KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

TAIKOMOSIOS MATEMATIKOS KATEDRA

Jolita Kazakevičiūtė

EUROPOS VALSTYBIŲ ŠVIETIMO DUOMENŲ STATISTINĖS ANALIZĖS MODELIAI IR PROGRAMINĖ ĮRANGA

Magistro darbas

Vadovas
dr. T. Ruzgas

KAUNAS, 2011

KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

TAIKOMOSIOS MATEMATIKOS KATEDRA

TVIRTINU
Katedros vedėjas
doc. dr. N. Listopadskis
2011 06 02

EUROPOS VALSTYBIŲ ŠVIETIMO DUOMENŲ STATISTINĖS ANALIZĖS MODELIAI IR PROGRAMINĖ ĮRANGA

Taikomosios matematikos magistro baigiamasis darbas

Vadovas
dr. T. Ruzgas
2011 06 01

Recenzentas
prof. habil. dr. M. Radavičius
2011 06 01

Atliko
FMMM-9 gr. stud.
J. Kazakevičiūtė
2011 05 30

KAUNAS, 2011

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Leonas Saulis, profesorius (VGTU)

Sekretorius: Eimutis Valakevičius, docentas (KTU)

Nariai: Algimantas Jonas Aksomaitis, profesorius (KTU)

Vytautas Janilionis, docentas (KTU)

Vidmantas Povilas Pekarskas, profesorius (KTU)

Rimantas Rudzkis, habil. dr., vyriausiasis analitikas (DnB NORD Bankas)

Zenonas Navickas, profesorius (KTU)

Arūnas Barauskas, dr., vice-prezidentas projektams (UAB „Baltic Amadeus“)

Kazakevičiūtė J. Europos valstybių švietimo duomenų statistinės analizės modeliai ir programinė įranga: Taikomosios matematikos magistro baigiamasis darbas / vadovas dr. T. Ruzgas; Taikomosios matematikos katedra, Fundamentaliųjų mokslų fakultetas, Kauno technologijos universitetas. – Kaunas, 2011. – 81 p.

SANTRAUKA

Magistro baigiamajame darbe apžvelgiama Tarptautinės asociacijos duomenų bazė ir statistinė programinė įranga. Pateikiama darbe analizuojamų statistinės analizės metodų apžvalga: skaitinės charakteristikos, dispersinės analizės, logistinės regresijos analizės, klasterinės analizės, pagrindinių komponenčių analizės, grafinės analizės, požymių priklausomumo lentelių tyrimo, suderinamumo hipotezių metodai. Statistinės analizės metodų pagalba galima daugiau sužinoti apie duomenis: kaip vieni faktoriai priklauso nuo kitų, priklausomybes tarp dviejų ar daugiau kintamųjų. Tai yra svarbu, norint priimti svarbius sprendimus švietimo srityje. Programai realizuoti pasirinkta SAS®9 programinė įranga, kadangi šioje sistemoje plačiai išvystytos statistinės duomenų analizės galimybės ir galima sukurti vartotojui patogią interaktyvią sąsają. Dauguma žmonių nežino, kaip dirbti su šia programa, taigi vienas iš mano darbo tikslų buvo sukurti vartotojo sąsają, kurios pagalba vartotojas galėtų atlikti pasirinktų duomenų statistinę analizę turėdamas minimalias žinias apie SAS sistemą.

Kazakevičiūtė J. The statistical analysis models and software of education data for European countries: Master's work in applied mathematics / supervisor dr. T. Ruzgas; Department of Applied mathematics, Faculty of Fundamental Sciences, Kaunas University of Technology. – Kaunas, 2011. – 81 p.

SUMMARY

Master's thesis provides an overview of the IEA (6), software for statistical analysis and statistical analysis methods. There is written about how it is important to know more than descriptive statistics. When we know about analysis of variance, correlation, hypothesis testing, descriptive statistics, graphical visualization, logistic regression, cluster, principal component models we can find out more about our data. It is good to know how one factor depends from others, relationships among two or more variables, distributions of variables. This paper represents statistical analysis models which help in making statistically reasonable decisions for organizing educational system in European countries. SAS®9 system is used in this master's work for statistical data analysis because it is one of the leaders in the world for the big data amount analysis and has a lot of statistical models implemented in it. Many people do not know how to work with this system, so my purpose was to create the simple user interface for analyzing data without having knowledge and experience in SAS programming language. From the results users can make statistically reasonable decisions.

TURINYS

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

ĮVADAS

Šiame magistro baigiamajame darbe analizuojami duomenys gauti iš Tarptautinės asociacijos (angl. – International Association for the Evaluation of Educational Achievement (IEA)) duomenų bazės (6). Tarptautinė asociacija kaupia daug ir įvairių duomenų elektroninėje formoje. Tarptautinėje duomenų bazėje yra daug vertingų duomenų, kurie gali padėti išsiaiškinti pasiekimus matematikoje ir moksle. Sukaupti duomenys saugomi keleto tipų duomenų šaltiniuose, o tiesioginiai jų analizei reikalingos SAS (9) (angl. – Statistical Analysis System*)* arba SPSS (angl. – Statistical Package for the Social Science*)* programos, kurios galėtų iš duomenų išgauti naudingą informaciją, padedančią priimti strateginius sprendimus.

Darbo tikslas – sukurti programinę įrangą, kuri palengvintų sprendimų priėmimą ir pagrindimą sprendžiant švietimo valdymo organizavimo klausimus. Magistro baigiamojo *darbo uždaviniai*: realizuoti sukurtus statistinės analizės modelius programiškai; panaudojant SAS sistemos objektinio programavimo galimybes, sukurti švietimo duomenų statistinės analizės posistemę; panaudojant sukurtus modelius ir programines priemones, atlikti realių duomenų statistinę analizę.

Programai realizuoti pasirinkta SAS programinė įranga, kadangi šioje sistemoje plačiai išvystytos statistinės duomenų analizės galimybės ir ji turi nesudėtingą programavimo kalbą, taip pat galima kurti vartotojui patogią interaktyvią sąsają, pateikti duomenis grafikuose, diagramose, lentelėse. SAS programavimo kalba yra paprasta ir lengvai įsimenama – daug gerų pavyzdžių ir patarimų galima rasti SAS pagalboje arba internete.

Sukurta sistema analizuoja Europos valstybių švietimo duomenis įvairiais pjūviais, skaičiuoja skaitines charakteristikas, atlieka koreliacinę analizę, dispersinę analizę, logistinės regresijos analizę, pagrindinių komponenčių analizę ir grafinę analizę, požymių priklausomumo lentelių tyrimą, tikrina suderinamumo hipotezes.

Pagrindiniai analizuojami rodikliai yra mokinio lytis, amžius, matematikos mokytojo amžius, darbo stažas ir kt., kurie yra gauti iš 2007 metais atliktų mokinių ir jų matematikos mokytojų apklausos anketų rezultatų iš įvairių Europos valstybių. Duomenys analizuojami pagal kelis požymius: valstybės pavadinimą, moksleivių ir jų matematikos mokytojų lytį. Visus duomenų sąrašus vartotojas gali sudaryti pagal savo poreikius, t.y. vartotojas gali suformuoti norimus duomenų pjūvius.

Darbas šia tematika buvo pristatytas VIII studentų konferencijoje „Taikomoji matematika“ (7).

1. TEORINĖ DALIS

1.1. APŽVALGA

1.1.1. TARPTAUTINĖS ASOCIACIJOS DUOMENŲ BAZĖ

Tarptautinės asociacijos duomenų bazėje (6) galime rasti duomenis ir dokumentaciją iš CivED, PIRLS, RL II, SITES ir TIMSS (angl. – Trends in International Mathematics and Science Study). CivED pateikiami tyrimai apie mokinių pilietinį išsimokslinimą; PIRLS ir RL II – apie mokinių raštingumą; SITES – apie tai, kokią įtaką turi informacija ir komunikacijos technologijos mokinių mokymuisi mokyklose. Duomenys yra pateikiami SPSS, SAS ir RAW formatu. Magistro darbe analizuojama informacija iš TIMSS. TIMSS duomenų bazėje pateikti metaduomenys apie kintamųjų tipus, galimas jų reikšmes, naudojimosi instrukcijos, aprašomosios analizės apie kai kuriuos kintamuosius rezultatai, išvados. Tarptautinėje duomenų bazėje yra daug vertingų duomenų iš įvairių šalių. Duomenys buvo gauti iš apklausų, kurios buvo pateiktos įvairioms šalims. Apklausos anketos buvo skirtos ir mokiniams, ir mokytojams, kad būtų galima kuo tiksliau įvertinti mokymo ir mokymosi kokybę tam tikroje šalyje.

TIMSS analizuoja duomenis apie ketvirtų ir aštuntų klasių mokinius ir jų mokytojus. Tyrimai apie ketvirtas klases buvo atlikti 1995, 2003 ir 2007 metais, o apie aštuntas klases – 1995, 1999, 2003 ir 2007 metais. Magistro darbe analizuojami duomenys yra gauti iš 2007 metais Europos valstybėse atliktų aštuntos klasės mokinių ir matematikos mokytojų apklausų rezultatų.

Mokiniams skirtoje anketoje klausiama: kokioje šalyje gyvena, kokioje mokykloje ir klasėje mokosi, kokia lytis, ar turi namuose kalkuliatorių, kompiuterį, rašomąjį stalą, žodyną, internetą, kiek knygų turi namuose, kaip jiems sekasi matematika, ar patinka ją mokytis, kiek valandų per dieną žiūri TV, žaidžia kompiuterinius žaidimus, būna su draugais, dirba namie, sportuoja ar daro namų darbus, kokioje šalyje gimęs, koks amžius ir t.t.

Matematikos mokytojams skirtoje anketoje klausiama: kokioje šalyje gyvena, kokioje mokykloje dirba, koks amžius, kokia lytis, kiek laiko dirba mokytoju, ar turi mokytojo kvalifikacijos dokumentą, koks mokinių skaičius klasėje, kiek laiko per savaitę vyksta matematikos pamokos, ar užduoda namų darbus ir t.t.

1.1.2. TIMSS TYRIMŲ REZULTATAI

TIMSS 2007 tarpautiniame matematikos pranešime (8) yra TIMSS tyrimų rezultatai iš 2007 metais atliktų ketvirtų ir aštuntų klasės mokinių ir jų mokytojų apklausų anketų. Apklausos anketos,

skirtos ketvirtų klasių mokiniams ir mokytojams, buvo pateiktos 37 šalims, o anketos, skirtos aštuntų klasių mokiniams ir mokytojams, - 50 šalių.

Aukščiausi vidutiniai matematikos pasiekimai ketvirtų klasių mokinių buvo Honkonge ir Singapūre. Aštuntų klasių – Kinijoje, Korėjoje ir Singapūre. Ketvirtų klasių mergaičių ir berniukų vidutiniai matematikos pasiekimai buvo beveik panašūs, tačiau aštuntų klasių mergaitės turėjo aukštesnį vidutinį matematikos pasiekimą nei aštuntų klasių berniukai. Geriausi pasiekimai buvo tų mokinių, kurie namuose turėjo daugiau knygų, negu tų, kurie namuose knygų turėjo mažiau. Aštuntų klasių mokiniai, kurių tėvai buvo labiau išsimokslinę, turėjo aukštesnius vidutinius matematikos pasiekimus beveik visose šalyse. Abiejų klasių mokiniai, kurie namuose turėjo kompiuterius, turėjo aukštesnius pasiekimus, negu tie, kurie neturėjo namuose nė vieno kompiuterio. Tą patį galima pasakyti ir apie tuos mokinius, kurie namuose turėjo internetą.

Ketvirtoms ir aštuntoms klasėms matematiką daugiausiai dėstė mokytojai, sulaukę trečio ir ketvirto dešimtmečio. 70 procentų ketvirtos klasės mokiniams ir 78 procentams aštuntos klasės mokiniams dėstė mokytojai su universitetiniu išsilavinimu. Dauguma aštuntos klasės mokinių turėjo mokytojus, kurie studijavo matematiką (70 %) ir/ar turėjo matematikos išsilavinimą (54 %). Daugelyje šalių ketvirtų klasių mokiniams nebuvo leidžiama naudotis kalkuliatoriais per matematikos pamokas, tačiau aštuntų klasių mokiniams daugelyje šalių buvo leidžiama naudotis kalkuliatoriais. Aštuntų klasių mokiniai, kurių mokytojai užduodavo daugiau namų darbų, turėjo didesnius matematikos pasiekimus. Mokytojai duodavo matematikos testus bent jau kas mėnesį 85 procentams aštuntos klasės mokinių.

1.1.3. STATISTINĖS PROGRAMINĖS ĮRANGOS APŽVALGA

Šiuolaikinė statistika yra neatsiejama nuo kompiuterinės duomenų analizės, padedančios greitai ir efektyviai spręsti įvairius statistikos uždavinius. Taikomosios statistikos metodai plačiai taikomi priimant svarbius sprendimus įvairiose srityse, pvz.: medicinoje, versle, valstybės valdyme, gamyboje ir kt.

Egzistuojančią duomenų analizės programinę įrangą galima suskirstyti į kelias grupes:

- duomenų analizės uždavinių programų bibliotekos universaliose programavimo kalbose (*Pascal*, *C* ir kt.);
- universalios matematinių uždavinių sprendimo sistemos (*MathCad*, *Maple*, *MatLab*, *Mathematica* ir kt.);
- universalios duomenų analizės sistemos (*SAS*, *SPSS*, *Statistica* ir kt.);

- ekspertinės duomenų analizės sistemos, skirtos konkrečiai analizei (*TABLE CURVE* – vieno kintamojo regresinė analizė, *ABP* – laiko eilučių analizė ir kt.);
- kitos paskirties sistemos (*Excel* ir kt.).

Ne visos duomenų analizei taikomos programinės priemonės yra pakankamai efektyvios ir patikimos, ypač kai duomenų yra labai daug, pavyzdžiui MS Excel, SPSS. SPSS leidžia analizuoti duomenis ir pavaizduoti analizės rezultatus. Pagrindinis SPSS programinio paketo privalumas — didelė šiuolaikinių statistinių analizės metodų pasirinktis ir duomenų analizės rezultatų vizualizavimo priemonių (duomenų pateikimo lentelių, diagramų, skirstinių kreivių) įvairovė, lengvai įvaldoma dialoginė sąsaja. SPSS programinis paketas dažniausiai taikomas sociologijos, psichologijos, biologijos, medicinos, rinkodaros, kokybės valdymo procese.

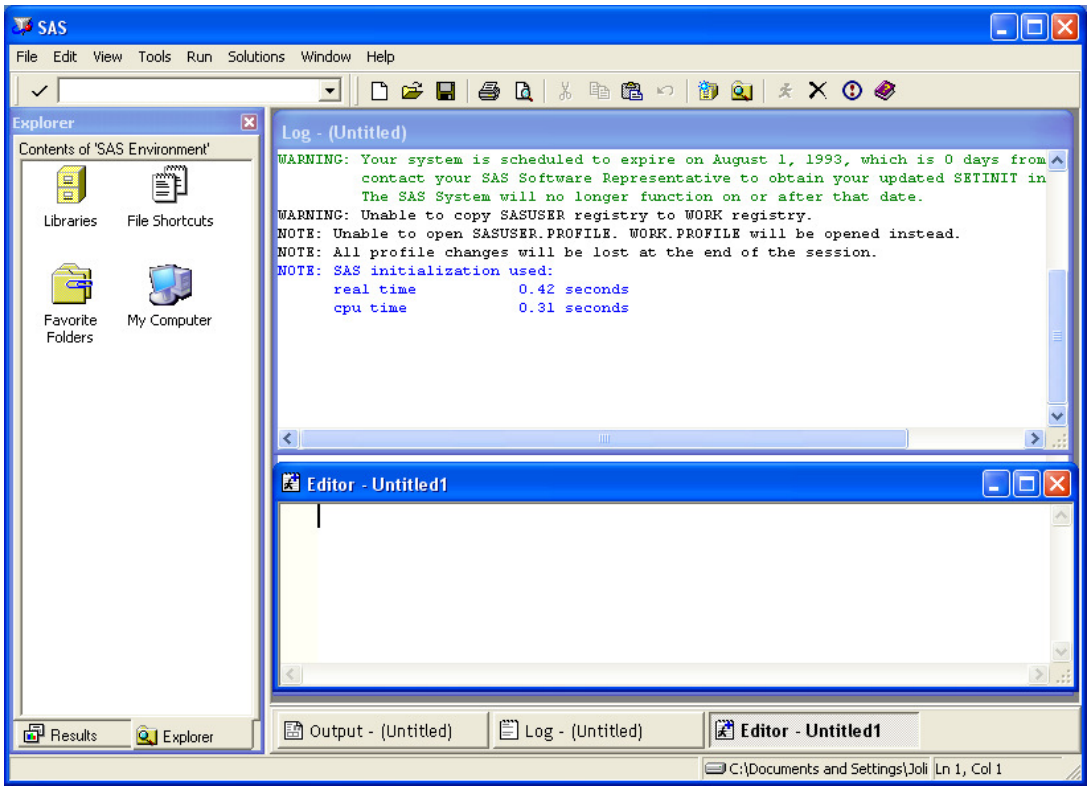
Šiame darbe pateikiama statistinės analizės vartotojo sąsaja, skirta analizuoti Europos valstybių švietimo duomenis. Sąsaja realizuota panaudojus programų paketo SAS 9 instrumentus.

Programinės įrangos kūrimui pasirinkta sistema SAS dėl to, kad tai yra viena iš populiariausių pasaulyje universalių duomenų analizės sistemų, galinti atlikti įvairias funkcijas ir netgi turinti labai geras taikomųjų programų kūrimo priemones. Kitose populiariose sistemose taikomųjų programų kūrimo priemonių arba nėra, arba jos yra labai elementarios. Be to, kitos sistemos atlieka mažiau funkcijų bei yra prastesnė vartotojo sąsaja.

1.1.4. STATISTINĖS ANALIZĖS SISTEMOS SAS GALIMYBIŲ APŽVALGA

SAS sistema yra viena iš plačiausias galimybes turinčių duomenų analizės sistemų. Ji dažnai naudojama dideliems duomenų kiekiams apdoroti. Be to, SAS sistema gali apdoroti daugelio populiarių formatų duomenų failus. SAS sistemos statistinės galimybės yra labai plačios: didžiulis rinkinys procedūrų pradedant nuo skaitinių charakteristikų apskaičiavimo ir baigiant specialiais taikomosios statistikos metodais. SAS sistema užtikrina duomenų pateikimą lentelių, grafikų, diagramų pavidalu. SAS programavimo kalba turi gerai išvystytą SAS/Macro (10), kurios neturi kai kurie kiti statistinės analizės paketai, pvz. Mathematica. SAS sistema veikia plačiausiai naudojamose operacinėse sistemose, tokiose kaip Microsoft Windows, Unix, Linux ir kt.

Sistema gali būti valdoma komandų ir meniu pagalba. SAS sistemos darbo aplinką sudaro: programos redaktoriaus langas (angl. *Editor*), rezultatų langas (angl. *Output*), sisteminių pranešimų langas (angl. *Log*), SAS failų naršyklė (angl. *Explorer*) ir rezultatų turinio langas (angl. *Results*) (1.1 pav.).



1.1 pav. SAS sistemos darbo aplinka

SAS programavimo kalba turi daug įvairių komandų, funkcijų, operatorių ir kitokių programavimo priemonių, kurios užtikrina norimo rezultato pasiekimą. Sistemos SAS programavimo kalba nėra sudėtinga, jos sintaksė yra labai panaši į kitų programavimo kalbų sintaksę. SAS programa susideda iš duomenų žingsnio DATA ir procedūrų žingsnio PROC. Duomenų žingsnis DATA yra skirtas SAS formato duomenų failo sukūrimui iš tekstinio failo ar kitų SAS formato failų. Procedūrų žingsnis PROC yra skirtas spręsti įvairiems probleminės srities (statistikos, tiesinio programavimo ir pan.) uždaviniams. Darbe naudojamus statistikos metodus realizuoja šios procedūros:

- Aprašomoji statistinė analizė (MEANS, UNIVARIATE ir FREQ). Procedūros MEANS ir UNIVARIATE naudojamos, kai dirbama su kiekybiniais kintamaisiais intervalų ir santykių skalėse. Procedūra FREQ naudojama, kai dirbama su kokybiniais kintamaisiais vardų ir tvarkos skalėse.
- Koreliacinė analizė (CORR ir FREQ). Procedūra CORR skaičiuoja koreliacijos koeficientus intervalų ir santykių skalėse, o procedūra FREQ skaičiuoja koreliacijos koeficientus vardų ir tvarkos skalėse.
- Logistinė regresinė analizė (LOGISTIC).
- Dispersinė analizė (ANOVA ir GLM).
- Klasterinė analizė (CLUSTER). Hierarchinis medis brėžiamas naudojant procedūrą TREE.
- Pagrindinių komponenčių analizė (PRINCOMP ir FACTOR).

