



KAUNO TECHNOLOGIJOS UNIVERSITETAS
FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS
TAIKOMOSIOS MATEMATIKOS KATEDRA

Laurynas Pabarčius

STUDIJŲ PROCESO DUOMENŲ ANALIZĖS
MODELIAI IR PROGRAMINĖS PRIEMONĖS

Magistro darbas

Vadovas
doc. dr. V. Janilionis

KAUNAS, 2011



KAUNO TECHNOLOGIJOS UNIVERSITETAS
FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS
TAIKOMOSIOS MATEMATIKOS KATEDRA

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

STUDIJŲ PROCESO DUOMENŲ ANALIZĖS
MODELIAI IR PROGRAMINĖS PRIEMONĖS

Taikomosios matematikos magistro baigiamasis darbas

Vadovas
doc. dr. V. Janilionis
2011 05 27

Recenzentas
prof. J. Augutis
2011 06 01

Atliko
FMMM-9 gr. stud.
L. Pabarčius
2011 05 27

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“).

Pabarčius L. Models of statistical analysis and software of study process : Master's work in applied mathematics / supervisor dr. assoc. prof. V. Janilionis; Department of Applied mathematics, Faculty of Fundamental Sciences, Kaunas University of Technology. – Kaunas, 2011. – 85 p.

SUMMARY

The problem of the statistical analysis and software of study process considered in this paper is a common problem arising in academic communities in Lithuania and foreign countries. New technologies are developing very fastly and it is very important that universities would be able to supply students with versatile knowledge and skills. Only then they would have the ability to do successful researches and create competitive future technologies. In order to achieve this goal, universities must have study process data analysis system which results would be used to improve the quality of overall study system. In this paper, analysing different statistical methods were created six statistical models for the data analysis of study process. These models were implemented in analysis software, developed using statistical analysis system SAS, its macroprogramming tools, SAS web development environment AppDev Studio webAF and Javascript, JSP, Java Servlet and HTML technologies. Users can use the system online and such architecture is the first in Lithuania and second in the world. Developed software was applied to the data of the study process of Kaunas University of Technology, which was collected in 2009 and 2010. The results showed that using developed tool it is possible to perform quick and high quality statistical analysis and make appropriate conclusions and proposals how the quality of study process should be improved. There is an overview of statistical models and software tools commonly used for this problem at the beginning of this paper. Other sections describe the structure of the databases of Kaunas University of Technology and proposed statistical models. There is also a description of developed software and explanation how to use it properly. At the end of this paper we tell the opinion about the results, generated by software tool, and provide recommendations for further statistical models and software improvement.

TURINYS

LENTELIŲ SĄRAŠAS	7
PAVEIKSLŲ SĄRAŠAS.....	8
ĮVADAS.....	9
1. BENDROJI DALIS.....	10
1.1. STUDIJŲ PROCESO DUOMENŲ ANALIZEI TAIKOMŲ STATISTIKOS METODŲ IR PROGRAMINIŲ PRIEMONIŲ APŽVALGA	10
1.2. SISTEMOS SAS OBJEKTINIO PROGRAMAVIMO TECHNOLOGIJOS IR JŲ GALIMYBIŲ ANALIZĖ.....	13
1.2.1. JAVA APLETAI IR PROGRAMOS.....	14
1.2.2. SAS webAF IR webEIS.....	14
1.2.3. JAVA SERVLET IR JSP.....	16
1.3. DARBE SPRENDŽIAMŲ UŽDAVINIAI.....	19
2. TIRIAMOJI DALIS	20
2.1. KTU STUDIJŲ PROCESO DUOMENŲ BAZIŲ STRUKTŪRA.....	20
2.2. STUDIJŲ PROCESO DUOMENŲ ANALIZĖS MODELIAI.....	21
2.2.1. APRAŠOMOJI STATISTIKA	21
2.2.2. KORELIACINĖ ANALIZĖ	24
2.2.3. DISPERSINĖ ANALIZĖ.....	26
2.2.4. LOGISTINĖ REGRESINĖ ANALIZĖ	32
2.2.5. KLASTERINĖ ANALIZĖ.....	36
2.3. MODELIŲ PROGRAMINĖ REALIZACIJA.....	38
2.4. MODELIŲ TAIKYMO PAVYZDŽIAI.....	45
2.4.1. APRAŠOMOSIOS STATISTIKOS MODELIO TAIKYMAS.....	46
2.4.2. RYŠIŲ TARP STUDENTŲ MOKYMOŠI RODIKLIŲ TYRIMAS.....	47
2.4.3. STOJANČIŲJŲ METINIŲ PAŽYMIŲ VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ SRITIES IR PAGEIDAVIMO GRUPĖS TYRIMAS	49
2.4.4. STOJANČIŲJŲ KONKURSINIO BALO VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ PROGRAMOS IR PAGEIDAVIMO GRUPĖS TYRIMAS.....	50
2.4.5. STUDENTŲ PAJĖGUMO INDEKSO VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ PROGRAMOS IR PAGEIDAVIMO GRUPĖS TYRIMAS	50
2.4.6. PIRMO KURSO BAIGIMO TIKIMYBĖS PROGNOZAVIMAS	55
2.4.7. MODULIO „MATEMATIKA 1“ ĮVERTINIMO PROGNOZAVIMAS	58
2.4.8. STUDIJŲ PROGRAMŲ KLASIFIKAVIMAS.....	61

2.5. REKOMENDACIJOS DĖL TOLIMESNIO SUKURTŲ MODELIŲ IR PROGRAMINIŲ PRIEMONIŲ TOBULINIMO	64
IŠVADOS	65
LITERATŪRA.....	66
1 PRIEDAS. STUDIJŲ PROCESO DUOMENŲ MATRICA	68
2 PRIEDAS. MODELIŲ IR VARTOTOJO SĄSAJOS PROGRAMŲ TEKSTAI.....	69
3 PRIEDAS. MODELIŲ TAIKYMO REZULTATAI.....	83

