ir antros grupės vidurkių sumos.

Dispersijų homogeniškumo kriterijus

	Statistika (Levene)	skirt1	skirt2	Reikšmingumo lygmuo
Kiekis	1,52	2	57	,23

ANOVA

		Kvadratų suma	skirt.	Vidutinis kvadratas	F	Reikšmingumo lygmuo
Kiekis	Tarp grupių	359589,5	2	179794,7	1,13	,33
	Vidinė	9044696	57	158678,9		
	Iš viso	9404285	59			

Kontrastų koeficientai

		Amžius		
		1	2	3
Kontrastas	1	-1	-1	2

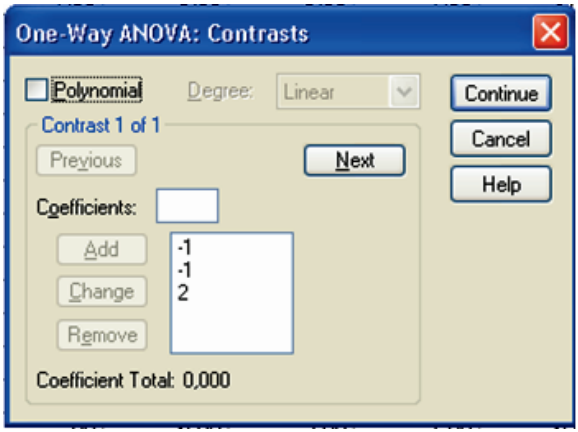
Kontrastų kriterijus

		Kontrastas	Kontrasto reikšmė	Std. paklaida	t	skirt.	p-reikšmė (2-pusė)
Kiekis	Manant, kas dispersijos yra lygios	1	238,98	218,87	1,09	57	,28
	Nemanant, kad lygios	1	238,98	238,68	1,00	36,56	,32

8.17 pav. Kontrastų tikrinimo tezultatas PSPP terpėje

SPSS terpė

Atliekant komandas *Analyze→Compare Means→One-way ANOVA,* atsiveria dispersinės analizės dialogo langas *One-Way ANOVA*. Priklausomą kintamąjį *Kiekis* perkeliame į langelį (*Dependent list*), į nepriklausomų kintamųjų langelį (*Factor*) perkeliame kintamąjį *Amžius.* Dispersijų lygybę nustato *Livyno* kriterijus. Paspaudus mygtuką *Contrasts,* atsiveria langas *One-Way ANOVA: Contrasts* (žr. 8.18 pav.). Šiame lange langelyje *Coefficients* iš eilės įvedami kontrasto koeficientai, kiekvieno įvedimas patvirtinamas mygtuko *Add* paspaudimu. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 8.19 pav.).



8.18 pav. Dialogo langas *One-Way ANOVA:Contrasts*

Test of Homogeneity of Variances

Kiekis

Levene Statistic	df1	df2	Sig.
1,517	2	57	,228

ANOVA

Kiekis

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	359589,476	2	179794,738	1,133	,329
Within Groups	9044695,524	57	158678,869		
Total	9404285,000	59			

Contrast Coefficients

Contrast	Amžius		
	1	2	3
1	-1	-1	2

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Kiekis	Assume equal variances	1	238,9810	218,87418	1,092	57	,279
	Does not assume equal variances	1	238,9810	238,67811	1,001	36,563	,323

8.19 pav. Kontrastų tikrinimo rezultatas SPSS terpėje

Livyno kriterijus (*Test of Homogeneiti of Variances)* rodo, kad nagrinėjamu atveju dispersijos lygios, $p\text{-level} = 0{,}228 > \alpha = 0{,}05$. Lentelėje *Contrasts Coefficients* pateikti pasirinkti koeficientai. Lentelėje *Contrast Tests p-level* (*Sig. (2-tailed))(Assume equal variances)*, yra didesnė už pasirinktą reikšmingumo lygmenį $p\text{-level} = 0{,}279 > \alpha = 0{,}05$, negalima

atmesti nulinės hipotezės apie kontrasto reikšmę, t. y. dviguba trečios grupės vidurkio reikšmė statistiškai reikšmingai nesiskiria nuo pirmos ir antros grupės vidurkių sumos.

Atliktos procedūros sintaksė pateikta 8.7 lentelėje.

8.7 lentelė. Kontrastų tikrinimo sintaksė

PSPP	SPSS
ONEWAY ONEWAY/VARIABLES=Kiekis BY Amžius /STATISTICS=HOMOGENEITY /CONTRAST= -1 -1 2.	ONEWAY Kiekis BY Amžius /CONTRAST= -1 -1 2 /STATISTICS HOMOGENEITY /MISSING ANALYSIS.

8.7. SKIRTUMŲ ANALIZĖ *MS EXCEL* TERPĖJE

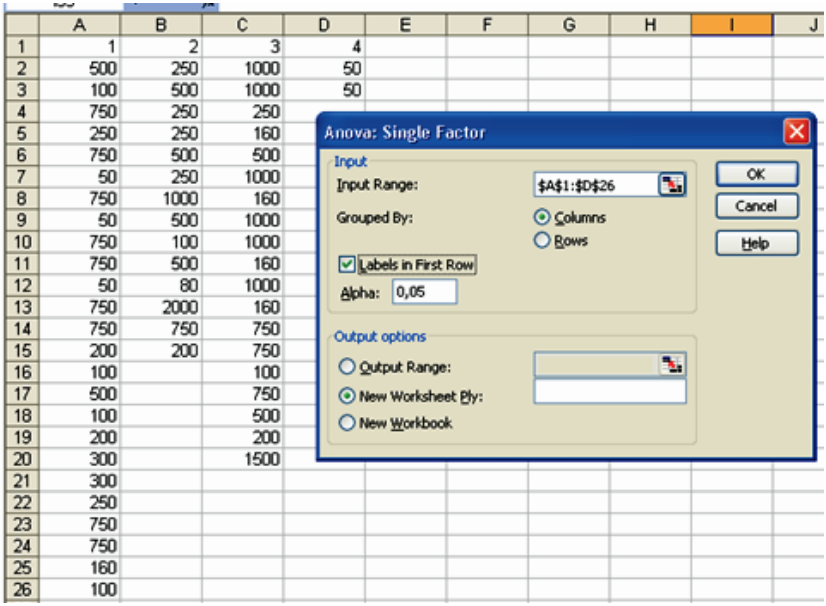
MS Excel paketas leidžia atlikti vienfaktorinę dispersinę analizę (*Anova: Single Factor), T* kriterijus dviem priklausomoms (porinėms) imtims *(t-Test: Paired Two Sample for Means), T* kriterijus dviem nepriklausomoms imtims *(t-Test: Two-Sample Assuming Equal Variances, t-Test: Two-Sample Assuming Unequal Variances).* Komanda *Data Analysis* atveria dialogo langą *Data Analysis* (žr. 6.4 pav.). Šiame lange pateiktas galimų statistinių operacijų sąrašas *Analysis Tools.*

8.7.1. Vienfaktorinė dispersinė analizė

Išnagrinėsime amžiaus poveikį perkamos kavos kiekiui. Respondentai pagal amžių paskirstyti į keturias grupes: 15–25 m., 26–35 m., 36–45 m., daugiau kaip 45 m. Reikia nustatyti, ar vienodą kavos kiekį jie perka. Kintamojo *Kiekis* reikšmes reikia surašyti į keturis stulpelius (surūšiuoti pagal amžiaus grupes): 15–25 m. (1), 26–35 m. (2), 36–45 m. (3), daugiau kaip 45 m. (4). Lange *Data Analysis* pasirenkame komandą *Anova: Single Factor.* Atsiveria dialogo langas *Anova: Single Factor* (žr. 8.20 pav.). Langelyje *Input Range:* reikia įvesti nagrinėjamo duomenų masyvo koordinates. Jei buvo taikomi stulpelių pavadinimai, reikia pažymėti langelį *Labels in First Row.*

Dispersinės analizės rezultatai pateikti 8.7 lentelėje. Jie yra analogiški pateiktiems 8.16 pav.

Lentelėje *SUMMARY* (aprašomoji statistika) pateiktos kiekvienos grupės skaitinės charakteristikos: stebėjimų skaičius (*Count*), vidurkis (*Average*), dispersija (*Variante)*.



	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4						
2	500	250	1000	50						
3	100	500	1000	50						
4	750	250	250							
5	250	250	160							
6	750	500	500							
7	50	250	1000							
8	750	1000	160							
9	50	500	1000							
10	750	100	1000							
11	750	500	160							
12	50	80	1000							
13	750	2000	160							
14	750	750	750							
15	200	200	750							
16	100		100							
17	500		750							
18	100		500							
19	200		200							
20	300		1500							
21	300									
22	250									
23	750									
24	750									
25	160									
26	100									

8.20 pav. Vienfaktorinės dispersinės analizės langas

Nepriklausomų kintamųjų grupių skirtumų reikšmingumą rodo *F* kriterijaus *p-level* (žr. stulpelį *P-value*) lentelėje *ANOVA*. Nagrinėjamu atveju $p\text{-level} = 0{,}106 > \alpha = 0{,}05$, skirtumas nereikšmingas. Negalima atmesti H_0 hipotezės dėl vidurkių lygybės. Galima analizuoti *Fišerio F* kriterijaus apskaičiuotą (*F*) ir kritinė (*F crit.*) reikšmę. $F = 2{,}134 < F\ crit = 2{,}769$, negalima atmesti H_0 hipotezės dėl vidurkių lygybės.

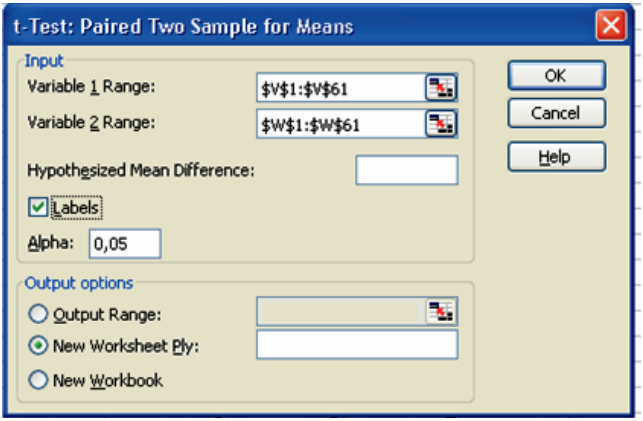
8.8 lentelė. Vienfaktorinės dispersinės analizės rezultatas

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
1	25	9960	398,4	85522,33		
2	14	7130	509,2857	249053,3		
3	19	11940	628,4211	174947,4		
4	2	100	50	0		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	965003,5	3	321667,8	2,134471	0,106088	2,769431
Within Groups	8439281	56	150701,5			
Total	9404285	59				

8.7.2. *T* kriterijus dviem priklausomoms (porinėms) imtims

Vartotojai nurodė A ir B jogurtų priimtiną 500 gr pakuotės kainą. Reikia **nustatyti, ar priimtinos kainos įvertinimas statistiškai reikšmingas.**

Lange *Data Analysis* pasirenkame komandą *t-Test: Paired Two Sample for Means.* Atsiveria dialogo langas *t-Test: Paired Two Sample for Means* (žr. 8.21 pav.). Langelyje *Variable 1 Range:* reikia įvesti kintamojo *A_kaina* reikšmių koordinates, langelyje Variable 1 Range – kintamojo *B_kaina* reikšmių koordinates.



8.21 pav. Dialogo langas *t-Test: Paired Two Sample for Means*

Analizės rezultatai pateikti 8.9 lentelėje. Jie yra analogiški pateiktiems 8.13 pav.

8.9 lentelė. *T* kriterijaus skaičiavimo rezultatas

t-Test: Paired Two Sample for Means		
	A_kaina	*B_kaina*
Mean	3	3,4
Variance	1,830508	1,294915
Observations	60	60
Pearson Correlation	-0,29724	
Hypothesized Mean Difference	0	
df	59	
t Stat	-1,54138	
P(T<=t) one-tail	0,064286	
t Critical one-tail	1,671093	
P(T<=t) two-tail	0,128572	
t Critical two-tail	2,000995	

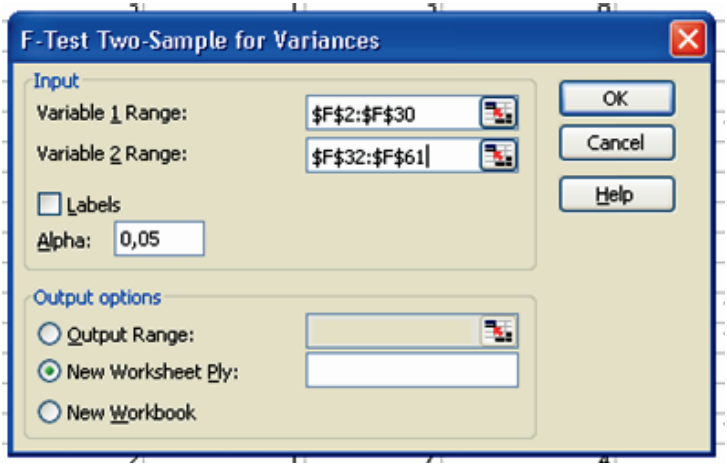
Lentelėje pateikti apskaičiuoti kintamųjų vidurkiai (*Mean*), dispersijos (*Variance),* kintamųjų skaičius *(Observation),* Pirsono koreliacijos koeficientas (*Pearson Correlation).* Išvada dėl nagrinėjamų kintamųjų vidurkių skirtumų daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*P(T<=t) two-tail*) reikšmę. Jei $p\text{-level} > \alpha$, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju $p\text{-level} = 0{,}129 > \alpha = 0{,}05$, vidurkių skirtumas 0,4 Lt (žr. *Mean*) yra statistiškai nereikšmingas. Negalima atmesti H_0 hipotezės. Išvadą dėl nagrinėjamų kintamųjų vidurkių skirtumų galima daryti ir lyginant Stjudento *t* kriterijaus apskaičiuotos (*t Stat*) ir kritinės (*t Critical two-tail)* reikšmės modulį, $|t\ Stat| = 1{,}54 < t\ Critical\ two\text{-}tail = 2{,}00$. Negalima atmesti H_0 hipotezės.

8.7.3. *T* kriterijus dviem nepriklausomoms imtims

Pavyzdžiui, **analizuojamas vyrų ir moterų suvartojamo per mėnesį kavos kiekio skirtumas**.

Turime kintamuosius *Kiekis* ir *Lytis*. Kintamąjį *Kiekis* reikia surūšiuoti pagal lytį. Turime du kintamųjų masyvus. MS Excel terpėje skirtingai skaičiuojama kintamųjų vidurkio lygybę, jei jų dispersijos lygios (*t-Test: Two-Sample Assuming Equal Variances),* arba nelygios (*t-Test: Two-Sample Assuming Unequal Variances).* Todėl prieš vidurkių lygybės tikrinimą, reikia nustatyti, ar nagrinėjamų duomenų masyvų dispersijos yra vienodos.

Lange *Data Analysis* pasirenkame komandą *F-test Two-Sample for Variances.* Atsiveria dialogo langas *F-test Two-Sample for Variances* (žr. 8.22 pav.). Langelyje *Variable 1 Range:* reikia įvesti kintamojo *Kiekis* (vyrai) reikšmių koordinates, langelyje Variable 1 Range – kintamojo *Kiekis* (moterys) reikšmių koordinates. Analizės rezultatai pateikti 8.10 lentelėje.



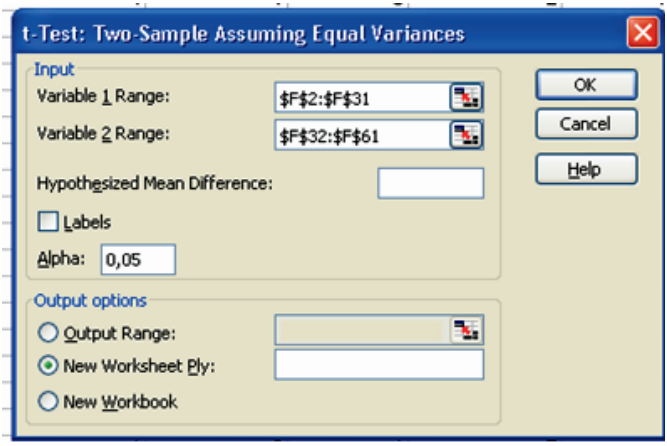
8.22 pav. Dialogo langas *F-test Two-Sample for Variances*

Išvada dėl nagrinėjamų kintamųjų dispersijų lygybę daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*P(F<=f) one-tail*) reikšmę. Jei *p-level* > α, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju *p-level* = *0,112* > α = *0,05*, negalima atmesti H_0 hipotezės, dispersijos lygios.

8.10 lentelė. Dispersijų lygybės tikrinimas

F-Test Two-Sample for Variances		
	Variable 1	*Variable 2*
Mean	513,4483	473
Variance	124094,8	196987,2
Observations	29	30
df	28	29
F	0,629964	
P(F<=f) one-tail	0,112498	
F Critical one-tail	0,53328	

Tirdami vidurkių lygybę, lange *Data Analysis* pasirenkame komandą *t-Test: Two-Sample Assuming Equal Variances.* Atsiveria dialogo langas *t-Test: Two-Sample Assuming Equal Variances* (žr. 8.23 pav.). Langelyje *Variable 1 Range:* reikia įvesti kintamojo *Kiekis* (vyrai) reikšmių koordinates, langelyje Variable 1 Range – kintamojo *Kiekis* (moterys) reikšmių koordinates. Analizės rezultatai pateikti 8.11 lentelėje.



8.23 pav. Dialogo langas *t-Test: Two-Sample Assuming Equal Variances*

8.11 lentelė. *T* kriterijaus skaičiavimo rezultatas

t-Test: Two-Sample Assuming Equal Variances		
	Variable 1	*Variable 2*
Mean	498	473
Variance	126975,2	196987,2
Observations	30	30
Pooled Variance	161981,2	
Hypothesized Mean Difference	0	
df	58	
t Stat	0,240577	
P(T<=t) one-tail	0,405366	
t Critical one-tail	1,671553	
P(T<=t) two-tail	0,810732	
t Critical two-tail	2,001717	

Lentelėje pateikti apskaičiuoti kintamųjų vidurkiai (*Mean*), dispersijos (*Variance),* kintamųjų skaičius *(Observation),* Pirsono koreliacijos koeficientas (*Pearson Correlation).* Išvada dėl nagrinėjamų kintamųjų vidurkių skirtumų daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*P(T < = t) two-tail*) reikšmę. Jei *p-level* > α, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju *p-level = 0,811* > α *= 0,05*, vidurkių skirtumas yra statistiškai nereikšmingas. Negalima atmesti H_0 hipotezės. Išvadą dėl nagrinėjamų kintamųjų vidurkių skirtumų galima daryti ir lyginant Stjudento *t* kriterijaus apskaičiuotos (*t Stat*) ir kritinės (*t Critical two-tail)* reikšmės modulį, | *t Stat* | = *0,24 < t Critical two-tail = 2,00.* Negalima atmesti H_0 hipotezės.