- Grafinė analizė (GPLOT ir GCHART). GPLOT skirta taškų sklaidos diagramų ir funkcijų grafikų braižymui. GCHART skirta juostinių, stulpelinių, skritulinių diagramų braižymui. (11)

Praktiškai dirbant su duomenų analizės modeliais, labai svarbi yra vartotojo sąsaja, kurios pagalba vartotojas lengviau gali naudoti programą, jam reikia mažiau žinių programavimo srityje. SAS sistemoje vartotojo sąsajos kūrimui yra skirta posistemė SAS/AF, kurios taikomosios programos leidžia sukurti interaktyvią vartotojo sąsają sąveikai su duomenimis, duomenų vadybai, analizei ir pristatymui (3).

SAS/AF taikomosios programos yra SAS katalogo elementai. Galima naudoti procedūrą BUILD (arba komandą BUILD), kuriant šiuos katalogo elementų tipus:

- FRAME elementas, kuris pateikia paruoštas, objektiškai orientuotas komponentes grafinės vartotojo sąsajos kūrimui;
- PROGRAM ir MENU elementai, skirti taikomųjų programų kūrimui aplinkoje, kurioje nėra grafinės vartotojo sąsajos;
- SCL (*SAS Component Language*) kalbos elementai;
- CBT ir HELP elementai, skirti informacijos ir pagalbos sistemos kūrimui.

SCL programavimo kalba turi elementus, kurie yra daugelio programavimo kalbų pagrindas: sakiniai, funkcijos, *CALL* paprogramės, operatoriai, išraiškos, kintamieji. Ji naudoja sudėtinius teiginius, funkcijas ir kitas galimybes, kurių pagalba lengva sukurti objektiškai orientuotas, interaktyvias taikomąsias programas. *SCL* programa taip pat leidžia naudoti makro kintamuosius ir makrokomandas.

1.2. METODAI

1.2.1. APRAŠOMOJI STATISTIKA

Prieš atliekant nuodugnesnę surinktų duomenų statistinę analizę, pirmiausia reikia nagrinėti kiekvieną kintamąjį (požymį) atskirai. Šiai analizei atlikti naudojami aprašomosios statistikos metodai. Aprašomosios statistikos metodus sudaro duomenų grupavimas, dažnių lentelės, statistinių charakteristikų (duomenų padėties ir sklaidos charakteristikų) skaičiavimas ir grafinis stebėjimų vaizdavimas. Šiame poskyryje aptariamos magistro baigiamajame darbe naudojamos aprašomosios statistikos skaitinės charakteristikos (1).

Imtyje nustatomos mažiausia ir didžiausia reikšmės, imties plotis. Imties plotis yra didžiausios ir mažiausios stebėtos reikšmės skirtumas.

Vidurkis (aritmetinis) – tai pirmasis empirinis pradinis momentas. Imties vidurkis žymimas $\overline{X}$.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j ,$$

čia *n* – imties dydis.

Dispersija – antros eilės empirinis centrinis momentas, žymima S^2. Ji apibūdina atsitiktinio dydžio X reikšmių išsisklaidymo apie vidurkį laipsnį.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 .$$

Standartinis nuokrypis gaunamas ištraukus kvadratinę šaknį iš dispersijos:

$$S = \sqrt{S^2} .$$

Asimetrijos ir eksceso koeficientai apibūdina skirstinio funkcijos formą. Asimetrijos koeficientu vadinamas dydis

$$A_s = \frac{\mu_3}{S^3},$$

čia μ_k - k-tosios eilės centrinis momentas. Asimetrijos koeficientas apibūdina empirinio skirstinio simetriškumą vidurkio atžvilgiu.

Eksceso koeficientu vadinamas dydis

$$E_k = \frac{\mu_4}{S^4} - 3.$$

Jis apibūdina empirinio skirstinio viršūnės smailumą normaliojo skirstinio atžvilgiu.

Realizuoti aprašomąją statistiką naudota SAS procedūra PROC MEANS. Ši procedūra skaičiuoja įverčius: stebėjimų, praleistų stebėjimų skaičių, imties plotį, sumą, vidurkį, mažiausią, didžiausią reikšmes, dispersiją, standartinį nuokrypį, asimetrijos ir eksceso koeficientus.

1.2.2. HIPOTEZIŲ TIKRINIMO METODAI

Statistine hipoteze vadinama bet kuri prielaida apie nežinomą atsitiktinio dydžio tikimybių skirstinį (1). Tikrinamoji hipotezė vadinama nuline ir žymima H_0. Kartu nagrinėjama ir jai priešinga hipotezė H_a. Ji vadinama alternatyviąja. Hipotezės skirstomos į parametrines ir neparametrines:

1) Parametrinės:

- H_0: $\mu = \mu_0$; Ha: $\mu \neq \mu_0$;
- H_0: $\mu_1 = \mu_2$; Ha: $\mu_1 < \mu_2$;
- H_0: $\sigma_1 = \sigma_2$; Ha: $\sigma_1 \neq \sigma_2$;

ir pan.

2) Neparametrinės:

- suderinamumo;
- atsitiktinumo;
- nepriklausomumo;
- homogeniškumo.

Parametrinėms hipotezėms tikrinti praktikoje dažnai naudojamas Stjudento kriterijus. Neparametrinėms suderinamumo hipotezėms tikrinti taikomi Pirsono χ^2, Kolmogorovo ir kiti suderinamumo kriterijai. Pirmuoju kriterijumi galime tikrinti hipotezę apie bet kurį skirstinį, antruoju kriterijumi galime tikrinti hipotezę, jei pasiskirstymo funkcija yra tolydžioji ir visiškai apibrėžta.

Šiame magistro baigiamajame darbe nagrinėjamos suderinamumo hipotezės.

Panaudojus SAS sistemos procedūrą PROC UNIVARIATE, buvo gauti rezultatai dėl skirstinių tapatumo bei vizualiai sulyginti stebimo atsitiktinio dydžio ir pasirinkto dėsnio skirstiniai.

1.2.3. DISPERSINĖS ANALIZĖS METODAI

Stebimų atsitiktinių dydžių skirstinių priklausomybės nuo kiekybinių ar kokybinių faktorių tyrimas vadinamas dispersine analize. Dispersinės analizės tikslas – nustatyti, ar priklausomojo kintamojo vidurkiai skirtingose populiacijose skiriasi. (2) Vienfaktorinei dispersinei analizei žymėti vartojama santrumpa ANOVA (angl. – ANalysis Of VAriance). Vienfaktorinė dispersinė analizė naudojama tada, kai imtys vieną nuo kitos tyrėjas skiria tik pagal vieną požymį. Jei yra atsižvelgiama į daugiau požymių, naudojama daugiafaktorinė dispersinė analizė.

ANOVA modelio prielaidos:

- stebėjimai pasiskirstę pagal normalųjį dėsnį;
- stebėjimų dispersijos lygios;
- stebėjimai tarpusavyje nepriklausomi.

SAS programinėje įrangoje dispersinė analizė realizuota procedūrose ANOVA ir GLM. ANOVA procedūra naudojama subalansuotiems duomenims, o GLM – nesubalansuotiems.

Tiriant faktoriaus *A* įtaką grupių vidurkiams tikrinama hipotezė: H_0: visi vidurkiai tarpusavyje lygūs, H_a: bent du vidurkiai skiriasi.

ANOVA hipotezei tikrinti naudojama Fišerio statistika, kurios radimui skaičiuojama nuokrypių kvadratų suma (1.1 lentelė). Faktoriai skirstomi į dvi grupes: pastovūs ir nepastovūs. Remiantis fiksuotu modeliu, lyginami visų turimų imčių vidurkiai. Taikant nepastovų modelį, priklausomas kintamasis matuojamas daugelyje imčių, o vidurkiams lyginti atsitiktinai parenkama tik dalis imčių. Darbe naudojama dispersinė analizė su pastoviais faktoriais.

1.1 lentelė

Vienfaktorinės dispersinės analizės rezultatų lentelė

Nuokrypių šaltinis	Nuokrypių kvadratų suma	Laisvės laipsniai ν	Nuokrypių kvadratų vidurkis	Fišerio statistika	$a_{imt} = P(F > F_{imt})$
Faktorius *A*	SS_A	I - 1	$\overline{SS_A}$	$\frac{\overline{SS_A}}{\overline{SS_e}}$	a_{imt}
Atsitiktinių klaidų faktorius *e*	SS_e	n - I	$\overline{SS_e}$		
Visi faktoriai	SS_p	n - 1			

Nuokrypių kvadratų suma, apibūdinanti faktoriaus *A* poveikį stebimo atsitiktinio dydžio *Y* vidurkiui:

$$SS_A = \sum_{i=1}^{I} (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 n_i .$$

Nuokrypių kvadratų suma, apibūdinanti atsitiktinių klaidų faktoriaus *e* poveikį stebimo atsitiktinio dydžio *Y* vidurkiui:

$$SS_e = \sum_{i=1}^{I} \sum_{j=1}^{n} (Y_{ij} - \overline{Y}_{i\cdot})^2 .$$

Bendroji nuokrypių kvadratų suma:

$$SS_p = \sum_{i=1}^{I} \sum_{j=1}^{n} (Y_{ij} - \overline{Y}_{\cdot\cdot})^2 ,$$

$$SS_p = SS_A + SS_e .$$

Nuokrypių kvadratų vidurkiai:

$$\overline{SS}_A = \frac{1}{I-1} SS_A , \ \overline{SS}_e = \frac{1}{n-I} SS_e .$$

Hipotezės tikrinimui naudojamas Fišerio kriterijus su dešine kritine sritimi $F_{H_0} = (0, \quad F_{1-\alpha;\, \nu_1;\, \nu_2})$, $F_K = [F_{1-\alpha;\, \nu_1;\, \nu_2}, \quad \infty)$. H_0 atmetama, kai stebėtoji kriterijaus statistikos reikšmė patenka į kritinę sritį $F_{imt} \in F_K$, priešingu atveju stebėjimų duomenys H_0 hipotezei neprieštarauja.

Dvifaktorinė dispersinė analizė leidžia formuluoti tris statistines hipotezes: apie faktoriaus *A* įtaką, apie faktoriaus B įtaką ir apie faktorių *A* ir *B* tarpusavio sąveikos įtaką.

Jeigu dispersinės analizės tiesinio modelio prielaidos netenkinamos, tuomet tiriamo faktoriaus įtakai nustatyti taikoma vienfaktorinė ranginė dispersinė analizė. SAS programinėje įrangoje ranginė dispersinė analizė realizuota procedūroje NPAR1WAY.

i-ojo lygmens rangų suma:

$$T_i = \sum_{j=1}^{n} c_{ij} a(Y_j) .$$

čia $a(Y_j)$ yra *j*-asis rangas, *n* – stebinių skaičius ($j = \overline{1;n}$), o c_{ij} – indikatorius, parodantis, ar *j*-ąjį stebinį veikia *i*-tasis faktoriaus *A* – lygmuo.

Rangų vidurkis:

$$\bar{a} = \sum_{j=1}^{n} \frac{a(Y_j)}{n},$$

dispersija:

$$S^2 = \sum_{j=1}^{n} \frac{(a(Y_j) - \bar{a})^2}{n-1}.$$

Tiriant faktoriaus *A* įtaką mokytojų skaičiaus svertiniam vidurkiui tikrinama hipotezė apie vidurkių lygybę: H_0: vidurkiai visuose faktoriaus A lygmenyse tarpusavyje lygūs, H_a: ne visuose faktoriaus A lygmenyse vidurkiai tarpusavyje lygūs.

Hipotezės tikrinimui naudojama Kruskalo-Voliso (angl. – Kruskal-Wallis) statistika:

$$C = \frac{\sum_{i=1}^{r} \frac{(T_i - E_0(T_i))^2}{n_i}}{S^2},$$

čia n_i yra faktoriaus *A* *i*-tojo lygmens stebėjimų skaičius, o $E_0(T_i)$ – tikėtina rangų suma, jei nulinė hipotezė teisinga:

$$E_0(T_i) = n_i \bar{a} .$$

Kruskalo-Voliso statistikos skirstinys yra χ^2 su *r*-1 laisvės laipsniais.

Daugialypio palyginimo metodu sugrupuojami faktoriaus *A* lygmenys į homogenines grupes, kad grupių viduje vidurkiai nesiskirtų statistiškai reikšmingai. Palyginimui įvedama palyginimo funkcija:

$$\psi = \sum_{i=1}^{I} c_i \beta_i ,$$

čia c_i – žinomi pastovūs koeficientai, kurie tenkina sąlygą $\sum_{i=1}^{I} c_i = 0$. Klasifikuojant pagal vieną požymį, palyginimo funkcijos taškinis įvertis:

$$\hat{\psi} = \sum_{i=1}^{I} c_i \overline{Y}_{i\cdot} .$$

Taškinio įverčio dispersija lygi:

$$\sigma_{\hat{\psi}}^2 = \sum_{i=1}^{I} c_i^2 \mathbf{D}(\overline{Y}_{i\cdot}) = \sigma_e^2 \sum_{i=1}^{I} \frac{c_i^2}{n_i} .$$

Dispersijos įvertis:

$$\hat{\sigma}_{\hat{\psi}}^2 = \overline{SS}_e \sum_{i=1}^{I} \frac{c_i^2}{n_i},$$

čia $\sigma_e^2 = \overline{SS}_e$.

Tjukio daugialypio palyginimo metodo vidurkio pasikliautinasis intervalas apskaičiuojamas pagal formulę:

$$P\left(\hat{\psi} - T_{n,I,\alpha}\hat{\sigma}_{\hat{\psi}} \le \psi \le \hat{\psi} + T_{n,I,\alpha}\hat{\sigma}_{\hat{\psi}}\right) = 1 - \alpha ,$$

čia $T_{n,I,\alpha}$ – Tjukio skirstinio kvantilis.

1.2.4. KORELIACINĖS ANALIZĖS METODAI

Koreliacinė analizė naudojama nustatant ryšį tarp stebimų kintamųjų ir šio ryšio stiprumo vertinimui. Vertinant įvairius reiškinius, dažnai sutinkami tarpusavyje priklausomi kintamieji. Kokią įtaką vienas kitam jie turi, kaip keičiantis vienam keičiasi kitas, nagrinėja koreliacinė analizė. Koreliacinėje analizėje statistinio ryšio stiprumas tarp stebėtų kintamųjų yra išreiškiamas tam tikru koeficientu. Pagal gautą koeficiento reikšmę galima spręsti koks ryšys sieja kintamuosius: labai stiprus, stiprus, vidutinio stiprumo, silpnas ar labai silpnas ryšys.

Pirsono koreliacijos koeficientas vertina tiesinio ryšio stiprumą. Kintamųjų reikšmės turi būti išmatuotos intervalų ar santykių skalėje (vartotojas pats turi spręsti kokioje skalėje išmatuoti kintamieji, nes programinė įranga visus skaitinius kintamuosius priskiria vienam tipui). Dviejų atsitiktinių dydžių X ir Y tiesinio ryšio stiprumą įvertina kovariacija:

$$K_{XY} = M(X - M(X))M(Y - M(Y)) .$$

Atsitiktiniai dydžiai X ir Y vadinami nekoreliuotais, jeigu $K_{XY} = 0$. Jei atsitiktiniai dydžiai yra nepriklausomi, tai jie yra ir nekoreliuoti, tačiau atvirkščias teiginys neteisingas. Praktikoje vietoje kovariacijos koeficiento patogiau naudoti bedimensinį dydį ρ – vadinamą Pirsono koreliacijos koeficientu. Imties Pirsono koeficientas yra apskaičiuojamas pagal formulę:

$$r_p = \frac{K_{xy}}{S_x S_y} = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sqrt{\overline{X^2} - \left(\overline{X}\right)^2} \cdot \sqrt{\overline{Y^2} - \left(\overline{Y}\right)^2}} .$$

Sprendžiant praktinius uždavinius retai kada kintamųjų reikšmės yra pasiskirsčiusios pagal normalųjį skirstinį. Tokiu atveju reikia naudotis koeficientais, kurie nėra susiję su skirstiniais. Jeigu stebimus dydžius X ir Y galima išrikiuoti didėjimo tvarka (intervalų, santykių arba tvarkos skalės), tai galima naudoti Spirmeno ranginį koreliacijos koeficientą r_s. Jis apibūdina ryšio tarp stebimų dydžių X ir Y stiprumą monotoniškumo prasme, t.y. didėjant *X* reikšmėms, *Y* monotoniškai didėja arba mažėja (nebūtinai tiesiškai). Skaičiuojant Spirmeno ranginį koreliacijos koeficientą naudojamos ne stebėtos (X,Y)reikšmės, bet jų rangai (rX_i, rY_i). Spirmeno koreliacijos koeficientas gali įgyti $-1 \le r_s \le 1$ reikšmes. Imties Spirmeno ranginės koreliacijos koeficientui apskaičiuoti naudojama formulė:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(rX_i - rY_i)^2}{n(n^2-1)}.$$

Hipotezei apie koreliacijos koeficiento reikšmingumą tikrinti naudojama statistika:

$$t = \frac{(n-2)^{1/2} r}{(1-r^2)^{1/2}},$$

kurios skirstinys yra Stjudento su *n*-2 laisvės laipsniais, čia *r* – Pirsono arba Spirmeno koreliacijos koeficientas.

Kadangi požymių priklausomumo lentelė yra braižoma iš dviejų kintamųjų, todėl yra naudojamas χ^2 kriterijus (12).

Panaudojus SAS procedūrą PROC FREQ, nubraižoma požymių priklausomumo lentelė.

1.2.5. GRAFINĖS ANALIZĖS METODAI

Stebimus dydžius grafiškai galima vaizduoti skirtingais būdais. Empiriniams pasiskirstymams grafiškai vaizduoti dažniausiai naudojamos histogramos bei kvantilių diagramos. Dažniams plačiai taikomos stulpelinės, juostinės bei skritulinės diagramos. Darbe naudojamos skaitinės charakteristikos vaizduojamos stulpelinėmis diagramomis (vertikaliai). Jos vaizduojamos susumuotos arba vaizduojami tų reikšmių vidurkiai.

Duomenims grafiškai pavaizduoti naudojama SAS procedūra PROC CHART. Jos įvykdymui naudojami skaitiniai ir neskaitiniai kintamieji.

1.2.6. LOGISTINĖS REGRESIJOS ANALIZĖS METODAI

Logistinė regresinė analizė – tai sąryšio funkcijos tarp kintamųjų radimas, kur priklausomi kintamieji *Y* yra dichotominiai (dvireikšmiai). Analizė atliekama taip pat kaip regresinė analizė, tik čia naudojama logistinė funkcija:

$$f(y)=\frac{e^{y}}{1+e^{y}}.$$

Ieškoma priklausomybė:

$$P(Y=1|X)=F(\Theta^{T}X),$$

čia Θ – nežinomi parametrai, o funkcija $F(z)$ tenkina šias sąlygas:

- $F(z)$ monotoniškai didėja pagal *z*;
- $0\le F(z)\le 1$;
- $F(z)\to 1$, kai $z\to\infty$;
- $F(z)\to 0$, kai $z\to 0$.

Logistinės regresinės analizės atveju funkcija $F(z)$ yra logistinė funkcija, t. y.

$$P(y_i=1|X_i)=\frac{e^{\Theta^{T}X_i}}{1+e^{\Theta^{T}X_i}}.$$

Ši funkcija tenkina aukščiau minėtas sąlygas ir yra simetrinė $\Theta^{\mathrm{T}}\mathrm{X}=0$ atžvilgiu, t. y. $\Lambda(-z)=1-\Lambda(z)$, kur

$$\Lambda(z)=\frac{e^{z}}{1+e^{z}}.$$

Parametrų Θ ieškome mažiausių kvadratų arba didžiausio tikėtinumo metodais. Kai $\pi_i=P(y=1|X)$, didžiausio tikėtinumo metodu sudaroma funkcija:

$$L=\prod_{i=1}^{n}\pi_i^{y_i}(1-\pi_i)^{1-y_i},$$

$$\ln L=\sum_{i=1}^{n}\left(y_i\ln\pi_i+(1-y_i)\ln(1-\pi_i)\right),$$

$$\frac{\partial\ln L}{\partial\theta_j}=\sum_{i=1}^{n}\frac{(y_i-\pi_i)x_{ij}}{\pi_i(1-\pi_i)}.$$

Išvestinę prilyginę nuliui, t. y. išsprendę lygtį

$$\sum_{i=1}^{n}(y_i - \pi_i)x_{ij} = 0,$$

gauname koeficientų Θ įverčius.

Tikrinamos hipotezės apie koeficientų reikšmingumą: $H_0 : \theta_i = 0$, $H_a : \theta_i \neq 0$. Hipotezių tikrinimui naudojama Voldo statistika, kurios skirstinys yra χ^2:

$$W = \left(\frac{\theta_i}{s_e}\right)^2 \sim \chi^2(1),$$

čia $s_e - \theta_i$ standartinis nuokrypis (2).

Kai koeficientai labai dideli, Voldo statistika gali klaidingai neatmesti nulinės hipotezės, nors turėtų. Tada reikia pasikliauti modelio korektiškumo hipotezės tikrinimu.

Modelio korektiškumo įvertinimui tikrinamos hipotezės: $H_0 : P(y = 1|X) = 1$, $H_a : P(y = 1|X) \neq 1$.