LENTELIŲ SĄRAŠAS

2.1 lentelė. Koreliacinėje analizėje naudojami žymėjimai.....	24
2.2 lentelė. Vienfaktorinėje dispersinėje analizėje naudojami žymėjimai.....	26
2.3 lentelė. Dvifaktorinėje dispersinėje analizėje naudojami žymėjimai.....	29
2.4 lentelė. Logistinėje regresinėje analizėje naudojami žymėjimai.....	32
2.5 lentelė. Modelyje naudojamos SAS ryšio funkcijos.....	33
2.6 lentelė. Klasterinėje analizėje naudojami žymėjimai.....	36
2.7 lentelė. Klasterių atstumo matai.....	37
2.8 lentelė. Aprašomosios statistikos modelio taikymas.....	46
2.9 lentelė. Ryšių tarp KTU studentų mokymosi rodiklių tyrimas.....	48
2.10 lentelė. Stojančiųjų metinių pažymių vidurkio priklausomybės nuo studijų srities ir pageidavimo grupės tyrimas.....	51
2.11 lentelė. Stojančiųjų konkursinio balo vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimas.....	52
2.12 lentelė. Studentų pajėgumo indekso vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimas.....	54
2.13 lentelė. Pirmo kurso baigimo tikimybės prognozavimas.....	56
2.14 lentelė. Pirmo kurso baigimo tikimybės prognozavimas (be nereikšmingų kintamųjų).....	57
2.15 lentelė. Modulio „Matematika 1“ įvertinimo prognozavimas.....	59
2.16 lentelė. Modulio „Matematika 1“ įvertinimo prognozavimas (be nereikšmingų kintamųjų).....	60
2.17 lentelė. KTU studijų programų klasifikavimas.....	63
1.1 priedo lentelė. Duomenų analizėje naudojamų kintamųjų matrica.....	68

PAVEIKSLŲ SĄRAŠAS

2.1 pav. Priėmimo duomenų struktūrinė schema	20
2.2 pav. Stojimo duomenų struktūrinė schema	22
2.3 pav. Studijų duomenų struktūrinė schema	23
2.4 pav. Studijų proceso duomenų analizės sistemos veikimo schema	40
2.5 pav. Duomenų analizės sistemos bylų struktūra ir tarpusavio sąveika	41
2.6 pav. Pradinis duomenų analizės sistemos langas	39
2.7 pav. Duomenų pjūvio formavimas	39
2.8 pav. Statistinių modelių pasirinkimas	42
2.9 pav. Aprašomosios statistikos modelio vartotojo sąsaja	42
2.10 pav. Koreliacinės analizės modelio vartotojo sąsaja	43
2.11 pav. Dispersinės analizės modelių vartotojo sąsaja	43
2.12 pav. Logistinės regresinės analizės modelio vartotojo sąsaja	44
2.13 pav. Klasterinės analizės modelio vartotojo sąsaja	44
2.14 pav. Fizinių mokslų studentų stojimo pageidavimo numerių pasiskirstymas	47
2.15 pav. Technologijos mokslų studentų stojimo pageidavimo numerių pasiskirstymas	47
2.16 pav. Technologijos mokslų studijų programų klasifikacija pagal Kauno mokyklas baigusių studentų 2009-2010 m. metinius mokyklos matematikos, fizikos ir informatikos įvertinimus	62
2.17 pav. Technologijos mokslų studijų programų klasifikacija pagal ne Kauno mokyklas baigusių studentų 2009-2010 m. metinius mokyklos matematikos, fizikos ir informatikos įvertinimus	62

IVADAS

Studijų procesas užima svarbią vietą šiuolaikinėje akademinėje visuomenėje. Norint gerinti studijų kokybę būtina nuolat tobulinti studijų procesą. Šiam tikslui pasiekti yra reikalinga kokybiška studijų proceso duomenų analizė.

Studijų kokybės tobulinimui didelę reikšmę turi studijų proceso duomenų analizės modeliai, bet jų sukūrimas reikalauja didelių darbo ir laiko sąnaudų. Šiuo metu Lietuvoje ir pasaulyje studijų proceso uždavinių sprendimui taikomi tik pavieniai aprašomosios statistikos, išlikimo analizės, logistinės regresinės analizės ir neparametriniai tikimybių branduolio tankio įvertinimo metodai, todėl studijų proceso duomenų analizei reikia kurti duomenų analizės modelių sistemas ir patogiai jas realizuojančias programines priemones.

Šio darbo tikslas – sukurti studijų proceso duomenų analizės modelius bei juos realizuojančias interaktyvias programines priemones ir prisidėti prie studijų kokybės tobulinimo. Modeliams sudaryti buvo naudojami aprašomosios statistikos, koreliacinės analizės, vienfaktorinės ir dvifaktorinės dispersinės analizės, logistinės regresinės analizės ir klasterinės analizės metodai. Modeliai programiškai realizuoti panaudojant statistinės analizės paketo SAS makroprogramavimo kalbą, posistemas SAS/BASE, SAS/STAT, SAS/GRAPH, internetinių duomenų analizės sistemų kūrimo įrankio SAS AppDev Studio posistemę webAF ir internetines Java technologijas.

Sukurti modeliai ir programinė įranga, leidžianti studijų proceso duomenis sugrupuoti įvairiais pjūviais ir, gautoms grupėms pritaikius statistinius modelius, atlikti duomenų analizę, rezultatus vartotojui pateikiant patogiu formatu (HTML ir MS Word RTF). Sukurta duomenų analizės sistema yra pirma tokio pobūdžio sistema Lietuvoje. Modeliai ir programinės priemonės buvo taikytos realių KTU 2009 ir 2010 metų studijų proceso duomenų analizei.

Šio darbo tematika buvo pristatyti du pranešimai konferencijose [31, 32].

1. BENDROJI DALIS

1.1. STUDIJŲ PROCESO DUOMENŲ ANALIZEI TAIKOMŲ STATISTIKOS METODŲ IR PROGRAMINIŲ PRIEMONIŲ APŽVALGA

Viename iš lietuvių autorių darbų [1] apie studijų proceso duomenų analizę yra naudojami išlikimo analizės metodai. Taikant išlikimo analizės metodus siekiama išsiaiškinti, kokie veiksniai (konkursinis balas, lytis, sesijų rezultatai) lemia studento išlikimą pasirinktame universitete ir pasirinktoje studijų programoje, o kas skatina nutraukti studijas. Darbe siekiama ištirti studentų išlikimo pasirinktame universitete ir pasirinktoje studijų programoje galimybes, įvertinti, kurie studijų semestrai yra rizikingiausi, t. y. kada būna didžiausia tikimybė iškristi. Taip pat ieškoma pagrindinių riziką lemiančių veiksnių.

Darbe yra nagrinėjamos sąlyginės tikimybės, t.y. tikimybės išlikti tam tikrą laiko tarpą veikiant tam tikriems veiksniams. Tyrime yra svarbi tikimybė sėkmingai baigti studijas veikiant tokiems veiksniams kaip konkursinis balas, sesijų rezultatai. Analizės metodai yra realizuojami naudojant išlikimo, intensyvumo, tikimybių tankio bei rizikos funkcijas.

Duomenys yra apdorojami ir rezultatai gaunami naudojant SPSS programų paketą. Rezultatai yra pateikiami linijiniais grafikais ir lentelėmis.