KONTROLINIAI KLAISIMAI

8.1. Išvardykite pagrindinius *t* kriterijaus tipus.

8.2. Kokie kintamieji analizuojami *t* kriterijaus metodu?

8.3. Koks kriterijus nustato imčių dispersijų lygybę?

8.4. Kokiais atvejais taikoma dispersinė analizė?

8.5. Kokios dispersinės analizės prielaidos?

8.6. Kaip dispersinės analizės atlikimo metu parenkamas priklausomas ir nepriklausomas kintamasis?

8.7. Kokie kintamieji vadinami veiksniais?

8.8. Kam taikomas kontrastų tikrinimas?

UŽDUOTYS

8.1. Prekybos centras akcijos lipduką išduoda už pirkiniams išleistą 30 Lt sumą. Atlikta analizė, norint nustatyti, ar realiai per vieną kartą išleidžiamų pinigų vidurkis reikšmingai skiriasi nuo šios sumos. Skaičiavimo rezultatas pateiktas 8.24 pav. Paaiškinkite gautą rezultatą.

8.2. Atlikta analizė, ar skirtingo amžiaus respondentai per vieną kartą pirkiniams išleidžia vienodai pinigų. Analizės rezultatas pateiktas 8.25 pav. Paaiškinkite gautą rezultatą.

Vienos imties statistika

	N	*Vidurkis*	*Std. nuokrypis*	*Vidur. s.pakl.*
Pinigai	36	9,00	3,28	,55

Kriterijus vienai imčiai

	Kriterijaus reikšmė = 30.000000					
					Skirtumo 95% pasikliautinasis intervalas	
	t	*skirt.*	*p-reikšmė (2-pusė)*	*Vidurkių skirtumas*	*Apačia*	*Viršus*
Pinigai	-38,44	35	,00	-21,00	-22,11	-19,89

8.24 pav. *T* kriterijaus skaičiavimo rezultatas

Grupių statistika

Amžius	*N*	*Vidurkis*	*Std. nuokrypis*	*Vidur. s.pakl.*
Pinigai >=35	20	7,95	2,84	,63
<35	16	10,31	3,40	,85

Kriterijus nepriklausomoms imtims

	Levene kriterijus dispersijų lygybei		t kriterijus vidurkių lygybei						
								Skirtumo 95% pasikliautinasis intervalas	
	F	*p-reikšmė*	*t*	*skirt.*	*p-reikšmė (2-pusė)*	*Vidurkių skirtumas*	*Vidurkių skirtumo std. paklaida*	*Apačia*	*Viršus*
PinigaiJei dispersijos būtų lygios	9,79	,00	2,27	34,00	,03	2,36	1,06	,21	4,52
Jei dispersijos būtų nelygios			2,23	29,21	,03	2,36	1,06	,19	4,53

8.25 pav. *T* kriterijaus skaičiavimo rezultatas

8.3. Nagrinėjant amžiaus poveikį pinigų, išleidžiamų per mėnesį šampūnui pirkti, kiekiui, gauti rezultatai pateikti 8.26 pav. Paaiškinkite gautus rezultatus.

Dispersijų homogeniškumo kriterijus

	Statistika (Levene)	skirt1	skirt2	Reikšmingumo lygmuo
Pinigai	3,79	22	67	,00

ANOVA

		Kvadratų suma	skirt.	Vidutinis kvadratas	F	Reikšmingumo lygmuo
Pinigai	Tarp grupių	597,73	22	27,17	9,96	,00
	Vidinė	182,77	67	2,73		
	Iš viso	780,50	89			

8.26 pav. Vienfaktorinės dispersinės analizės rezultatas

9. FAKTORINĖ ANALIZĖ

Faktorinė analizė leidžia kintamųjų masyve skirti keletą grupių, kurios vadinamos veiksniais. Grupavimas atliekamas apskaičiuojant kintamųjų koreliacija. Į vieną veiksnį jungiami kintamieji, stipriai koreliuojantys tarpusavyje ir silpnai koreliuojantys arba nekoreliuojantys su kitais kintamaisiais, kurie sudaro kitus veiksnius. Tokiu būdu sumažinamas kintamųjų skaičius.

Faktorinė analizė plačiai taikoma ekonomikoje, rinkodaroje ir kituose taikomuosiuose moksluose, kurie nagrinėja didelius duomenų kiekius. Faktorinė analizė pasižymi sudėtingais matematiniais skaičiavimais. Pateikiant faktorinės analizės aprašus, tenka įvertinti ir pagrįsti šių metodų taikymo galimybes ir ypatybes, matematinių modelių veikimo principus, o rezultatų interpretacija nėra paprasta ir vienareikšmė.

Faktorinė analizė neturi vienintelio sprendinio. Koreliacinę matricą galima pavaizduoti veiksniais skirtingais būdais. Be to, faktorinę analizę sudaro skirtingi modeliai, kurie paremti veiksnių svorių matricos apskaičiavimu. Dažniausiai faktorinės analizės programos iškart siūlo keletą skirtingų metodų. Geriausio metodo pasirinkimą dažniausiai lemia skaičiavimo patogumas, taip pat tyrėjo nuomonė.

Faktorinės analizės metodais sprendžiami dviejų tipų uždaviniai: tiriamieji ir patvirtinamieji. Pagal šiuos uždavinius ir faktorinė analizė vadinama tiriamąja ir patvirtinamąja.

Tiriamoji faktorinė analizė taikoma tada, kai turima kintamųjų aibė ir neaišku, nei kiek gali būti veiksnių, nei kokie kintamieji juos sudaro. Ir netgi neaišku, ar apskritai kintamieji yra kolinearūs.

Patvirtinančioji faktorinė analizė taikoma tada, kai iš anksto žinomi teoriniai ar anksčiau atlikti empiriniai tyrimai ir norima jau turimas žinias patvirtinti ar išplėsti.

Tiriamoji faktorinė analizė taikoma dažniau nei patvirtinančioji. Pagrindinė tiriamosios faktorinės analizės taikymo sritis – kintamųjų tarpusavio ryšių ir šablonų išskyrimas ir apibrėžimas.

Faktorinė analizė leidžia sumažinti duomenų kiekį, kuris dažnai apsunkina rinkos tyrimų analizę. Taikant faktorinės analizės metodus skiriama keletas bendrųjų veiksnių, kuriuose sutelkta informacija be nuostolių pakeičia informaciją, užfiksuotą dešimtyse kintamųjų.

Faktorinės analizės metodai leidžia priskirti stebimus kintamuosius veiksniams, o kiekvienas veiksnys atvaizduoja kintamųjų empirinius ryšius. Be to, veiksniai kiekvienam kintamajam suteikia svorį, kuris gali būti naudojamas jungiant kintamuosius į integruotus rodiklius.

Analizuojant socialinių grupių ar atskirų asmenų požiūrius, elgesį ar konfliktiškumą, dažnai formuluojamos pradinės hipotezės apie šių savybių tarpusavio ryšius. Faktorinės analizės metodai leidžia įvertinti vienai ar kitai grupei būdingas savybes ir jų ryšius.

Faktorinė analizė gali būti pritaikyta pradiniams duomenims transformuoti tokiu būdu, kad jie atitiktų kitų statistikos metodų keliamus reikalavimus. Pavyzdžiui, taikant daugiamatės regresijos metodus reikalaujama, kad nepriklausomi kintamieji tarpusavyje būtų nekoreliuoti. Kai ši sąlyga netenkinama, pradžioje gali būti taikomi faktorinės analizės metodai, kuriais nepriklausomų kintamųjų skaičius sumažinamas iki mažesnės veiksnių reikšmių imties, kurią galima panaudoti regresinei analizei, neprarandant duomenyse esančios informacijos.

Ryšių šablonai gali būti nagrinėjami dviem būdais.

Taikant pirmąjį būdą, tarp kintamųjų ieškoma latentinių (lot. *latens* (kilm. *latentis*) – paslėptas, nematomas) veiksnių, kurie lemia kintamųjų pokyčius. Galima stebėti šių požymių kitimo reguliarumą ir aprašyti jį kaip kitimo šabloną. Faktorinė analizė, kuri aprašo kintamųjų kitimo šablonus, vadinama ***R*** faktorine analize.

Tačiau galima išnagrinėti kintamųjų pokyčius ir stebėjimus sugrupuoti pagal jų panašumą. Pritaikę pastebėtus panašumo šablonus stebėjimams (individams, grupėms, miestams, valstybėms), gautume vadinamąją ***Q*** faktorine analizę. Šios analizės rezultatas – remiantis nustatytų veiksnių skaičiumi sudarytos stebėjimų grupės (klasteriai). Pastaruoju metu šis metodas retai taikomas, nes jau sukurta nemažai specialių daugiamačių kintamųjų sudalijimo į klasterius metodų.

Faktorinės analizės taikymo principai: absoliutus ir euristinis mąstymo būdai

Dažnai statistiniai metodai taikomi siekiant nustatyti ryšį tarp priklausomo ir nepriklausomų kintamųjų. ***Faktorinės analizės tikslas kitas: ištirti daugelio priklausomų kintamųjų ryšius. Šio tyrimo tikslas atrasti ir paaiškinti juos veikiančio nepriklausomo kintamojo (kuris nėra matuojamas tiesiogiai) prigimtį***. Todėl atsakymai, gauti taikant faktorinės analizės metodus, visuomet yra negalutiniai ir daugiau nuspėjami, palyginti su tais, kurie gaunami tada, kai nepriklausomas kintamasis yra matuojamas tiesiogiai. Numanomi nepriklausomi kintamieji yra vadinami veiksniais (*faktoriais*). Taikant faktorinę analizę siekiama nustatyti skirtingų veiksnių, kurie paaiškina tiriamų priklausomų kintamųjų ryšius, skaičių bei šių veiksnių prigimtį. Faktorinė analizė leidžia patikrinti, kaip gerai šie hipotetiniai veiksniai paaiškina turimus stebėjimus (duomenis) ir kurią dalį dispersijos paaiškina kiekvienas veiksnys.

Jeigu teigiame, kad tik konkretus veiksnių skaičius gali paaiškinti tiriamą reiškinį, tai absoliutaus mąstymo būdo pavyzdys. Jeigu manome, kad geriausiai reiškinį paaiškina du (trys, keturi, penki) veiksniai, nors jų gali būti ir daugiau, tai gauname euristinio mąstymo būdo pavyzdį.

Euristinio mąstymo būdo samprata padeda suprasti faktorinės duomenų analizės savybes. Keletas mokslininkų gali pritaikyti faktorinę analizę tiems patiems arba labai panašiems kintamiesiems ir nustatyti skirtingą veiksnių skaičių. Toks nevienareikšmiškumas yra būdingas daugeliui sričių. Jeigu norime gauti paprastesnius teorinius samprotavimus, renkamės mažesnį veiksnių skaičių; jei norime išsamesnių rezultatų – veiksnių skaičių tenka padidinti. Faktorinė analizė skiriasi nuo kitų metodų tuo, kad nėra tikslaus kriterijaus, leidžiančio vienareikšmiškai nustatyti veiksnių skaičių. Taikomi kriterijai dažniausiai pateikia daugiau veiksnių, nei galima tinkamai interpretuoti.

Apribojimai: *kintamųjų ir stebėjimų (atvejų) skaičius*

Stebėjimų skaičius turi būti toks, kad būtų galima gauti patikimus kintamųjų koreliacijos įverčius. Jeigu imtis nedidelė ir dar yra nutolusių reikšmių, gaunami netikslūs koreliacijos koeficientų įverčiai. Be to, reikiamas stebėjimų skaičius priklauso nuo tikrojo veiksnio struktūros. Kuo aiškesnė ir paprastesnė veiksnio struktūra, tuo mažesnės duomenų imties užtenka šiai struktūrai atskleisti. Remiantis bendra nuomone, faktorinės analizės metodų taikyti netikslinga, jeigu stebėjimų yra mažiau nei 50.

Taisyklė, susijusi su ***kintamųjų*** skaičiumi, yra priešinga regresinės analizės nuostatoms. Akivaizdu, kad vienas kintamasis negali apibūdinti bendrojo veiksnio. Tuo tikslu reikia turėti ne mažiau kaip tris kintamuosius. Apskritai kalbant, kuo daugiau kintamųjų, tuo geriau, jeigu šie kintamieji yra susiję su juos veikiančiais veiksniais. Kartais kintamųjų gali būti daugiau nei stebėjimų (atvejų). Tačiau apskritai rekomenduojama, kad imties dydis būtų dukart didesnis nei kintamųjų skaičius, t. y. kad stebėjimų būtų dukart daugiau nei kintamųjų, ir stebėjimai būtų heterogeniniai taikomų matų atžvilgiu. Faktorinei analizei tinka kiekybiniai ir dichotominiai kintamieji (jei dichotominis kintamasis koduojamas skaičiais 1 ir 2). Dėl ranginių kintamųjų galimi tik euristiniai faktoriniai modeliai be rezultatų statistinės interpretacijos. Tai reiškia, kad galima atlikti ranginių kintamųjų faktorinę analizę, tačiau negalima apskaičiuoti veiksnių reikšmių ir nustatinėti veiksnių reikšmingumo.

Be šių apribojimų, faktorinės analizės metodais gaunami patikimesni rezultatai, jei kintamieji atitinka šias ***prielaidas***:

- kintamieji turi daugiamatį normalųjį skirstinį;
- kintamieji susiję tiesinės formos ryšiais;
- nėra ryškiai išsiskiriančių (nutolusių) reikšmių.

Faktorinės analizės etapai:

1. Apskaičiuoti stebimų duomenų tarpusavio koreliacinę matricą ir nustatyti, kaip duomenys tinka faktorinei analizei.

2. Rasti pradinį galimų veiksnių sprendinį.

3. Iš pradinio sprendinio nustatyti tinkamiausią veiksnių kiekį.

4. Jeigu reikalinga išryškinti veiksnių šabloną, pritaikyti veiksnių sukimo metodą.

5. Atsižvelgiant į sprendžiamą uždavinį, kiekvienam tiriamam objektui apskaičiuoti veiksnių reikšmes (*factor scores*).

Faktorinės analizės duomenimis gali būti pradiniai stebėjimai arba jų koreliacinė matrica. Jeigu skaičiavimams naudojami pradiniai stebėjimai, tai pirmuoju faktorinės analizės žingsniu šie stebėjimai standartizuojami (z transformacija) ir tarp jų reikšmių apskaičiuojami Pirsono koreliacijos koeficientai. Pagrindiniu tolesnių skaičiavimų elementu tampa koreliacinė matrica. Jeigu koreliacinė matrica gali būti paaiškinta bendruoju veiksniu g, tai tada turi būti tokia stebimų kintamųjų

koreliacijos su kintamuoju *g* aibė, kad bet kurių dviejų šių koreliacijos koeficientų sandauga būtų lygi koreliacijai tarp stebimų kintamųjų. Jeigu mes įvertinome veiksnio koreliaciją su stebimais kintamaisiais, tai galime pasakyti, su kuriais kintamaisiais veiksnys koreliuoja stipriai, o su kuriais – silpnai. Taikant faktorinės analizės metodus negalima pasakyti, kokią įtaką veiksniui turi atskiras stebimas asmuo (respondentas). Tačiau, jei norima įvertinti atskiro asmens įtaką veiksniui, tai galima padaryti taikant daugiamatės regresinės analizės metodus.

Koreliacinės matricos gali nesiderinti su vieno bendrojo veiksnio hipoteze, tačiau gali pasitvirtinti hipotezė apie du, tris ar dar daugiau bendrųjų veiksnių. Kuo mažiau veiksnių, tuo paprastesnės hipotezės. Kadangi paprastos hipotezės turi loginį mokslinį pranašumą prieš sudėtingesnes hipotezes, mažiau veiksnių turinčios hipotezės yra vertesnės nei apimančios daug veiksnių.

Nesigilinant į matematines išraiškas, galima pasakyti, kad faktorinė analizė mėgina išreikšti kiekvieną kintamąjį kaip dviejų dalių – bendrosios ir specifinės – sumą. Bendrąją kintamųjų dalį išsamiai paaiškina bendrieji veiksniai, o specifinės dalys yra visiškai nekoreliuotos tarpusavyje. Kiek turimi duomenys atitinka šią sąlygą, sprendžiama iš vadinamosios liekanų kovariacinės matricos. Šios matricos įstrižainės elementai nėra lygūs vienetui, tačiau tipiniais metodais ši matrica gali būti transformuota į koreliacinę matricą (kiekvieną neįstrižinį elementą dalijant iš kvadratinės šaknies iš jį atitinkančių įstrižinių elementų sandaugos). Iš taip apskaičiuotų koreliacinės matricos elementų sprendžiama, ar tiriami kintamieji suderinami su *m* bendrųjų veiksnių hipoteze. Jeigu apskaičiuotos koreliacijų reikšmės yra didelės, jos yra nesuderinamos su hipoteze, kad populiacijoje jos lygios nuliui, tada hipotezę, kad yra *m* bendrųjų veiksnių, tenka atmesti. Didinant veiksnių skaičių *m* šios koreliacijos mažėja ir hipotezė apie veiksnių skaičių geriau derinasi su duomenimis.

Visada norima surasti mažiausią veiksnių skaičių *m*, kuris derintųsi su duomenimis. Tuo tikslu skaičiavimai pradedami nuo mažiausio veiksnių skaičiaus (paprastai $m = 1$) ir tikrinama hipotezė, kad šis skaičius atitinka turimus duomenis. Jei reikia, skaičius *m* didinamas. Kiekvienąkart didinant veiksnių skaičių *m,* hipotezė tampa sudėtingesnė, bet geriau atitinka duomenis. Sustojama, kai randama, kad hipotezė suderinama su duomenimis.

Centrinė faktorinės analizės teorema yra kiek panaši į dispersinės analizės centrinę teoremą, kuri teigia, kad kintamojo *X* dispersija gali būti išskirstyta arba, kitaip sakant, padalyta į atskirus dėmenis *(komponentus)*. Analogiškai dispersinei analizei, faktorinės analizės kovariacijų matrica R gali būti padalyta į bendrąją dalį *C*, kurią paaiškina veiksnių aibė, ir į specifinę dalį *U*, kurios šie veiksniai nepaaiškina. Kalbant matricų terminais, $R = C + U$. Kaip ir dispersinėje analizėje, paaiškinta dalis *C* gali būti dalijama toliau į komponentų matricas c_1, c_2, …, kurias paaiškina atskiri veiksniai. Kiekvienas iš šių vieno veiksnio komponentų c_j lygus veiksnių svorių stulpelio sandaugai. Kiekvienas diagonalinis šios matricos elementas yra veiksnio paaiškintas dispersijos kiekis.