Hipotezių tikrinimui naudojamos statistikos:

$$-2\ln L = -2\sum_{i=1}^{n}(y_i \ln \pi_i + (1 - y_i)\ln(1 - \pi_i)),$$

$$AIC = -2\ln L + 2(p+1),$$

$$SC = -2\ln L + (p+1)\ln n.$$

Šios statistikos yra apskaičiuojamos pilnam modeliui ir modeliui, kuriame yra tik koeficientas θ_0. Tuomet imamas jų skirtumas. Jei imame statistiką $-2\ln L$, tai turime tikėtinumo santykio kriterijų:

$$LR = -2\ln\frac{L_0}{L_1} \sim \chi^2(p-1),$$

čia L_0 – tikėtinumo funkcija, kai modelyje yra tik koeficientas θ_0; L_1 – tikėtinumo funkcija, kai modelyje yra visi koeficientai.

Jei imame *AIC* statistiką, tai turime Akaike informacijos kriterijų, ir modelis nėra korektiškas, jei $AIC_0 - AIC_1 << 1$, čia AIC_0 – AIC statistika, kai modelyje yra tik koeficientas θ_0; AIC_1 – AIC statistika, kai modelyje yra visi koeficientai.

Jei imame *SC* statistiką, tai turime Švarco kriterijų, ir modelis nėra korektiškas, jei $SC_0 - SC_1 << 1$, čia SC_0 – SC statistika, kai modelyje yra tik koeficientas θ_0; SC_1 – SC statistika, kai modelyje yra visi koeficientai.

Modelis realizuojamas panaudojant SAS procedūrą LOGISTIC.

1.2.7. KLASTERINĖS ANALIZĖS METODAI

Taikydami klasterinę analizę, nustatome objektų panašumą ir skirstome juos į klasterius (2). Klasterinės analizės tikslas – suskirstyti objektus taip, kad skirtumai klasterių viduje būtų kuo mažesni, o tarp klasterių – kuo didesni.

Artumo tarp objektų metrikos parinkimas yra vienas iš pagrindinių klasterinės analizės uždavinių. Nuo jos pasirinkimo priklauso ir galutinis objektų klasifikavimo variantas. Metrikos parinkimo uždavinys sprendžiamas kiekvienam atvejui atskirai ir priklauso nuo klasifikavimo tikslo, fizinės ir statistinės stebimų objektų prigimties, pradinių duomenų apie objektus pilnumo ir X tikimybinio skirstinio.

Klasterinė analizė taikoma, kai reikia klasifikuoti objektus neturint mokomųjų imčių. Yra duota n objektų O_1, O_2, …, O_n, kurių savybes nusako matrica "objektai – savybės"

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ \dots & \dots & \dots & \dots \\ x_1^{(p)} & x_2^{(p)} & \dots & x_n^{(p)} \end{pmatrix}$$

arba atstumų matrica

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2n} \\ \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \dots & \rho_{nn} \end{pmatrix},$$

čia $x_i^{(j)}$ – objekto O_i j-toji savybė.

Jei objektų savybes nusako matrica "objektai – savybės", tai ji transformuojama į atstumų matricą nurodžius metriką, pagal kurią apskaičiuojamas atstumas tarp objektų.

Klasterinės analizės uždavinys yra analizuojamą objektų aibę $O = \{O_i,\ i = \overline{1,n}\}$, užduota matrica X arba ρ, apjungti į homogeninę klasę (klasterį).

Apibrėžti objektų homogeniškumo sąvoką yra labai sunkus uždavinys, nes jam apibrėžti nėra formalių metodų. Paprastai atstumų matricoje elementai ρ_{ij} apibūdina atstumą tarp objektų $d(O_i, O_j)$ arba įvertina jų panašumo laipsnį $r(O_i, O_j)$.

Atstumo tarp objektų metrikos parinkimas yra vienas iš pagrindinių klasterinės analizės uždavinių. Nuo jos pasirinkimo priklauso ir galutinis objektų apjungimo variantas. Metrikos parinkimo uždavinys sprendžiamas kiekvienam atvejui atskirai ir priklauso nuo klasifikavimo tikslo, fizinės ir

statistinės stebimų objektų prigimties, pradinių duomenų apie objektus pilnumo ir X tikimybinio skirstinio.

Viena iš dažniausiai taikomų metrikų yra Euklido kvadratinis atstumas:

$$d_{E2}\left(X_i, X_j\right) = \sum_{s=1}^{p}\left(x_i^{(s)} - x_j^{(s)}\right)^2 ,$$

čia X_i – i-tojo objekto savybių vektorius.

Euklido metrika taikoma duomenims, išmatuotiems intervalinėje ir santykių skalėje.

Įvedame pažymėjimus: S_i – i-toji objektų klasė (klasteris); n_i – objektų skaičius klasėje; $\rho(S_l, S_m)$ – atstumas tarp klasių S_l ir S_m. Tuomet galime užrašyti atstumo tarp klasių metriką. Viena iš dažniausiai naudojamų yra vidutinis atstumas tarp visų galimų dviejų klasių objektų porų:

$$\rho_{vid}(S_l, S_m) = \frac{1}{n_l n_m}\sum_{X_i \in S_l}\sum_{X_j \in S_m} d\left(X_i, X_j\right).$$

Dažnai naudojamas hierarchinis apjungimo algoritmas. Pradinis skaidinys yra $S^{(0)} = \left(S_1^{(0)},...,S_n^{(0)}\right)$, čia $S_i^{(0)} = \{X_i\}$. k – lygio skaidinys yra $S^{(k)} = \left(S_1^{(k)},...,S_{n-k}^{(k)}\right)$. Šis k – lygio skaidinys gaunamas iš $S^{(k-1)}$, $k \geq 1$ skaidinio, apjungus klasių porą $\left(S_1^*, S_2^*\right)$:

$$\left(S_1^*, S_2^*\right) = \arg \min_{\substack{S_1 \neq S_2 \\ S_1, S_2 \in S^{(k-1)}}} \rho\left(S_1, S_2\right).$$

Galutinę hierarchiją sudaro įdėtų skaidinių sistema $S^{(0)} \subset S^{(1)} \subset ... \subset S^{(n-1)} \equiv X$, kurią galima atvaizduoti grafiškai.

Modelis realizuojamas panaudojant SAS procedūrą CLUSTER.

1.2.8. PAGRINDINIŲ KOMPONENČIŲ ANALIZĖS METODAI

Pagrindinių komponenčių analizės tikslas – sumažinti požymių erdvės skaičių, pakeičiant požymius mažesniu apibendrintų požymių skaičiumi. Ji naudojama sprendžiant šiuos pagrindinius uždavinius: požymių erdvės matavimo sumažinimo, grafinio vaizdavimo, ortogonalizacijos, informacijos suspaudimo.

Analizuojami kintamieji turi būti vienodos prigimties ir matuojami tais pačiais vienetais. Jei ši sąlyga neišpildyta, tai pereinama prie centruotų ir normuotų dydžių.

Pagrindinių komponenčių analizės metu yra apskaičiuojamas informatyvumo kriterijus, kuris parodo, kiek komponentė yra reikšminga lyginant ją su kitomis komponentėmis.

Pagrindinė savybė, kuria pasižymi pagrindinių komponenčių analizė – prognozuojant padaroma mažiausia paklaida iš visų galimų, nes imamos tos komponentės, kurios apima daugiausiai informacijos.

Požymių vektorius $X=\left(x^{(1)},...,x^{(p)}\right)^T$, čia $a=\left(a^{(1)},...,a^{(p)}\right)$ – parametrų vidurkiai, $\Sigma=\left(\sigma_{ij}\right)$, $i,j=1,2,...,p$ – kovariacinė matrica. Reikia rasti tokį $\tilde{Z}(X)$ rinkinį klasėje $F(X)$, kad

$$I_{p'}\left(\tilde{Z}(X)\right)=\max_{Z(X)\in F}\left\{I_{p'}(Z(X))\right\},$$

$$F=\left\{Z: z^{(j)}=\sum_{\nu=1}^{p}c_{j\nu}\left(x^{(\nu)}-a^{(\nu)}\right),\ j=1,2,...,p\right\},\ \text{kur}$$

$$\sum_{\nu=1}^{p}c_{j\nu}^2=1,\ \sum_{\nu=1}^{p}c_{j\nu}c_{k\nu}=0,\ j=1,2,...,p,\ k=1,2,...,p,\ j\neq k.$$

Tuomet informatyvumo kriterijus:

$$I_{p'}\left(\tilde{Z}(X)\right)=\frac{\mathbf{D}z^{(1)}+...+\mathbf{D}z^{(p')}}{\mathbf{D}x^{(1)}+...+\mathbf{D}x^{(p')}}.$$

Ieškomos pagrindinės komponentės yra požymių tiesinės kombinacijos $\tilde{Z}=LX$, kur

$$L=\begin{pmatrix}l_{11} & ... & l_{1p}\\ ... & ... & ...\\ l_{p1} & ... & l_{pp}\end{pmatrix},$$

čia matricos eilutės yra ortogonalios, t. y. $LL^T=L^TL=1$.

Tiriamos sistemos požymių $X=\left(x^{(1)},...,x^{(p)}\right)^T$ pirmąja pagrindine komponente $\tilde{z}^{(1)}(X)$ vadinama tokia požymių normuota ir centruota tiesinė kombinacija, kuri tarp visų kitų kintamųjų $x^{(1)},...,x^{(p)}$ normuotų ir centruotų tiesinių kombinacijų turi didžiausią dispersiją.

Tiriamos sistemos požymių $X=\left(x^{(1)},...,x^{(p)}\right)^T$ k-tąja pagrindine komponente $\tilde{z}^{(k)}(X)$, $k=2,3,...,p$ vadinama tokia požymių normuota ir centruota tiesinė kombinacija, kuri nekoreliuota su $k-1$ prieš tai einančiomis pagrindinėmis komponentėmis ir tarp visų kitų kintamųjų $x^{(1)},...,x^{(p)}$ normuotų ir centruotų bei nekoreliuotų su prieš tai einančiomis $k-1$ pagrindinėmis komponentėmis tiesinių kombinacijų turi didžiausią dispersiją.

Ieškant pagrindinių komponenčių kintamieji turi būti centruoti. Jei kintamieji nėra vienodos prigimties ir matuojami skirtingais vienetais, tai jie turi būti ir normuoti.

Norint rasti pirmąją pagrindinę komponentę, reikia išspręsti tokį optimizavimo uždavinį:

$$\begin{cases} \mathbf{D}(l_1 X) \to \max\limits_{l_1} \\ l_1 l_1^T = 1 \end{cases},$$

kur l_1 – pirma matricos L eilutė.

Kadangi $\mathbf{E}X = 0$, tai $\mathbf{E}\left(XX^T\right) = \Sigma$, tai

$$\mathbf{D}(l_1 X) = \mathbf{E}(l_1 X)^2 = \mathbf{E}\left(l_1 X X^T l_1^T\right) = l_1 \Sigma l_1^T,$$

$$\begin{cases} l_1 \Sigma l_1^T \to \max\limits_{l_1} \\ l_1 l_1^T = 1 \end{cases}.$$

Įvedama Lagranžo funkcija ir gaunama lygtis:

$$(\Sigma - \lambda I) l_1^T = 0.$$

Išsprendus charakteringąją lygtį $|\Sigma - \lambda I| = 0$, gaunama p tikrinių reikšmių $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$.

Kadangi

$$\mathbf{D}\tilde{z}^{(1)} = \mathbf{D}(l_1 X) = l_1 \Sigma l_1^T,$$

o $l_1 \Sigma l_1^T = \lambda$, tai $\mathbf{D}\tilde{z}^{(1)}(X) = \lambda$.

Kadangi dispersija turi būti didžiausia, tai $\mathbf{D}\tilde{z}^{(1)}(X) = \lambda_1$.

Išsprendus, gaunamas tikrinis vektorius l_1. Analogiškai ieškoma kitų tikrinių vektorių:

$$\tilde{z}^{(k)}(X) = l_k X.$$

Tuomet informatyvumo kriterijus:

$$I_{p'}\left(\tilde{Z}(X)\right) = \frac{\lambda_1 + ... + \lambda_{p'}}{\lambda_1 + ... + \lambda_p}$$

Svarbi charakteristika pagrindinių komponenčių analizėje yra svorių matrica $A = (a_{ij})$, $i, j = 1,2,..., p$, kuri aprašo tiesinę koreliaciją:

$$A = L^T \Lambda^{1/2},$$

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & ... & 0 \\ 0 & \sqrt{\lambda_2} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sqrt{\lambda_p} \end{pmatrix}$$

Iš svorių matricos galima spręsti apie pagrindinių komponenčių interpretaciją.

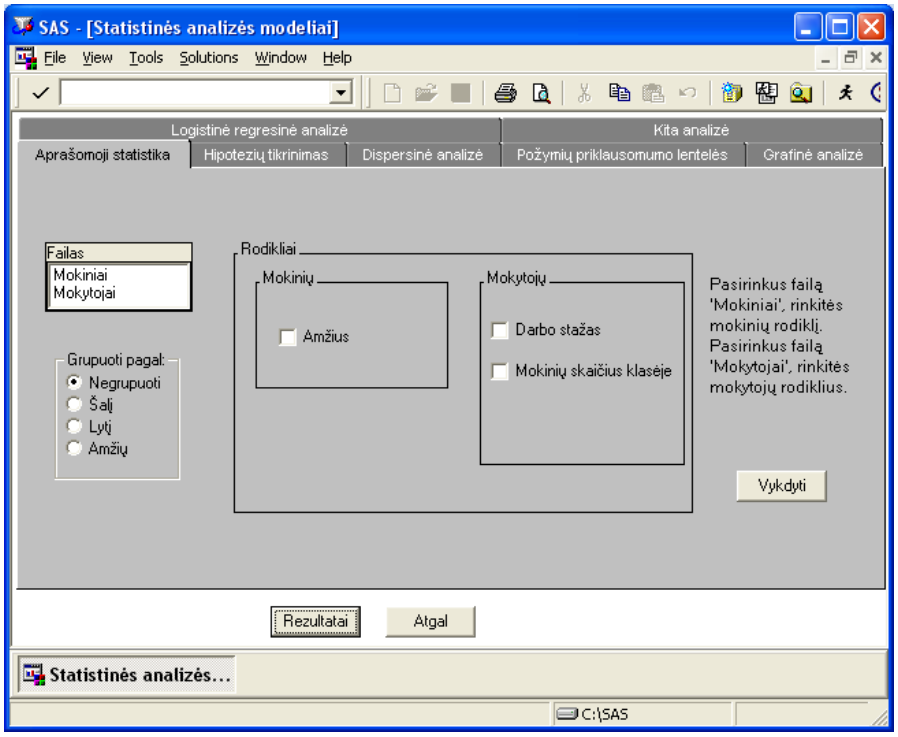
Modelis realizuojamas panaudojant SAS procedūrą FACTOR (5).

2. TIRIAMOJI DALIS

2.1. APRAŠOMOJI STATISTIKA

Realizuoti aprašomąją statistiką naudota SAS procedūra PROC MEANS. Ši procedūra skaičiuoja įverčius: stebėjimų, praleistų stebėjimų skaičių, vidurkį, dispersiją, standartinį nuokrypį, mažiausią, didžiausią reikšmes, imties plotį, sumą, asimetriją ir ekscesą. Šias statistikas galima apskaičiuoti visai imčiai arba atskirai kiekvienai grupei, kurias galima sudaryti, panaudojant grupavimo kintamąjį.

Sukurtos programinės įrangos interaktyvios sąsajos aprašomosios statistikos langas pavaizduotas 2.1 paveiksle.



2.1 pav. Aprašomosios statistikos langas

Analizuojamas kintamasis *Darbo stažas*. Rezultatai, kai duomenys gauti iš visų šalių, matomi 2.1 lentelėje, o kai duomenys gauti tik iš Lietuvos, matomi 2.2 lentelėje.

Iš 2.1 lentelės matome, kad mokytojų darbo stažo vidurkis yra 18.98 metai, didžiausias mokytojo darbo stažas yra 50 metų. Iš 2.2 lentelės matome, kad Lietuvoje dirbančių mokytojų darbo stažo vidurkis yra 18.24 metai, didžiausias mokytojo darbo stažas yra 47 metai.

Palyginus Lietuvos ir visų šalių kintamojo *Darbo stažas* skaitines charakteristikas, matome, kad vidurkiai, didžiausia ir mažiausia reikšmės skiriasi nežymiai. Asimetrijos koeficientas analizuojant Lietuvos duomenis yra neigiamas palyginus, o analizuojant visas šalis – teigiamas. Eksceso

koeficientai abiem atvejais yra neigiami, tai tankio funkcijos yra bukesnės lyginant su normaliojo tankio funkcija.

2.1 lentelė

Kintamojo *Darbo stažas* skaitinės charakteristikos

Stebėjimų skaičius	**Trūkstamų stebėjimų skaičius**	**Imties plotis**	**Suma**	**Vidurkis**	**Mažiausia reikšmė**
5020	126	50	95282	18.98	0
Didžiausia reikšmė	**Dispersija**	**Standartinis nuokrypis**	**Asimetrijos koeficientas**	**Ekscesо koeficientas**	
50	133.68	11.562	0.119	-1.115	

2.2 lentelė

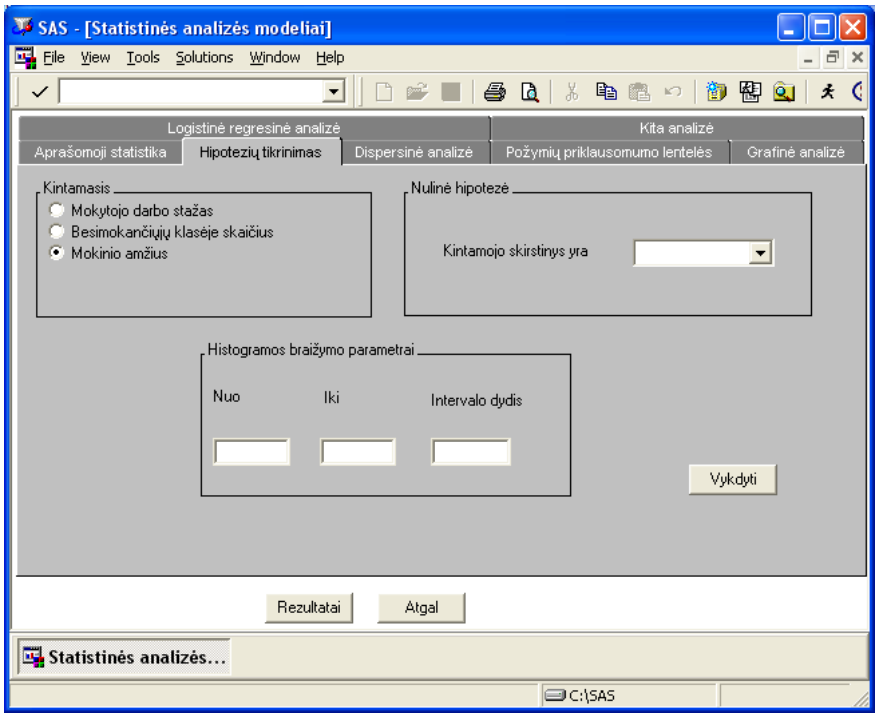
Kintamojo *Darbo stažas* skaitinės charakteristikos

Stebėjimų skaičius	**Trūkstamų stebėjimų skaičius**	**Imties plotis**	**Suma**	**Vidurkis**	**Mažiausia reikšmė**
256	1	46	5677	18.24	1
Didžiausia reikšmė	**Dispersija**	**Standartinis nuokrypis**	**Asimetrijos koeficientas**	**Ekscesо koeficientas**	
47	96.906	9.844	-0.127	-0.319	

Kitų kintamųjų skaitinės reikšmės yra pateiktos 1 priede.

2.2. HIPOTEZIŲ TIKRINIMAS

Sukurtos programinės įrangos interaktyvios sąsajos hipotezių tikrinimo langas pavaizduotas 2.2 pav. Tikrinamos hipotezės apie skirstinių lygybę. Vartotojas gali pasirinkti nulinę hipotezę, t.y. prilyginti kintamojo skirstinį su pasirinktuoju.



2.2 pav. Hipotezių tikrinimo langas

Su PROC UNIVARIATE procedūra patikriname tokią hipotezę:

H_0 : imtys pasiskirsčiusios pagal F(x) skirstinį,

H_a : imtys nėra pasiskirsčiusios pagal F(x) skirstinį,

kur F(x) – pasirinktas skirstinys iš sąrašo.