Kitame lietuvių autoriaus darbe [2] nagrinėjamos aktualios matematikos dalyko dėstymo, vertinimo strategijos problemos, aptariami studijų proceso tobulinimo aspektai, kompiuterių taikymo matematikos paskaitose ir pratybose tikslingumas. Tyrimo metu buvo atlikta Vilniaus Kolegijos I kurso studentų matematikos dalyko pažangumo lyginamoji analizė bei studentų nuomonės apklausa. Analizuoti studentų matematikos dalyko pažangumo rodikliai, panaudojus aprašomosios statistikos metodus sudaryti duomenų analizės modeliai ir atlikta studentų pažangumo mokykloje ir Kolegijoje duomenų lyginamoji analizė.

Tiriant studentų matematikos dalyko pažangumą, taip pat bandyta atsakyti į klausimą, ar būsimasis studentas, turintis gerą matematikos pažymį atestate, gali tikėtis, kad neturės problemų mokydamasis matematikos Kolegijoje, todėl į tyrimą buvo įtraukta ir studentų brandos atestatų matematikos dalyko pažangumo analizė.

Tyrimo rezultatai darbe yra pateikiami stulpelinėmis, linijinėmis diagramomis ir lentelėmis.

Viename iš Lietuvos matematikų draugijos darbų [3] yra tiriami Vilniaus Gedimino Technikos Universiteto studentų pažangumo duomenys. Yra naudojami duomenys apie VGTU inžinerinės informatikos (II) specialybės studentų 2000–2005 metų egzaminų pažymius bei bendra nekonfidenciali informacija apie egzaminavusius dėstytojus.

Tyrimui buvo pasirinkti du analizės modeliai: sąlyginės logistinės regresijos modelis, leidžiantis palyginti skirtingų egzaminų sudėtingumą bei identifikuoti susijusius veiksnius ir pakelinės logistinės

regresijos su fiksuotais veiksniais modelis, kuriuo siekiama išsiaiškinti egzaminuojančių dėstytojų vertinimo sistemų ypatumus. Darbe nagrinėjami du binariniai požymiai, nusakyti egzamino pažymiu ir parodantys, ar sėkmingai, ar nesėkmingai buvo atlikta atitinkama užduotis. Vienas jų yra pažangumas, o kitas - egzamino išlaikymo rodiklis (išlaikė arba neišlaikė). Naudojant sąlyginės logistinės regresijos modelį buvo tirtas minėtų požymių statistinis ryšys su studijų metais, egzaminavusių dėstytojų ypatybėmis bei bendromis jų savybėmis (dėstytojo mokslo vardas, mokslo laipsnis, lytis), bei dėstytojų savybių sąveika su studentų savybėmis.

Tyrimas buvo atliekamas naudojant SAS programų paketą ir rezultatai pateikiami lentelėmis.

Kitame Lietuvos matematikų draugijos darbe [4] nagrinėjama matematinių gebėjimų įvertinimo problema. Jame panaudoti Vilniaus Universiteto informacinės sistemos duomenys apie įstojusius 2003–2006 m. į Matematikos ir Informatikos Instituto bakalauro studijų programas. Laikoma, kad studento matematinės analizės pažymys yra vienas iš pagrindinių matematinių gebėjimų rodiklių, leidžiantis gana patikimai prognozuoti jo tolesnių matematinės specialybės studijų eigą. Tyrimo tikslas yra nustatyti matematinės analizės įvertinimus įtakojančius faktorius. Pagrindinis tiriamas faktorius yra skirtingų dėstytojų vertinimo sistemos, kuris yra tiesiogiai susijęs su matematinių gabumų vertinimo kokybe.

Naudojant VU matematinių specialybių studentų duomenis ir rankinį logistinės regresijos su kintančiu masteliu modelį tiriama įvairių veiksnių, tarp jų ir egzaminuojančio dėstytojo vertinimo sistemos, įtaka matematinės analizės pažymiams.

Skaičiavimai atlikti su PASW Statistics 18 programa ir logistinės regresijos modelių analizei naudota PLUM procedūra. Rezultatai pateikti lentelėmis.

Užsienio autoriaus darbe [5] yra nagrinėjami realūs Pavia universiteto (Italija) e-mokymo sistemos duomenys. Darbo tikslas yra surasti ir pateikti e-mokymo sistemos testų santykinės svarbos matą kiekvieno mokomojo modulio pabaigoje, sudaryti prognozuojančius studentų veiklos modelius ir personalizuoti e-mokymo sistemą. Šiam tikslui pasiekti yra taikomi neparimetriniai tikimybių branduolio tankio įvertinimai, apibendrinti tiesiniai modeliai ir apibendrinti adityvieji modeliai prognozavimo tikslams.

Analizė susideda iš dviejų pagrindinių dalių: pirmojoje dalyje randamas kiekvieno testo santykinio svarbumo matas, o antrojoje yra prognozuojama kiekvieno studento veikla. Pirmajai daliai įvykdyti yra pasitelkiami aprašomieji metodai, tokie kaip tikimybių branduolio tankio įvertinimas. Antroje dalyje bandoma įvertinti kiekvieno studento įgytas žinias ir tuo remiantis prognozuoti jo tolimesnę veiklą. Šiam tikslui pasiekti yra naudojami prognozuojantys statistiniai metodai: logistinė regresija ir tiesinė regresija.

Tyrimo rezultatai yra pateikiami lentelėmis ir grafikais.

Toliau pateikiama minėtų metodų realizacija SAS statistinės analizės sistemoje.

Tikimybių branduolio tankio įvertinimas

KDE procedūra vykdo vienmatį arba dvimatį tikimybių branduolio tankio įvertinimą [11, 12]. Statistinis tankio įvertinimas apima tikėtinos tikimybės tankio funkcijos įvertinimą turint konkrečius surinktus duomenis. Tikimybių branduolio tankio įvertinimas yra neparаметrinis būdas tankiui įvertinti, pagal kurį visuose surinktų duomenų taškuose yra apskaičiuojamas tankio funkcijos vidurkis, taip sukuriant tolygią aproksimaciją. KDE procedūra naudoja Gauso tankį kaip branduolį, ir jo parinkta dispersija nustato galutinio įverčio tolygumą. Ji išveda branduolio tankio įvertį į SAS duomenų rinkinį, kurį galima vėliau panaudoti su kitomis procedūromis braižymui ar analizei. KDE procedūra taip pat skaičiuoja daugybę bendrų statistikų, įskaitant tikėtinos tikimybės tankio funkcijos procentilių įverčius.

Logistinė regresija

LOGISTIC procedūra pritaiko tiesinius logistinius regresijos modelius dvireikšmiams arba daugiareikšmiams duomenims naudojant didžiausio tikėtimumo metodą [13, 14]. Didžiausio tikėtimumo įvertinimas yra atliekamas su Fišerio taškiniu algoritmu arba su Newton-Raphson algoritmu. Parametro įverčiams galima apibrėžti pradines reikšmes. Logit ryšio funkcija logistinės regresijos modeliuose gali būti pakeista probit funkcija ar papildoma kartotinio logaritmo (log-log) funkcija. LOGISTIC procedūra leidžia arba pilno rango parametrizaciją, arba mažesnę parameterizaciją. Pilno rango parameterizacija siūlo keturis kodavimo metodus: padarinys (angl. effect), nuoroda (angl. reference), daugianaris (angl. polynomial), ir ortogonalus daugianaris (angl. orthogonal polynomial). Padarinio kodavimas yra tas pats metodas, kuris yra naudojamas CATMOD procedūroje. Mažesnė, negu pilno rango parameterization yra tas pats kodavimas, kuris yra naudojamas GLM ir GENMOD procedūrose.