Liekanų matrica *U* yra kovariancinė matrica, kurios reikšmės atspindi veiksnio nepaaiškintą kiekvieno kintamojo dispersijos dalį. Dažnai *C* yra kelių matricų c_j suma, o ne atskira matrica. Komponentų matricų c_1, c_2, …, kurių suma sudaro matricą *C*, skaičius yra vadinamos matricos *C* rangu. Matricos *C* rangas yra modelio bendrųjų veiksnių skaičius. Jeigu apibrėžiamas veiksnių skaičius *m*, matricos *C* ir *U* (kurių suma yra pradinė kovariacijų matrica *R*) sudaromos taip, kad matricos *C* rangas būtų *m*. Kuo didesnis *m*, tuo geriau *C* aproksimuoja *R*. Jeigu $m = k$, kai *k* yra matricos kintamųjų skaičius, tai kiekvienas matricos *C* elementas bus lygus matricos *R* elementui, o matricos *U* elementai bus nuliai. Tikslas – rasti kuo mažesnę veiksnių aibę *m* ir pasiekti, kad *C* vis dar tinkamai aproksimuotų *R*.

Veiksnių skaičius. Kartais veiksnių skaičius *m* gali būti spėjamas iš anksto. Jeigu taip nėra, tai taikomi matematiniai metodai. Dažniausiai naudojami metodai pagrįsti koreliacinės matricos tikrinėmis reikšmėmis. Populiariausi iš jų – *Kaiser* tikrinių reikšmių taisyklė ir *Cattel scree* kriterijus. *Kaiser* taisyklė veiksnių skaičiui nustatyti: veiksnių skaičius *m* turėtų būti lygus matricos tikrinių reikšmių skaičiui minus vienetas. Bendra šios taisyklės taikymo logika yra tokia: tikrinė reikšmė parodo dispersijos kiekį, kurį paaiškina vienas veiksnys, todėl nėra prasmės įtraukti veiksnį, kuris paaiškina mažiau dispersijos, negu jos yra viename kintamajame. Ši taisyklė buvo sukurta tais laikais, kai faktorinei analizei nebuvo taikomos kompiuterinės programos. Šiuo metu galima nesunkiai atlikti ir palyginti analizės rezultatus su skirtingais veiksnių rinkiniais.

Kaiser metodo alternatyva – *Scree* kriterijus. Taikant šį metodą, iš eilės, nuosekliai braižoma matricos tikrinių reikšmių diagrama. Tada pagal brėžinį sprendžiama, kurioje vietoje brėžinys išsilygina (nusistovi). Šio metodo trūkumas yra tas, kad iš brėžinio galima gauti skirtingas išvadas.

Kitas taikomas metodas panašus į *Scree* kriterijų, tačiau remiasi papildomais skaičiavimais. Kiekvienai tikrinei reikšmei T yra apskaičiuojama visų tolesniųjų tikrinių reikšmių suma S, kartu įtraukiant ir pačią reikšmę T. Santykis T/S yra T nepaaiškintos variacijos dalis. Pavyzdžiui, tarkime, kad devynių kintamųjų modelyje gauname keturias paskutiniąsias tikrines reikšmes, lygias 0,7, 0,3, 0,15 ir 0,1. Jų suma yra 1,25, o santykis $T/S = 0{,}7/1{,}25 = 0{,}56$. Taigi, pridėjus dar vieną veiksnį, bendras modelis paaiškintų 0,56 arba 56 proc. anksčiau nepaaiškintos dispersijos. Skaičiuodami toliau, gautume, kad didinant veiksnių skaičių toliau, dar kitas veiksnys paaiškintų $T/S = 0{,}3/(0{,}3 + 0{,}15 + 0{,}1) = 0{,}54$, taigi modelis su dar vienu veiksniu paaiškintų 54 proc. anksčiau nepaaiškintos dispersijos.

Be šių taisyklių, bendrųjų veiksnių skaičiui nustatyti taikomi ir kiti metodai. Kadangi, naudojant šiuolaikines programas, skaičiavimai yra greiti, kartu su kitais metodais galima taikyti ir patį paprasčiausią logiškai pagrįstą metodą. Skaičiavimai atliekami įvairiems veiksnių skaičiams m ir parenkamas toks veiksnių skaičius, kuris geriausiai paaiškina tiriamą reiškinį.

Nustatant didžiausią veiksnių skaičių, derėtų laikytis šios taisyklės: jeigu kintamųjų yra p, tai *didžiausią* galimą veiksnių skaičių m galima rasti iš formulės:

$$(p + m) \leq (p + m)^2. \tag{9.1}$$

Nesilaikant šios taisyklės, uždavinys gali išsigimti ir tapti trivialus. Taikant įvairias taisykles, visada būtina atminti, kad faktorinės analizės tikslas – paaiškinti daugiamačius kintamuosius su kiek galima mažesniu veiksnių skaičiumi.

Sukimas. Pradinius kintamuosius galima pakeisti jų tiesinėmis funkcijomis neprarandant informacijos, nes tiesinės funkcijos turi tą patį informacijos kiekį kaip ir pradiniai kintamieji. Tiesinės kintamųjų funkcijos gali užtikrinti paprastesnę ir suprantamesnę modelio struktū-

rą nei pradiniai kintamieji. Šį metodą galima pritaikyti faktorinei analizei. Faktorinėje analizėje manoma, kad stebėti kintamieji X_1, X_2,...,X_k sudaro priklausomų kintamųjų aibę, o veiksniai F_1, F_2,...,F_m yra nepriklausomi kintamieji. Tada gauname k lygčių sistemą:

$$\begin{aligned} X_1 &= \lambda_{11}F_1 + \lambda_{12}F_2 + \ldots + \lambda_{1m}F_m + \varepsilon_1, \\ X_2 &= \lambda_{21}F_1 + \lambda_{22}F_2 + \ldots + \lambda_{2m}F_m + \varepsilon_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_k &= \lambda_{k1}F_1 + \lambda_{k2}F_2 + \ldots + \lambda_{km}F_m + \varepsilon_k. \end{aligned} \qquad (9.2)$$

Daugikliai λ_{ij} vadinami veiksnių svoriais, ε_i – specifinis (charakteringasis) veiksnys.

Reikia atsiminti, kad *veiksniai F_1, F_2,...,F_m yra kintamųjų funkcija, o ne kintamieji*. Nežinomi kintamieji, kurie sudaro kiekvieną funkciją *F*, yra susieti nežinomais būdais. Žinodama kintamuosius X_1, X_2,...,X_k faktorinė analizė nustato nežinomas funkcijas F_1, F_2,...,F_m. Standartizuoti šios algebrinės lygties koeficientai (konstantos) sudaro matricą $k \times m$, kuri vadinama veiksnių svorių (*factor loadings*) matrica. Be to, faktorinė analizė suskaičiuoja veiksnių reikšmes (*factor scores*) ir veiksnių reikšmių koeficientus.

Jeigu originalūs veiksniai bus pakeisti jų tiesinėmis funkcijomis, gausime tuos pačius priklausomų kintamųjų įverčius kaip ir anksčiau, tačiau veiksnių svorių matricos skirsis. Išnagrinėję daugelį galimų tiesinių funkcijų, rastume vieną, kurios veiksnių svorių matrica būtų paprasčiausia. Paprastumas suprantamas kaip matricos nulinių elementų kiekis: kuo daugiau nulių, tuo paprastesnė struktūra. Sukimo būdas, kuris reikalauja, kad veiksniai liktų nekoreliuoti, vadinamas ortogonaliu sukimu, o kiti sukimo būdai – pasvirusiu sukimu. Tipiniai ortogonalaus sukimo metodai yra *Varimax, Quartimax, Equamax* ir *Orthomax*. Dažniausiai naudojamas *Varimax* sukimo būdas. Pasvirę sukimai dažnai struktūrą daro paprastesnę, tačiau, interpretuojant rezultatus, kartu būtina nagrinėti veiksnių tarpusavio koreliacijas. Reikia pažymėti, kad matematinė veiksnių sukimo procedūra paremta iteraciniais skaičiavimais, todėl kartais veiksnių gali nepavykti pasukti.

Pagrindinių komponenčių analizė (nagrinėjama suminė dispersija) ir bendrųjų veiksnių analizė (kai nagrinėjama bendroji dispersija) yra du svarbiausi faktorinės analizės metodai. Tarp šių dviejų metodų yra konceptualus skirtumas: bendrųjų veiksnių modelius galima įvertinti statistiniais kriterijais, o pagrindinių komponentų modeliams to padaryti negalima. Taip yra dėl to, kad bendrųjų veiksnių modelių specifiniai (charakteringieji) veiksniai privalo būti nekoreliuoti, o pagrindinių komponentų modelių liekanos tarpusavyje koreliuoja. Pagaliau komponentai nėra latentiniai kintamieji, jie yra tiesinės pradinių kintamųjų kombinacijos, taigi determinuoti. Tuo tarpu veiksniai yra latentiniai kintamieji, taigi nedeterminuoti.

Nors pagrindinių komponenčių analizė (PKA) sprendžia panašias problemas kaip ir bendrųjų veiksnių analizė, tačiau skirtingomis priemonėmis. PKA formuluoja ir sprendžia tiksliai statistiškai apibrėžtas problemas ir gaunamas sprendinys, išskyrus keletą specialių atvejų, yra vienintelis su aiškiomis matematinėmis savybėmis. Tačiau, PKA taikant praktiškai, susiduriama su problemomis. Todėl dažnai PKA taikoma kaip bendrųjų veiksnių analizės aproksimacija. Jeigu analizės tikslas yra duomenų kiekio sumažinimas, reikia taikyti pagrindinių komponenčių analizės metodiką. Jeigu norima paaiškinti pradinių duomenų tarpusavio koreliacijas ir sudaryti testuojamą modelį, reikia taikyti faktorinės analizės metodus.

Visa koreliacinė matrica pateikiama tais atvejais, kai kintamųjų skaičius nėra labai didelis. Paprastai koreliacinė matrica pateikiama be komentarų, nes ji nėra analizės tikslas. Tačiau koreliacinė matrica aprėpia daug naudingos informacijos, padedančios suvokti ryšius tarp kintamųjų porų.

Kad *duomenys tiktų faktorinei analizei*, jie turėtų būti koreliuoti tarpusavyje. Jeigu duomenys nesusiję, tai nėra ir bendrumų. Todėl prieš pradedant faktorinę analizę reikia nustatyti, ar šio metodo taikymas turi prasmę. Ar kintamieji susiję tarpusavyje, galima spręsti iš pradinės koreliacinės matricos koeficientų. Jeigu kintamieji nesusiję, jie negali būti grupuojami ir gali sudaryti tik atskirus veiksnius. Tokius kintamuosius iš pradinių kintamųjų sąrašo derėtų pašalinti.

Kaip duomenys (kintamieji) tinka faktorinei analizei, tikrinama Kaizerio–Mejerio–Olkino matu *KMO* ir *Bartletto* sferiškumo kriterijumi. Šie du metodai sudaro mažiausią standartą, kurį būtinas prieš atliekant faktorinę ar pagrindinių komponentų analizę. *KMO* matas tikrina, ar daliniai kintamųjų koreliacijos koeficientai yra maži. *KMO* reikšmės kinta nuo 0 iki 1 ir, kuo reikšmės artimesnės vienetui, tuo geriau. *KMO* matas apskaičiuojamas pagal formulę:

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}}{\sum\sum_{i \neq j} r_{ij} + \sum\sum_{i \neq j} \tilde{r}_{ij}}, \qquad (9.3)$$

čia r_{ij} – koreliacijos koeficientas; $\tilde{r}_{ij}$ – dalinės koreliacijos koeficientas. Jei *KMO* reikšmė maža, tai nagrinėjami rodikliai koreliacinei analizei netinka. *KMO* reikšmių interpretavimas pateiktas 9.1 lentelėje.

9.1 lentelė. Kaizerio–Mejerio–Olkino mato interpretacija

KMO reikšmė	Interpretacija
0,9 < KMO ≤ 1,0	Faktorinė analizė tinka idealiai
0,8 < KMO ≤ 0,9	Faktorinė analizė tinka puikiai
0,7 < KMO ≤ 0,8	Faktorinė analizė tinka vidutiniškai
0,6 < KMO ≤ 0,7	Faktorinė analizė tinka pakenčiamai
0,5 < KMO ≤ 0,6	Faktorinė analizė tinka blogai
0,0 ≤ KMO ≤ 0,5	Faktorinė analizė netinka

Kitas kintamųjų tarpusavio ryšio stiprumo rodiklis yra *Bartletto sferiškumo testas*. Šis testas tikrina nulinę hipotezę, kad populiacijos koreliacinės matricos kintamieji yra nekoreliuoti, t. y. kad koreliacinė matrica yra vienetinė, jos įstrižainės elementai yra vienetai, o visi kiti – nuliai (kitaip sakant, kintamieji yra nekolinearūs).

Skaičiavimuose pateikiami standartiniai pagrindinių komponenčių analizės rezultatai. Įvedamos naujos komponentės paaiškina kaskart vis mažiau ir mažiau dispersijos. Visa matricos dispersija priklauso nuo kintamųjų skaičiaus.

Pasirinkus veiksnių skaičių, galima apskaičiuoti veiksnių svorius. Veiksnio svoris – tai koreliacija tarp kintamojo ir veiksnio, kuris buvo

gautas (kitaip – ištrauktas) iš stebėtų duomenų. Skaičiavimo programos pateikia keletą veiksnių matricos variantų. Pirmasis – nepasukta veiksnių matrica, antras – pasukta veiksnių svorių matrica.

Veiksnių interpretavimas ir sukimas

Vienas iš pagrindinių faktorinės analizės uždavinių – suprasti ir aprašyti, ką matuoja veiksniai, o svarbiausias rodiklis, padedantis suprasti veiksnių prigimtį, yra ***veiksnių svoriai***. Idealiu atveju kintamojo svoris vienam veiksniui turėtų būti didelis (arti vieneto), o kitiems veiksniams – arti nulio. Praktiškai pirmiausia reikėtų nustatyti, kuris kintamojo svoris veiksniui yra didžiausias. Interpretuojant veiksnius galima taikyti žymių metodiką, t. y. nustatytiems veiksniams priskirti atitinkamas žymes. Šios žymės palengvina rezultatų aptarimą ir tolesnę jų analizę. Interpretacijai naudojamos trijų tipų žymės: *simbolinės*, *aprašomosios* ir *priežastinės*.

Simbolinės žymės yra paprasčiausi simboliai be jokių savarankiškų prasmių. Jų tikslas – pažymėti šablonus. Trijų veiksnių šablonai gali būti žymimi, pavyzdžiui, A, B ir C arba A1, A2 ir A3. Šios žymės tiesiog priskiriamos veiksniams be jokios papildomos prasmės ar interpretacijos.

Nors simbolinės žymės padeda išvengti nesusipratimų, jos nepadeda aiškinti tyrimų rezultatų ir lyginti tyrimų. Tokių trūkumų neturi ***aprašomosios*** žymės, kai veiksniams priskiriami raktiniai žodžiai. Aprašomosioms žymėms naudojamos sąvokos atspindi tiriamo reiškinio esmę ir turi padėti rezultatus suskirstyti į grupes.

Priežastinių žymių tikslas – atskleisti šablonus sukeliančias priežastis. Priežastinė žymė yra glaustas paaiškinimas, kodėl šablonas įtraukia tam tikrus kintamuosius.

Veiksnių šablonų interpretavimo būdas priklauso nuo asmeninio požiūrio ir patirties. Tyrėjas gali vartoti sąvokas, kurios skatina mąstyti apie gautus rezultatus ir padaro juos lengviau suprantamus. Tačiau šiuo atveju kyla klaidingos interpretacijos ar konkretumo praradimo pavojus. Kartais suteikta simbolinė etiketė yra tikslesnė ir aiškesnė nei išsamus aprašas to, ką ji reiškia.

Jeigu vienas ar keli kintamieji turi beveik vienodus svorius veiksniams, kartu veiksnių aiškinimą padarydami nevienareikšmiu, skaičiuojant veiksnių svorius taikomos įvairios, dažniausiai ortogonalios, veiksnių sukimo strategijos. Šių strategijų tikslas yra gauti aiškius veiks-

nių svorius, t. y. kad vieni kintamieji turėtų didelius veiksnių svorius (idealiu atveju – vienetus), o kiti – mažus (idealiu atveju –nulius). Taip būtų išryškinti svarbiausi šablonai.

Veiksnių reikšmių (*factor scores*) skaičiavimas yra vienas iš pagrindinių tyrimo etapų. Iš tikrųjų neužtenka nustatyti vien veiksnių, paaiškinančių tiriamų požymių tarpusavio koreliacijas ir pagrindinę jų dispersijos dalį, egzistavimą. Pageidautina tiesiogiai apibrėžti šiuos bendrus veiksnius, aprašyti juos pasitelkus kintamuosius ir pateikti patogią jų interpretaciją. Tam taikomi įvairūs metodai, pavyzdžiui, *Bartletto, Thomsono* arba *Andersono ir Rubino*. Skaičiuojant veiksnių reikšmes, dažniausiai veiksniai interpretuojami kaip nežinomi regresijos lygties koeficientai.

Praktiškai taikantys faktorinę analizę tyrėjai teigia, kad sąlygos, kurias teoriškai tenkina veiksnių reikšmės (pavyzdžiui, jų dispersija lygi vienetui, jeigu veiksniai yra ortogonalūs), ne visada tenkinamos. Apskaičiuotos veiksnių reikšmės retai kada turi vienetinę dispersiją, ir jos dažnai koreliuoja tarpusavyje, nors atliekant analizę veiksniai buvo ortogonalūs.

Veiksnių reikšmėms apskaičiuoti taikomi metodai, kuriuos galima suskirstyti į dvi grupes. Pirmąją grupę sudaro vadinamieji *tikslūs* metodai (žr. 9.2 lentelę), pasižymintys skirtingomis savybėmis. Vienas iš pirmųjų, *Thurstone* regresijos metodas minimizuoja neapibrėžtumą, tačiau juo apskaičiuoti veiksnių reikšmių įverčiai koreliuoja tarpusavyje net ir tada, kai veiksniai yra ortogonalūs. *Andersono ir Rubino* metodu apskaičiuoti veiksnių reikšmių įverčiai yra visiškai ortogonalūs, tačiau neminimizuoja neapibrėžtumo.