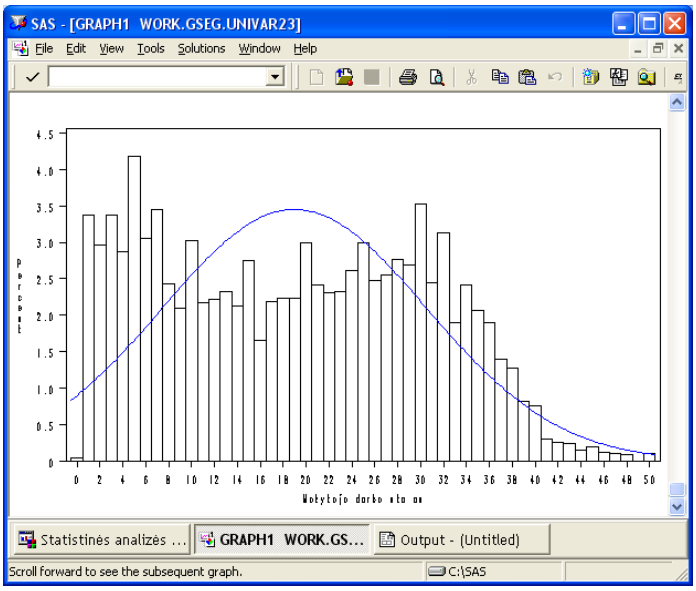
Analizuojamas kintamasis *Mokytojo darbo stažas,* nulinė hipotezė – kintamojo skirstinys yra normalusis. Rezultatai, kai duomenys gauti iš visų šalių, matomi 2.3 lentelėje, o kai duomenys gauti tik iš Lietuvos, matomi 2.4 lentelėje.

2.3 lentelė

Hipotezės tikrinimo rezultatai

	STATISTIKA	**P-REIKŠMĖ**
KOLMOGOROVO-SMIRNOVO	D 0.0897054	Pr > D <0.010
KRAMERIO	W-Sq 9.7027675	Pr > W-Sq <0.005
ANDERSONO-DARLINGO	A-Sq 64.9967544	Pr > A-Sq <0.005

Matome, jog hipotezė H_0 atmetama, kadangi *p*-reikšmė < 0.05. Vadinasi skirstiniai nesuderinami (2.3 pav.).



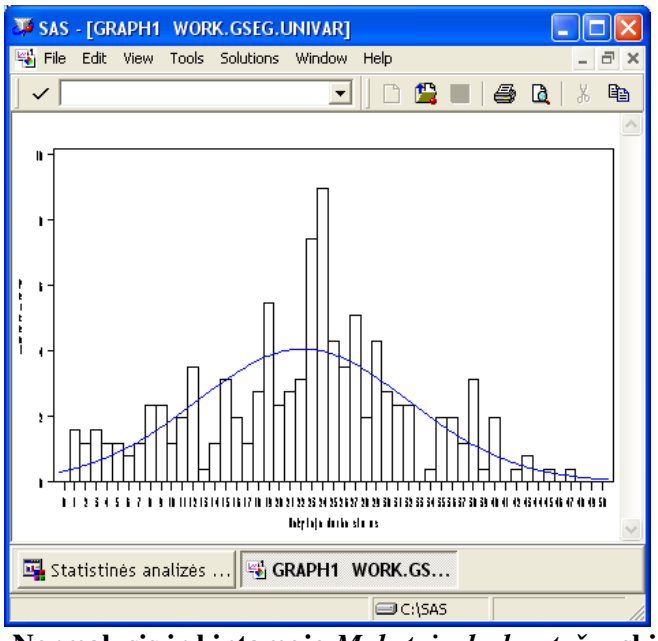
2.3 pav. Normalusis ir kintamojo *Mokytojo darbo stažas* skirstiniai

2.4 lentelė

Hipotezės tikrinimo rezultatai

	STATISTIKA	P-REIKŠMĖ
KOLMOGOROVO-SMIRNOVO	D 0.09195709	Pr > D <0.010
KRAMERIO	W-Sq 0.27002066	Pr > W-Sq <0.005
ANDERSONO-DARLINGO	A-Sq 1.47431479	Pr > W-Sq <0.005

Matome, jog hipotezė H_0 atmetama, kadangi *p*-reikšmė < 0.05, vadinasi skirstiniai nesuderinami (2.4 pav.).



2.4 pav. Normalusis ir kintamojo *Mokytojo darbo stažas* skirstiniai

Hipotezių tikrinimas su kitais kintamaisiais pateiktas 2 priede.

2.3. DISPERSINĖ ANALIZĖ

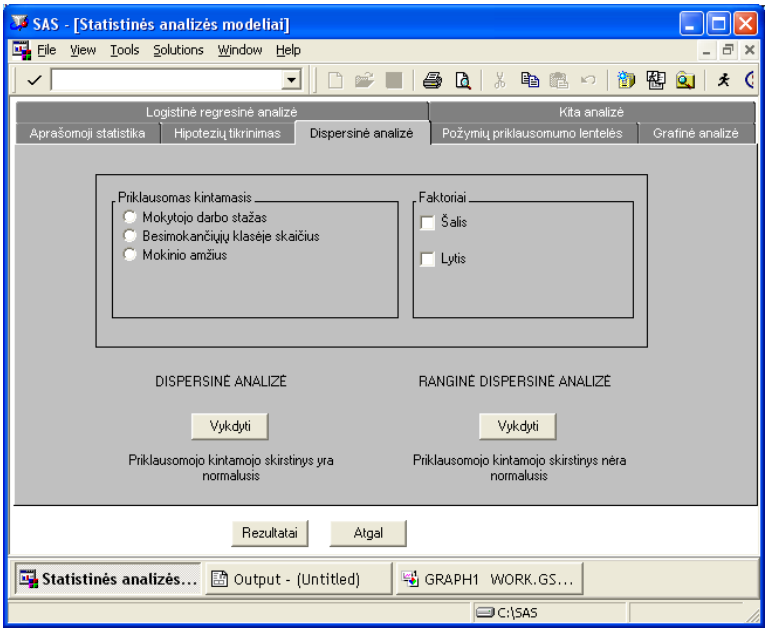
Programinio modelio sąsaja su vartotoju meniu pateikta 2.5 paveiksle. Vartotojas čia gali pasirinkti faktorius ir priklausomąjį kintamąjį, kurio vidurkį veikia pasirinkti faktoriai. Gali būti atlikta vienfaktorinė arba dvifaktorinė dispersinė analizė.

2.5 lentelė

Vienfaktorinės dispersinės analizės rezultatų lentelė

Nuokrypių šaltinis	Laisvės laipsnių skaičius	Nuokrypių kvadratų sumos	Nuokrypių kvadratų vidurkiai	Fišerio statistika	Reikš-mingumo lygmuo
Faktorius šalis	19	87331.2558	4596.3819	39.38	<.0001
Atsitiktiniai faktoriai	5000	583608.8311	116.7218		
Visi faktoriai	5019	670940.0869			

Tikrinama, ar faktorius *Šalis* neturi įtakos priklausomam kintamajam *Mokytojo darbo stažas*. Duomenys gauti iš visų šalių. Kaip matome iš rezultatų (2.5 lentelė), faktorius *Šalis* turi įtakos mokytojo darbo stažui, kadangi nulinė hipotezė, kad vidurkiai statistiškai reikšmingai nesiskiria kiekviename lygmenyje, yra atmetama.



2.5 pav. Dispersinės analizės langas

Kaip žinome iš ankščiau atliktų tyrimų rezultatų, kintamojo skirstinys nėra normalusis. Todėl taikyti Tjuko daugialypio palyginimo metodo negalima. Kai stebimo kintamojo vidurkio skirstinys nėra normalusis, ranginės dispersinės analizės modelis leidžia nustatyti faktoriaus *Šalis* įtaką kintamajam *Mokytojo darbo stažas* (2.6 lentelė).

2.6 lentelė

Vilkoksono rangų sumos

Šalis	N	Rangų suma	Tikėtina rangų suma	Std. nuokrypis	Rangų vidurkis
Bosnija ir Hercegovina	166	519922.50	416743.00	18354.7617	3132.06325
Rumunija	254	792978.50	637667.00	22497.7298	3121.96260
Bulgarija	231	718209.50	579925.50	21506.6728	3109.13203
Rusija	269	829380.50	675324.50	23116.0420	3083.19888
Italija	287	872997.00	720513.50	23831.6450	3041.80139
Ukraina	180	530178.50	451890.00	19085.5093	2945.43611
Gruzija	177	516074.50	444358.50	18931.6596	2915.67514
Lietuva	256	742887.00	642688.00	642688.00	2901.90234
Vengrija	268	764814.50	672814.00	23075.4635	2853.78545
Armėnija	247	669487.50	620093.50	22201.8423	2710.47571
Serbija	221	592407.50	554820.50	21057.9617	2680.57692
Čekija	204	513698.00	512142.00	20267.6382	2518.12745
Slovėnija	490	1146271.00	1230145.00	30464.3531	2339.32857
Norvegija	252	584125.50	632646.00	22413.6824	2317.95833
Švedija	425	936939.50	1066962.50	28574.7239	2204.56353
Škotija	274	546678.00	687877.00	23317.6065	1995.17518
Anglija	218	398293.50	547289.00	20921.0821	1827.03440
Kipras	231	392654.50	579925.50	21506.6728	1699.80303
Malta	235	339855.00	589967.50	21683.0176	1446.19149
Turkija	135	194857.50	338917.50	16605.1953	1443.38889

Iš gautų rezultatų matome, kad nulinė hipotezė atmestina (p-reikšmė < 0,0001) – šalis turi įtakos mokytojų darbo stažo vidurkiui (2.7 lentelė). Labiausiai išsiskiria Bosnija ir Hercegovina – rangų vidurkis yra didžiausias. Mažiausias rangų vidurkis yra Maltoje ir Turkijoje (2.6 lentelė).

Analizės rezultatai, ar mokinio amžiui turi įtakos šalis, pateikti 3 priede.

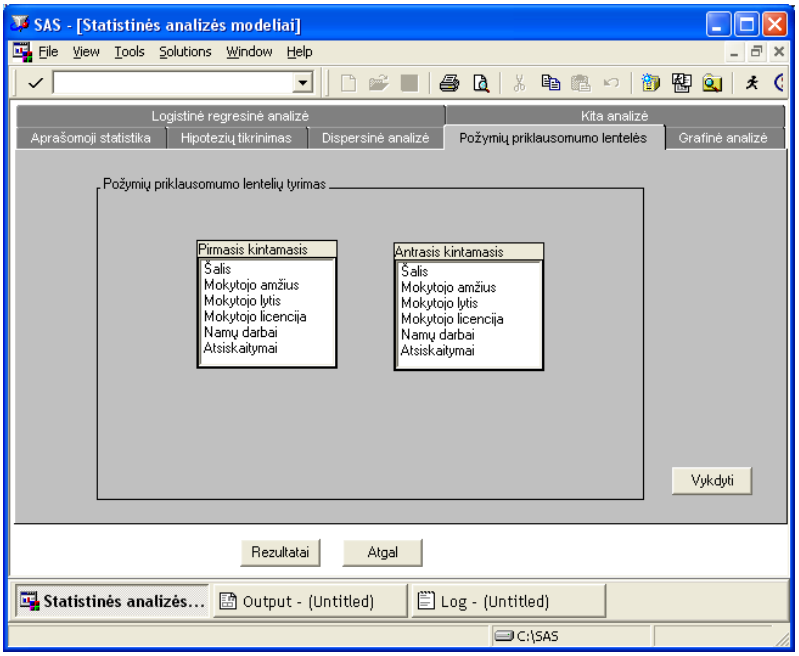
2.7 lentelė

Kruskalo-Voliso statistika

Statistika	Reikšmė
Chi kvadrato	653.2605
Laisvės laipsnių skaičius	19
p-reikšmė	<.0001

2.4. KORELIACINĖ ANALIZĖ

Požymių priklausomumo lentelių analizės langas matomas 2.6 paveiksle.



2.6 pav. Požymių priklausomumo lentelių tyrimo langas

Analizuojama požymių priklausomumo lentelė, kai požymiai yra *Mokytojo amžius* ir *Mokytojo lytis*. Duomenys yra gauti iš visų šalių.

Nulinė hipotezė atmesta ($p < 0.05$). Statistiškai įrodėme, kad požymiai *Mokytojo amžius* ir *Mokytojo lytis* yra priklausomi, tarp jų yra silpnas ryšys – Kramerio ryšio stiprumo koeficientas lygus 0.15 (2.9 lentelė). Kaip matome iš lentelės, daugiausiai matematikos mokytojų moterų yra 40-49 metų amžiaus, o vyrų – 50-59 metų amžiaus. Mažiausiai moterų ir vyrų mokytojų yra iki 25 metų. Europos šalyse daugiausiai aštuntos klasės matematikos mokytojų yra 50-59 metų amžiaus (30.46 %),

mažiausiai – iki 25 metų amžiaus (2.14 %). Net 71.48 procentai visų matematikos mokytojų yra moterys. (2.8 lentelė)

2.8 lentelė

Požymių priklausomumo lentelė

		Mokytojo lytis		Iš viso
		Moteris	Vyras	
Mokytojo amžius				
50-59	Dažnis	1118	448	1566
	Visumos procentai	21.75	8.71	30.46
	Eilutės procentai	71.39	28.61	100
	Stulpelio procentai	30.42	30.56	60.98
30-39	Dažnis	809	346	1155
	Visumos procentai	15.74	6.73	22.47
	Eilutės procentai	70.04	29.96	100
	Stulpelio procentai	22.01	23.60	45.61
40-49	Dažnis	1154	303	1457
	Visumos procentai	22.45	5.89	28.34
	Eilutės procentai	79.20	20.80	100
	Stulpelio procentai	31.40	20.67	52.07
60<	Dažnis	178	170	348
	Visumos procentai	3.46	3.31	6.77
	Eilutės procentai	51.15	48.85	100
	Stulpelio procentai	4.84	11.60	16.44
25-29	Dažnis	340	165	505
	Visumos procentai	6.61	3.21	9.82
	Eilutės procentai	67.33	32.67	100
	Stulpelio procentai	9.25	11.26	20.51
<25	Dažnis	76	34	110
	Visumos procentai	1.48	0.66	2.14
	Eilutės procentai	69.09	30.91	100
	Stulpelio procentai	2.07	2.32	4.39
Iš viso	Dažnis	3675	1466	5141
	Visumos procentai	71.48	28.52	100

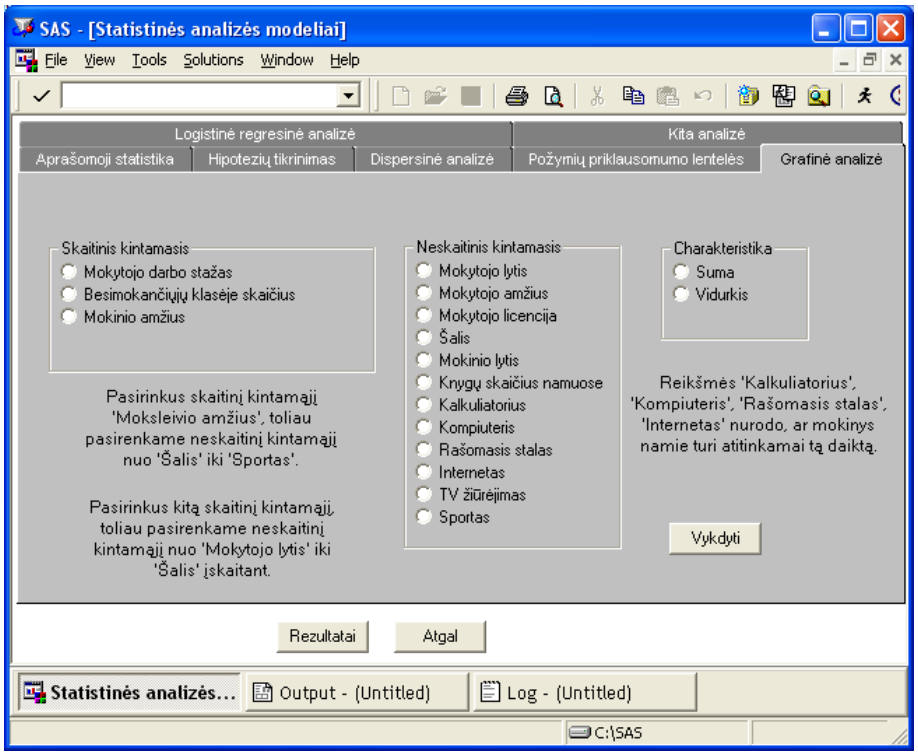
2.9 lentelė

Požymių ryšio stiprumo įvertinimo lentelė

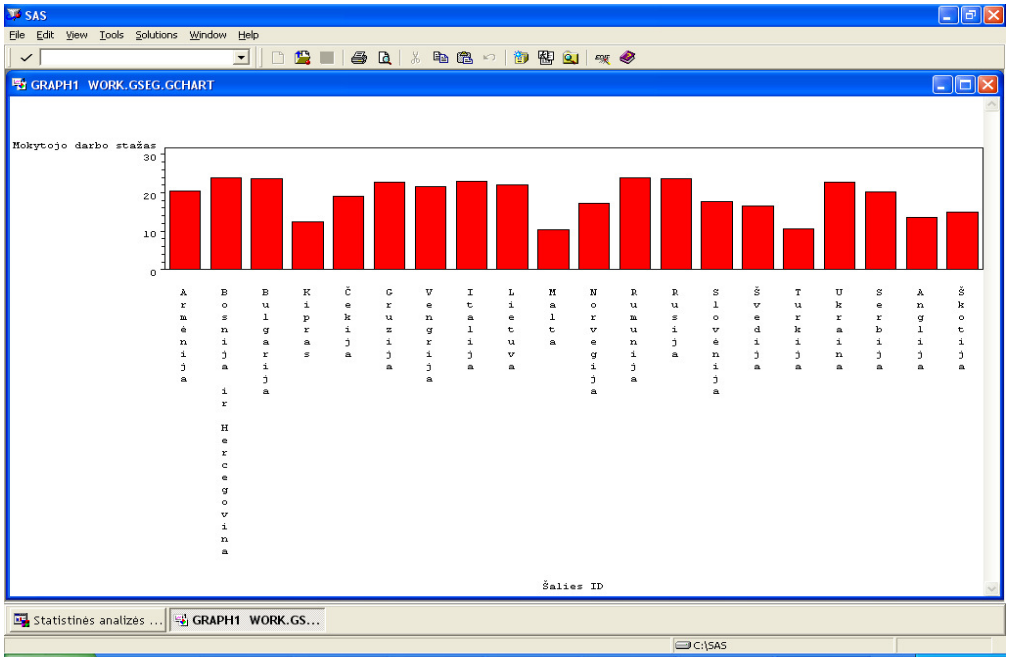
Statistika	Laisvės laipsnių skaičius	Stebėta statistikos reikšmė	p-reikšmė
Chi-kvadratu	5	118.9621	<.0001
Kramerio V		0.1521	

2.5. GRAFINĖ ANALIZĖ

Grafinės analizės langas matomas 2.7 paveiksle.



2.7 pav. Grafinės analizės langas

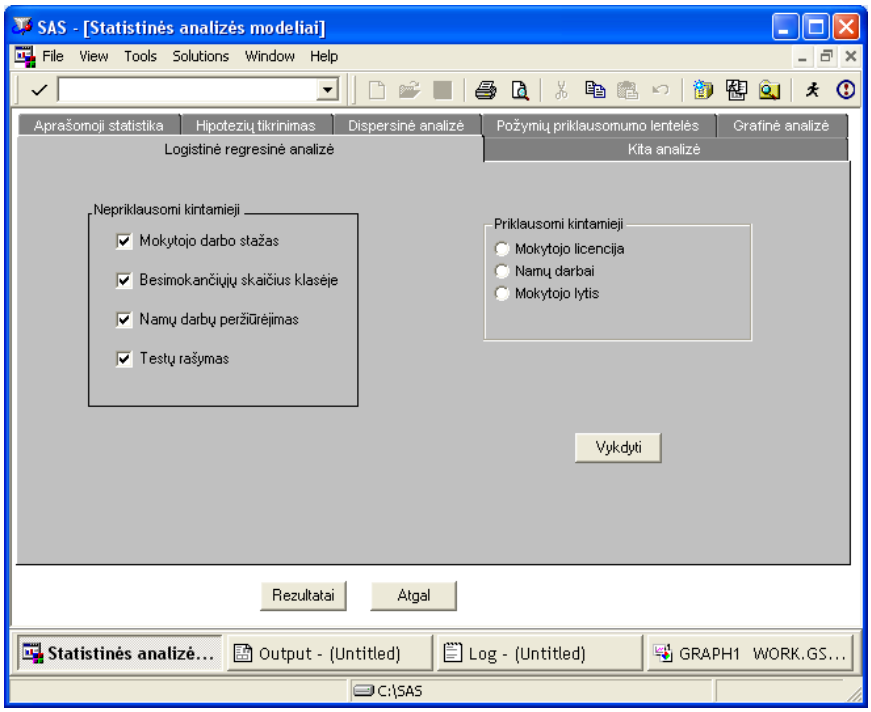


2.8 pav. Mokytojo darbo stažo vidurkio ir šalies priklausomybė

Grafinė analizė padeda vizualiai įvertinti duomenis. Vaizdavimui galime pasirinkti suminius kintamuosius arba jų vidurkius. Analizuojamas skaitinis kintamasis *Mokytojo darbo stažas* ir neskaitinis kintamasis *Šalis*. 2.8 paveiksle matome, kaip priklauso mokytojo darbo stažo vidurkis nuo šalies. Mažiausi matematikos mokytojų darbo stažo vidurkiai yra Kipre, Maltoje ir Turkijoje. Didžiausi vidurkiai yra Bosnijoje ir Hercegovinoje, Rumunijoje ir Bulgarijoje.