Kitos LOGISTIC procedūros savybės leidžia:

- valdyti atsakymo lygių rūšiavimą;
- skaičiuoti kriterijaus R^2 reikšmę pritaikytam modeliui;
- perklasifikuoti binarinių atsakymų stebėjimus remiantis jų atsakymų tikimybių prognozėmis;
- tikrinti tiesines hipotezes apie regresijos parametrus;
- kurti duomenų rinkinį, skirtą ROC kreivės braižymui kiekvienam pritaikytam modeliui;
- kurti duomenų rinkinį su apskaičiuotomis atsakymų tikimybėmis, liekanomis ir poveikio statistikomis.

Tiesinė regresija

REG procedūra yra viena iš daugelio regresijos procedūrų SAS sistemoje. Tai yra bendro pobūdžio regresijos procedūra. Kitos SAS regresijos procedūros atlieka labiau specializuotas funkcijas [15, 16]. Kitos SAS/STAT regresinės analizės procedūros yra CATMOD, GENMOD, GLM, LOGISTIC, MIXED, NLIN, ORTHOREG, PROBIT, RSREG ir TRANSREG. SAS/ETS procedūros yra taikomos programoms, dirbančioms su laiko eilutėmis ar vienalaikėmis sistemomis.

PROC REG suteikia šias galimybes:

- daugialypius MODEL teiginius;
- devynis modelių pasirinkimo metodus;
- tiesinės lygybės apribojimus parametrui;
- tiesinių ir daugiamačių hipotezių tikrinimą;
- kolinearumo diagnostiką;
- prognozuotas reikšmes, liekanas, studentizuotas liekanas, pasikliautinumo intervalus ir poveikio statistikas;
- norimos statistikos gali būti įrašytos į išvesties duomenų rinkinius;
- įvairius grafikus.

Išlikimo analizė

SAS sistemoje yra trys procedūros išlikimo duomenų analizei: LIFEREG, LIFETEST ir PHREG [17, 18]. LIFETEST yra neparimetrinė procedūra išlikimo laiko skirstinio įvertinimui ir ryšio tarp išlikimo laiko ir kitų kintamųjų tikrinimui. Ją naudojant galima gauti Kaplano ir Mejerio gyvavimo lentelės (draudimo statistikos) skirstinio įverčius. Procedūra taip pat skaičiuoja ryšio tarp išlikimo laiko kintamojo ir kitų lydinčių kintamųjų rangų testus. LIFEREG ir PHREG yra regresijos procedūros išlikimo laiko skirstinio modeliavimui. LIFEREG procedūra pritaiko parametrinius padidintos rizikos laiko modelius išlikimo duomenims. PHREG procedūra pritaiko proporcingąjį rizikos modelį (Kokso modelį (angl. Cox model)) išlikimo duomenims. Modelis yra vadinamas pusiau parametriniu modeliu, kadangi viena modelio dalis apima nenustatytą bazinę funkciją nuo laiko (kuri turi begalinį matavimą), o kita dalis apima baigtinį regresijos parametru skaičių.

1.2. SISTEMOS SAS OBJEKTINIO PROGRAMAVIMO TECHNOLOGIJOS IR JŲ GALIMYBIŲ ANALIZĖ

Java technologijomis pagrįsti programiniai sprendimai tampa vis labiau ir labiau populiarūs dėl savo mobilumo, lankstumo ir gerų rezultatų. Java appletų ir Java programų kūrimui gali būti naudojamas SAS AppDev Studio programinis paketas. Tokie įrankiai kaip webAF ir webEIS

programinė įranga bei JSP ir Java Servlet technologijos leidžia sukurti appletus ir Java programas, kurios gali dirbti su įvairiais SAS duomenimis [19].

Jeigu reikia sukurti labai interaktyvias vartotojo sąsajas, yra naudojamos kliento pusės (angl. client-side) Java technologijos. Ir appletai, ir Java programos suteikia tą patį interaktyvumo lygmenį, bet pateikimu vartotojams.

Toliau šios technologijos apžvelgiamos plačiau.

1.2.1. JAVA APLETAI IR PROGRAMOS

Naudojant Java appletus, nereikia įdiegti programos vartotojo kompiuteryje. Vietoj to, kai appletas yra iškviečiamas (paprastai per HTML puslapį), reikalingos Java klasių bylos yra automatiškai atsiunčiamos iš web serverio [19]. Tuomet appletas yra įkeliamas į atmintį per naršyklės priedą, kuris interpretuoja ir vykdo atsiųstas Java klases. Dėl tokio pateikimo modelio, appletai yra priklausomi nuo saugumo apribojimų klientui ir serveriui.

Java programos gali būti sukurtos ir paleistos nepriklausomai, panašiai kaip ir kitos iškviečiamos programos. Šios programos turi būti įdiegtos kiekviename kliento kompiuteryje. Dėl šios priežasties, skirtingai nei naršyklės, Java programos neturi jokių saugumo apribojimų.

Neigiamas Java programų bruožas yra tai, kad joms vis dar reikalinga administracija: kai programa yra atnaujinama, tas atnaujinimas turi būti padarytas kiekviename kompiuteryje, kur ta programa yra. Priešingai nei Java programos, appletai naudoja tokį modelį, kuris yra artimesnis “nulinei administracijai”. Atnaujinimas appletui turi būti padarytas tik vieną kartą - į atvaizdį, kuris yra saugomas web serveryje ir automatiškai atsiunčiamas klientui, kai tik appletas yra iškviečiamas iš to kliento kompiuterio.

1.2.2. SAS webAF IR webEIS

webAF programinė įranga yra vizuali programų kūrimo aplinka, kuri leidžia greitai kurti Java programas, appletus, internetines programas ir klases, naudojant interaktyvią į objektus orientuotą vartotojo sąsają, kuri sumažina programavimo kiekį [7, 8, 20, 21].

webAF programinė įranga padeda kurti programas, kurias lengva valdyti ir kurios greitai susijungia su SAS programine įranga. Taip pat yra teikiamas palaikymas kuriant ir testuojant Java servlets ir Java ServerPages objektus. webAF programinės įrangos vizuali kūrimo aplinka suteikia lengvą prieigą prie SAS programinės įrangos per Java klases, nesudėtingą prieigą prie SAS/AF objektų, lentelių ir daigiamųjų duomenų bazių ir prie SAS skaičiavimų per procedūrinės operacijas.