9.2 lentelė. Veiksnių reikšmių įverčių skaičiavimo metodų savybės

Metodas	**Didina patikimumą** (*validity*)	**Vienareikšmiškumas esant ortogonaliems faktoriams**	**Koreliacijos išsaugojimas**
Thurstone'o regresijos metodas (1935)	Taip	Ne	Ne
Berge, Krijneno, Wansbeeko, ir Shapiro (1999)	Ne	Ne	Taip
Andersono ir Rubino (1956)	Ne	Ne	Taip
Bartletto (1937)	Ne	Taip	Ne
Harmano (1976)	Ne	Taip	Ne

Vienareikšmiškumas esant ortogonaliems veiksniams parodo, ar veiksnių reikšmių įverčiai gali perimti dispersiją (*užsikrėsti ja*) iš kitų ortogonalių veiksnių. Koreliacijos išsaugojimas parodo, ar koreliacija tarp veiksnių reikšmių įverčių atitinka koreliaciją tarp pačių veiksnių. Nors šių metodų skaičiavimo algoritmai sudėtingi, juos galima rasti populiariose statistinėse programose.

Antrąją metodų grupę sudaro vadinamieji *netikslūs*, *apytiksliai* metodai. Jais skaičiuojant veiksnių reikšmes, pradinės kintamųjų reikšmės yra standartizuojamos, dauginamos iš atitinkamų kintamųjų veiksnių svorių, pateikiamų veiksnių svorių matricoje, ir šios sandaugos susumuojamos. Veiksnio reikšmės parodo objekto (valstybės, miesto ar atskiro asmens) sąsają su nagrinėjamu veiksniu. Tai sudėtinis matas, kuris gali būti apskaičiuotas kiekvienam tiriamam objektui. Kintamųjų, menkai susietų su konkrečiu veiksniu, svoris artimas nuliui. Šie veiksnių reikšmių koeficientai gali būti pavaizduoti grafiškai ir naudojami tolesnei analizei. Taip apskaičiuotos veiksnių reikšmės dažnai pasitaiko literatūroje ir vadinamos suminiais koeficientais (*total*), indeksais (*index*), domenais (*domain*), aspektais (*facet*), skalės (*scale*) ar subskalės (*subscale*) reikšmėmis. Šios *apytikslės* veiksnių reikšmės tarpusavyje koreliuoja net ir tada, kai veiksniai yra ortogonalūs, jų dispersija nelygi vienetui, ir beveik visada šios reikšmės silpniau koreliuos su veiksniais nei gautos vadinamaisiais tiksliais metodais.

Pagerinti metodai gali užtikrinti tam tikras statistines savybes: patikimumo didinimą ar veiksnių reikšmių įverčių ortogonalumą. Apytiksliais metodais įverčiai lengviau apskaičiuojami ir apskritai teigiama, kad jie stabilesni. Pradžioje buvo manoma, kad abiejų šių grupių metodai smarkiai siejasi tarpusavyje ir kad jų skirtumai nėra esminiai. Tačiau po gausių tyrimų pasirodė, kad šie skirtumai praktiškai gali būti reikšmingi. Kol kas yra neaišku, ar tiksliųjų metodų nestabilumas yra didelis. Todėl pasirinkti vieną iš skirtingų veiksnių reikšmių skaičiavimo procedūrų nėra taip paprasta. Tai turėtų priklausyti nuo konkrečios tyrimo programos konteksto.

Galutinis faktorinės analizės rezultatas – tai veiksnių įverčiai. Užuot kūrus žemėlapius su atskirais erdviniais kintamųjų pavyzdžiais, galima pateikti diagramą su nustatytu veiksnių reikšme. Kiekvieno veiksnio įvertis gali būti pateikiamas linijinėje pradinių kintamųjų diagramoje.

9.1. FAKTORINĖ ANALIZĖ PSPP IR SPSS TERPĖJE

Faktorinei analizei parinksime kintamuosius: respondentų amžius (*amžius*), respondentų pajamos per mėnesį (*Pajamos*), suma, paskirta per mėnesį pirkiniams (*Pirkinių_suma*), vidutinė kaina, kurią respondentas sutinka mokėti už kosmetiką (*vid_kaina_kosmetika*), vidutinė kaina, kurią respondentas sutinka mokėti už aprangą (*vid_kaina_apranga*), vidutinė kaina, kurią respondentas sutinka mokėti už buitinę techniką (*vid_kaina_techn*).

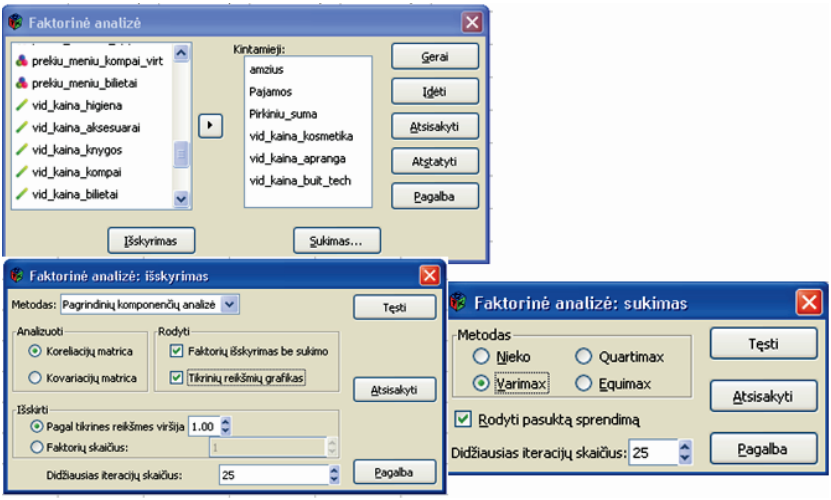
PSPP terpė

Faktorinės analizės dialogo langas *Faktorinė analizė* atveriamas komanda *Analizuoti → Faktorinė analizė*. Analizuojami kintamieji perkeliami į langelį *Kintamieji.*

Mygtuko *Išskyrimas* paspaudimas atveria dialogo langą *Faktorinė analizė: išskyrimas* (žr. 9.1 pav.). Šiame lange reikia pažymėti langelį: *Faktorių išskirimas be sukimo.* Pagal nutylėjimą nustatytas veiksnių nustatymo metodas (*Metodas → Pagrindinių komponenčių analizė*) geriausiai tinka daugumai rinkos tyrimų užduočių spręsti. Veiksnių skaičius nustatomas srityje *Išskirti.* Pagal nutylėjimą nustatomas veiksnių skaičiaus apskaičiavimas pagal tikrines reikšmes (*Pagal tikrines reikšmes viršija).* Tačiau galima nustatyti tam tikrą veiksnių skaičių (*Faktorių skaičius).* Langelio *Tikrinių reikšmių grafikas* žymėjimas leidžia nubraižyti tikrinių įverčių diagramą. Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą.

Mygtuko *Sukimas* paspaudimas atveria dialogo langą *Faktorinė analizė: sukimas* (žr. 9.1 pav.). Šiame lange reikia nustatyti koreliacijos koeficientų matricos sukimo būdą (dažniausiai pasirenkamas *Varimax* metodas). Langelio *Rodyti pasuktą sprendimą* žymėjimas leidžia pateikti pasuktą veiksnių matricą. Paspaudę *Tęsti* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai.

Kaizerio–Mejerio–Olkino mato KMO ir Bartletto sferiškumo kriterijaus PSPP paketas neskaičiuoja.



9.1 pav. Faktorinės analizės dialogo langai

Lentelėje *Visa paaiškinama dispersija* pateikiami standartiniai pagrindinių komponenčių (veiksnių) analizės rezultatai (žr. 9.2 pav.). Įvedamos naujos komponentės kaskart vis mažiau ir mažiau paaiškina dispersiją. Visa matricos dispersija priklauso nuo kintamųjų skaičiaus.

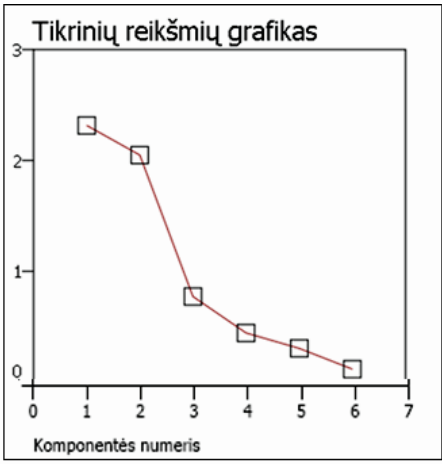
Visa paaiškinama dispersija

	Pradinės tikrinės reikšmės			*Pradinė svorių kvadratų suma*			*Svorių kvadratų suma po sukimo*		
Komponentė	*Iš viso*	*% dispersijos*	*Sukaupta %*	*Iš viso*	*% dispersijos*	*Sukaupta %*	*Iš viso*	*% dispersijos*	*Sukaupta %*
1	2,32	38,60	38,60	2,32	38,60	38,60	2,29	38,18	38,18
2	2,05	34,14	72,75	2,05	34,14	72,75	2,07	34,56	72,75
3	,77	12,86	85,61						
4	,44	7,38	92,99						
5	,30	5,06	98,06						
6	,12	1,94	100,00						

9.2 pav. Faktorinės analizės rezultatas PSPP terpėje

Pavyzdyje nagrinėjami šeši kintamieji, taigi visa dispersija gali būti išgauta per šešis kartus. Stulpelyje *Pradinės tikrinės reikšmės* pateiktos tikrinės reikšmės (*Iš viso)* parodo kiekvieno vertinamo veiksnio dispersiją. Stulpelyje *% dispersijos* yra apskaičiuoti bendros dispersijos procentai. Matome, kad pirmojo veiksnio tikrinė reikšmė yra 2,32. Kadangi ši reikšmė didesnė už 1,0 tai reiškia, kad veiksnys paaiškina daugiau dispersijos

nei atskiras kintamasis, jei tiksliau, 2,32 karto daugiau. Skaičiuodami procentinę išraišką, gautume, kad pirmasis veiksnys paaiškina 2,32/6 = 0,386 arba 38,60 proc. dispersijos. Antrojo veiksnio tikrinė reikšmė yra 2,05 ir ji irgi didesnė už vienetą. Kitų veiksnių tikrinės reikšmės mažesnės už vienetą ir jie paaiškina mažiau dispersijos nei atskiri kintamieji. Stulpelyje *Sukaupta %* pateikta sukauptoji dispersija. Pirmieji du veiksniai paaiškina 72,75 proc. visos šešių veiksnių dispersijos.



9.3 pav. Tikrinių verčių diagrama PSPP terpėje

Būtina nutarti, kokį veiksnių skaičių reikėtų nagrinėti. Iš esmės tai sutartinis pasirinkimas, tačiau pagal *Kaiserio* kriterijų siūloma nagrinėti tuos veiksnius, kurių tikrinės vertės viršija vienetą. Pagal *Cattello scree* kriterijų (žr. 9.3 pav.) reikėtų žiūrėti, kur nusistovi tikrinių verčių lygis. Abiem atvejais priimtiniausia – du veiksniai.

Lentelėje *Komponenčių (faktorių svorių) matrica* (žr. 9.4 pav.) pateikti nepasuktų veiksnių svoriai, t. y. kintamųjų ir veiksnių koreliacijos koeficientai, kurie rodo tarpusavio ryšių stiprumą ir kryptį. Pirmas kintamasis vidutiniškai koreliuoja su antru veiksniu, bet nekoreliuoja su pirmu veiksniu. Antras kintamasis labai gerai koreliuoja su antru veiksniu, bet nekoreliuoja su pirmu veiksniu ir t. t. Lentelėje *Pasukta komponenčių matrica* pateikti veiksnių svoriai, gauti pritaikius *Varimax* sukimo metodą. Galima pastebėti, kad šiuo atveju koreliaciniai ryšiai

geriau išryškinti. Po sukimo matome, kad veiksnys F_1 turi stiprų ryšį su vidutine kaina, kurią respondentas sutinka mokėti už kosmetiką (*vid_kaina_kosmetika*), vidutine kaina, kurią respondentas sutinka mokėti už aprangą (*vid_kaina_apranga*), vidutine kaina, kurią respondentas sutinka mokėti už buitinę techniką (*vid_kaina_techn*). Veiksnys F_2 turi stiprų ryšį su respondentų amžiumi (a*mžius*), respondentų pajamomis per mėnesį (*Pajamos*), suma, paskirta per mėnesį pirkiniams (*Pirkinių_suma*). Veiksnį F_1 galima apibūdinti kaip *tinkamų kainų pasirinkimą*, o veiksnį F_2 – kaip *asmenines respondento savybes.*

Komponenčių (faktorių svorių) matrica

	Komponentė	
	1	2
amzius	,23	,64
Pajamos	,22	,85
Pirkiniu_suma	,34	,80
vid_kaina_kosmetika	,89	-,32
vid_kaina_apranga	,83	-,42
vid_kaina_buit_tech	,79	,03

Pasukta komponenčių matrica

	Komponentė	
	1	2
amzius	,02	,69
Pajamos	-,05	,88
Pirkiniu_suma	,08	,86
vid_kaina_kosmetika	,95	-,03
vid_kaina_apranga	,92	-,14
vid_kaina_buit_tech	,74	,27

9.4 pav. Nepasukta ir pasukta veiksnių svorių matrica PSPP terpėje

SPSS terpė

Faktorinės analizės dialogo langas *Factor Analysis* atveriamas komanda *Analyze →Dimension Reduction → Factor*. Analizuojami kintamieji perkeliami į langelį *Variables.* Mygtuko *Descriptives* paspaudimas atveria dialogo langą *Factor Analysis:Descriptives.* Šiame lange reikia pažymėti langelį *KMO and Bartlett's tests of sphericity.* Šis kriterijus leidžia nustatyti kintamųjų tinkamumą faktorinei analizei. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Mygtuko *Extraction* paspaudimas atveria dialogo langą *Factor Analysis:Extraction*. Pagal nutylėjimą nustatytas veiksnių nustatymo metodas (*Method → Principal components*) geriausiai tinka daugumos rinkos tyrimų užduočių sprendimui. Veiksnių skaičius nustatomas srityje *Extract.* Pagal nutylėjimą nustatomas veiksnių skaičiaus apskaičiavimas pagal tikrines reikšmes (*Based on Eigenvalue)*. Tačiau galima nustatyti tam tikrą veiksnių skaičių (*Fixed number of factor).* Langelio *Scree plot* žymėjimas leidžia nubraižyti tikrinių įverčių (*Scree plot*) diagramą. Paspaudę *Continue* mygtuką grįžtame į pradinį langą.

Mygtuko *Rotation* paspaudimas atveria dialogo langą *Factor Analysis:Rotation*. Šiame lange reikia nustatyti koreliacijos koeficientų matricos sukimo būdą (dažniausiai pasirenkamas *Varimax* metodas). Paspaudę *Continue* mygtuką grįžtame į pradinį langą.

Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai.

Kaip duomenys (kintamieji) tinka faktorinei analizei, tikrinama Kaizerio–Mejerio–Olkino matu KMO ir Bartletto sferiškumo kriterijumi (žr. 9.5 pav.). Lentelėje *KMO and Bartlet's Test KMO = 0,618* reikšmė rodo, kad duomenys faktorinei analizei tinka patenkinamai. Bartletto sferiškumo kriterijus (*Bartlet's test of signification*) parodo, ar tarp analizuojamų kintamųjų yra statistiškai reikšmingai koreliuojančių, t. y. tikrina tokią hipotezę:

H_0*: visi kintamieji yra nekoreliuoti;*

H_1*: kintamieji koreliuoti.*

Pasirenkame reikšmingumo lygmenį $\alpha = 0{,}05$, $p - level = 0{,}000 < \alpha$. Patvirtinta hipotezė H_1, Turimiems duomenims faktorinė analizė turi prasmę.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,618
Bartlett's Test of Sphericity	Approx. Chi-Square	134,728
	df	15
	Sig.	,000

9.5 pav. Faktorinės analizės rezultatas SPSS terpėje

Total Variance Explained

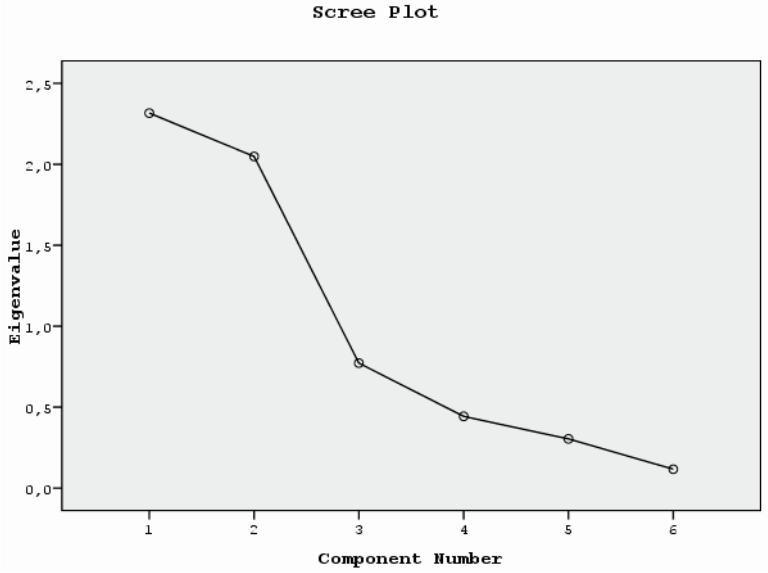
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,316	38,603	38,603	2,316	38,603	38,603	2,291	38,182	38,182
2	2,049	34,142	72,745	2,049	34,142	72,745	2,074	34,563	72,745
3	,772	12,865	85,610						
4	,443	7,384	92,994						
5	,304	5,062	98,057						
6	,117	1,943	100,000						

Extraction Method: Principal Component Analysis.

9.6 pav. Faktorinės analizės rezultatas SPSS terpėje

Lentelėje *Total Variance Explained* pateikiami standartiniai pagrindinių komponenčių (veiksnių) analizės rezultatai (žr. 9.6 pav.).

Įvedamos naujos komponentės paaiškina kaskart vis mažiau ir mažiau dispersijos. Visa matricos dispersija priklauso nuo kintamųjų skaičiaus.