2.6. LOGISTINĖ REGRESIJOS ANALIZĖ

Kaip matome iš logistinės regresijos analizės lango (2.9 pav.), vartotojas gali pasirinkti vieną iš trijų priklausomų kintamųjų. Pasirinkus visus nepriklausomus kintamuosius ir priklausomąjį kintamąjį *Mokytojo lytis*, gaunami žemiau pateikti rezultatai. Analizuojami visų šalių duomenys.



2.9 pav. Logistinės regresijos analizės langas

2.10 lentelė

Regresijos lygties koeficientų reikšmingumas

H_0 : BETA=0			
Statistika	**Chi-kvadratu**	**Laisvės laipsnių skaičius**	**p-reikšmė**
Valdo	355.42	179	<.0001

2.11 lentelė

Kintamųjų reikšmingumo analizės rezultatai

Kintamasis	Laisvės laipsnių skaičius	Valdo Chi-kvadratu statistika	p-reikšmė
Mokytojo darbo stažas	49	123.77	<.0001
Besimokančiųjų skaičius klasėje	58	55.53	<.0001
Namų darbų peržiūrėjimas	34	83.89	0.5675
Testų rašymas	38	95.29	<.0001

2.12 lentelė

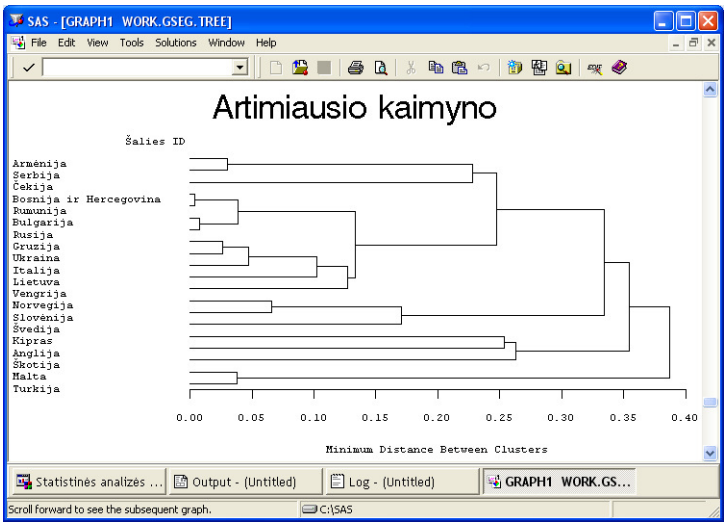
Ryšio stiprumo tarp prognozuojamų ir stebėtų tikimybių nustatymas

Statistika	Stebėta reikšmė	Statistika	Stebėta reikšmė
Suderintų porų procentas	70.0	Samerio D	0.404
Nesuderintų porų procentas	29.6	Gudmeno-Kruskalo gama	0.406
Surištų porų procentas	0.4	Kendalo Tau-a	0.161
Porų skaičius	4149824	Hanley ir Makveilo c	0.702

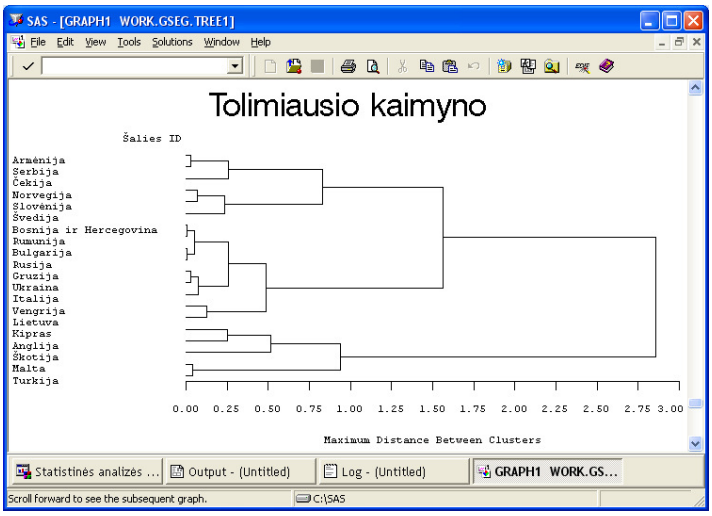
Logistinės regresijos analizės modelis yra korektiškas, nes *p*-reikšmė < 0.05 (2.10 lentelė). Kaip matome iš 2.11 lentelės, visi nepriklausomi kintamieji, išskyrus kintamąjį *Besimokančiųjų skaičius klasėje*, yra reikšmingi. Taigi tikimybė, kad mokytoja bus moteris, priklauso nuo kintamųjų *Mokytojo darbo stažas, Namų darbų peržiūrėjimas* ir *Testų rašymas,* nes jų *p*-reikšmė < 0.05 (2.11 lentelė). Tačiau prognozė yra vidutiniškai patikima, nes ryšio stiprumo laipsnis gama yra lygus 0.406 (2.12 lentelė).

2.7. KLASTERINĖ ANALIZĖ

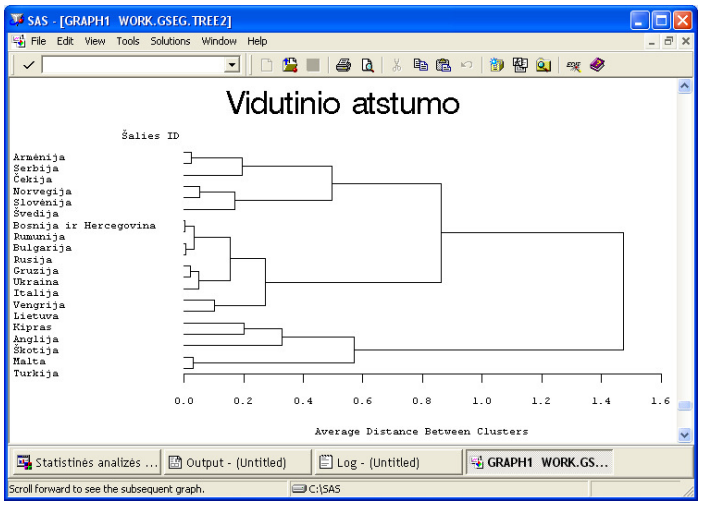
Klasterinės analizės langas matomas 2.14 paveiksle. Kintamasis – *Mokytojo darbo stažas,* kuris klasifikuojamas pagal šalis keturiais būdais: artimiausio kaimyno, tolimiausio kaimyno, vidutinio atstumo ir atstumo tarp centrų. Kaip matome iš 2.10, 2.11, 2.12 ir 2.13 paveikslų, visais, išskyrus 2.10 pav., atvejais klasterizavimas yra toks pats, kai išsiskiria 2 pagrindiniai klasteriai: vieną grupę sudaro Kipras, Anglija, Škotija, Malta ir Turkija, o kitą – likusios analizuojamos Europos šalys.



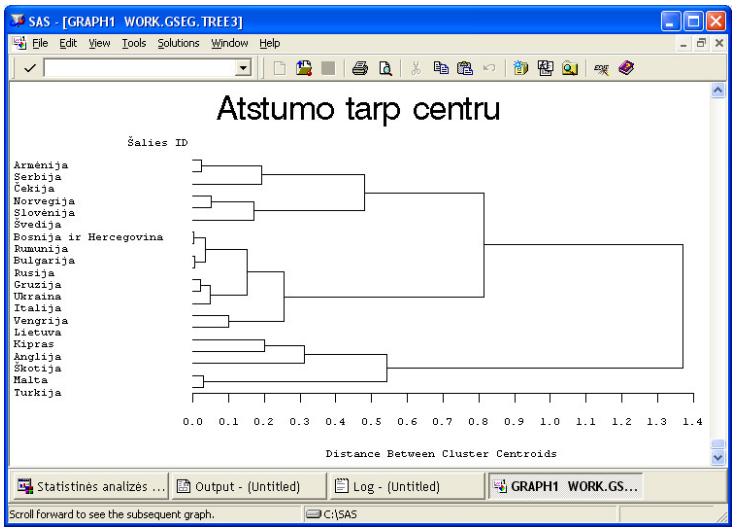
2.10 pav. Dendograma pagal artimiausio kaimyno metodą



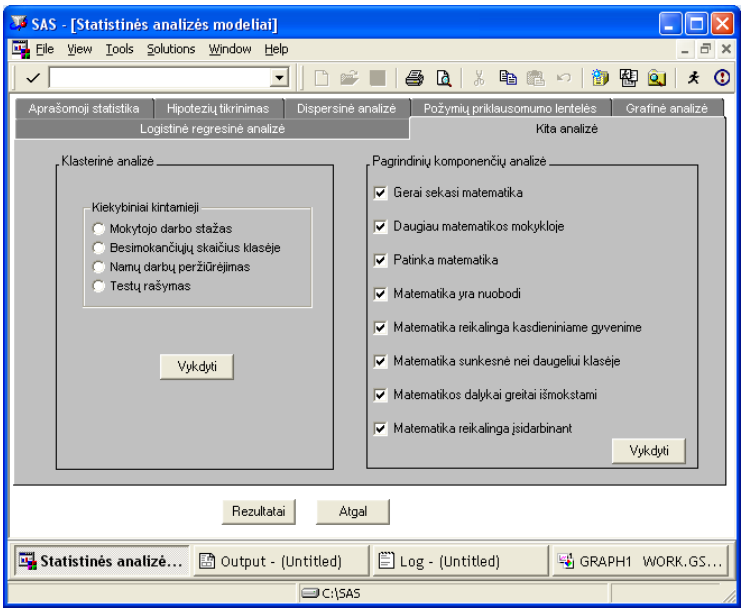
2.11 pav. Dendograma pagal tolimiausio kaimyno metodą



2.12 pav. Dendograma pagal vidutinio atstumo metodą



2.13 pav. Dendograma pagal atstumo tarp centrų metodą



2.14 pav. Klasterinės ir pagrindinių komponenčių analizių langas

2.8. PAGRINDINIŲ KOMPONENČIŲ ANALIZĖ

Sukurtos programinės įrangos interaktyvios sąsajos pagrindinių komponenčių analizės langas pavaizduotas 2.14 paveiksle. Analizuojami kintamieji (*Gerai sekasi matematika, Daugiau matematikos mokykloje, Matematika yra nuobodi, Matematika reikalinga kasdieniniame gyvenime, Matematika sunkesnė nei daugeliui klasėje, Matematikos dalykai greitai išmokstami, Matematika reikalinga įsidarbinant*) iš mokinių failo. Kiekvieną kintamąjį sudaro 4 galimos skirtingos reikšmės (labai

sutinka, truputį sutinka, truputį nesutinka, visiškai nesutinka), kurios nurodo, ką mokiniai galvoja apie šiuos teiginius. Analizuojami visų šalių duomenys. Iš 2.13 lentelės matome, kad gaunamos dvi komponentės, kadangi jų tikrinės reikšmės yra didesnės už vienetą ir jose yra 56 procentai informacijos.

2.13 lentelė

Tikrinės reikšmės

	Tikrinė reikšmė	**Skirtumas**	**Proporcija**	**Kaupiamasis dydis**
1	2.63283737	1.34418484	0.3761	0.3761
2	1.28865253	0.37989867	0.1841	0.5602
3	0.90875386	0.24947288	0.1298	0.6900
4	0.65928097	0.02516818	0.0942	0.7842
5	0.63411279	0.14448338	0.0906	0.8748
6	0.48962941	0.10289634	0.0699	0.9448
7	0.38673307		0.0552	1.0000

2.14 lentelė

Faktorių pasiskirstymas

Kintamasis	**Faktorius Nr. 1**	**Faktorius Nr. 2**
Gerai sekasi matematika	78*	24
Daugiau matematikos mokykloje	21	66*
Matematika yra nuobodi	-39	-53*
Matematika reikalinga kasdieniniame gyvenime	5	72*
Matematika sunkesnė nei daugeliui klasėje	-79*	9
Matematikos dalykai greitai išmokstami	80*	22
Matematika reikalinga įsidarbinant	2	70*

Iš 2.14 lentelės matome, kad pirmai komponentei priskiriami kintamieji: *Gerai sekasi matematika, Matematika sunkesnė nei daugeliui klasėje, Matematikos dalykai greitai išmokstami,* o antrai komponentei: *Daugiau matematikos mokykloje, Matematika yra nuobodi, Matematika*

reikalinga kasdieniniame gyvenime, Matematika reikalinga įsidarbinant. 2.15 lentelėje pateiktos komponenčių koeficientų reikšmės.

2.15 lentelė

Koeficientų reikšmės

Kintamasis	Faktorius Nr. 1	Faktorius Nr. 2
Gerai sekasi matematika	0.37870	-0.00756
Daugiau matematikos mokykloje	-0.01307	0.36353
Matematika yra nuobodi	-0.10583	-0.25171
Matematika reikalinga kasdieniniame gyvenime	-0.11464	0.43156
Matematika sunkesnė nei daugeliui klasėje	-0.44526	0.20679
Matematikos dalykai greitai išmokstami	0.39434	-0.02204
Matematika reikalinga įsidarbinant	-0.14754	0.43278

Pagrindinių komponenčių išraiškos:

$$Z_1 = 0.38X_1 - 0.01X_2 - 0.11X_3 - 0.11X_4 - 0.45X_5 + 0.39X_6 - 0.15X_7,$$

$$Z_2 = -0.01X_1 + 0.36X_2 - 0.25X_3 + 0.43X_4 + 0.21X_5 - 0.02X_6 + 0.43X_7.$$

3. PROGRAMINĖ REALIZACIJA IR INSTRUKCIJA VARTOTOJUI

3.1. VARTOTOJO SĄSAJOS STRUKTŪRA

Duomenys IEA duomenų bazėje yra pateikti SAS duomenų failo formatu *.sas7bdat. Kiekvienos šalies duomenys saugomi atskirai. Failų pavadinimai ir trumpi aprašymai pateikiami 3.1 lentelėje.

3.1 lentelė

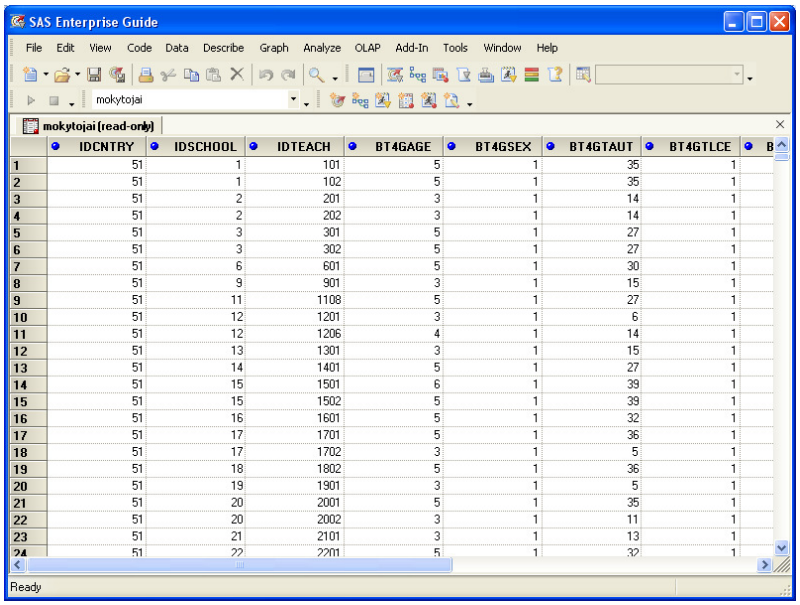
Duomenų failų aprašymas

NR.	Failo pavadinimas	Aprašymas
1.	BSGAUTM4	Austrijos mokyklų mokinių anketų rezultatai
2.	BSGCZEM4	Čekijos mokyklų mokinių anketų rezultatai
3.	BSGDNKM4	Danijos mokyklų mokinių anketų rezultatai
4.	BSGENGM4	Anglijos mokyklų mokinių anketų rezultatai
5.	BSGDEUM4	Vokietijos mokyklų mokinių anketų rezultatai
6.	BSGHUNM4	Vengrijos mokyklų mokinių anketų rezultatai
7.	BSGITAM4	Italijos mokyklų mokinių anketų rezultatai
	<...>	
21.	BTGAUTM4	Austrijos mokyklų matematikos mokytojų anketų rezultatai
22.	BTGCZEM4	Čekijos mokyklų matematikos mokytojų anketų rezultatai
23.	BTGDNKM4	Danijos mokyklų matematikos mokytojų anketų rezultatai
24.	BTGENGM4	Anglijos mokyklų matematikos mokytojų anketų rezultatai
25.	BTGDEUM4	Vokietijos mokyklų matematikos mokytojų anketų rezultatai
26.	BTGHUNM4	Vengrijos mokyklų matematikos mokytojų anketų rezultatai
27.	BTGITAM4	Italijos mokyklų matematikos mokytojų anketų rezultatai
	<...>	

Sukurta sistema yra programinių priemonių paketas skirtas statistinei analizei atlikti. Ši programinė sistema yra parašyta SAS (angl. – Statistical Analysis System) paketo aplinkoje. Sistema suprantama kiekvienam vartotojui.

Dėl didelies apimties duomenų failų, paleidus vartotojo sąsają, yra iš karto nuskaitomos tik tos reikšmės, kurios yra naudojamos darbe, Europos valstybių duomenys yra sujungiami į vieną failą. Taip

sukuriami du failai, kur viename saugomi duomenys iš mokinių apklausos anketų, o kitame – iš matematikos mokytojų apklausos anketų (3.1 pav.).

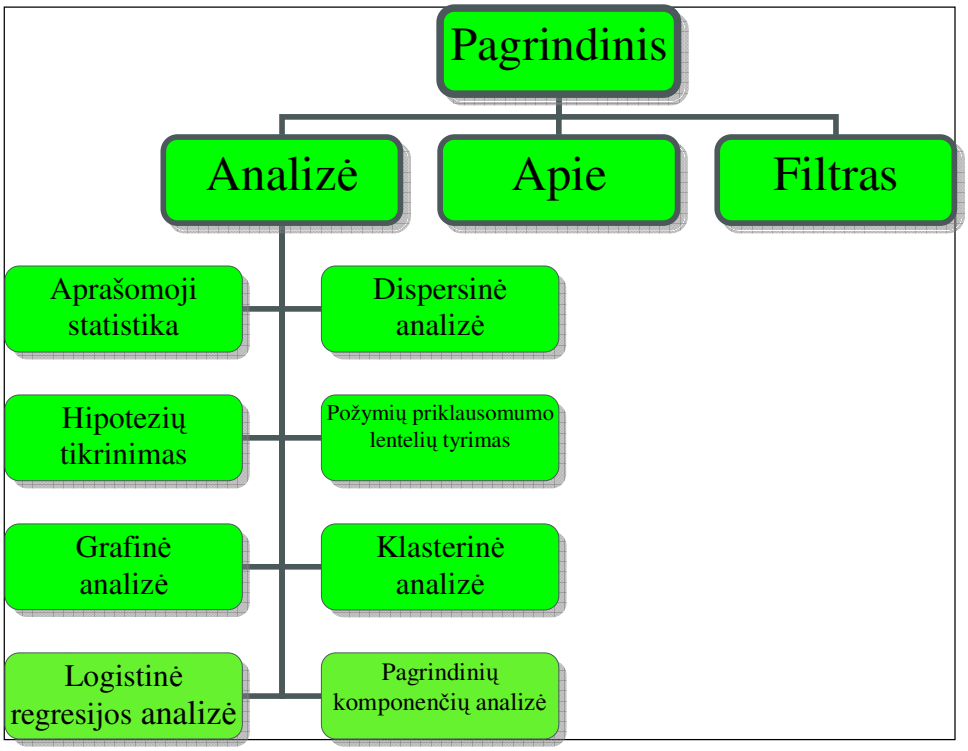


3.1 pav. Mokytojai.sas7bdat duomenų failas

Atlikus duomenų analizę ir įvertinus jose saugomų kintamųjų reikšmingumą sprendimų priėmimui, atrinkti kintamieji statistinei analizei. Sukurtą programą nesunkiai galima papildyti naujais duomenų failais ir juos analizuoti, taip pat nesunkiai galima sistemą papildyti naujais statistinės analizės modeliais.

Sistemos struktūra pateikta 3.2 paveiksle. Programą sudaro trys posistemės. „Apie“ lange pateikiama informacija apie magistro baigiamojo darbo temos pavadinimą, jos autorių ir darbo vadovą. Iš šio lango grįžtama į pagrindinį langą. „Filtras“ - duomenų paruošimo statistinei analizei posistemė, kuri formuoja duomenų pjūvius ir paruošia duomenis statistinių metodų taikymui. Statistinės analizės posistemę sudaro įvairūs duomenų analizės modeliai.