Taigi SAS webAF programinė įranga suteikia:

- **lengvą prieigą prie įvairios informacijos per internetą.** Ji naudojama kurti patogias, Java technologijomis paremtas programas, kurios gali naudotis visomis SAS programinės įrangos galimybėmis – nuo duomenų saugojimo iki pagalbos darant sprendimus. webAF taip pat palengvina vientisą prieigą prie plataus duomenų šaltinių diapazono, apimant pagrindines duomenų bazių valdymo sistemas (Oracle, SYBASE, Informix, DB2) visose pagrindinėse operacinėse sistemose, įskaitant HP-UX, Solaris, AIX/6000, COMPAQ Tru64 UNIX, Windows 95/98/NT/2000 ir OS/390;
- **greitą programų kūrimą ir ryšius su SAS.** Su webAF, yra išvengiamas kodo perrašymas norint suderinamumo su įvairiomis platformomis. Vietoj to, galima į programą įkelti iš anksto supakuotus SAS komponentus ir sujungti kuriamą programą su visomis SAS programinės įrangos galimybėmis. Taip pat galima pakartotinai panaudoti jau sukurtą kodą kuriant nestandartinius objektus. Objektai, kurie yra sukuriami naudojant webAF veikia identišškai ant bet kokios palaikomos platformos;
- **centralizuotą programos valdymą.** webAF programinės įrangos unikali n-lygių aplinka siunčia užklausas į tinkamiausią SAS serverį, šalindama nereikalingus serverio seansus ir daug laiko reikalaujančias programų valdymo užduotis daugialypėse platformose ir tinkluose. Java apletai ir programos, kurios yra sukurtos su webAF programine įranga yra lengvai platinamos, atnaujinamos ir valdomos internetu iš vienos centralizuotos vietos;
- **galimybę sukurti tarpininkus, kurie yra suderinami su JavaBeans, norint prisijungti prie nuotolinių SAS/AF objektų, esančių SAS serveryje.** Tas pats gebėjimas įgalina sukurti ryšius prisijungimui prie nuotolinių objektų per CORBA ir Remote Method Invocation (RMI).

webAF programinė įranga sukuria kokybiškus Java apletus ir programas be paslėptų priedų ar makrokomandų, taigi sukurtas programos yra lengva integruoti su programiniu kodu, gautu naudojant kitus programų kūrimo įrankius. webAF programinė įranga taip pat integruoja Java programas, kurios nėra susijusios su SAS. webAF suteikia paprastą kūrimo aplinką, naudodama standartines grafines vartotojo sąsajas, kurios sumažina apmokymo išlaidas ir organizacijos informacijos prieigai prideda nuoseklumo.

webEIS programinė įranga yra internetinė OLAP sistema. Ji leidžia vartotojams greitai ir lengvai sukurti interaktyvius dokumentus ir paskui juos publikuoti internete kaip Java apletus, arba naudojant Java ServerPage (JSP) technologiją. webEIS platformoje galingi analizės įrankiai yra sujungiami su draugiška ir patogia naudoti aplinka [14].

OLAP dokumentai yra pagaminami naudojant aukšto lygio vartotojo sąsają. Aplinka leidžia verslo analitikams susikurti savus internetinius dokumentus informacijos peržiūrėjimui ir saugojimui. Jie gali prieiti prie duomenų, peržvelgti norimas sritis ir pateikti ataskaitas įvairiais formatais. Duomenys, kurie yra saugomi daigiamatėje duomenų struktūroje yra peržiūrimi ir naršomi naudojant interneto naršyklę, tokią kaip Microsoft Internet Explorer ar Netscape.

Kūrėjai gali naudoti webEIS norėdami sukurti nestandartinį OLAP peržiūros įrankį InformationBeans, kuris yra JavaBeans technologija paremta programa, kuri pasinaudoja pažangiu SAS programinės įrangos duomenų prieigos ir sprendimo palaikymo funkcionalumu. Be to, kūrėjai gali pakartotinai panaudoti InformationBeans kituose Java apletuose ar programose. Naudojant šią technologiją sutrumpėja programų kūrimo ciklo trukmė ir daugiau laiko lieka strateginiams projektams.

1.2.3. JAVA SERVLET IR JSP

Java Servlet – Sun Microsystems sukurta technologija, skirta dinaminiam turiniui generuoti. Java Servlet yra programuojamas Java programavimo kalba [7, 22]. Pagrindinė sąvoka šioje technologijoje – servletas. Tai serveryje esantis komponentas, gaunantis iš kliento (pvz. naršyklės) užklausą bei generuojantis atsaką. Servletas gali būti generuojamas automatiškai, pasitelkiant Java ServerPages technologiją ar kitą šablonų generavimo variklį.

Servleto naudojimas

Servletas gali būti naudojamas keliais būdais:

- dviejų lygių architektūra – servletas pats vienas generuoja vartotojui reikalingą puslapį. Šiuo atveju vartotojo naršyklė yra pirmas lygis, o servletas – antras. Ši architektūra pasirenkama, kuomet servleto užduotis yra loginiu požiūriu paprasta;
- trijų lygių architektūra – servletas darbo metu susisiekiama arba su atskirai parašytu ir veikiančiu ir sprendimus priimančiu moduli, arba tiesiai su duomenų baze. Tiesioginis susisiekimas su duomenų baze, nors ir techniškai galimas, paprastai nerekomenduojamas;
- keturių lygių architektūra – servletas darbo metu susiekiama su atskirai parašytu ir veikiančiu sprendimus priimančiu moduli. Susiekiama dažniausiai per CORBA ar RMI protokolus. Šis modulis savo ruožtu susiekiama su viena ar keliomis duomenų bazėmis.

Servleto gyvavimo ciklas

Servletas gali būti vienoje iš šių būsenų:

- inicializuojamas: šiuo metu arba vykdomas servleto konstruktoriaus kodas, arba `init()` metodas. Servletą kuria konteineris, naudodamas konstruktorių be parametrų. Sukūrus servletą, vykdomas `init()` metodas. Viso servleto gyvavimo metu šis `init()` iškviečiamas tiktai vieną kartą;
- veikia (inicializuotas): vykdomas servleto metodas `service()`. Būtent šioje būsenoje servletas gali apdoroti kliento užklausas bei generuoti atsakus. Kiekvieną atsaką apdoroja atskira gija;
- sunaikintas: šiuo metu vykdomas metodas `destroy()`. Jis yra iškviečiamas vieną kartą ir yra skirtas atlikti veiksams, kurių gali prireikti prieš sunaikinant servleto objektą;
- neegzistuoja: servleto objektas nesukurtas arba sunaikintas.

Java ServerPages (JSP) yra technologija, leidžianti dinamiškai generuoti HTML, XML, ar kito tipo puslapius. Ši technologija suteikia galimybę į statinį puslapį įterpti Java kodą ir kitą dinaminį turinį [7, 23].