9.7 pav. Tikrinių verčių diagrama (*Cattell scree* kriterijus) SPSS terpėje

Pavyzdyje nagrinėjami šeši kintamieji, taigi visa dispersija gali būti išgauta per šešis kartus. Stulpelyje *Initial Eigenvalues* pateiktos tikrinės reikšmės (*Total*) parodo kiekvieno vertinamo veiksnio dispersiją. Stulpelyje *% of Variance* yra apskaičiuoti bendros dispersijos procentai. Matome, kad pirmojo veiksnio tikrinė reikšmė yra 2,316. Kadangi ši reikšmė didesnė už 1,0, tai reiškia, kad veiksnys paaiškina daugiau dispersijos nei atskiras kintamasis, jei tiksliau, 2,316 karto daugiau. Skaičiuodami procentinę išraišką, gautume, kad pirmasis veiksnys paaiškina 2,316/6 = 0,38603, arba 38,603 proc. dispersijos. Antrojo veiksnio tikrinė reikšmė yra 2,049 ir ji irgi didesnė už vienetą. Kitų veiksnių tikrinės reikšmės mažesnės už vienetą ir jie paaiškina mažiau dispersijos nei atskiri kintamieji. Stulpelyje *Cumulative %* pateikta sukauptoji dispersija. Pirmieji du veiksniai paaiškina 72,745 proc. visos šešių veiksnių dispersijos. Būtina nutarti, kokį veiksnių skaičių reikėtų nagrinėti. Iš esmės tai sutartinis pasirinkimas, tačiau pagal *Kaiserio* kriterijų siūloma nagrinėti tuos veiksnius, kurių tikrinės vertės viršija vienetą. Pagal

Cattello scree kriterijų (žr. 9.7 pav.), reikėtų žiūrėti, kur tikrinių verčių lygis nusistovi. Abiem atvejais priimtiniausias skaičius – du veiksniai.

Lentelėje *Component Matrix* (žr. 9.8 pav.) pateikti nepasuktų veiksnių svoriai, t. y. kintamųjų ir veiksnių koreliacijos koeficientai, kurie rodo tarpusavio ryšių stiprumą ir kryptį. Pirmas kintamasis vidutiniškai koreliuoja su antru veiksniu, tačiau nekoreliuoja su pirmu veiksniu. Antras kintamasis labai gerai koreliuoja su antru veiksniu, tačiau nekoreliuoja su pirmu veiksniu ir t. t. Lentelėje *Rotated Component Matrix* pateikti veiksnių svoriai, gauti pritaikius *Varimax* sukimo metodą. Galima pastebėti, kad šiuo atveju koreliaciniai ryšiai geriau išryškinti. Po sukimo matome, kad veiksnys F_1 turi stiprų ryšį su vidutine kaina, kurią respondentas sutinka mokėti už kosmetiką (*vid_kaina_kosmetika*), vidutine kaina, kurią respondentas sutinka mokėti už aprangą (*vid_kaina_apranga*), vidutine kaina, kurią respondentas sutinka mokėti už buitinę techniką (*vid_kaina_techn*). Veiksnys F_2 turi stiprų ryšį su respondentų amžiumi (a*mžius*), respondentų pajamomis per mėnesį (*Pajamos*), suma, paskirta per mėnesį pirkiniams (*Pirkinių_suma*). Veiksnį F_1 galima apibūdinti kaip *tinkamų kainų pasirinkimą*, o veiksnį F_2 – kaip *asmenines respondento savybes.*

Component Matrix[a]

	Component	
	1	2
amzius	,232	,645
Pajamos	,223	,848
Pirkiniu_suma	,344	,796
vid_kaina_kosmetika	,889	-,322
vid_kaina_apranga	,828	-,419
vid_kaina_buit_tech	,786	,026

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
amzius	,023	,685
Pajamos	-,048	,875
Pirkiniu_suma	,083	,864
vid_kaina_kosmetika	,945	-,034
vid_kaina_apranga	,917	-,144
vid_kaina_buit_tech	,740	,266

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

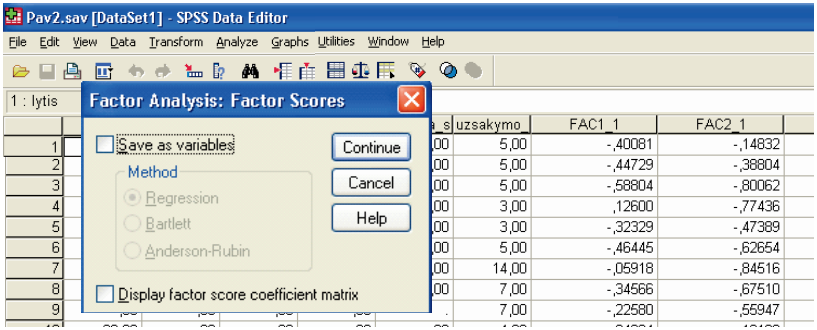
9.8 pav. Nepasukta ir pasukta veiksnių svorių matrica

Žinant veiksnių svorius, galima apskaičiuoti, kiek kiekvieno kintamojo dispersijas paaiškina parinkti veiksniai.

Kiekvienas veiksnys – tai yra naujas kintamasis. Šių kintamųjų reikšmės galima išsaugoti. Dialogo lange *Factor Analysis* reikia paspausti mygtuką *Scores,* atsiveria dialogo langas *Factor Anglysis: Factor Scores* (žr. 9.9 pav.). Šiame lange reikia pažymėti *Save as variables*

langelį. Kiekvieno veiksnio (faktoriaus) reikšmės bus išsaugotos (kintamieji FAC1_1 ir FAC2_1) SPSS duomenų lange. Gauti kintamieji yra kiekybiniai, jie tinka tolesnei statistinei analizei.

Pastaba: PSPP paketas veiksnių reikšmių išsaugojimo galimybės nesuteikia.



9.9 pav. Veiksnių reikšmių išsaugojimas SPSS terpėje

Atliktos procedūros sintaksė pateikta 9.3 lentelėje.

9.3 lentelė. Faktorinės analizės sintaksė

PSPP	SPSS
FACTOR FACTOR VARIABLES=amzius Pajamos Pirkiniu_suma vid_kaina_kosmetika vid_kaina_apranga vid_kaina_buit_tech /CRITERIA=MINEIGEN (1) ITERATE (25) /EXTRACTION=PC /METHOD=CORRELATION /PLOT=EIGEN /PRINT=INITIAL EXTRACTION ROTATION /CRITERIA=ITERATE (25) /ROTATION=VARIMAX.	FACTOR /VARIABLES amzius Pajamos Pirkiniu_suma vid_kaina_kosmetika vid_kaina_apranga vid_kaina_buit_tech /MISSING LISTWISE /ANALYSIS amzius Pajamos Pirkiniu_suma vid_kaina_kosmetika vid_kaina_apranga vid_kaina_buit_tech /PRINT INITIAL KMO EXTRACTION ROTATION FSCORE /PLOT EIGEN ROTATION /CRITERIA MINEIGEN(1) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /METHOD=CORRELATION.

KONTROLINIAI KLAUSIMAI

9.1. Kokiais atvejais taikoma faktorinė analizė?
9.2. Kokie uždaviniai sprendžiami faktorinės analizės metodais?
9.3. Koks faktorinės analizės tikslas?
9.4. Kokie kintamieji vadinami veiksniais?
9.5. Koks kintamųjų ir stebėjimų skaičius tinkamas faktorinei analizei?
9.6. Išvardykite faktorinės analizės prielaidas.
9.7. Kaip nustatomas veiksnių skaičius?
9.8. Kam reikalingi sukimo metodai?
9.9. Kaip nustatomas duomenų tikimas faktorinei analizei?

UŽDUOTYS

9.1. Atliekant faktorinę analizę SPSS terpėje gautas KMO ir Bartletto sferiškumo kriterijų skaičiavimo rezultatas (žr. 9.10 pav.). Ar faktorinė analizė turi prasmę?

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,383
Bartlett's Test of Sphericity	Approx. Chi-Square	176,580
	df	21
	Sig.	,000

9.10 pav. Faktorinės analizės rezultatas

9.2. Atliekant faktorinę analizę gautas pagrindinių komponenčių (veiksnių) analizės rezultatas (žr. 9.11 pav.). Nustatykite, kiek veiksnių reikia pasirinkti ir kokią bendros dispersijos dalį jie paaiškina.

Visa paaiškinama dispersija

Komponentė	*Pradinės tikrinės reikšmės*			*Pradinė svorių kvadratų suma*			*Svorių kvadratų suma po sukimo*		
	Iš viso	*% dispersijos*	*Sukaupta %*	*Iš viso*	*% dispersijos*	*Sukaupta %*	*Iš viso*	*% dispersijos*	*Sukaupta %*
1	3,24	46,29	46,29	3,24	46,29	46,29	2,48	35,43	35,43
2	1,64	23,49	69,77	1,64	23,49	69,77	2,03	28,93	64,36
3	1,17	16,79	86,56	1,17	16,79	86,56	1,55	22,20	86,56
4	,88	12,51	99,07						
5	,05	,69	99,76						
6	,01	,20	99,96						
7	,00	,04	100,00						

9.11 pav. Faktorinės analizės rezultatas

9.3. Atliekant gyvybės draudimo paslaugų tyrimą faktorinei analizei buvo pasirinkti kintamieji: *draudimo sutarties laikotarpis, mėnesinė draudimo įmoka, gyvybės draudimo suma, pajamos per mėnesį, apdraustų šeimos narių skaičius.* Analizės rezultatai pateikti 9.12 ir 9.13 pav. Paaiškinkite gautus rezultatus.

Visa paaiškinama dispersija

Komponentė	Pradinės tikrinės reikšmės			Pradinė svorių kvadratų suma			Svorių kvadratų suma po sukimo		
	Iš viso	% dispersijos	Sukaupta %	Iš viso	% dispersijos	Sukaupta %	Iš viso	% dispersijos	Sukaupta %
1	2,29	45,75	45,75	2,29	45,75	45,75	1,27	25,33	25,33
2	1,24	24,79	70,55	1,24	24,79	70,55	1,07	21,38	46,71
3	,86	17,22	87,77						
4	,56	11,28	99,04						
5	,05	,96	100,00						

9.12 pav. Faktorinės analizės rezultatas

Pasukta komponenčių matrica

	Komponentė	
	1	2
Draudimo sutarties laikotarpis	,554	,645
Mėnesinė draudimo įmoka	,823	,169
Gyvybės draudimo suma	,477	-,542
Pajamos per mėnesį	-,300	,679
Apdraustų šeimos narių skaičius	,601	-,057

9.13 pav. Faktorinės analizės rezultatas. Pasukta veiksnių svorių matrica

10. KLAUSIMINŲ PATIKIMUMO ANALIZĖ

Klausimyno patikimumas (angl. *reliability*) yra suprantamas kaip koreliacija tarp gautų testo rezultatų ir hipotetinių idealių rezultatų. Klausimyno patikimumui įvertinti dažniausiai yra naudojamos šios pagrindinės charakteristikos:

- klausimyno skalės vidinis nuoseklumas, kuris remiasi atsakymų į atskirus klausimyną sudarančius klausimus koreliacija;
- vertinimo patikimumas, kuris remiasi koreliacija tarp dviejų (ar daugiau) ekspertų vertinimų.

Klausimyno skalės vidiniam nuoseklumui įvertinti dažniausiai yra naudojamas Kronbacho *alfa* (*Cronbach's alpha*) koeficientas, kuris įvertina, ar visi skalės klausimai pakankamai atspindi tiriamąjį dydį, ir leidžia patikslinti reikiamų klausimų skaičių skalėje. Jeigu atskirų klausimų dispersijų suma yra artima visos skalės dispersijai, vadinasi, atskiri klausimai tarpusavyje nekoreliuoja, t. y. jie neatspindi to paties dalyko. Šiuo atveju klausimyno skalė yra sudaryta iš atsitiktinių klausimų ir Kronbacho *alfa* koeficientas yra artimas 0. Jeigu visos skalės dispersija yra ženkliai didesnė už atskirų klausimų dispersijų sumą, vadinasi, atskiri klausimai tarpusavyje koreliuoja, t. y. jie atspindi tą patį dalyką. Šiuo atveju, Kronbacho *alfa* koeficientas yra artimas 1.

Kronbacho *alfa* koeficientas didėja didinant klausimų, sudarančių klausimyną, skaičių. Tai reiškia, kad klausimynai su didesniu klausimų skaičiumi yra patikimesni. Negalima lyginti pagal Kronbacho *alfa* koeficientą dviejų skalių su skirtingu elementų skaičiumi. Kronbacho *alfa* koeficientas apskaičiuojamas pagal formulę:

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} S_i^2}{S_p^2}\right), \qquad (10.1)$$

čia α – Kronbacho *alfa* koeficientas, k – elementų skaičius, S_i^2 – i-tojo elemento dispersija, S_p^2 – bendra dispersija.

Vertinimo patikimumas nustato vertinimo homogeniškumą ir taikomas vertintojų suderinamumo laipsniui nustatyti, kai du ar daugiau vertintojų taiko tuos pačius vertinimo kriterijus tiems patiems žmonėms ar reiškiniams vertinti. Duomenų lentelėje recenzuojamieji (vertinami) subjektai sudaro atskirus kintamuosius (stulpelius), o ekspertai — atskirus stebėjimus (eilutes). Kiekvienas ekspertas pagal nustatytus kriterijus suteikia rangą kiekvienam vertinamam subjektui. Ranginiams duomenims skaičiuojamas Kendalo konkordancijos koeficientas *W*. Konkordancijos koeficientas *W* kinta nuo 0 iki 1 ($0 < W < 1$); 0 reiškia visišką nesuderinamumą; 1 – visišką suderinamumą.

Jeigu nėra sutampančių reikšmių, tai konkordancijos koeficientas apskaičiuojamas pagal formulę:

$$W = \frac{12S^2}{m^2(k^3 - k)}, \tag{10.2}$$

čia *m* – ekspertų skaičius; *k* – ekspertizės objektų skaičius.

Nuokrypio nuo *rangų* vidurkio kvadratų suma lygi:

$$S^2 = \sum_{j=1}^{k}\left(\sum_{i=1}^{m} x_{ij} - a\right)^2 \tag{10.3}$$

Rangų sumų vidurkį *a* galima apskaičiuoti pagal formulę

$$a=0{,}5m\,(k + 1) \tag{10.4}$$

Jeigu vertinimuose yra sutampančių rangų, tai nuokrypio nuo rangų vidurkio kvadratų suma apskaičiuojama pagal formulę:

$$S^2 = \frac{m^2\left(k^3 - k\right) - m\sum_{l=1}^{r} T_l}{12} \tag{10.5}$$

čia *r* – eilučių, turinčių sutampančių rangų, skaičius; T_l – sutampančių rangų *l*-tojoje rangų eilutėje skaičius:

$$T_l = \sum_{q=1}^{u}\left(t_q^3 - t_q\right), \tag{10.6}$$

čia *t* – skaičius *q* tipo vienodų rangų grupių *l*-tojoje rangų eilutėje; *u* – susijusių rangų tipų skaičius eilutėje.

Konkordancijos koeficientas, kai vertinimuose yra sutampančių rangų, skaičiuojamas pagal kitą formulę:

$$W = \frac{12S^2}{m^2(k^3 - k) - m\sum_{l=1}^{r} T_l}. \qquad (10.7)$$

Sprendimui priimti taikomas χ^2 kriterijus:

$\chi^2 = mfW$ turi χ^2 skirstinį su $f = k - 1$ laisvės laipsniu.

Tikrinamos hipotezės:

H_0: konkordancijos koeficientas lygus nuliui,

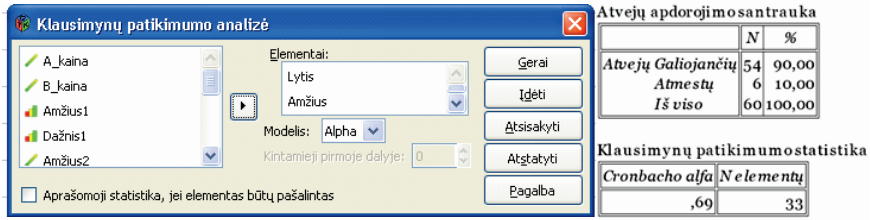
H_1: konkordancijos koeficientas nelygus nuliui.

H_0 hipotezė yra atmetama, kai apskaičiuotos statistikos χ^2 reikšmė yra didesnė už kritinę reikšmę arba stebimasis reikšmingumo lygmuo *p-level* < α (pasirenkamas patikimumo lygmuo).

10.1. KRONBACHO *ALFA* SKAIČIAVIMAS

PSPP terpė

Dialogo langas *Klausimynų patikimumo analizė* atidaromas komandomis *Analizuoti* → *Klausimynų patikimumas* (žr. 10.1 pav.). Analizuojamus kintamuosius reikia perkelti į langelį *Elementai,* langelyje *Modelis:* palikti nustatytą *Alpha.* Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 10.1 pav.).

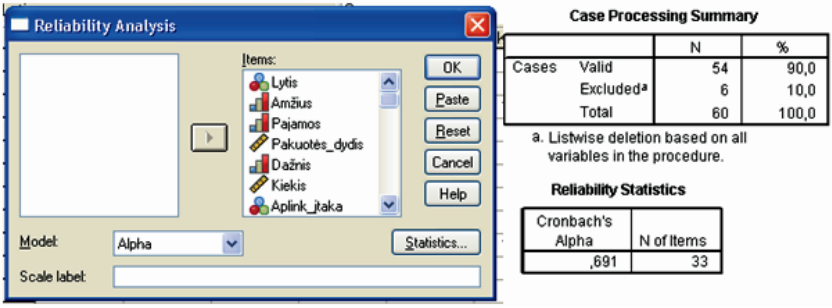


10.1 pav. Kronbacho *alfa* skaičiavimas PSPP terpėje

Lentelėje *Klausimynų patikimumo statistika pateikta* Kronbacho *alfa* reikšmė (*Cronbacho alfa)* $\alpha = 0{,}69$ rodo gerą klausimų suderinamumą.

SPSS terpė

Dialogo langas *Reliability Analysis* atidaromas komandomis *Analyze → Scale → Reliabiliti Analysis*. Analizuojamus kintamuosius reikia perkelti į langelį *Items:*, langelyje *Model:* palikti nustatytą *Alpha*. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 10.2 pav.).



10.2 pav. Kronbacho alfa skaičiavimas SPSS terpėje

Atliktos procedūros sintaksė pateikta 10.1 lentelėje.

10.1 lentelė. Kronbacho alfa skaičiavimo sintaksė

PSPP	SPSS
RELIABILITY /VARIABLES=Lytis Amžius Pajamos Pakuotės_dydis Dažnis Kiekis Aplink_įtaka Kita_įtaka Apsisprendimas Pirkimo_vieta Informacija R_A_vertinimas R_B_vertin R_C_vertin R_D_vertin R_E_vertin R_F_vertin R_G_vertin R_H_vertin R_I_vertin A_kaina B_kaina Amžius1 Dažnis1 Amžius2 Pajamos2 Pakuotė /MODEL=ALPHA.	RELIABILITY /VARIABLES=Lytis Amžius Pajamos Pakuotės_dydis Dažnis Kiekis Aplink_įtaka Kita_įtaka Apsisprendimas Pirkimo_vieta Informacija R_A_vertinimas R_B_vertin R_C_vertin R_D_vertin R_E_vertin R_F_vertin R_G_vertin R_H_vertin R_I_vertin A_kaina B_kaina Amžius1 Dažnis1 Amžius2 Pajamos2 Pakuotė /SCALE(,ALL VARIABLES') ALL/MODEL=ALPHA.