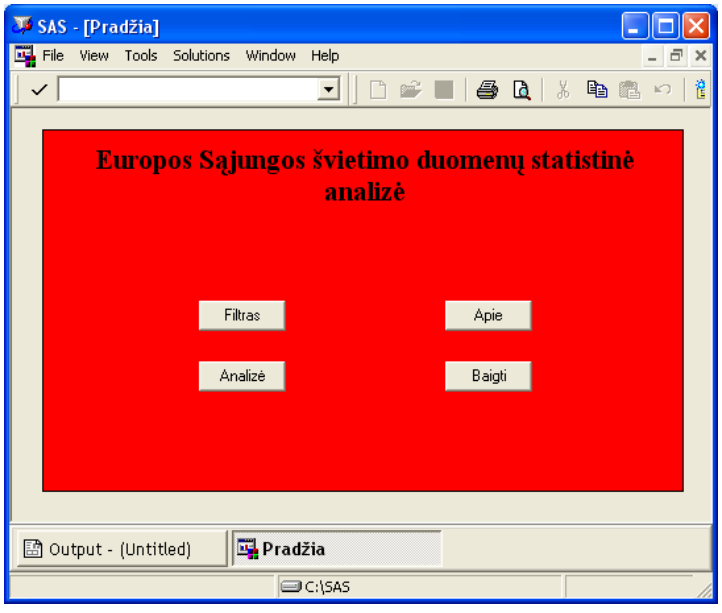
Sisteminiame faile _config.cfg nurodytas programos sesijos metu naudojamo šrifto tipas ir jo dydis, paleidimo metu atliekami veiksmai, o taip pat išvardinti naudojami SAS moduliai bei nurodyta jų fizinė vieta kompiuteryje. Faile bibliotekos.sas programiškai aprašyti sukurtos analizės programinės įrangos loginių jungčių fiziniai adresai kompiuterio katalogų struktūroje. Sistemos paleidimo failas yra SAS.lnk. Jame nurodyta, kuriame kataloge programa turi pradėti darbą, nurodytas SAS sistemos paleidimo failas sas.exe, konfigūracinio failo _config.cfg vieta ir nurodyta komanda, kuri aktyvuoja pagrindinę sąsajos su vartotoju posistemę.



3.2 pav. Programos struktūra

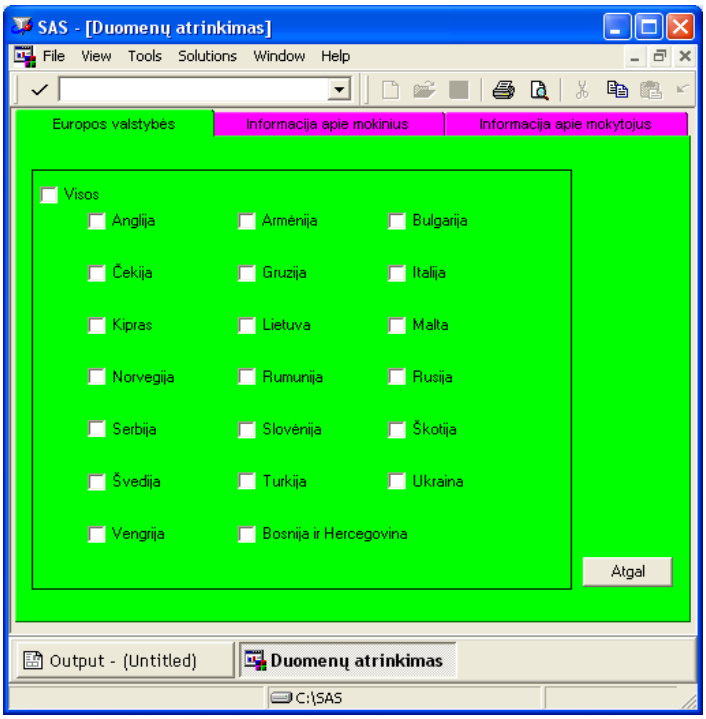
3.2. VARTOTOJO SĄSAJA

Paleidus programą, iškviečiamas pagrindinis meniu langas (3.3 pav.).



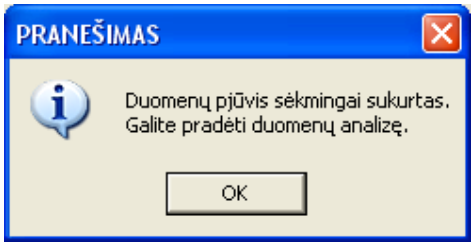
3.3 pav. Vartotojo sąsajos pagrindinis langas

Europos valstybių švietimo duomenų atrinkimas atliekamas formavimo meniu languose (3.4 pav.).



3.4 pav. Pjūvių formavimo langas

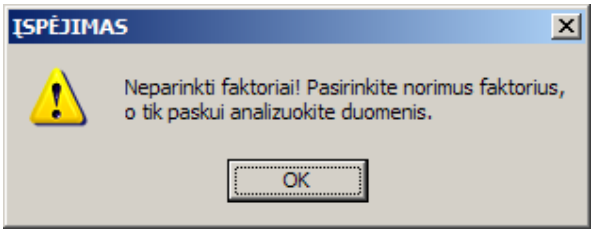
Pjūvių formavimo languose vartotojas turi nurodyti parametrus, pagal kuriuos nori formuoti duomenų pjūvį. Pjūvių formavimas susideda iš trijų blokų: Europos valstybės, Informacija apie mokinius, Informacija apie mokytojus, kur atitinkamai suformuojamos kintamųjų *Šalis, Moksleivio lytis, Mokytojo lytis* reikšmės. Sužymėjus parametrus, spaudžiamas mygtukas „Filtruoti", kuris pagal nurodytus parametrus suformuoja duomenų pjūvį skirtą statistinei analizei. „Filtruoti" spaudžiame, kai sužymime pasirinktus parametrus apie mokinius, ir dar kartą spaudžiame, kai sužymime pasirinktus parametrus apie mokytojus. Nenurodžius nė vieno parametro, pagal kurį formuojamas duomenų bazės pjūvis, programa analizuoja visus tuo metu duomenų bazėje esančius duomenis. Kai pjūvis atliktas, rodomas pranešimas (3.5 pav.).



3.5 pav. Pranešimas apie atliktą duomenų pjūvį

SAS programų įvykių žurnalai, bei duomenų grafinio vaizdavimo rezultatai pateikiami atskiruose languose: rezultatai lentelių pavidalu išvedami į rezultatų langą (angl. – Output), o grafinės analizės rezultatai pateikiami GRAPH1 lange.

Vartotojo sąsaja sukurta naudojant SCL kalbą, puslapio antraštes, mygtukus, linijas, rėmelius, paveikslėlius, akutes, žymimuosius langelius, pranešimų langus. FRAME langui aktyvuoti, naudojama ekrano valdymo kalba SCL, kurios pagalba aprašomos objektų, esančių FRAME lange, funkcijos (4).



3.6 pav. Įspėjimas apie klaidą

Baigęs filtravimą, vartotojas turi spausti mygtuką „Atgal", kad uždarytų pjūvių formavimo langą ir sugrįžtų į pagrindinį meniu. Statistinė duomenų analizė atliekama po to, kai sudarytas duomenų pjūvis. Pagrindiniame lange vartotojas jau gali paspaudęs mygtuką „Analizė" atlikti statistinę sudaryto duomenų pjūvio analizę. Analizei atlikti vartotojas turi pasirinkti rodiklius, kintamuosius, grupes. Nepasirinkus jų, programa apie tai informuos išvesdama įspėjimą apie negalimą atlikti analizę ir pateiks priežastį, kodėl vartotojas negali atlikti analizės. Pavyzdys pateiktas 3.6 pav.

4. DISKUSIJA

Sukurta sistema analizuoja Europos valstybių švietimo duomenis įvairiais pjūviais, atlieka aprašomąją statistiką, hipotezių tikrinimo, požymių priklausomumo lentelių tyrimo, dispersinę, regresinę, klasterinę, pagrindinių komponenčių ir faktorinę analizes.

Kintamasis *Mokytojo darbo stažas*, kai duomenys yra gauti iš visų analizuojamų šalių, nėra pasiskirstęs pagal normalųjį skirstinį. Todėl yra atliekama ranginė dispresinė analizė. Remiantis gautais rezultatais, faktorius *Šalis* turi įtakos kintamajo *Mokytojo darbo stažas* vidurkiui. Labiausiai išsiskiria Bosnija ir Hercegovina, Rumunija ir Bulgarija – rangų vidurkis yra didžiausias. Mažiausias rangų vidurkis yra Maltoje, Turkijoje, Anglijoje ir Škotijoje. Iš grafinės analizės rezultatų matome, kad mažiausi darbo stažo vidurkiai yra Kipre, Maltoje, Turkijoje, Anglijoje ir Škotijoje. Didžiausi vidurkiai yra Bosnijoje ir Hercegovinoje, Rumunijoje ir Bulgarijoje. Atlikus klasterizavimą, išskiriami 2 pagrindiniai klasteriai: vieną grupę sudaro Kipras, Anglija, Škotija, Malta ir Turkija, o kitą – likusios analizuojamos Europos šalys. Taigi, atlikus dispresinę, grafinę ir klasterinę analizę pagal mokytojų darbo stažo vidurkius kiekvienoje Europos šalyje, galima išskiriamos dvi pagrindinės grupės. Pirmai grupei (su mažesniu mokytojų darbo stažo vidurkiu) priklauso Kipras, Anglija, Škotija, Malta ir Turkija, o antrai grupei (su didesniu mokytojų darbo stažo vidurkiu) priklauso Bosnija ir Hercegovina, Rumunija, Bulgarija, Rusija, Italija, Ukraina, Gruzija, Lietuva, Vengrija, Armėnija, Serbija, Čekija, Slovėnija, Norvegija ir Švedija. Taip pat galimos 3 pagrindinės grupės (išdėstyta nuo grupės tų šalių, kurių aštuntų klasių matematikos mokytojų darbo stažo vidurkiai yra mažiausi, iki grupės, kurios šalių mokytojų darbo stažo vidurkiai yra didžiausi):

1. Škotija, Anglija, Kipras, Malta ir Turkija.
2. Armėnija, Serbija, Čekija, Slovėnija, Norvegija ir Švedija.
3. Bosnija ir Hercegovina, Rumunija, Bulgarija, Rusija, Italija, Ukraina, Gruzija, Lietuva ir Vengrija.

Daugiausiai matematikos mokytojų moterų yra 40-49 metų amžiaus, o vyrų – 50-59 metų amžiaus. Mažiausiai moterų ir vyrų mokytojų yra iki 25 metų. Europos šalyse daugiausiai aštuntos klasės matematikos mokytojų yra 50-59 metų amžiaus (30.46 %), mažiausiai – iki 25 metų amžiaus (2.14 %). Net 71.48 procentai visų matematikos mokytojų yra moterys.

Kadangi amžius ir darbo stažas yra tiesiogiai proporcingas, tai galima teigti, kad atitinkamų šalių matematikos mokytojų amžiaus vidurkis yra mažiausias pirmoje grupėje, o didžiausias trečioje grupėje. Vadinasi trečios grupės šalyse mokytojai turėjo daugiausiai patirties.

IŠVADOS

1. Panaudojus SAS objektinio programavimo priemones ir įvertinus TIMSS 2007 duomenų bazės struktūrą bei galimybes, sukurta statistinės analizės sistema, kuri leidžia vartotojui formuoti duomenų pjūvius ir juos analizuoti taikant duomenų statistinės analizės metodus.
2. Vartotojui, dirbančiam su sistema, pakanka minimalių darbo su SAS paketu pagrindų. Net nepatyręs vartotojas gali nesunkiai formuluoti uždavinį ir atlikti analizę. Sukurta sąsajos su vartotoju sistema įspėja apie klaidas, informuoja apie jų atsiradimo priežastis, o tai ženkliai palengvina analizės procesą ir sumažina reikalavimus vartotojo pasirengimui informatikos bei SAS sistemos naudojimo srityje.
3. Sukurtos sistemos testavimas, atliktas naudojant realius Europos valstybės švietimo duomenis, parodė, kad sistema yra pajėgi spręsti statistinės analizės užduotis. Gauti šie pagrindiniai rezultatai:
    1) Išskiriamos trys pagrindinės grupės šalių pagal aštuntų klasių matematikos mokytojų darbo stažo vidurkius didėjimo tvarka:
        a. Škotija, Anglija, Kipras, Malta ir Turkija.
        b. Armėnija, Serbija, Čekija, Slovėnija, Norvegija ir Švedija.
        c. Bosnija ir Hercegovina, Rumunija, Bulgarija, Rusija, Italija, Ukraina, Gruzija, Lietuva ir Vengrija.
    2) Mokytojų darbo stažo vidurkis nagrinėjamose Europos šalyse yra 18.98 metai.
    3) Europos šalyse daugiausiai aštuntos klasės matematikos mokytojų yra 50-59 metų amžiaus (30.46 %), mažiausiai – iki 25 metų amžiaus (2.14 %). Net 71.48 % visų matematikos mokytojų yra moterys.

REKOMENDACIJOS

Darbo privalumas yra tas, kad didžiąją dalį gautų rezultatų nesunkiai galima pritaikyti kuriant sprendimų paramos sistemas kitose probleminėse srityse arba toliau gilintis į švietimo sistemos problemas. Rekomenduočiau analizuoti ne tik Europos, bet ir kitų žemynų duomenis, aprėpiant didesnį kiekį svarbių analizuoti kintamųjų. Taip pat išsiaiškinti, kas lemia aštuntų klasių matematikos mokytojų darbo stažo vidurkių skirtumus kiekvienoje šalyje.

PADĖKOS

Dėkoju dr. Tomui Ruzgui už pagalbą rašant magistro darbą, taip pat šeimos nariams už kantrybę ir palaikymą.

ŠALTINIAI IR LITERATŪRA

1. Čekanavičius V., Murauskas G. Statistika ir jos taikymai. I. – Vilnius: TEV, 2003. – 240 psl.
2. Čekanavičius V., Murauskas G. Statistika ir jos taikymai, II. – Vilnius: TEV, 2004. – 272 psl.
3. Elliott R.J. Learning SAS in the Computer Lab. – Belmont, California, USA: Duxbury Press, 1995. – 175 p.
4. Getting Started with the FRAME Entry: Developing Object-Oriented Applications, Second Edition. – Cary, NC, USA: SAS Institute Inc., 1997. – 72 p.
5. Hatcher L. A Step-by-Step Approach to Using the SAS. System for Factor Analysis and Structural Equation Modeling. – Cary, NC, USA: SAS Institute Inc., 1994. – 588 p.
6. International Association for the Evaluation of Educational Achievement (IEA) – [žiūrėta 2011-05-26]. Prieiga per internetą: <http://rms.iea-dpc.org/>.
7. Kazakevičiūtė J., Ruzgas T., Europos sąjungos švietimo duomenų statistinė analizė. Taikomoji matematika / Applied mathematics. VIII studentų konferencijos pranešimų medžiaga. – Kaunas: Technologija, 2010. – p. 68-69.
8. Mullis, I.V.S., Martin, M.O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. – Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2008. – 476 p.
9. SAS® – [žiūrėta 2011-05-26]. Prieiga per internetą: <http://www.sas.com>.
10. SAS OnlineDoc®. CD documentation. – Cary, NC, USA: SAS Institute Inc. 1999.
11. SAS User's Guide: Statistics. – Cary, NC, USA: SAS Institute Inc., 1985. – 956 p.
12. Stokes M.E. Categorical Data Analysis Using The SAS System./ Stokes M.E., Davis C.S., Koch G.G. – Cary, NC, USA: SAS Institute Inc., 1995. – 499 p.

1 PRIEDAS. APRAŠOMOSIOS STATISTIKOS REZULTATAI

Analizuojamas kintamasis *Amžius*. Rezultatai, kai duomenys gauti iš visų šalių mokinių anketų, matomi 1 lentelėje, o kai duomenys, gauti tik iš Lietuvos, matomi 2 lentelėje.

1 lentelė

Kintamojo *Amžius* skaitinės charakteristikos

Stebėjimų skaičius	Trūkstamų stebėjimų skaičius	Imties plotis	Suma	Vidurkis	Mažiausia reikšmė
87065	82	8.083	1250874.67	14.367	10.5
Didžiausia reikšmė	Dispersija	Standartinis nuokrypis	Asimetrijos koeficientas	Eksceso koeficientas	
18.583	0.404	0.635	0.372	1.082	

Iš 1 lentelės matome, kad mokinių amžiaus vidurkis yra 14.38 metai, didžiausia amžiaus reikšmė yra 18.58 metai, o mažiausia – 10.5 metai.

2 lentelė

Kintamojo *Amžius* skaitinės charakteristikos

Stebėjimų skaičius	Trūkstamų stebėjimų skaičius	Imties plotis	Suma	Vidurkis	Mažiausia reikšmė
3991	0	5	59431.08	14.891	13.167
Didžiausia reikšmė	Dispersija	Standartinis nuokrypis	Asimetrijos koeficientas	Eksceso koeficientas	
18.167	0.176	0.419	0.623	3.748	

Iš 2 lentelės matome, kad Lietuvoje besimokančiųjų vidurkis yra 14.89 metai, didžiausia amžiaus reikšmė yra 18.17 metai, o mažiausia – 13.17 metai. Palyginus Lietuvos ir visų šalių kintamojo *Amžius* skaitines charakteristikas, matome, kad vidurkiai ir didžiausios reikšmės skiriasi nežymiai. Tačiau Lietuvos mokinių amžiaus mažiausia reikšmė, asimetrijos ir eksceso koeficientai yra didesni palyginus su visomis šalimis.

2 PRIEDAS. HIPOTEZIŲ TIKRINIMO REZULTATAI

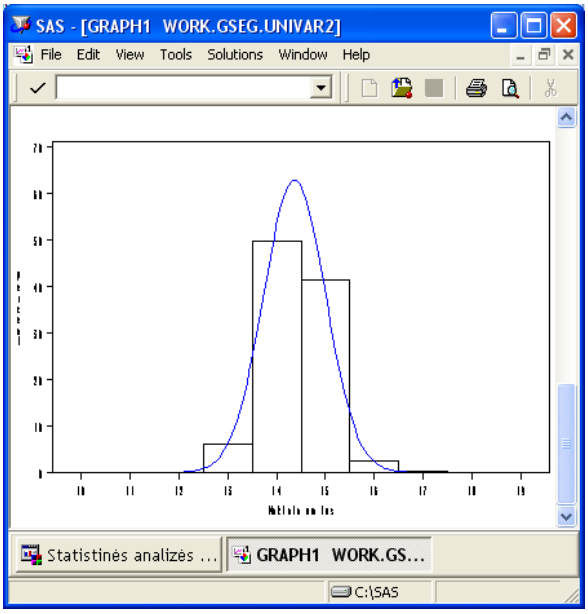
Analizuojamas kintamasis *Mokinio amžius,* nulinė hipotezė – kintamojo skirstinys yra normalusis. Rezultatai, kai duomenys gauti iš visų šalių, matomi 3 lentelėje, o kai duomenys gauti tik iš Lietuvos, matomi 4 lentelėje.

3 lentelė

Hipotezės tikrinimo rezultatai

	STATISTIKA	**P-REIKŠMĖ**
KOLMOGOROVO-SMIRNOVO	D 0.055767	Pr > D <0.010
KRAMERIO	W-Sq 38.911878	Pr > W-Sq <0.005
ANDERSONO-DARLINGO	A-Sq 279.218661	Pr > A-Sq <0.005

Matome, jog hipotezė H_0 atmetama, kadangi *p*-reikšmė < 0.05. Vadinasi skirstiniai nesuderinami (1 pav.).



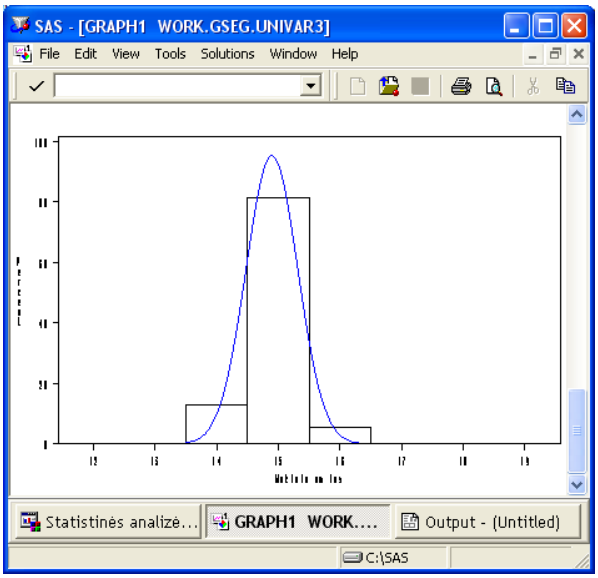
1 pav. Normalusis ir kintamojo *Mokinio amžius* skirstiniai

4 lentelė

Hipotezės tikrinimo rezultatai

	STATISTIKA	**P-REIKŠMĖ**
KOLMOGOROVO-SMIRNOVO	D 0.0592234	Pr > D <0.010
KRAMERIO	W-Sq 2.7747049	Pr > W-Sq <0.005
ANDERSONO-DARLINGO	A-Sq 19.8017287	Pr > W-Sq <0.005

Matome, jog hipotezė H_0 atmetama, kadangi *p*-reikšmė < 0.05, vadinasi skirstiniai nesuderinami (2.4 pav.).