JSP puslapyje kaitaliojasi įprastine HTML kalba parašyti fragmentai bei Java kalba parašytos dalys. J2EE serveris vykdo puslapyje esantį Java kodą kuris iš anksto specialia JSP direktyva deklaruotiems kintamiesiems priskiria reikiamas reikšmes. Kita JSP direktyva šias reikšmes galima reikiamoje vietoje įterpti į dokumentą.

Geroje sistemoje JSP puslapio kodas atsakingas tik už paties dokumento generavimą. Sąveika su duomenų bazėmis, sudėtingesni skaičiavimo algoritmai, sprendimų priėmimas bei kiti panašūs algoritmai turi būti vykdomi aukštesniame lygyje.

Parametrai

JSP Java programos tekstas gali pasiekti keletą iš anksto apibrėžtų kintamųjų, iš kurių svarbiausias yra `request` (angl. užklausa). Iš `request` objekto galima sužinoti, kokie buvo puslapio kvietimo parametrai bei kokie yra anksčiau toje naršyklėje įdiegti slapukai. Slapukams įdiegti naudojamas kitas iš anksto apibrėžtas kintamasis `response`. Slapukai dažniausiai reikalingi prisijungimui prie slaptažodžiu apsaugotos sistemos.

JSP puslapio dalys

JSP puslapis paprastai turi tokias dalis:

- statiniai duomenys, pavyzdžiui, HTML;

Statiniai duomenys yra tiesiogiai perrašomi į HTTP atsakymą. Taigi paprasta HTML byla yra suprantama kaip korektiškas Java ServerPage puslapis be jokių specifinių elementų. Tokiu atveju į HTTP atsakymą visą laiką būtų siunčiama ta pati informacija. Bet JSP tikslas yra papildyti puslapį dinaminio turiniu, todėl vien tik paprasto HTML pateikimas JSP kompiliatoriui yra neprasmingas.

- JSP direktyvos;

JSP direktyvos kontroliuoja kaip JSP kompiliatorius generuoja servletą. Žemiau pateiktas galimų direktyvų sąrašas:

- `include` – ši direktyva informuoja JSP kompiliatorių įterpti pilną bylą į redaguojamą. Įterpimą galima išivaizduoti kaip veiksmą „įterpti“ toje vietoje kur įrašyta direktyva. Include direktyva funkcionuoja panašiai kaip `include` komanda C preprocesoriuje;
- `page` – yra keletas `page` direktyvos variantų:

<code>import</code>	jei norima į galutinę bylą įterpti „ <code>java import</code> “ sakinį nustato generuojamo bylos turinio tipą. Reiktų naudoti, jei
<code>contentType</code>	generuojama byla ne HTML arba „character set“ nėra standartinis
<code>errorPage</code>	nurodo puslapį kuris bus parodomas įvykus generavimo klaidai
<code>isErrorPage</code>	jeigu nustatyta „true“ reikšmė, šis byla laikoma klaidos puslapiu
<code>isThreadSafe</code>	nurodo ar servletas, kuris bus sugeneruotas, yra saugus gijų atžvilgiu

- `taglib` direktyva nurodo kad bus naudojama JSP žymių biblioteka. Direktyva reikalauja kad būtų nurodytas prefiksas ir žymės biblioteką aprašantis URI;

- JSP skripto elementai ir kintamieji;

skripte visada yra prieinami šie kintamieji:

- `out` – `JSPWriter`, naudojamas perduoti duomenis į HTTP atsakymo srautą;
- `page` – pats servletas;
- `pageContext` – `PageContext` objekto elementas, laikantis su visu puslapiu susijusią informaciją. Tokiu būdu keliems JSP gali būti perduodamas tas pats HTML puslapis;
- `request` – HTTP užklausos objektas;
- `response` – HTTP atsako objektas;

- session – HTTP session objektas kuris gali būti naudojamas sekti vartotojo sesijai tarp atskirų puslapių.
- JSP veiksmai;
- papildomi elementai.

Apibendrinant galima teigti, kad apžvelgtos technologijos yra naudingos kuriant ir diegiant duomenų analizės sistemas, tačiau turi savų privalumų ir trūkumų. Pagrindinis SAS webAF, webEIS, Java Applets ir JSP servlets technologijų privalumas yra tai, kad jos suteikia aukštą interaktyvumo lygį, mobilumą bei lankstų bendravimą tarp programos ir vartotojo. Trūkumas yra tai, kad šios technologijos yra pilnai priklausomos nuo vartotojo/programų kūrėjo kompiuteryje įdiegtos Java sistemos versijos. Nėra vieningos universalios versijos, kuri tiktų visoms Java programoms.

1.3. DARBE SPRENDŽIAMŲ UŽDAVINIAI

Pagrindinis darbo tikslas yra sukurti studijų proceso duomenų analizės modelius panaudojant statistikos metodus, programų paketą SAS ir interneto technologijas. Sukurtos priemonės turi veikti interneto aplinkoje realiu laiku, turėti interaktyvią vartotojo sąsąką bei spręsti uždavinius, kurių rezultatai naudojami priimant sprendimus dėl studijų kokybės tobulinimo.

Darbe yra sprendžiami šie uždaviniai:

1. remiantis literatūros šaltiniais išanalizuoti, kokie statistiniai metodai, modeliai ir programinės priemonės yra taikomos studijų proceso duomenų analizei Lietuvoje ir užsienyje;
2. atlikti sistemos SAS objektinio programavimo technologijų galimybių analizę;
3. atlikti KTU studijų proceso duomenų bazių struktūros analizę;
4. panaudojant statistikos metodus sukurti statistinės analizės modelius pasirinktų studijų proceso duomenų analizės uždavinių sprendimui;
5. panaudojus sistemą SAS sukurti studijų proceso duomenų analizės programines priemones, veikiančias interneto aplinkoje realaus laiko režime, parengti sukurtų priemonių taikymo metodiką ir aprašymą;
6. parengti rekomendacijas dėl tolimesnio sukurtų modelių ir programinių priemonių tobulinimo.