10.2. KONKORDANCIJOS KOEFICIENTO SKAIČIAVIMAS

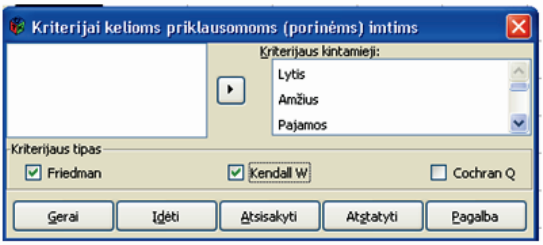
PSPP terpė

Dialogo langas *Kriterijai kelioms priklausomoms (porinėms) imtims* atidaromas komandomis *Analizuoti →Neparametrinė statistika → K priklausomų imčių* (žr. 10.3 pav.). Analizuojamus kintamuosius reikia perkelti į langelį *Kriterijaus kintamieji:* ir pažymėti langelį *Kendall W*. Procedūra baigiama mygtuko *Gerai* paspaudimu. Lange *Rezultatai – PSPPIRE Rezultatų peržiūra* rodomi analizės rezultatai (žr. 10.3 pav.).

Kendalo konkordancijos koeficiento reikšmė pateikta lentelėje *Kriterijaus statistika. W = 0,78* rodo gerą ekspertų nuomonių suderinamumą. *P-level (Asimpt. P-reikšmė) = 0,00 < α = 0,05,* patvirtinta hipotezė H_1, konkordancijos koeficientas reikšmingas.

SPSS terpė

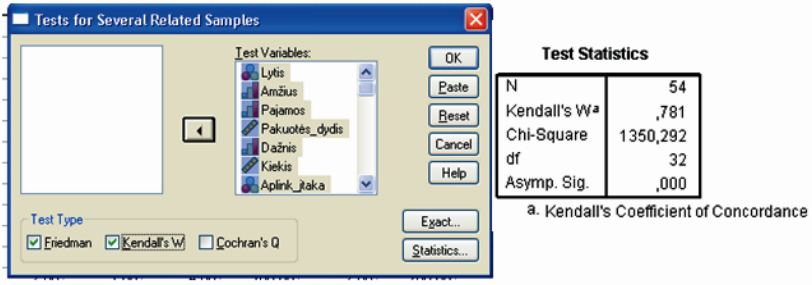
Dialogo langas *Tests for Several Related Samples* atidaromas komandomis *Analyze → Nonparametric Tests → K related Samples* (žr. 10.4 pav.). Analizuojamus kintamuosius reikia perkelti į langelį *Test Variables::* ir pažymėti langelį *Kendall's W* Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 10.3 pav.).



Kriterijaus statistika

N	54
Kendall W	,78
Chi-kvadratas	1350,29
skirt.	32
Asimt. p-reikšmė	,00

10.3 pav. Konkordancijos koeficiento *W* skaičiavimas PSPP terpėje



10.4 pav. Konkordancijos koeficiento *W* skaičiavimas SPSS terpėje

Kendalo konkordancijos koeficiento reikšmė pateikta lentelėje *Test Statistics*. $W = 0{,}78$ rodo gerą ekspertų nuomonių suderinamumą. *P-level (Asimp. Sig.)* $= 0{,}00 < \alpha = 0{,}05$, patvirtinta hipotezė H_1, konkordancijos koeficientas reikšmingas.

Atliktos procedūros sintaksė pateikta 10.2 lentelėje.

10.2 lentelė. Konkordancijos koeficiento *W* skaičiavimo sintaksė

PSPP	SPSS
NPAR TEST /KENDALL=Pakuotės_dydis Dažnis Aplink_įtaka Kita_įtaka Apsisprendimas Pirkimo_vieta Informacija R_A_vertinimas R_B_vertin R_C_vertin R_D_vertin R_E_vertin R_F_vertin R_G_vertin R_H_vertin R_I_vertin A_kaina B_kaina.	NPAR TESTS /KENDALL = Pakuotės_dydis Dažnis Aplink_įtaka Kita_įtaka Apsisprendimas Pirkimo_vieta Informacija R_A_vertinimas R_B_vertin R_C_vertin R_D_vertin R_E_vertin R_F_vertin R_G_vertin R_H_vertin R_I_vertin A_kaina B_kaina /MISSING LISTWISE.

KONTROLINIAI KLAUSIMAI

10.1. Kokiais atvejais taikomas Kronbacho alfa koeficientas?

10.2. Kokiais atvejais taikomas konkordancijos koeficientas?

10.3. Kaip tikrinamas konkordancijos koeficiento reikšmingumas?

11. LOGISTINĖ REGRESIJA IR *ROC* ANALIZĖ

11.1. LOGISTINĖ REGRESIJA

Tais atvejais, kai reikia nustatyti vieno priklausomo dichotominio ir vieno ar daugiau nepriklausomų kintamųjų ryšį, taikomas logistinės regresijos modelis. Logistine regresija naudojamasi tada, kai priklausomas kintamasis gali įgauti tik dvi reikšmes. Pavyzdžiui, pirkėjas gali pirkti prekę, bet gali jos nepirkti ir pan. Šiuo atveju netinka standartiniai regresijos modeliai, nes jie neįvertina, kad priklausomo kintamojo pobūdis yra dichotominis, todėl sudarytu modeliu galima prognozuoti reikšmes, didesnes už 1 ir mažesnes už 0. Tačiau regresijos modelį galima sudaryti taip, kad vietoje dvejetainio kintamojo bus prognozuojamas tolydusis kintamasis, kuris natūraliai kinta nuo 0 iki 1. Dažniausiai naudojama logistinė regresija (*logit regression*).

Logistinės regresijos modelis

Naudojant logistinę regresiją galima tirti ir prognozuoti priklausomo dichotominio kintamojo bei nepriklausomų kintamųjų, išmatuotų bet kuria skale (nepriklausomi kintamieji ir liekanos gali neturėti normaliojo skirstinio), ryšius. Prognozuojamos reikšmės niekada nebus didesnės (ir nebus lygios) už 1 ir nebus mažesnės (ir lygios) už 0. Tai pasiekiama taikant logistinės regresijos modelį:

$$y = \frac{\exp(a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}{1 + \exp(a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)} \qquad (11.1)$$

Akivaizdu, kad ir kokio dydžio būtų x reikšmės, modelio prognozuojamos y reikšmės visada bus nuo 0 iki 1. Kadangi priklausomas kintamasis y kinta nuo 0 iki 1, jis gali būti nagrinėjamas kaip tolydi tikimybė p, kuri kinta nuo 0 iki 1. Ši tikimybė gali būti transformuojama:

$$p' = \ln\left(\frac{p}{1-p}\right). \quad (11.2)$$

Ši transformacija vadinama logistine transformacija. Kadangi p' teoriškai gali įgyti bet kurią reikšmę nuo minus iki plius begalybės, tai šias transformuotas reikšmes jau galima naudoti įprastame tiesinės regresijos modelyje:

$$p' = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n + \text{e}. \quad (11.3.)$$

Pastaba: PSPP terpėje logistinės regresijos modelis nerealizuotas.

11.1.1. Logistinė regresija SPSS terpėje

Išnagrinėsime, kaip logistinė regresija gali būti taikoma binarinio kintamojo priklausomybei nuo kiekybinio kintamojo nustatyti. Šiuo atveju taikoma binarinė logistinė regresija. Nustatykime, kaip tikimybė pirkti gėles ar jų nepirkti priklauso nuo amžiaus. Dialogo langas *Logistic Regression* atveriamas komanda *Analyze →Regression → Binary Logistic*.... Priklausomas binarinis kintamąjį *Pirko* perkeliamas į langelį *Dependent,* kiekybinis kintamasis *Amžius* perkeliamas į langelį *Covariates*. Paspaudę *Continue* mygtuką grįžtame į pradinį langą. Procedūra baigiama mygtuko *OK* paspaudimu. Lange *Output-SPSS Viewer* rodomi analizės rezultatai (žr. 11.1 pav.).

Tikrinant, ar logistinės regresijos modelis tinka prognozėms, būtina patikrinti, ar dažnai žinomos reikšmės sutampa su reikšmėmis, kurias remdamiesi logistine regresija prognozuotume. Jeigu duomenų, iš kurių gauti parametrų įverčiai, prognozės netikslios, tada logistinės regresijos taikyti negalima, net jei kiti rodikliai ir statistinės išvados rodytų, kad ji tinkama.

Lentelėje *Omnibus Tests of Model Coefficients* pateiktas Pirsono χ^2 modelio suderinamumo kriterijus, kuris yra regresijos *ANOVA* analogas. Formuluojamos hipotezės:

H_0: $b = 0$,
H_1: $b \neq 0$.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24,110	1	,000
	Block	24,110	1	,000
	Model	24,110	1	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	104,546(a)	,228	,305

a Estimation terminated at iteration number 4 because parameter estimates changed by less than, 001.

Classification Table(a)

Observed			Predicted		
			Pirko		Percentage Correct
			ne	taip	
Step 1	Pirko	ne	32	12	72,7
		taip	16	33	67,3
	Overall Percentage				69,9

a The cut value is, 500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Amzius	,081	,019	18,898	1	,000	1,084
	Constant	-3,116	,759	16,852	1	,000	,044

a Variable(s) entered on step 1: Amzius.

11.1 pav. Pagrindiniai logistinės regresijos skaičiavimo rezultatai

Hipotezė H_0 atmetama, jeigu *p-level* < α. Hipotezė H_0 neatmetama jeigu *p-level* ≥ α, čia α — reikšmingumo lygmuo. Šio pavyzdžio *p-level* (*Sig.*) = *0,000* < *0,05*, todėl nulinę hipotezę atmetame.

Logistinė regresija turi net keletą tiesinės regresijos determinacijos koeficientų analogų. Visiems jiems galioja ta pati taisyklė: kuo didesnis koeficientas, tuo geriau logistinė regresija suderinta su duomenimis, nors patys koeficientai yra nelyginami dėl skirtingo jų apibrėžimo. Šie koeficientai *Cox & Snell R Square* pseudodeterminacijos koeficientas ir normuotas šio koeficiento variantas — *Nagelkerke R Square* pateikiami lentelėje *Model Summary*. Koeficientai rodo tą dispersijos dalį, kurią paaiškina logistinės regresijos modelis. Šioje lentelėje taip pat pateikiama dviguba tikėtinumo funkcijos logaritmo reikšmė *–2 Log likelihood.*

Lentelėje *Variables in the Equation* pateikiamos apskaičiuotų koeficientų *a* ir *b* reikšmės (*B* stulpelyje) ir jų nelygybės nuliui reikšmingumas. *Wald* kriterijus yra Stjudento kriterijaus tiesinėje regresijoje analogas, t. y. jis atsako į klausimą, ar konkretus koeficientas lygus nuliui. Formuluojamos hipotezės:

H_0: $b = 0$,

H_1: $b \neq 0$.

Hipotezė H_0 atmetama, jeigu *p-level* < α. Hipotezė H_0 neatmetama jeigu *p-level* ≥ α, čia α – reikšmingumo lygmuo. Visi šio pavyzdžio regresijos lygties koeficientai reikšmingai nelygūs nuliui, nes *p-level* (*Sig.*) < *0,05*. Šios lentelės *Exp(B)* stulpelyje pateikiama galimybių santykio reikšmė. Galimybių santykis rodo, kaip kinta Y galimybė įgyti reikšmę 1. Šiuo pavyzdžiu galime teigti, kad kiekvieni papildomi metai amžiaus padidina tikimybę pirkti gėles 1,084 karto.

Rezultatų interpretavimas

Gauti koeficientai interpretuojami taip pat, kaip ir tiesinės regresijos. Galima parašyti regresijos lygtį:

$$\hat{p}' = -3{,}116 + 0{,}081 \cdot amžius. \qquad (11.4)$$

Šios lygties parametrai susiję su $\ln\left(p/(1-p)\right)$, o ne su pirkimo tikimybėmis. Norint apskaičiuoti pirkimo tikimybę konkrečiu atveju, pritaikoma (6.12) formulė. Pavyzdžiui, kai amžius 30 metų, tikimybė pirkti apskaičiuojama taip:

$\hat{p}' = -3{,}116 + 0{,}081 \cdot 30 = -0{,}686;$

$\hat{p}'(30) = exp(-0{,}686)/(1 + exp(-0{,}686)) = (0{,}504/1{,}504) = 0{,}335.$