2 pav. Normalusis ir kintamojo *Mokinio amžius* skirstiniai

3 PRIEDAS. DISPERSINĖS ANALIZĖS REZULTATAI

Analizuojama faktoriaus *Šalis* įtaka kintamajo *Mokinio amžius* vidurkiui.

5 lentelė

Vilkoksono rangų sumos

Šalis	N	Rangų suma	Tikėtina rangų suma	Std. nuokrypis	Rangų vidurkis
Rumunija	4198	290440353	182751534	1587491.11	69185.4103
Bulgarija	4018	267108662	174915594	1554770.23	66478.0144
Serbija	4044	266021423	176047452	1559548.30	65781.7564
Lietuva	3991	262492439	173740203	1549789.45	65771.0947
Armėnija	4689	304983261	204126237	1672783.53	65042.2821
Švedija	5215	334810748	227024595	1758473.28	64201.4857
Bosnija ir Hercegovina	4220	249869433	183709260	1591434.08	59210.7660
Vengrija	4110	221660658	178920630	1571598.01	53932.0336
Rusija	4472	221133092	194679576	1635768.33	49448.3657
Čekija	4845	226259827	210917385	1698771.25	46699.6546
Anglija	4025	155428289	175220325	1556058.39	38615.7239
Gruzija	4178	151944803	181880874	1583896.16	36367.8322
Ukraina	4424	146363592	192589992	1627438.62	33083.9945
Malta	4660	135597372	202863780	1667896.19	29098.1485
Turkija	4498	129785899	195811434	1640258.34	28854.1349
Italija	4407	100836885	191849931	1624475.81	22881.0722
Kipras	4331	91409676	188541423	1611147.79	21105.9053
Norvegija	4627	90244157	201427191	1662312.80	19503.8161
Slovėnija	4043	74372763.5	176003919	1559364.86	18395.4399
Škotija	4070	69437315.5	177179310	1564308.65	17060.7655

Iš gautų rezultatų matome, kad nulinė hipotezė atmestina ($p < 0{,}0001$) – šalis turi įtakos mokinių amžiaus vidurkiui (6 lentelė). Labiausiai išsiskiria Rumunija – rangų vidurkis yra didžiausias. Mažiausias rangų vidurkis yra Škotijoje (5 lentelė).

6 lentelė

Kruskalo-Voliso statistika

Statistika	Reikšmė
Chi kvadrato	47260.9201
Laisvės laipsnių skaičius	19
p-reikšmė	<.0001

4 PRIEDAS. PROGRAMOS TEKSTAS

PAGRINDINIS.SCL

```
init:
return;

pushbutton1:

    call display('SAS.programa.Apie.frame');

return;

pushbutton3:

      call display('SAS.programa.Filtras.frame');

return;

pushbutton4:

      call display('SAS.programa.Analize.frame');

return;
```

APIE.SCL

```
init:
return;
```

FILTRAS.SCL

```
init:

submit continue;
proc format;
value salis 051='Armėnija' 070='Bosnija ir Hercegovina' 100='Bulgarija'
196='Kipras'
203='Čekija' 268='Gruzija' 348='Vengrija' 380='Italija' 440='Lietuva' 470='Malta'
578='Norvegija' 642='Rumunija' 643='Rusija' 705='Slovėnija' 752='Švedija'
792='Turkija'
804='Ukraina' 891='Serbija' 926='Anglija' 927='Škotija' 99999='Nenurodyta';
value mokykla 9999='Nenurodyta';
value lytisa 1='Mergaitė' 2='Berniukas' 8,9='Nenurodyta';
value lytisb 1='Moteris' 2='Vyras' 8,9='Nenurodyta';
value knygos 1='0-10' 2='11-25' 3='26-100' 4='101-200' 5='200<' 8,9='Nenurodyta';
value turi 1='Turi' 2='Neturi' 8,9='Nenurodyta';
value sutinka 1='Labai sutinka' 2='Truputį sutinka' 3='Truputį nesutinka'
4='Visiškai nesutinka' 8,9='Nenurodyta';
value val 1='0' 2='(0;1)' 3='[1;2]' 4='(2;4)' 5='4 ir daugiau' 8,9='Nenurodyta';
value amziusa 99,98='Nenurodyta';
value amziusb 1='<25' 2='25-29' 3='30-39' 4='40-49' 5='50-59' 6='60<'
8,9='Nenurodyta';
value stazas 99,98='Nenurodyta';
value klase 998,999='Nenurodyta';
value nd 1='Užduoda' 2='Neužduoda' 8,9='Nenurodyta';
value kontr 1='1 k./sav.' 2='Kas 2 sav.' 3='1 k./mėn.' 4='Keletą kartų per metus'
5='Niekada' 8,9='Nenurodyta';
run;
```

```
      data mokiniai (keep=IDCNTRY IDSCHOOL IDSTUD ITSEX BS4GBOOK BS4GTH01 BS4GTH02
BS4GTH03 BS4GTH05 BS4MAWEL
                        BS4MAMOR BS4MALIK BS4MABOR BS4MAHDL BS4MACLM BS4MAQKY
BS4MAGET BS4GWATV BS4GPLCG BS4GPLSP
                        BS4GDOHW BSDAGE) label;
            set Duomenys.bsgarmm4 Duomenys.bsgbihm4 Duomenys.bsgbgrm4
Duomenys.bsgcypm4 Duomenys.bsgczem4
                Duomenys.bsgengm4 Duomenys.bsggeom4 Duomenys.bsghunm4
Duomenys.bsgitam4 Duomenys.bsgltum4
                Duomenys.bsgmltm4 Duomenys.bsgnorm4 Duomenys.bsgromm4
Duomenys.bsgrusm4 Duomenys.bsgscom4
                Duomenys.bsgscgm4 Duomenys.bsgsvnm4 Duomenys.bsgswem4
Duomenys.bsgturm4 Duomenys.bsgukrm4;

      label
      IDCNTRY = 'Šalies ID' IDSCHOOL = 'Mokyklos ID' IDSTUD = 'Mokinio ID'
      ITSEX = 'Mokinio lytis' BS4GBOOK = 'Knygų skaičius namuose' BS4GTH01 =
'Kalkuliatorius' BS4GTH02 = 'Kompiuteris'
      BS4GTH03 = 'Rašomasis stalas' BS4GTH05 = 'Internetas' BS4MAWEL = 'Gerai
sekasi matematika'
      BS4GWATV = 'Žiūri TV (val./d.)' BS4GPLCG = 'Žaidžia komp. žaidimus
(val./d.)' BS4GPLSP = 'Sportuoja (val./d.)'
      BS4GDOHW = 'Daro namų darbus (val./d.)' BSDAGE = 'Mokinio amžius' BS4MAMOR =
'Daugiau matematikos mokykloje'
      BS4MALIK = 'Patinka matematika' BS4MABOR = 'Matematika yra nuobodi' BS4MAHDL
= 'Matematika reikalinga kasdieniniame gyvenime'
      BS4MACLM = 'Matematika sunkesnė nei daugeliui klasėje' BS4MAQKY =
'Matematikos dalykai greitai išmokstami'
      BS4MAGET = 'Matematika reikalinga įsidarbinant';

      format
      IDCNTRY salis. IDSCHOOL mokykla. ITSEX lytisa. BS4GBOOK knygos. BS4GTH01
turi. BS4GTH02 turi.
      BS4GTH03 turi. BS4GTH05 turi. BS4GWATV val. BS4GPLCG val. BS4GPLSP val.
BS4GDOHW val.
      BSDAGE amziusa. BS4MAWEL sutinka. BS4MAMOR sutinka. BS4MALIK sutinka.
BS4MABOR sutinka. BS4MAHDL sutinka.
      BS4MACLM sutinka. BS4MAQKY sutinka. BS4MAGET sutinka.;

      run;

      data mokytojai (keep=IDCNTRY IDSCHOOL IDTEACH BT4GAGE BT4GSEX BT4GTAUT
BT4GTLCE BT4MSTUD BT4MHMWO BT4MTEEX BT4MPTRH BT4MPTTQ) label;
          set Duomenys.btmarmm4 Duomenys.btmbihm4 Duomenys.btmbgrm4
Duomenys.btmcypm4 Duomenys.btmczem4
                Duomenys.btmengm4 Duomenys.btmgeom4 Duomenys.btmhunm4
Duomenys.btmitam4 Duomenys.btmltum4
                Duomenys.btmmltm4 Duomenys.btmnorm4 Duomenys.btmromm4
Duomenys.btmrusm4 Duomenys.btmscom4
                Duomenys.btmscgm4 Duomenys.btmsvnm4 Duomenys.btmswem4
Duomenys.btmturm4 Duomenys.btmukrm4;

      label
      IDCNTRY = 'Šalies ID' IDSCHOOL = 'Mokyklos ID' BT4GAGE = 'Mokytojo amžius'
BT4GSEX = 'Mokytojo lytis'
      BT4GTAUT = 'Mokytojo darbo stažas' BT4GTLCE = 'Mokytojo licencija' BT4MSTUD
= 'Mokinių skaičius klasėje'
      BT4MHMWO = 'Nam7 darbai' BT4MTEEX = 'Atsiskaitymai' BT4MPTRH= 'Namų darbų
peržiūrėjimas klasėje, %'
      BT4MPTTQ= 'Testų rašymas klasėje, %' IDTEACH = 'Mokytojo ID';

      format
```

```
      IDCNTRY salis. BT4GAGE amziusb. BT4GSEX lytisb. BT4GTAUT stazas. BT4GTLCE
turi. BT4MSTUD klase. BT4MHMWO nd.
      BT4MTEEX kontr. BT4MPTRH klase. BT4MPTTQ klase.;

      run;


endsubmit;

return;


CheckBox1:

if    CheckBox1.selected='Yes' then do;
      CheckBox2.selected='Yes';
      CheckBox3.selected='Yes';
      CheckBox4.selected='Yes';
      CheckBox5.selected='Yes';
      CheckBox6.selected='Yes';
      CheckBox7.selected='Yes';
      CheckBox8.selected='Yes';
      CheckBox9.selected='Yes';
      CheckBox10.selected='Yes';
      CheckBox11.selected='Yes';
      CheckBox12.selected='Yes';
      CheckBox13.selected='Yes';
      CheckBox14.selected='Yes';
      CheckBox15.selected='Yes';
      CheckBox16.selected='Yes';
      CheckBox19.selected='Yes';
      CheckBox20.selected='Yes';
      CheckBox21.selected='Yes';
      CheckBox22.selected='Yes';
      CheckBox25.selected='Yes';
end;
else do;
      CheckBox2.selected='No';
      CheckBox3.selected='No';
      CheckBox4.selected='No';
      CheckBox5.selected='No';
      CheckBox6.selected='No';
      CheckBox7.selected='No';
      CheckBox8.selected='No';
      CheckBox9.selected='No';
      CheckBox10.selected='No';
      CheckBox11.selected='No';
      CheckBox12.selected='No';
      CheckBox13.selected='No';
      CheckBox14.selected='No';
      CheckBox15.selected='No';
      CheckBox16.selected='No';
      CheckBox19.selected='No';
      CheckBox20.selected='No';
      CheckBox21.selected='No';
      CheckBox22.selected='No';
      CheckBox25.selected='No';
end;
return;

PushButton3:

      id = open('Work.mokiniai','I');
```

```
Salis_num=varnum(id,'IDCNTRY');
Lytis_num=varnum(id,'ITSEX');
rc = close(id);

if  Salis_num=0 or Lytis_num = 0 then do;
   commandlist=makelist();
   commandlist=insertc(commandlist,"Nėra duomenų!",1);
   command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
   commandlist=dellist(commandlist);
end;
else do;
   /* Sukuriami tarpiniai failai */
   submit continue;
      data salis01; run;
      data salis02; run;
      data salis03; run;
      data salis04; run;
      data salis05; run;
      data salis06; run;
      data salis07; run;
      data salis08; run;
      data salis09; run;
      data salis10; run;
      data salis11; run;
      data salis12; run;
      data salis13; run;
      data salis14; run;
      data salis15; run;
      data salis16; run;
      data salis17; run;
      data salis18; run;
      data salis19; run;
      data salis20; run;
      data lytis1; run;
      data lytis2; run;
   endsubmit;

   if checkbox2.selected='Yes' then
   submit continue;
      data salis01;
         set mokiniai;
         if IDCNTRY='51';
      run;
   endsubmit;

   if checkbox3.selected='Yes' then
   submit continue;
      data salis05;
         set mokiniai;
         if IDCNTRY='203';
      run;
   endsubmit;

   if checkbox4.selected='Yes' then
   submit continue;
      data salis04;
         set mokiniai;
         if IDCNTRY='196';
      run;
   endsubmit;

   if checkbox5.selected='Yes' then
   submit continue;
```

```
   data salis19;
      set mokiniai;
      if IDCNTRY='926';
   run;
endsubmit;

if checkbox6.selected='Yes' then
submit continue;
   data salis02;
      set mokiniai;
      if IDCNTRY='70';
   run;
endsubmit;

if checkbox7.selected='Yes' then
submit continue;
   data salis07;
      set mokiniai;
      if IDCNTRY='348';
   run;
endsubmit;

if checkbox8.selected='Yes' then
submit continue;
   data salis08;
      set mokiniai;
      if IDCNTRY='380';
   run;
endsubmit;

if checkbox9.selected='Yes' then
submit continue;
   data salis03;
      set mokiniai;
      if IDCNTRY='100';
   run;
endsubmit;

if checkbox10.selected='Yes' then
submit continue;
   data salis09;
      set mokiniai;
      if IDCNTRY='440';
   run;
endsubmit;

if checkbox11.selected='Yes' then
submit continue;
   data salis06;
      set mokiniai;
      if IDCNTRY='268';
   run;
endsubmit;

if checkbox12.selected='Yes' then
submit continue;
   data salis20;
      set mokiniai;
      if IDCNTRY='927';
   run;
endsubmit;

if checkbox13.selected='Yes' then
```

```
submit continue;
   data salis18;
      set mokiniai;
      if IDCNTRY='891';
   run;
endsubmit;

if checkbox14.selected='Yes' then
submit continue;
   data salis14;
      set mokiniai;
      if IDCNTRY='705';
   run;
endsubmit;

if checkbox15.selected='Yes' then
submit continue;
   data salis15;
      set mokiniai;
      if IDCNTRY='752';
   run;
endsubmit;

if checkbox16.selected='Yes' then
submit continue;
   data salis10;
      set mokiniai;
      if IDCNTRY='470';
   run;
endsubmit;

if checkbox19.selected='Yes' then
submit continue;
   data salis11;
      set mokiniai;
      if IDCNTRY='578';
   run;
endsubmit;

if checkbox20.selected='Yes' then
submit continue;
   data salis12;
      set mokiniai;
      if IDCNTRY='642';
   run;
endsubmit;

if checkbox21.selected='Yes' then
submit continue;
   data salis13;
      set mokiniai;
      if IDCNTRY='643';
   run;
endsubmit;

if checkbox22.selected='Yes' then
submit continue;
   data salis16;
      set mokiniai;
      if IDCNTRY='792';
   run;
endsubmit;
```

```
         if checkbox25.selected='Yes' then
         submit continue;
            data salis17;
               set mokiniai;
               if IDCNTRY='804';
            run;
         endsubmit;

         submit continue;
         data Duomenys.Mokiniai;
            set salis01 salis02 salis03 salis04 salis05 salis06 salis07 salis08
salis09 salis10
                salis11 salis12 salis13 salis14 salis15 salis16 salis17 salis18
salis19 salis20;
         run;
         endsubmit;

         submit continue;
            data salis01; run;
            data salis02; run;
            data salis03; run;
            data salis04; run;
            data salis05; run;
            data salis06; run;
            data salis07; run;
            data salis08; run;
            data salis09; run;
            data salis10; run;
            data salis11; run;
            data salis12; run;
            data salis13; run;
            data salis14; run;
            data salis15; run;
            data salis16; run;
            data salis17; run;
            data salis18; run;
            data salis19; run;
            data salis20; run;
         endsubmit;

         if checkbox17.selected='Yes' then
         submit continue;
            data lytis1;
               set Duomenys.Mokiniai;
               if ITSEX='1';
            run;
         endsubmit;

         if checkbox18.selected='Yes' then
         submit continue;
            data lytis2;
               set Duomenys.Mokiniai;
               if ITSEX='2';
            run;
         endsubmit;

         submit continue;
            data  Duomenys.Mokiniai;
               set lytis1 lytis2;
            run;
         endsubmit;

         submit continue;
```

```
            data lytis1; run;
            data lytis2; run;
         endsubmit;


         commandlist=makelist();
         commandlist=insertc(commandlist,"Duomenų pjūvis sėkmingai sukurtas.",1);
         commandlist=insertc(commandlist,"Galite pradėti duomenų analizę.",2);
         command=messagebox(commandlist,'i','O','PRANEŠIMAS');
         commandlist=dellist(commandlist);
      end;

return;

PushButton1:

      id = open('Work.mokytojai','I');
      Salis_num=varnum(id,'IDCNTRY');
      Lytis_num=varnum(id,'BT4GSEX');
      rc = close(id);

      if  Salis_num=0 or Lytis_num = 0 then do;
         commandlist=makelist();
         commandlist=insertc(commandlist,"Nėra duomenų!",1);
         command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
         commandlist=dellist(commandlist);
      end;
      else do;
         /* Sukuriami tarpiniai failai */
         submit continue;
            data salis01; run;
            data salis02; run;
            data salis03; run;
            data salis04; run;
            data salis05; run;
            data salis06; run;
            data salis07; run;
            data salis08; run;
            data salis09; run;
            data salis10; run;
            data salis11; run;
            data salis12; run;
            data salis13; run;
            data salis14; run;
            data salis15; run;
            data salis16; run;
            data salis17; run;
            data salis18; run;
            data salis19; run;
            data salis20; run;
            data lytis3; run;
            data lytis4; run;
         endsubmit;

         if checkbox2.selected='Yes' then
         submit continue;
            data salis01;
               set mokytojai;
               if IDCNTRY='51';
            run;
         endsubmit;

         if checkbox3.selected='Yes' then
```

```
submit continue;
   data salis05;
      set mokytojai;
      if IDCNTRY='203';
   run;
endsubmit;

if checkbox4.selected='Yes' then
submit continue;
   data salis04;
      set mokytojai;
      if IDCNTRY='196';
   run;
endsubmit;

if checkbox5.selected='Yes' then
submit continue;
   data salis19;
      set mokytojai;
      if IDCNTRY='926';
   run;
endsubmit;

if checkbox6.selected='Yes' then
submit continue;
   data salis02;
      set mokytojai;
      if IDCNTRY='70';
   run;
endsubmit;

if checkbox7.selected='Yes' then
submit continue;
   data salis07;
      set mokytojai;
      if IDCNTRY='348';
   run;
endsubmit;

if checkbox8.selected='Yes' then
submit continue;
   data salis08;
      set mokytojai;
      if IDCNTRY='380';
   run;
endsubmit;

if checkbox9.selected='Yes' then
submit continue;
   data salis03;
      set mokytojai;
      if IDCNTRY='100';
   run;
endsubmit;

if checkbox10.selected='Yes' then
submit continue;
   data salis09;
      set mokytojai;
      if IDCNTRY='440';
   run;
endsubmit;
```

```
if checkbox11.selected='Yes' then
submit continue;
   data salis06;
      set mokytojai;
      if IDCNTRY='268';
   run;
endsubmit;

if checkbox12.selected='Yes' then
submit continue;
   data salis20;
      set mokytojai;
      if IDCNTRY='927';
   run;
endsubmit;

if checkbox13.selected='Yes' then
submit continue;
   data salis18;
      set mokytojai;
      if IDCNTRY='891';
   run;
endsubmit;

if checkbox14.selected='Yes' then
submit continue;
   data salis14;
      set mokytojai;
      if IDCNTRY='705';
   run;
endsubmit;

if checkbox15.selected='Yes' then
submit continue;
   data salis15;
      set mokytojai;
      if IDCNTRY='752';
   run;
endsubmit;

if checkbox16.selected='Yes' then
submit continue;
   data salis10;
      set mokytojai;
      if IDCNTRY='470';
   run;
endsubmit;

if checkbox19.selected='Yes' then
submit continue;
   data salis11;
      set mokytojai;
      if IDCNTRY='578';
   run;
endsubmit;

if checkbox20.selected='Yes' then
submit continue;
   data salis12;
      set mokytojai;
      if IDCNTRY='642';
   run;
endsubmit;
```

```
         if checkbox21.selected='Yes' then
         submit continue;
            data salis13;
               set mokytojai;
               if IDCNTRY='643';
            run;
         endsubmit;

         if checkbox22.selected='Yes' then
         submit continue;
            data salis16;
               set mokytojai;
               if IDCNTRY='792';
            run;
         endsubmit;

         if checkbox25.selected='Yes' then
         submit continue;
            data salis17;
               set mokytojai;
               if IDCNTRY='804';
            run;
         endsubmit;

         submit continue;
         data Duomenys.Mokytojai;
            set salis01 salis02 salis03 salis04 salis05 salis06 salis07 salis08
salis09 salis10
                salis11 salis12 salis13 salis14 salis15 salis16 salis17 salis18
salis19 salis20;
         run;
         endsubmit;

         submit continue;
            data salis01; run;
            data salis02; run;
            data salis03; run;
            data salis04; run;
            data salis05; run;
            data salis06; run;
            data salis07; run;
            data salis08; run;
            data salis09; run;
            data salis10; run;
            data salis11; run;
            data salis12; run;
            data salis13; run;
            data salis14; run;
            data salis15; run;
            data salis16; run;
            data salis17; run;
            data salis18; run;
            data salis19; run;
            data salis20; run;
         endsubmit;

         if checkbox23.selected='Yes' then
         submit continue;
            data lytis3;
               set Duomenys.Mokytojai;
               if BT4GSEX='1';
            run;
```

```
         endsubmit;

         if checkbox24.selected='Yes' then
         submit continue;
            data lytis4;
               set Duomenys.Mokytojai;
               if BT4GSEX='2';
            run;
         endsubmit;

         submit continue;
            data  Duomenys.Mokytojai;
               set lytis3 lytis4;
            run;
         endsubmit;

         submit continue;
            data lytis3; run;
            data lytis4; run;
         endsubmit;


         commandlist=makelist();
         commandlist=insertc(commandlist,"Duomenų pjūvis sėkmingai sukurtas.",1);
         commandlist=insertc(commandlist,"Galite pradėti duomenų analizę.",2);
         command=messagebox(commandlist,'i','O','PRANEŠIMAS');
         commandlist=dellist(commandlist);
      end;

return;
```