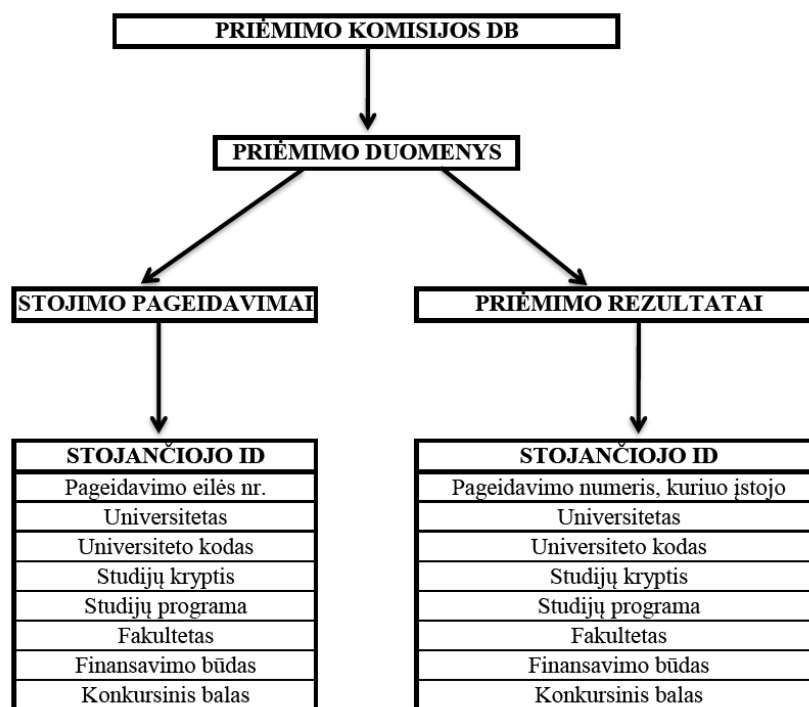
2. TIRIAMOJI DALIS

2.1. KTU STUDIJŲ PROCESO DUOMENŲ BAZIŲ STRUKTŪRA

Studijų procesas – tai pagrindinis universiteto veiklos tipas, kurio metu yra paruošiami kvalifikuoti specialistai ir mokslininkai. Jis susideda iš dviejų pagrindinių dalių: priėmimo į universitetą ir studijų jame. Priėmimo metu visi su studento stojimu ir mokymusi susiję duomenys yra išsaugomi priėmimo komisijos duomenų bazėje ir yra naudojami stojančiųjų konkursinei eilei sudaryti.

Studijas sudaro studijų programos ir moduliai. Jų eigoje studentai atsiskaitinėja už modulius ir įvertinimų rezultatai bei studentų mokymosi roikliai (skolininkas ar ne; pirmą kursą baigė ar ne) išsaugomi studijų rezultatų duomenų bazėje. Taigi studijų proceso duomenų bazių struktūra susideda iš trijų pagrindinių dalių: stojimo duomenų, priėmimo duomenų ir studijų rezultatų duomenų. Pirmosios dvi priklauso priėmimo komisijos duomenų bazei, o trečiosios duomenys saugomi studijų rezultatų duomenų bazėje. Abi duomenų bazės sujungiamos naudojant vidinį stojančiojo-studento kodą.

Išanalizavus Kauno technologijos universiteto priėmimo komisijoje saugomų duomenų apie stojančiuosius tipus ir panagrinėjus studijų rezultatų duomenų bazėje saugomų duomenų apie studentų studijų pasiekimus struktūrą, buvo pasirinkti kintamieji, kuriais remiantis sukurti studijų proceso statistinės analizės modeliai. 2.1 pav., 2.2 pav. ir 2.3 pav. yra pateikiamos visų trijų duomenų dalių struktūrinės schemas, atspindinčios pasirinktus kintamuosius.



2.1 pav. Priėmimo duomenų struktūrinė schema

2.2. STUDIJŲ PROCESO DUOMENŲ ANALIZĖS MODELIAI

Aprašytiems kintamiesiems tirti sudaryti studijų proceso duomenų statistinės analizės modeliai. Buvo naudojami aprašomosios statistikos, koreliacinės, vienfaktorinės ir dvifaktorinės dispersinės, logistinės regresinės ir klasterinės analizės metodai. Pagrindinis modelių tikslas – pateikti statistiškai reikšmingas KTU studijų proceso prognozes ir išvadas. Statistinė analizė yra atliekama duomenų pjūviui, kuris yra gaunamas grupuojant pradinius duomenis pagal 8 grupuojančius kintamuosius:

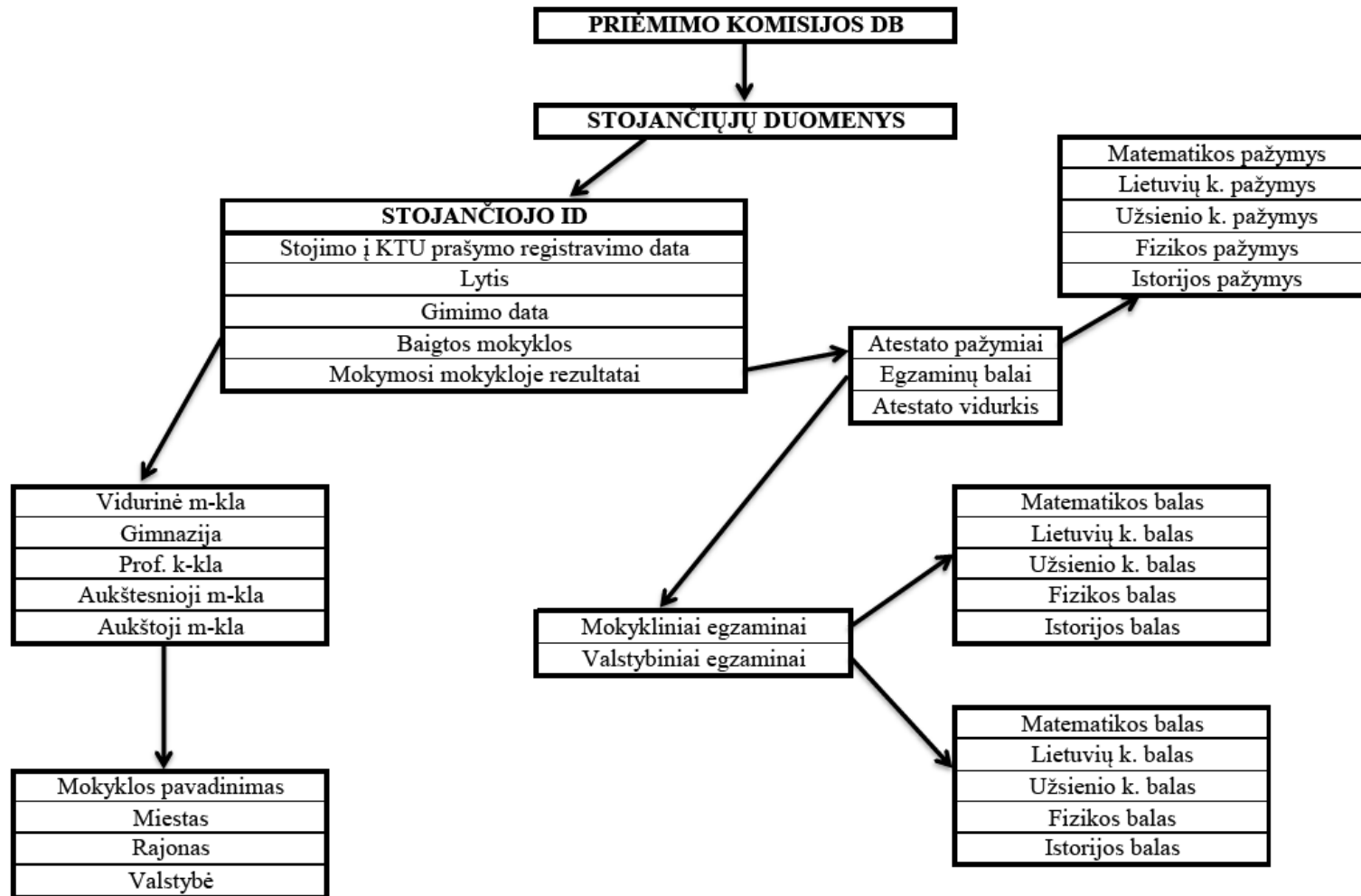
- studijų sritį,
- studijų kryptį,
- universitetą,
- fakultetą,
- įstojimo į universitetą metus,
- studijuojamą studijų programą,
- miestą, kuriame baigta vidurinė mokykla arba gimnazija,
- lytį.