Atliktos procedūros sintaksė:

```
LOGISTIC REGRESSION VARIABLES Pirko
 /METHOD = ENTER Amžius
 /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

11.2. OPERATORIAUS CHARAKTERINGA KREIVĖ

Logistinė regresija leidžia įvertinti pasirinkto modelio parametrus. Tačiau dažnai reikia atlikti skirtingų modelių veiksmingumo lyginimą. Operatoriaus charakteringa kreivė (*ROC kreivė* – angl. *receiver operator characteristics*) – tai kreivė, kuri dažniausiai naudojama rezultatų tikslumo įvertinimui sprendžiant klasifikavimo į dvi grupes uždavinius, kai turima bandomoji imtis.

ROC kreivė rodo teisingai kvalifikuotų teigiamų rezultatų skaičiaus priklausomybę nuo neteisingai kvalifikuotų neigiamų rezultatų skaičiaus. Numatoma, kad yra parametras-klasifikatorius, kurio keitimas nustato skirtingą tyrimo rezultatų padalijimą į dvi grupes. Šis parametras vadinamas slenksčiu arba atskirties tašku (*cut-off value*). Nuo jo priklauso skirtingos pirmosios ir antrosios rūšies klaidos reikšmės.

Vertinant modelių veiksmingumą sudaroma tyrimo rezultatų klasifikavimo matrica (*confusion matrix*) (žr. 11.1 lentelę), kurioje lyginami modelio analizės rezultatai su tikrosiomis (faktinėmis) tyrimo reikšmėmis. Matricoje informacija apie tyrimo rezultatus užkoduota taip: „0“ – neigiamas rezultatas; „1“ – teigiamas rezultatas.

11.1 lentelė. Tyrimo rezultatų klasifikavimo matrica (*confusion matrix*)

	Tikrosios reikšmės	
Modelio rezultatai	**1**	**0**
1	TP	FP (II rūšies klaida)
0	FN (I rūšies klaida)	TN

- TP (*True Positives*) – teisingai kvalifikuoti teigiami rezultatai;
- TN (*True Negatives*) – teisingai kvalifikuoti neigiami rezultatai;
- FN (*False Negatives*) – teigiami rezultatai, kvalifikuoti kaip neigiami (I rūšies klaida);
- FP (*False Positives*) – neigiami rezultatai, kvalifikuoti kaip teigiami (II rūšies klaida).

Teigiamas ir neigiamas rezultatas nustatomas kiekvienam konkrečiam atvejui. Pavyzdžiui, prognozuojant prekės pirkimo galimybę, teigiamas rezultatas: „perka“, neigiamas: „neperka“.

Dažniausiai modelių veiksmingumui įvertinti skaičiuojami santykiniai koeficientai:

- teisingai kvalifikuotų teigiamų rezultatai dalis (*True Positives Rate*): $TPR = \frac{TP}{TP + FN}100\%$;
- neigiamų rezultatų, kvalifikuotų kaip teigiami, dalis (*False Positives*): $FPR = \frac{FP}{TN + FP}100\%$.

ROC kreivė yra jautrumo (*sensitivity*) ir (1-specifiškumas (*specificity*)) priklausomybės grafikas.

Jautrumas $Se = TPR$, specifiškumas $Sp = \frac{TN}{TN + FP}100\%$, FPR=100-Sp. Jautrumas atidedamas Y ašyje, (1 – specifiškumas) – X ašyje.

Nubraižius *ROC* grafikus galima vizualiai įvertinti ir palyginti kelių modelių veiksmingumą. Kuo *ROC* kreivė yra arčiau kairiojo viršutiniojo kampo, tuo geresnis modelis. Tačiau jeigu kelių modelių *ROC* kreivės persikerta ir yra labai arti viena kitos, vizualiai įvertinti modelių veiksmingumą sudėtinga. Tokiu atveju lyginamas plotas po *ROC* kreivėmis (AUC – angl. *area under curve*). Teoriškai jis yra nuo 0 iki 1, tačiau kadangi modelis laikomas prasmingu tik tada, kai jo kreivė yra aukščiau $y = x$ įstrižainės, paprastai modelio veiksmingumas vertinamas nuo 0,5 (bevertis modelis) iki 1 (idealus modelis). Šios reikšmės gali būti gautos įvertinus plotą figūros, kuri apačioje ir dešinėje ribojasi su grafiko kraštinėmis, o viršuje ir kairėje pusėje – su *ROC* kreive.

Rekomenduojama taikyti AUC reikšmių interpretavimą, pateiktą 11.2 lentelėje.

11.2 lentelė. AUC reikšmės

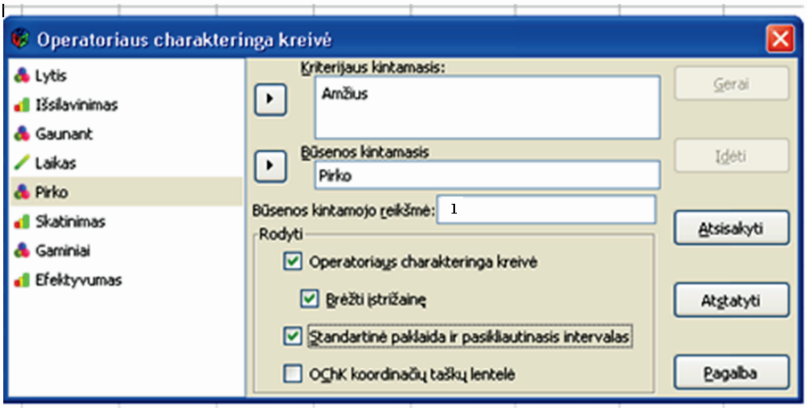
AUC	Modelio kokybė
0,9–1	puiki
0,8–0,9	labai gera
0,7–0,8	gera
0,6–0,7	patenkinama
0,5–0,6	nepatenkinama

Išnagrinėsime **reklaminės spaudos poveikio sprendimui pirkti reklamuojamą prekę priklausomybę nuo amžiaus ir nuo reklaminės spaudos skaitymo laiko**.

Tikrinamos hipotezės:
H_0: reklama nepaveikė pirkėjo (nepirko);
H_1: reklama paveikė pirkėją (pirko).

PSPP terpė

Komandos *Analizuoti* → *Operatoriaus charakteringa kreivė* atveria dialogo langą *Operatoriaus charakteringa kreivė* (žr. 11.2 pav.). Nepriklausomą kintamąjį *Amžius* reikia perkelti į langelį *Kriterijaus kintamasis:*, priklausomą kintamąjį *Pirko* – perkelti į langelį *Būsenos kintamasis.* Langelyje *Būsenos kintamojo reikšmė* reikia nurodyti teigiamą nagrinėjamo priklausomo kintamojo reikšmę (*1*).



11.2 pav. Dialogo langas *Operatoriaus charakteringa kreivė*

Analizės rezultatai pateikti 11.3 pav. ir 11.4 pav.

Plotas po kreive (Amžius)

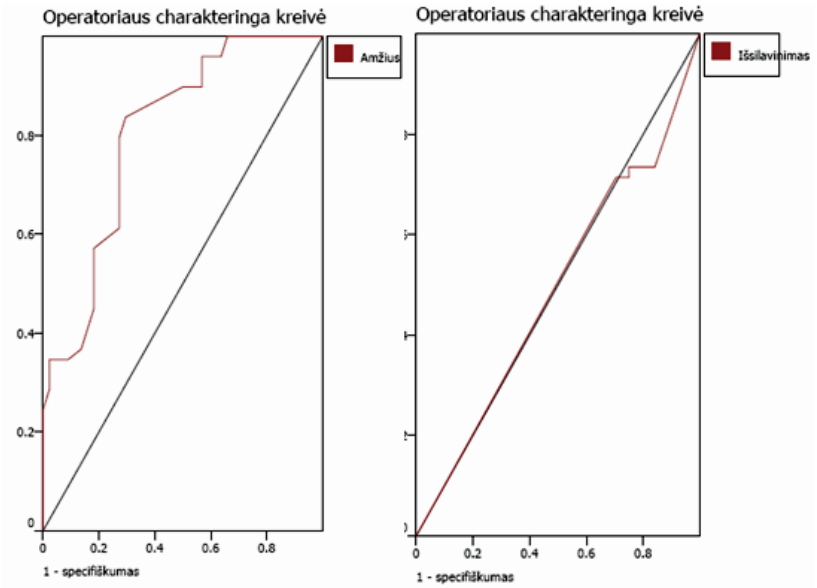
Plotas	Std. paklaida	Asimptotinė p-reikšmė	Asimp. 95% pasikliautinasis intervalas	
			Apačia	Viršus
,80	,05	,00	,73	,88

Plotas po kreive (Išsilavinimas)

Plotas	Std. paklaida	Asimptotinė p-reikšmė	Asimp. 95% pasikliautinasis intervalas	
			Apačia	Viršus
,51	,06	,87	,41	,61

11.3 pav. *ROC* analizės rezultatas PSPP terpėje

Lentelėse *Plotas po kreive* (žr. 11.3 pav.) pateiktos *AUC* reikšmės. Sprendimą dėl modelio tinkamumo galima priimti analizuojant *p-level (Asimptotinė p-reikšmė)* reikšmes. Galima teigti, kad tinkamas yra reklamos poveikio priklausomybės nuo amžiaus modelis (*p-level = 0,00 < α = 0,05).* Plotas po kreive (*Plotas = 0,80),* modelio kokybė labai gera. Reklamos poveikio priklausomybės nuo išsilavinimo modelis yra netinkamas. Analogišką išvadą galima padaryti ir analizuojant *ROC* grafikus (žr. 11.4 pav.).



11.4 pav. *ROC* kreivės PSPP terpėje

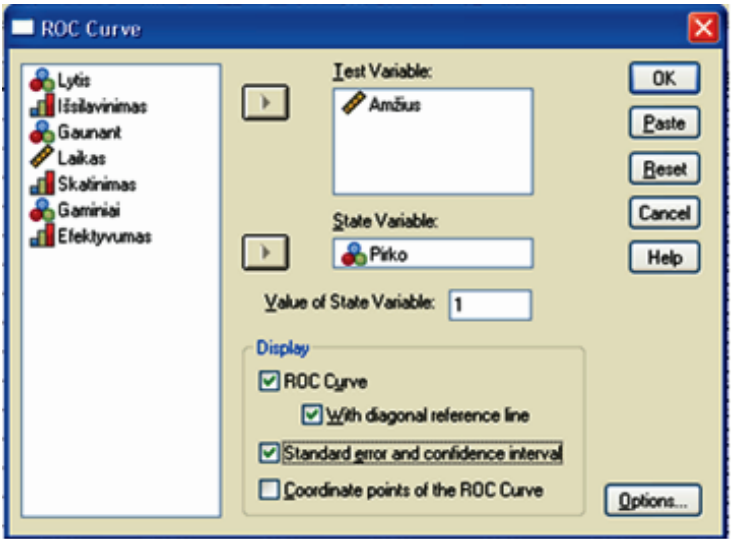
SPSS terpė

Komandos *Analyze → ROC Curve* atveria dialogo langą *ROC Curve* (žr. 11.5 pav.). Nepriklausomą kintamąjį *Amžius* reikia perkelti į langelį *Test Variable:,* priklausomą kintamąjį *Pirko* – perkelti į langelį *State Variable.* Langelyje *Value of State Variable:* reikia nurodyti teigiamą nagrinėjamo priklausomo kintamojo reikšmė (*1*).

Analizės rezultatai pateikti 11.6 pav. ir 11.7 pav.

Lentelėse *Area Under the Curve* (žr. 11.6 pav.) pateiktos *AUC* reikšmės. Sprendimą dėl modelio tinkamumo galima priimti analizuojant *p-level (Asymptotic Sig.(b))* reikšmes. Galima teigti, kad tinkamas yra reklamos poveikio priklausomybės nuo amžiaus modelis (*p-level = 0,000 < α=0,05).*

Plotas po kreive (*Area = 0,804),* modelio kokybė labai gera. Reklamos poveikio priklausomybės nuo išsilavinimo modelis yra netinkamas. Analogišką išvadą galima padaryti ir analizuojant *ROC* grafikus (žr. 11.7 pav.).



11.5 pav. Dialogo langas *ROC Curve*

Area Under the Curve

Test Result Variable(s): Amžius

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Upper Bound	Lower Bound
,804	,045	,000	,715	,893

The test result variable(s): Amžius has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.
a Under the nonparametric assumption
b Null hypothesis: true area = 0.5

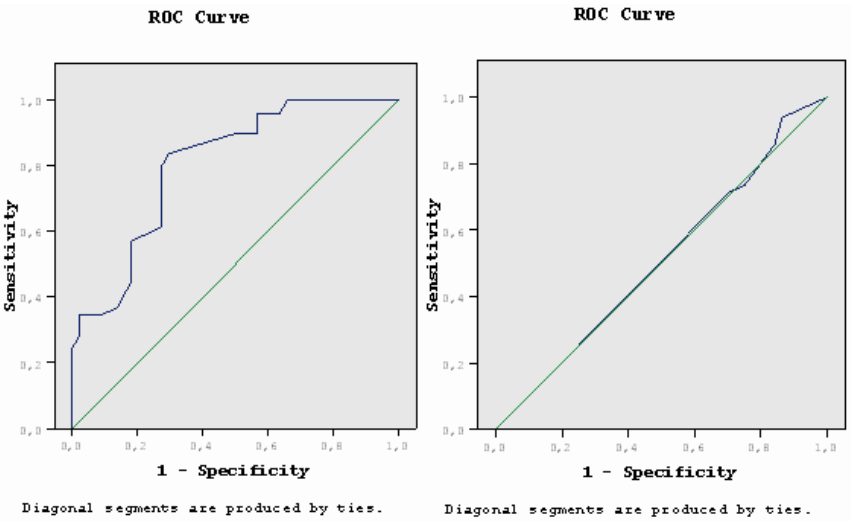
Area Under the Curve

Test Result Variable(s): Išsilavinimas

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Upper Bound	Lower Bound
,510	,060	,875	,391	,628

The test result variable(s): Išsilavinimas has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.
a Under the nonparametric assumption
b Null hypothesis: true area = 0.5

11.6 pav. *ROC* analizės rezultatas SPSS terpėje



11.7 pav. *ROC* kreivės SPSS terpėje

Atliktos procedūros sintaksė pateikta 11.3 lentelėje.

11.3 lentelė. ROC analizės sintaksė

PSPP	SPSS
ROC ROC Amžius BY Pirko(1) /PLOT CURVE (REFERENCE) /PRINTSE.	ROC Amžius BY Pirko (1) /PLOT = CURVE(REFERENCE) /PRINT = SE /CRITERIA = CUTOFF(INCLUDE) T E S T P O S (L A R G E) DISTRIBUTION(FREE) CI(95) /MISSING = EXCLUDE .

KONTROLINIAI KLAUSIMAI

11.1. Koks paprastos ir logistinės regresijos skirtumas?
11.2. Ar logistinės regresijos modelis tinka prognozei?
11.3. Kokiais atvejais taikoma logistinė regresija?
11.4. Koks tiesinės ir logistinės regresijos ryšys?
11.5. Koks operatoriaus charakteringos kreivės taikymo tikslas?
11.6. Kaip ROC analizėje nustatomas modelio tinkamumas?

Literatūra

1. Berkman, E. T. *A conceptual guide to statistics using SPSS*. Los Angeles (Calif.): Sage Publications, 2012, 296 p.
2. Bilevičienė, T.; Jonušauskas S. *Statistinių metodų taikymas rinkos tyrimuose*. Vadovėlis. Vilnius: Mykolo Romerio universiteto Leidybos centras, 2011, 307 p.
3. Blunch, N. J. *Introduction to structural equation modeling using SPSS and AMOS*. Los Angeles (Calif.): SAGE Publications, 2008, 270 p.
4. Brace, N.; Kemp, R.; Snelgar, R. *SPSS for psychologists: a guide to data analysis using SPSS for Windows: (version 12 and 13)*. 3rd ed. Basingstoke: Palgrave Macmillan, 2006, 450 p.
5. Buračas, A. *Statistikos terminai: anglų–lietuvių–prancūzų–rusų–vokiečių kalbos žodynas*. Vilnius: Mykolo Romerio universiteto Leidybos centras, 2008, 878 p.
6. Burns, R. A.; Burns, R. B. *Business research method and statistics using SPSS*. Los Angeles (Calif.); London: Sage Publications, 2008, 544 p.
7. Connolly, P. *Quantitative data analysis in education: a critical introduction using SPSS*. London; New York (N.Y.): Routledge: Taylor & Francis Group, 2006, 268 p.
8. Cooper, D. R. *Business Research Methods*. Boston (Mass.): McGraw-Hill, 2006, 744 p.
9. Curan, F. M. *Introduction to data analysis with R for forensic scientists*. Boca Raton (Fla.): CRC Press: Taylor & Francis Group, 2011, 309 p.
10. Curwin, J.; Slater R. *Quantitative methods for business decisions*. 6th ed. London: Thomson Learning, 2008, 790 p.
11. Čekanavičius, V.; Murauskas, G. *Statistika ir jos taikymai*. 1 tomas. Vilnius: TEV, 2003, 237 p.
12. Čekanavičius, V.; Murauskas, G. *Statistika ir jos taikymai*. 2 tomas. Vilnius: TEV, 2004, 272 p.
13. Dancey, C. P.; O'Riada, S. *Statistics without maths for psychology: using SPSS for Windows*. 3rd ed. Harlow: Pearson: Prentice Hall, 2004, 612 p.
14. Doane, D.; Seward, L. W. *Applied statistics in business and economics*. 3rd ed. New York (N.Y.): McGraw-Hill: Irwin, 2011, 831 p.
15. Europos Parlamento ir Tarybos reglamentas (EB) Nr. 295/2008, 2008 m. kovo 11 d. dėl verslo struktūros statistikos (nauja redakcija). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:097:0013:0059:LT:PDF>.

16. Faherty, V. E. *Compassionate statistics: applied quantitative analysis for social servises: with exercises and instructions in SPSS*. Los Angeles (Calif.): Sage Publications, 2008, 319 p.
17. Falcon, T. *Logit Models to Assess Credit Risk. Credit Risk Assessment Revisited: Methodological Issues and Practical Implications*. 2007, p. 28–47.
18. Fogel, K. F. *Producing Open Source Software: How to Run a Successful Free Software Project*. - O'Reilly Media, 2005, 304 p.
19. Gill, J.; Johnson, P. *Research methods for managers*. London: Sage Publication, 2010, 270 p.
20. Greasley, P. *Quantative data analysis using SPSS: an introduction for health and social science*. Maidenhead: Open University Press: McGrawHill, 2008, 138 p.
21. Howitt, D.; *Cramer, D. Introduction to SPSS in psychology: with supplements for releases 10, 11, 12 and 13. 3rd ed*. Harlow: Pearson: Prentice Hall, 2005, 237 p.
22. Howitt, D.; Cramer, D. *Introduction to SPSS in psychology: for version 16 and earlier*. 4th ed. Harlow: Pearson: Prentice Hall, 2008, 377 p.
23. Krzanowski, W. J. *Statistical principles and techniques in scientific and social investigations*. Oxford; New York (N.Y.): Oxford University Press, 2007, 241 p.
24. Leonavičienė, T. *SPSS programų paketo taikymas statistiniuose tyrimuose*. Vilnius: Vilniaus pedagoginio universiteto leidykla, 2006, 125 p.
25. Lietuvos Respublikos statistikos įstatymo pakeitimo įstatymas. 1999 m. gruodžio 23 d. Nr. VIII-1511 (*Žin*. 1999, Nr. 114-3299).
26. Market, opinion and social research - Vocabulary and service requirements. ISO 20252:2012.
27. McClave, J. T.; Benson, P. G.; Sincich, T. *Statistics for business and economics*. 10th ed. Upper Saddle River (N. J.): Pearson: Prentice Hall, 2008, 904 p.
28. Meier, K. J.; Brudney, J. L.; Bohte, J. *Applied statistics for public and nonprofit administration*. 7th ed. Belmont (Calif.): Wadsworth/Thomson Learning, 2009, 572 p.
29. Palalant, J. *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows*. Maidenhead: Open University Press, 2007, 335 p.
30. Pranulis, V. *Marketingo tyrimai: teorija ir praktika*. Vilnius: Vilniaus universiteto leidykla, 2007, 352 p.
31. Pukėnas, K. *Kokybinių duomenų analizė SPSS programa*. Mokomoji knyga. Kaunas, 2009, 94 p.
32. Rudzkienė, V. *Socialinė statistika*. Vadovėlis. Vilnius, 2005, 260 p.
33. Rudzkienė, V.; Burinskienė, M. *Plėtros krypčių vertinimo ir valdymo informaciniai modeliai*. Monografija. Vilnius: Technika, 2007, 406 p.
34. Sarantakos, S. *A toolkit for quantitative data analysis: using SPSS*. Basingstoke: Palgrave Macmillan, 2007, 118 p.
35. Treiman, D. J. *Quantitative data analysis: doing social research to test ideas*. San Francisco (Calif.): Jossey-Bass: A Wiley Imprint, 2009, 443 p.

36. Vaitkevičius, R.; Saudargienė, A. *Statistika su SPSS psichologiniuose tyrimuose*. Kaunas: Vytauto Didžiojo universiteto leidykla, 2006, 223 p.
37. Дуброва, Т. А. *Прогнозирование социально-экономических процессов*. 2-е изд., испр. и доп. Москва: МФПА, 2010, 188 с.
38. *Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS*. Москва, 2009, 308 с.
39. Моосмюллер, Г.; Ребик, Н. *Маркетинговые исследования с SPSS*. Москва: ИНФРА-М, 2009, 158 с.
40. Соколов, Г. А.; Сагитов, Р. В. *Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике*. Москва: ИНФРА-М, 2010, 200 с.
41. *Социальная статистика*. Под редакцией И. И. Елисеевой. Москва: Финансы и статистика, 2003, 480 с.
42. Таганов, Д. Н. *SPSS статистический анализ в маркетинговых исследованиях*. Санкт-Петербург: Питер, 2005, 190 с.

1 priedas

χ^2 SKIRSTINIO A LYGMENS KRITINĖS REIKŠMĖS

n/α	α = 0,10	α = 0,05	α = 0,025	α = 0,01
1	2,71	3,84	5,02	6,63
2	4,61	5,99	7,38	9,21
3	6,25	7,81	9,35	11,34
4	7,78	9,49	11,14	13,28
5	9,24	11,07	12,83	15,09
6	10,64	12,59	14,45	16,81
7	12,02	14,07	16,01	18,48
8	13,36	15,51	17,53	20,09
9	14,68	16,92	19,02	21,67
10	15,99	18,31	20,48	23,21
12	18,55	21,03	23,34	26,22
13	19,81	22,36	24,74	27,69
14	21,06	23,68	26,12	29,14
15	22,31	25,00	27,49	30,58
16	23,54	26,30	28,85	32,00
18	25,99	28,87	31,53	34,81
19	27,20	30,14	32,85	36,19
20	28,41	31,41	34,17	37,57
21	29,62	32,67	35,48	38,93
22	30,81	33,92	36,78	40,29
23	32,01	35,17	38,08	41,64
24	33,20	36,42	39,36	42,98
25	34,38	37,65	40,65	44,31
26	35,56	38,89	41,92	45,64
27	36,74	40,11	43,19	46,96
28	37,92	41,34	44,46	48,28

29	39,09	42,56	45,72	49,59
30	40,26	43,77	46,98	50,89