ANALIZĖ.SCL

```
init:

/* Hipoteziu tikrinimas */
  if symget('analize03')='beta'
  then combobox1.selectedindex=1;
  else if symget('analize03')='gamma'
       then combobox1.selectedindex=2;
       else if symget('analize03')='lognormal'
            then combobox1.selectedindex=3;
            else if symget('analize03')='normal'
                 then combobox1.selectedindex=4;
                 else if symget('analize03')='weibull'
                      then combobox1.selectedindex=5;
                      else if symget('analize03')='exponential'
                           then combobox1.selectedindex =6;
                           else if symget('analize03')='kernel'
                                then combobox1.selectedindex=7;
                                else combobox1.selectedindex=0;

      if symget('analize05') = 'BT4GTAUT'
         then radiobox2.selectedindex = 1;
      else if symget('analize05') = 'BT4MSTUD'
              then radiobox2.selectedindex = 2;
           else radiobox2.selectedindex = 3;

      textentry1.text = symget('analize07');
      textentry2.text = symget('analize08');
      textentry3.text = symget('analize09');
```

```
submit continue;
data Duomenys.Mokiniai;
set Duomenys.Mokiniai;
run;
data Duomenys.Mokytojai;
set Duomenys.Mokytojai;
run;
endsubmit;

return;

pushbutton1:


if Listbox1.selectedindex=1 then do;

    if checkbox1.selected='No' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas faktorius!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
    end;

    else do;
      call symput('grupuoti', '');
      if radiobox1.selectedindex=1 then call symput('grupuoti', '');
      else if radiobox1.selectedindex=2 then call symput('grupuoti', 'IDCNTRY');
      else if radiobox1.selectedindex=3 then call symput('grupuoti', 'ITSEX');
      else if radiobox1.selectedindex=4 then do;
         commandlist=makelist();
         commandlist=insertc(commandlist,"Negalima grupuoti pagal amžių");
         command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
         command=command;
         commandlist=dellist(commandlist);
      end;
      call symput('kintamieji', '');
      if checkbox1.selected='Yes' then call symput('kintamieji', 'BSDAGE');

submit continue;
options mlogic mprint;
%macro aprasomoji;
%macro skip; %mend skip;

proc sql noprint;
      create table duom as
      select a.*
      from Duomenys.Mokiniai as a;
quit;

data duom;
 set duom;
 if IDCNTRY^='';
run;

%if &grupuoti ne %str() %then %do;
proc sort data=duom;
      by &grupuoti;
run;
%end;

proc means data=duom n nmiss range sum mean min max var std skew kurt;
     %if &grupuoti ne %str() %then %do;
```

```
      by &grupuoti;
     %end;
      var &kintamieji;
run;

%mend;
%aprasomoji;

endsubmit;

      end;
end;

if Listbox1.selectedindex=2 then do;

    if checkbox2.selected='No' and
      checkbox3.selected='No' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas faktorius!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
    end;

    else do;
      call symput('grupuoti', '');
      if radiobox1.selectedindex=1 then call symput('grupuoti', '');
      else if radiobox1.selectedindex=2 then call symput('grupuoti', 'IDCNTRY');
      else if radiobox1.selectedindex=3 then call symput('grupuoti', 'BT4GSEX');
      else if radiobox1.selectedindex=4 then call symput('grupuoti', 'BT4GAGE');

      call symput('kintamieji', '');
      call symput('kintamieji2', '');
      if checkbox2.selected='Yes' then call symput('kintamieji', 'BT4GTAUT');
      if checkbox3.selected='Yes' then call symput('kintamieji2', 'BT4MSTUD');

submit continue;
options mlogic mprint;
%macro aprasomoji;
%macro skip; %mend skip;

proc sql noprint;
      create table duom as
      select b.*
      from Duomenys.Mokytojai as b;
quit;

data duom;
 set duom;
 if IDCNTRY^='';
run;

%if &grupuoti ne %str() %then %do;
proc sort data=duom;
      by &grupuoti;
run;
%end;

proc means data=duom n nmiss range sum mean var min max std skew kurt;
     %if &grupuoti ne %str() %then %do;
      by &grupuoti;
     %end;
      var &kintamieji &kintamieji2;
```

```
run;

%mend;
%aprasomoji;

endsubmit;

      end;
end;

if Listbox1.selectedindex=0 then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas failas!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
end;

return;

combobox1:

      if combobox1.selectedindex = 1 then call symput('analize03','beta');
      if combobox1.selectedindex = 2 then call symput('analize03','gamma');
      if combobox1.selectedindex = 3 then call symput('analize03','lognormal');
      if combobox1.selectedindex = 4 then call symput('analize03','normal');
      if combobox1.selectedindex = 5 then call symput('analize03','weibull');
      if combobox1.selectedindex = 6 then call symput('analize03','exponential');
      if combobox1.selectedindex = 7 then call symput('analize03','kernel');
      if combobox1.selectedindex = 7
         then call symput('analize04' ,'amise="AMISE" bandwidth="Juostos
platumo"');
      else call symput('analize04' ,'ksd="Kolmogorovo (D)" ksdpval="p" chisq="Chi-
kvadratu" df="l.l.sk." pchisq="p"');

return;

radiobox2:

      if radiobox2.selectedindex = 1 then call symput('analize05','BT4GTAUT');
      if radiobox2.selectedindex = 2 then call symput('analize05','BT4MSTUD');
      if radiobox2.selectedindex = 3 then call symput('analize05','BSDAGE');
      if radiobox2.selectedindex = 3 then call symput('analize06','Mokiniai');
                                     else call symput('analize06','Mokytojai');

return;

textentry1:

      if textentry1.text > textentry2.text then textentry1.text = textentry2.text;
      call symputn('analize07',textentry1.text);

return;

textentry2:

      if textentry1.text > textentry2.text then textentry2.text = textentry1.text;
      call symputn('analize08',textentry2.text);

return;

textentry3:
```

```
      if textentry3.text < 1 then textentry3.text = 1;
      textentry3.text = ROUND(textentry3.text);
      call symputn('analize09',textentry3.text);

return;

pushbutton3:

    if combobox1.selectedindex = 0  then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinkta nulinė hipotezė!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
    end;
    else do;
    if textentry1.text='' and textentry2.text ='' and textentry3.text=''  then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinkti parametrai!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
    end;

submit continue;
options mlogic mprint;
%macro hipoteze;
%macro skip; %mend skip;

data duom;
 set Duomenys.&analize06;
 if IDCNTRY^='';
run;

proc univariate data=duom;
      var &analize05;
      histogram/&analize03  midpoints=&analize07 to &analize08 by &analize09;
run;

%mend;
%hipoteze;

endsubmit;
      end;
return;

radiobox3:

if radiobox3.selectedindex = 1 then call symput('dispanalize4', 'BT4GTAUT');
if radiobox3.selectedindex = 2 then call symput('dispanalize4', 'BT4MSTUD');
if radiobox3.selectedindex = 3 then call symput('dispanalize4', 'BSDAGE');

return;

Pushbutton5:

if checkbox5.selected='No' and checkbox6.selected='No' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas faktorius!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
end;
```

```
else do;
      if radiobox3.selectedindex in (1, 2) then call symput('dispanalize1',
'Duomenys.Mokytojai');
      if radiobox3.selectedindex = 3 then call symput('dispanalize1',
'Duomenys.Mokiniai');

if checkbox5.selected = 'Yes' and checkbox6.selected = 'No' then do;
      call symput('dispanalize2', 'IDCNTRY');
      call symput('dispanalize3', 'IDCNTRY');
end;
if checkbox5.selected = 'No' and checkbox6.selected = 'Yes' then do;
      if radiobox3.selectedindex in (1, 2) then do;
      call symput('dispanalize2', 'BT4GSEX');
      call symput('dispanalize3', 'BT4GSEX');
      end;
      else do;
      call symput('dispanalize2', 'ITSEX');
      call symput('dispanalize3', 'ITSEX');
      end;
end;
if checkbox5.selected = 'Yes' and checkbox6.selected = 'Yes' then do;
      if radiobox3.selectedindex in (1, 2) then do;
      call symput('dispanalize2', 'IDCNTRY BT4GSEX IDCNTRY*BT4GSEX');
      call symput('dispanalize3', 'IDCNTRY BT4GSEX');
      end;
      else do;
      call symput('dispanalize2', 'IDCNTRY ITSEX IDCNTRY*ITSEX');
      call symput('dispanalize3', 'IDCNTRY ITSEX');
      end;
end;

submit continue;
options mlogic mprint;
%macro dispanalize;
%macro skip; %mend skip;

data duom;
 set &dispanalize1;
 if IDCNTRY^='';
run;

proc glm data=duom;
      class &dispanalize3;
      model &dispanalize4=&dispanalize2;
      means &dispanalize3/tukey scheffe duncan lines;
run;

proc univariate normal plot data=&dispanalize1;
      var &dispanalize4;
run;

%mend;
%dispanalize;

endsubmit;
    end;
return;

pushbutton6:

if checkbox5.selected='No' and checkbox6.selected='No' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas faktorius!");
```

```
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
end;
else do;
      if radiobox3.selectedindex in (1, 2) then call symput('rangdispanalize1',
'Duomenys.Mokytojai');
      if radiobox3.selectedindex = 3 then call symput('rangdispanalize1',
'Duomenys.Mokiniai');

if checkbox5.selected = 'Yes' then do;
      call symput('rangdispanalize2', 'IDCNTRY');
end;
if checkbox6.selected = 'Yes' then do;
      if radiobox3.selectedindex in (1, 2) then do;
      call symput('rangdispanalize2', 'BT4GSEX');
      end;
      else do;
      call symput('rangdispanalize2', 'ITSEX');
      end;
end;

submit continue;
options mlogic mprint;
%macro rangdispanalize;
%macro skip; %mend skip;

data duom;
 set &rangdispanalize1;
 if IDCNTRY^='';
run;

proc npar1way data=duom wilcoxon;
      class &rangdispanalize2;
      var &dispanalize4;
run;
quit;

%mend;
%rangdispanalize;

endsubmit;
   end;
return;


pushbutton8:

      if Listbox2.selectedindex = 1
         then do;
            call symput('pozymis1','IDCNTRY');
         end;
      if Listbox2.selectedindex = 2
         then do;
            call symput('pozymis1','BT4GAGE');
         end;
      if Listbox2.selectedindex = 3
         then do;
            call symput('pozymis1','BT4GSEX');
         end;
      if Listbox2.selectedindex = 4
         then do;
            call symput('pozymis1','BT4GTLCE');
```

```
         end;
      if Listbox2.selectedindex = 5
         then do;
            call symput('pozymis1','BT4MHMWO');
         end;
      if Listbox2.selectedindex = 6
         then do;
            call symput('pozymis1','BT4MTEEX');
         end;

      if Listbox3.selectedindex = 1
         then do;
            call symput('pozymis2','IDCNTRY');
         end;
      if Listbox3.selectedindex = 2
         then do;
            call symput('pozymis2','BT4GAGE');
         end;
      if Listbox3.selectedindex = 3
         then do;
            call symput('pozymis2','BT4GSEX');
         end;
      if Listbox3.selectedindex = 4
         then do;
            call symput('pozymis2','BT4GTLCE');
         end;
      if Listbox3.selectedindex = 5
         then do;
            call symput('pozymis2','BT4MHMWO');
         end;
      if Listbox3.selectedindex = 6
         then do;
            call symput('pozymis2','BT4MTEEX');
         end;

      if Listbox2.selectedindex=0 or Listbox3.selectedindex=0 then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas kintamasis!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
      end;
      else do;

submit continue;
options mlogic mprint;
%macro koreliacija2;
%macro skip; %mend skip;

data duom;
set Duomenys.Mokytojai;
if IDCNTRY ^= '';
run;

proc freq data=duom order=data;
      tables &pozymis1*&pozymis2/chisq relrisk;
run;

%mend;
%koreliacija2;

endsubmit;
      end;
```

```
return;

radiobox7:

if radiobox7.selectedindex = 1 then call symput('grafika1', 'sum');
if radiobox7.selectedindex = 2 then call symput('grafika1', 'mean');

return;

radiobox5:

if radiobox5.selectedindex = 1 then call symput('grafika2', 'BT4GSEX');
if radiobox5.selectedindex = 2 then call symput('grafika2', 'BT4GAGE');
if radiobox5.selectedindex = 3 then call symput('grafika2', 'BT4GTLCE');
if radiobox5.selectedindex = 4 then call symput('grafika2', 'IDCNTRY');
if radiobox5.selectedindex = 5 then call symput('grafika2', 'ITSEX');
if radiobox5.selectedindex = 6 then call symput('grafika2', 'BS4GBOOK');
if radiobox5.selectedindex = 7 then call symput('grafika2', 'BS4GTH01');
if radiobox5.selectedindex = 8 then call symput('grafika2', 'BS4GTH02');
if radiobox5.selectedindex = 9 then call symput('grafika2', 'BS4GTH03');
if radiobox5.selectedindex = 10 then call symput('grafika2', 'BS4GTH05');
if radiobox5.selectedindex = 11 then call symput('grafika2', 'BS4GWATV');
if radiobox5.selectedindex = 12 then call symput('grafika2', 'BS4GDOHW');

return;

radiobox6:

if radiobox6.selectedindex = 1 then call symput('grafika3', 'BT4GTAUT');
if radiobox6.selectedindex = 2 then call symput('grafika3', 'BT4MSTUD');
if radiobox6.selectedindex = 3 then call symput('grafika3', 'BSDAGE');

return;

pushbutton9:

    if radiobox7.selectedindex='0' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinkta charakteristika!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
    end;
    else do;

submit continue;
options mlogic mprint;
%macro grafikai;
%macro skip; %mend skip;

data duom;
set Duomenys.Mokiniai Duomenys.Mokytojai;
if IDCNTRY^='';
run;

proc gchart data=duom;

      vbar &grafika2/discrete sumvar=&grafika3 type=&grafika1;

run;

%mend;
%grafikai;
```

```
endsubmit;
    end;
return;


pushbutton7:

      call symput('log_reg3','');

      if Checkbox8.selected = 'Yes'
         then do;
           call symput('log_reg3',symget('log_reg3')|| 'BT4GTAUT ');
         end;
      if Checkbox9.selected = 'Yes'
         then do;
           call symput('log_reg3',symget('log_reg3')|| 'BT4MSTUD ');
         end;
      if Checkbox10.selected = 'Yes'
         then do;
           call symput('log_reg3',symget('log_reg3')|| 'BT4MPTRH ');
         end;
      if Checkbox10.selected = 'Yes'
         then do;
           call symput('log_reg3',symget('log_reg3')|| 'BT4MPTTQ ');
         end;

      if Radiobox4.selectedindex = 1 then call symput('log_reg4','BT4GTLCE');
      if Radiobox4.selectedindex = 2 then call symput('log_reg4','BT4MHMWO');
      if Radiobox4.selectedindex = 3 then call symput('log_reg4','BT4GSEX');


submit continue;
options mlogic mprint;
%macro logist;
%macro skip; %mend skip;

data duom;
 set Duomenys.Mokytojai;
 if IDCNTRY^='';
run;

proc logistic data = duom;
    class &log_reg3;
    model &log_reg4= &log_reg3 / expb;
run;


%mend;
%logist;

endsubmit;

return;

pushbutton10:

      if radiobox9.selectedindex='0' then do;
       commandlist=makelist();
       commandlist=insertc(commandlist,"Neparinktas kiekybinis kintamasis!");
       command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
       command=command;
       commandlist=dellist(commandlist);
```

```
      end;
      else do;

      if Radiobox9.selectedindex = 1 then call symput('klast2','BT4GTAUT');
      if Radiobox9.selectedindex = 2 then call symput('klast2','BT4MSTUD');
      if Radiobox9.selectedindex = 3 then call symput('klast2','BT4MPTRH');
      if Radiobox9.selectedindex = 4 then call symput('klast2','BT4MPTTQ');


submit continue;
options mlogic mprint;
%macro klasterine;
%macro skip; %mend skip;

data duom;
 set Duomenys.Mokytojai;
 if IDCNTRY^='';;
run;

proc sort data=duom;
      by IDCNTRY;
run;

proc means data=duom noprint;
    var &klast2;
    by IDCNTRY;
    output out = Vidurkiai mean = Vidurkis;
run;

data Klasteriai;
      set duom (keep = IDCNTRY) ;
run;

data Klasteriai;
      merge Klasteriai Vidurkiai ;
run;

data Klasteriai;
      set Klasteriai;
      if Vidurkis ^=' ';
run;

title 'Artimiausio kaimyno';

proc cluster data=Klasteriai method=single out=Klasteriai1;
    var Vidurkis;
    id IDCNTRY;
run;
proc tree data=Klasteriai1 horizontal spaces=2;
    id IDCNTRY;
    copy Vidurkis;
run;


title 'Tolimiausio kaimyno';

proc cluster data=Klasteriai method=complete out=Klasteriai2;
    var Vidurkis;
    id IDCNTRY;
run;
proc tree data=Klasteriai2 horizontal spaces=3;
    id IDCNTRY;
    copy Vidurkis;
```

```
run;

title 'Vidutinio atstumo';

proc cluster data=Klasteriai method=average pseudo out=Klasteriai3;
    var Vidurkis;
    id IDCNTRY;
run;
proc tree data=Klasteriai3 horizontal spaces=3;
    id IDCNTRY;
    copy Vidurkis;
run;

title 'Atstumo tarp centru';

proc cluster data=Klasteriai method=centroid pseudo out=Klasteriai4;
    var Vidurkis;
    id IDCNTRY;
run;
proc tree data=Klasteriai4 horizontal spaces=3;
    id IDCNTRY;
    copy Vidurkis;
run;

%mend;
%klasterine;

endsubmit;
     end;
return;


pushbutton11:

if checkbox7.selected='No' and checkbox11.selected='No' and
checkbox12.selected='No' and
      checkbox13.selected='No' and checkbox14.selected='No' then do;
      commandlist=makelist();
      commandlist=insertc(commandlist,"Neparinktas kintamasis!");
      command=messagebox(commandlist,'!','O','ĮSPĖJIMAS');
      command=command;
      commandlist=dellist(commandlist);
end;
else do;
      call symput('faktor','');

      if Checkbox7.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MAWEL ');
         end;
      if Checkbox11.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MAMOR ');
         end;
      if Checkbox12.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MALIK ');
         end;
      if Checkbox13.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MABOR ');
         end;
      if Checkbox14.selected = 'Yes'
```

```
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MAHDL ');
         end;
      if Checkbox15.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MACLM ');
         end;
      if Checkbox16.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MAQKY ');
         end;
      if Checkbox17.selected = 'Yes'
         then do;
            call symput('faktor',symget('faktor')|| 'BS4MAGET');
         end;

submit continue;
options mlogic mprint;
%macro faktorine;
%macro skip; %mend skip;

data duom;
 set Duomenys.Mokiniai;
 if IDCNTRY^='';
run;

proc factor data=duom
         simple
         method=prin
         priors=one
         mineigen=1
         scree
         rotate=varimax
         round
         flag=0.40;
      var &faktor;
run;

proc factor data=duom
         simple
         method=prin
         priors=one
         mineigen=1
         nfact=2
         rotate=varimax
         round
         flag=0.40
         out=duom2;
      var &faktor;
run;

proc corr data=duom2;
      var factor1 factor2;
      with &faktor factor1 factor2;
run;

%mend;
%faktorine;

endsubmit;
    end;
return;
```