40	51,80	55,76	59,34	63,69
50	63,17	67,50	71,42	76,15
60	74,40	79,08	83,30	88,38
70	85,53	90,53	95,02	100,42
80	96,58	101,88	106,63	112,33
90	107,56	113,14	118,14	124,12
100	118,50	124,34	129,56	135,81

n – laisvės laipsnių skaičius
a – reikšmingumo lygmuo

2 priedas

STJUDENTO SKIRSTINIO α LYGMENS KRITINĖS REIKŠMĖS $t_{\alpha}(n)$

n/ α	α = 0,40	α = 0,10	α = 0,05	α = 0,025	α = 0,01	α = 0,005
1	0,325	3,078	6,314	12,706	31,821	63,657
2	0,289	1,886	2,920	4,303	6,965	9,925
3	0,277	1,638	2,353	3,182	4,541	5,841
4	0,271	1,533	2,132	2,776	3,747	4,604
5	0,267	1,476	2,015	2,571	3,365	4,032
6	0,265	1,440	1,943	2,447	3,143	3,707
7	0,263	1,415	1,895	2,365	2,998	3,499
8	0,262	1,397	1,860	2,306	2,896	3,355
9	0,261	1,383	1,833	2,262	2,821	3,250
10	0,260	1,372	1,812	2,228	2,764	3,169
11	0,260	1,363	1,796	2,201	2,718	3,106
12	0,259	1,356	1,782	2,179	2,681	3,055
13	0,259	1,350	1,771	2,160	2,650	3,012
14	0,258	1,345	1,761	2,145	2,624	2,977
15	0,258	1,341	1,753	2,131	2,602	2,947
16	0,258	1,337	1,746	2,120	2,583	2,921
17	0,257	1,333	1,740	2,110	2,567	2,898
18	0,257	1,330	1,734	2,101	2,552	2,878
19	0,257	1,328	1,729	2,093	2,539	2,861
20	0,257	1,325	1,725	2,086	2,528	2,845
21	0,257	1,323	1,721	2,080	2,518	2,831
22	0,256	1,321	1,717	2,074	2,508	2,819
23	0,256	1,319	1,714	2,069	2,500	2,807
24	0,256	1,318	1,711	2,064	2,492	2,797

25	0,256	1,316	1,708	2,060	2,485	2,787
30	0,256	1,310	1,697	2,042	2,457	2,750
40	0,255	1,303	1,684	2,021	2,423	2,704
60	0,254	1,296	1,671	2,000	2,390	2,660
120	0,254	1,289	1,658	1,980	2,358	2,617

n – laisvės laipsnių skaičius
a – reikšmingumo lygmuo

3 priedas

ATSAKYMAI

2.2 užduotis

n = 384.

4.1 užduotis

Vidurkis ir mediana padidės 100 vienetų, kitkas nesikeis.

4.2 užduotis

Prieš nustatant medianą, diametro reikšmės nebuvo išdėstytos didėjimo (mažėjimo) tvarka.

5.1 užduotis

Tikrinant respondentų amžiaus skirstinio normalumą gauti Kolmogorovo–Smirnovo ir Šapiro–Vilko kriterijų skaičiavimo rezultatai. Kriterijų reikšmingumo lygmuo *p-level (Sig.)* < $\alpha = 0,05$, kintamojo skirstinio skirtumas nuo normaliojo statistiškai reikšmingas. Šiam kintamajam negalima taikyti parametrinius kriterijus.

6.1 užduotis

Tikrinant priklausomybę pirkimo dažnio nuo lyties gautas rezultatas, pateiktas 6.31 pav. Suformuluokite hipotezes ir paaiškinkite gautą rezultatą.

Hipotezės:

H_0: pirkimo dažnis nepriklauso nuo lyties,

H_1: pirkimo dažnis priklauso nuo lyties.

Analizės rezultatai pateikti 6.31 pav. Lentelėje *Lytis*Dažnis1[kiekis]* pateikti kintamųjų dažniai. PSPP paketas neinformuoja, ar duomenys tinka analizei. Pirsono χ^2 kriterijus taikomas, kai bendras imties dydis $m \geq 30$ elementų bei 75 proc. porinių dažnių lentelės narvelių dažniai ≥ 5. Jei šios sąlygos netenkinamos, rezultatai gali būti nepatikimi. Nagrinėjami kintamieji tinkami analizei. Lentelėje *Chi-kvadrato kriterij.* pateikti Pirsono χ^2 suderinamumo kriterijaus skaičiavimo rezultatai (žr. 6.31 pav.). *P-level (Asimp p-reikšmė (2-pusė))* = $0,12 > \alpha = 0,05$, negalima atmesti hipotezės H_0, kad pirkimo dažnis nepriklauso nuo lyties.

Apskaičiavus χ^2 kriterijų nustatoma, ar tarp požymių yra statistinis ryšys. Lentelės dydis 2x3. Ryšio stiprumui nustatyti dėl *ranginių* kintamųjų taikomas *Cramer's V* testas. Lentelėje *Simetriniai matavimai* pateikta šio testo reikšmė (*Kramerio V = 0,26).*

6.2 užduotis

Hipotezės:

H_0: priimtinos jogurto kaina nepriklauso nuo amžiaus,

H_1: priimtinos jogurto kaina priklauso nuo amžiaus.

Analizės rezultatai pateikti 6.32 pav. Lentelėje *Lytis*Dažnis1[kiekis]* pateikti kintamųjų dažniai. PSPP paketas neinformuoja, ar duomenys tinka analizei. Pirsono χ^2 kriterijus taikomas, kai bendras imties dydis $m \geq 30$ elementų ir 75 proc. porinių

dažnių lentelės narvelių dažniai ≥ 5. Jei šios sąlygos netenkinamos, rezultatai gali būti nepatikimi. Nagrinėjami kintamieji tinkami analizei. Lentelėje *Chi-kvadrato kriterij.* pateikti Pirsono χ^2 suderinamumo kriterijaus skaičiavimo rezultatai (žr. 6.32 pav.). *P-level (Asimp p-reikšmė (2-pusė)) = 0,000 < α = 0,05,* pasitvirtina hipotezė H_1, kad priimtina jogurto kaina priklauso nuo amžiaus.

Eilutė *Dvitiesinė priklausomybė* leidžia padaryti išvadą dėl nenominalių kintamųjų tiesinio ryšio, *p-level = 0,000 < α = 0,05,* kintamųjų tiesinis ryšys reikšmingas.

Apskaičiavus χ^2 kriterijų nustatoma, ar tarp požymių yra statistinis ryšys. Abu kintamieji – kiekybiniai. *Kiekybinių* kintamųjų ryšio stiprumui nustatyti taikomas Pirsono koreliacijos koeficientas (*Pirsono R = 0,37,* ryšys silpnas), tačiau kintamųjų skirstinys nežinomas, todėl geriau taikyti Spirmeno koreliacijos koeficientą (*Spirmano koreliacija = 0,36,* ryšys silpnas). Matoma, kad šie koreliacijos koeficientai mažesni už *Kramerio V* koreliaciją.

6.3. užduotis

Lentelės dydis 2x5, todėl tam, kad būtų galima taikyti Pirsono χ^2 suderinamumo kriterijų, tik 25 proc. narvelių dažnis gali būti mažesnis kaip 5. Nagrinėjamu atveju tokių narvelių trys (30 proc.), Pirsono χ^2 suderinamumo kriterijų taikymas duos nepatikimą rezultatą.

6.4 užduotis

Galima apskaičiuoti:

a) amžiaus ir vidutinio pinigų kiekio, išleidžiamo per mėnesį prekei A, Pirsono koreliacijos koeficientą (kintamieji kiekybiniai, duomenų skirstinys yra normalusis);
b) respondentų vidutinių pajamų ir priimtinos prekės A kainos Spirmeno koreliacijos koeficientą (vidutinės pajamos – ranginis kintamasis, priimtina prekės A kaina – kiekybinis kintamasis, tačiau jo skirstinys neatitinka normalumo reikalavimų);
c) respondentų amžiaus ir priimtinos prekės A kainos Spirmeno koreliacijos koeficientą (amžius – kiekybinis kintamasis, priimtina prekės A kaina – kiekybinis kintamasis, tačiau jo skirstinys neatitinka normalumo reikalavimų).

6.5 užduotis

Statistinės hipotezės:

H_0: pajamų ir pirkimo dažnio koreliacijos koeficientas lygus nuliui,

H_1: pajamų ir pirkimo dažnio koreliacijos koeficientas nelygus nuliui.

Spirmeno koreliacijos koeficientas $r_s = -0,01$ rodo nereikšmingą pajamų ir pirkimo dažnio koreliaciją.

Hipotezės H_0 teisingumui patikrinti naudosime *t* kriterijų. Lentelėje pateikta t kriterijaus reikšmė (*Apytiksl. T*), $|t| = 0,14$.

Parenkamas reikšmingumo lygmuo α = 0,05. Hipotezė H_0 atmetama, jei absoliuti *t* reikšmė viršija Stjudento skirstinio su *(n–2)* laisvės laipsniais *α/2* lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

$t_{\alpha/2(n-2)} = t_{0,025/2(60-2)} = 2 > t = 0,14$. Hipotezės H_0 atmesti negalima, koreliacija nereikšminga.

6.6 užduotis

Statistinės hipotezės:

H_0: amžiaus ir pinigų, išleistų pirkiniams, koreliacijos koeficientas lygus nuliui,

H_1: amžiaus ir pinigų, išleistų pirkiniams, koreliacijos koeficientas nelygus nuliui.

Pirsono koreliacijos koeficientas $r = 0,66$ rodo vidutinę amžiaus ir pinigų, išleistų pirkiniams, koreliaciją.

Hipotezės H_0 teisingumui patikrinti naudosime t kriterijų. Lentelėje pateikta t kriterijaus reikšmė (*Apytiksl. T*), $t = 3,6$.

Parenkamas reikšmingumo lygmuo α = 0,05. Hipotezė H_0 atmetama, jei absoliuti t reikšmė viršija Stjudento skirstinio su $(n–2)$ laisvės laipsniais $\alpha/2$ lygmens kritinę reikšmę $t_{\alpha/2(n-2)}$.

$t_{\alpha/2(n-2)} = t_{0,025/2(51-2)} = 2 < t = 3,6$. Patvirtinta hipotezė H_1: amžiaus ir pinigų, išleistų pirkiniams, koreliacijos koeficientas nelygus nuliui, koreliacija reikšminga.

7.1 užduotis, 7.2 užduotis

Determinacijos koeficiento *R kvadratas* reikšmės rodo, kad išleidžiamų kavinėje pinigų priklausomybę nuo amžiaus galima aprašyti tiesinės regresijos modeliu (*R kvadratas >0,25*).

$y = a + bx$, arba *Pinigai = 12,41 – 0,08 Amžius,*

čia y – išleidžiami šampūnui pinigai, x – amžius.

Tiesinės regresijos modelio tinkamumo tikrinimas

Suformuluojame hipotezės:

H_0: $\alpha_1 = 0$,

H_1: $\alpha_1 \neq 0$.

Stebimasis reikšmingumo lygmuo *p – level = 0,02*, tai mažiau už pasirinktą patikimumo lygmenį $\alpha = 0,05$, pasitvirtino H_1, kad $\alpha_1 \neq 0$.

H_0: $\beta_1 = 0$,

H_1: $\beta_1 \neq 0$.

Stebimasis reikšmingumo lygmuo *p – level = 0,00*, tai mažiau už pasirinktą patikimumo lygmenį $\alpha = 0,05$, pasitvirtino H_1, kad $\beta_1 \neq 0$.

7.3 užduotis

Determinacijos koeficiento *R kvadratas* reikšmės rodo, kad tiesinės regresijos modelis netinka (*R kvadratas < 0,25).*

8.1 užduotis

Pirmoje lentelėje *Vienos imties statistika* pateikimas apskaičiuotas kintamojo vidurkis (*Vidurkis)* ir standartinis nuokrypis (*Std. nuokrypis)*. Lentelė *Kriterijus vienai imčiai* leidžia įvertinti statistiškai reikšmingus skirtumus. Išvada daroma vertinant stebimojo reikšmingumo lygmens *p-level* (*p-reikšmė (2-pusė)*) reikšmę. Jei *p-level* > α, tai kintamųjų vidurkių skirtumas statistiškai nereikšmingas. Nagrinėjamų atveju

p-level = 0,00 < α = 0,05, skirtumas 21 Lt (stulpelis *Vidurkių skirtumas*) yra statistiškai reikšmingas. Realiai per vieną kartą išleidžiamų pinigų vidurkis reikšmingai skiriasi nuo 30 Lt sumos, už kurią duoda akcijos lipduką.

8.2 užduotis

Pirmoje lentelėje *Grupių statistika* pateikti kintamojo *Pinigai* vidurkiai (*Vidurkis*) respondentų grupėms *>= 35 m. ir < 35 m*. Lentelė *Kriterijus nepriklausomoms imtims* leidžia nustatyti grupių statistinius skirtumus. Livyno kriterijus (*Levene kriterijus dispersijų lygybei*) taikomas hipotezės dėl analizuojamų kintamųjų dispersijų lygybei tikrinti. Jeigu stebimojo reikšmingumo lygmens *p-level* reikšmė (*p-reikšmė)* stulpelyje *Jei dispersijos būtų lygios* eilutėje rodo, kad rezultatas statistiškai nereikšmingas (*p-level>α,* čia *α* – pasirenkamas reikšmingumo lygmuo, pvz., *α = 0,05*) (nagrinėjamu atveju *p-level = 0,00*), tai kintamųjų dispersijos statistiškai lygios ir vidurkių skirtumas nustatomas iš eilutės *Jei dispersijos būtų lygios* stulpelio *p-reikšmė (2-pusė)*. Nagrinėjamu atveju Livyno kriterijus rodo statistinį reikšmingumą (*p-level< α),* dispersijos nelygios ir stulpelio *p-reikšmė (2-pusė)* reikšmė *0,03* (*p-level<α*) eilutėje *Jei dispersijos būtų nelygios* rodo, kad skirtingo amžiaus respondentų per vieną kartą pirkiniams išleidžiamų pinigų skirtumas statistiškai reikšmingas.

8.3 užduotis

Lentelėje *Dispersijų homogeniškumo kriterijus p-level* (stulpelis *Reikšmingumo lygmuo*) rodo Livyno kriterijaus (*Statistika (Levene)*) kriterijaus reikšmingumą. Jei *p-level > α = 0,05* (kriterijus nereikšmingas), dispersijos lygios. Nepriklausomų kintamųjų grupių skirtumų reikšmingumą rodo *F* kriterijaus *p-level* (žr. stulpelį *Reikšmingumo lygmuo*) lentelėje *ANOVA*. Jei *Livyno* kriterijus parodė dispersijų skirtumą, tai patikimumo lygmenį *α* reikia padidinti nuo *0,05* iki *0,01*. Nagrinėjamu atveju dispersijos nelygios, *p-level = 0,00 < α = 0,05*, skirtumas nereikšmingas.

9.1 užduotis

Duomenų (kintamųjų) tikimas faktorinei analizei tikrinamas Kaizerio–Mejerio–Olkino matu KMO ir Bartletto sferiškumo kriterijumi. Lentelėje *KMO and Bartlet's Test KMO = 0,383* reikšmė rodo, kad duomenys faktorinei analizei netinka. Bartleto sferiškumo kriterijus (*Bartlet's test of signification*) parodo, ar tarp analizuojamų kintamųjų yra statistiškai reikšmingai koreliuojančių, t. y. tikrina tokią hipotezę:

H_0*: visi kintamieji yra nekoreliuoti;*

H_1*: kintamieji koreliuoti.*

Pasirenkame reikšmingumo lygmenį $\alpha = 0,05$, $p - level = 0,000 < \alpha$. Patvirtinta hipotezė H_1,

9.2 užduotis

Galima pasirinkti tris veiksnius, kurie paaiškina 86,56 proc. bendros dispersijos.

9.3 užduotis

Galima pasirinkti du veiksnius, kurie paaiškina 70,55 proc. bendros dispersijos.

Pasukta veiksnių svorių matrica rodo, kad veiksnys F_1 turi stiprų ryšį su kintamaisiais *mėnesinė draudimo įmoka* ir *apdraustų šeimos narių skaičius*. Veiksnys F_2 turi stiprų ryšį su kintamaisiais *draudimo sutarties laikotarpis, gyvybės draudimo suma*.

Bilevičienė T., Jonušauskas S.

Bi238 ATVIROJO KODO PROGRAMŲ TAIKYMAS RINKOS TYRIMUOSE. Vadovėlis. – Vilnius: Mykolo Romerio universitetas, 2013. 302 p.

Bibliogr. 298–293 p.

ISBN 978-9955-19-513-9

Vadovėlyje pateikiama atvirojo kodo statistikos programų apžvalga ir rekomenduojamas tinkamo paketo pasirinkimas. Nagrinėjamas kiekybinių metodų taikymas rinkos tyrimų procesams: duomenų rinkimo, analizės ir interpretavimo metodologija. Glaustai, be sudėtingos matematikos nurodoma, kaip atlikti rinkos tyrimus, kokią statistiką naudoti pateikiant ir interpretuojant gautus rezultatus. Pateikiami tipinių uždavinių sprendimo būdai, statistinių išvadų pagrįstumo analizės ir rezultatų interpretavimo pavyzdžiai. Vadovėlyje naudojami studentų tyrimo darbų pavyzdžiai, kuriems spręsti pasirinkti skirtingi programų paketai: atvirojo kodo programų paketas PSPP – supaprastintas patentuoto ir plačiai žinomo statistikos programų paketo SPSS variantas; programų paketas SPSS (Statistical Package for the Social Science) (šio paketo licenciją turi nemažai universitetų, jį naudoja ir verslo įmonės), MS Excel. Nagrinėjama programinė įranga knygoje aprašoma tiek, kiek reikia, kad būtų galima išspręsti pagrindinius rinkos tyrimo uždavinius, t. y. vadovėlis nepretenduoja į išsamią PSPP, SPSS, MS Excel programinių paketų apžvalgą, juo labiau į išsamų jų mokymą. Turint omenyje, kad PSPP nerealizuoja visų SPSS taikomų statistinės analizės metodų, autoriai nusprendė atskleisti tik bendras šios programinės įrangos galimybes.

Vadovėlis skirtas Mykolo Romerio universiteto studentams bei dėstytojams, atliekantiems statistinius tyrimus. Pateikta medžiaga gali būti naudojama dėstant ir kitus dalykus, susijusius su statistiniais tyrimais, taip pat atliekant bakalauro bei magistro darbo tyrimus, rašant mokslinius straipsnius ir disertacijas. Dėstoma medžiaga bus naudinga ir kitų aukštųjų mokyklų studentams, ypač būsimiems vadybos, ekonomikos ir valdymo specialistams, taip pat rinkodaros specialistams, kurie norės atlikti rinkos tyrimus.

UDK 004.4:339.13(075.8)

Tatjana Bilevičienė, Steponas Jonušauskas

ATVIROJO KODO PROGRAMŲ TAIKYMAS RINKOS TYRIMUOSE

Vadovėlis

Redagavo *Ramutė Pinkevičienė*
Maketavo *Daiva Šepetauskaitė*

SL 585. 2013 01 29. 10,36 leidyb. apsk. l.
Tiražas 300 egz. Užsakymas 17 909
Išleido Mykolo Romerio universitetas
Ateities g. 20, Vilnius
Puslapis internete www.mruni.eu
El. paštas leidyba@mruni.eu
Parengė spaudai UAB „Baltijos kopija“
Kareivių g. 13B, Vilnius
Puslapis internete www.kopija.lt
El. paštas info@kopija.lt
Spausdino UAB „Vitae Litera“
Kurpių g. 5–3, Kaunas
Puslapis internete www.bpg.lt
El. paštas info@bpg.lt