Suformavus pjūvius ir pasirinkus statistinius modelius, atrinktiems studentams skaičiuojamos statistinės charakteristikos bei braižomi grafikai. Šiame skyrelyje pateikti statistinės analizės modeliai sudaryti remiantis [6, 9, 10, 27-30].

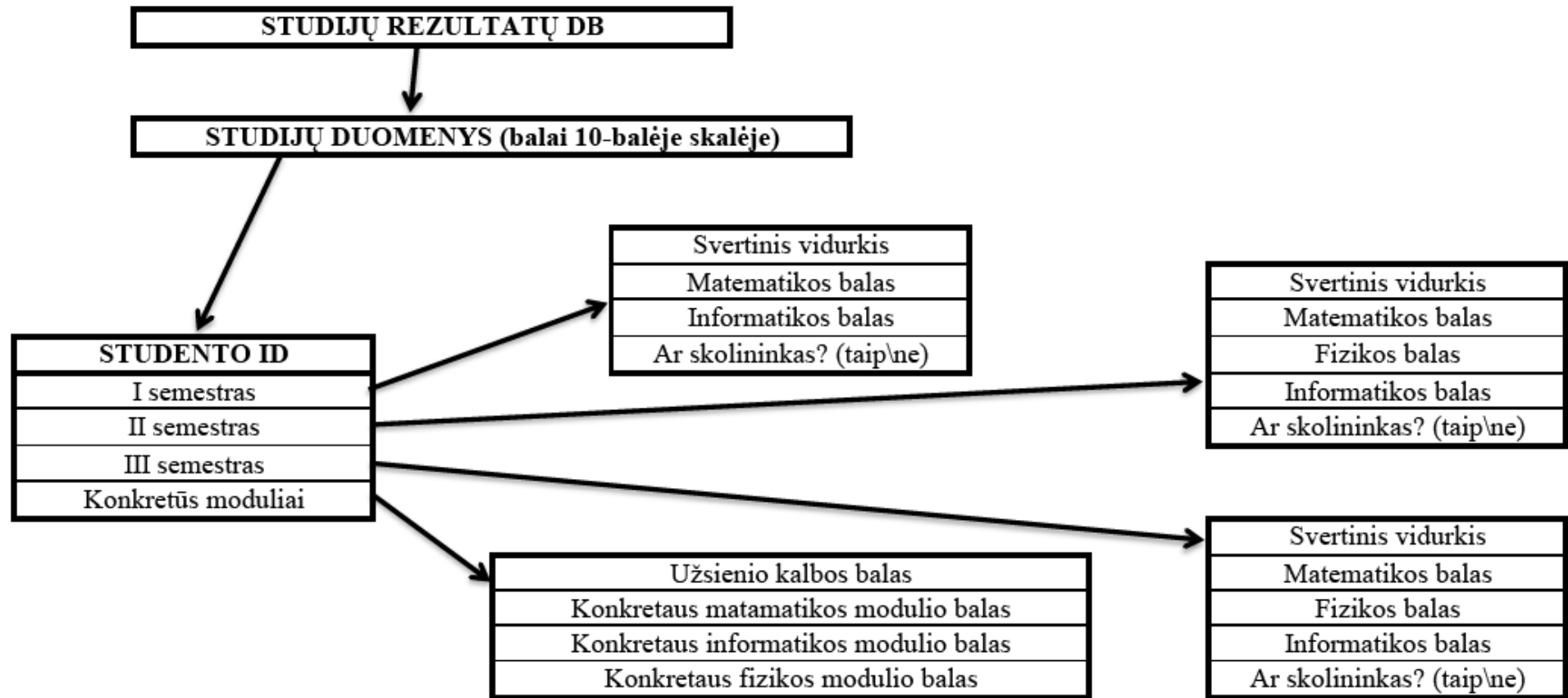
2.2.1. APRAŠOMOJI STATISTIKA

Vartotojui suformavus duomenų pjūvį pateikiamos šios aprašomosios statistikos:

- a) duomenų pjūvyje esančių studentų skaičius n ;
- b) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių (kuriuo studentas įstojo), studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių aritmetinis vidurkis (imties vidurkį žymėsime \bar{X});
- c) standartinis nuokrypis, kuris matuoja grupės abiturientų-studentų mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių (kuriuo studentas įstojo), studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba nurodytų studijų sričių skaičių sklaidą (imties standartinį nuokrypį žymėsime $\hat{\sigma}$);
- d) standartinis nuokrypis, kuris matuoja grupės abiturientų-studentų mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių (kuriuo studentas įstojo), studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba nurodytų studijų sričių skaičių sklaidą;



2.2 pav. Stojimo duomenų struktūrinė schema



2.3 pav. Studijų duomenų struktūrinė schema

- e) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių dispersija (imties dispersiją žymėsime $\hat{\sigma}^2$);
- f) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studento pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių aritmetinio vidurkio standartinė paklaida;
- g) minimali mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių reikšmė;
- h) maksimali mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių reikšmė;
- i) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių mediana;
- j) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių moda;
- k) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių sklaidos matas, lygus skirtumui tarp didžiausio ir mažiausio įvertinimo;
- l) kvartilių skirtumas IQR ;
- m) kvantiliai;
- n) mokyklos metinių pažymių vidurkių, konkursinių balų, pageidavimo numerių, studentų pajėgumo indeksų, stojimo pageidavimų skaičių arba prašyme nurodytų studijų sričių skaičių pasiskirstymas pateikiamas skritulinėse, stulpelinėse, žiedinėse ir žvaigždinėse diagramose.

2.2.2. KORELIACINĖ ANALIZĖ

2.1 lentelėje pateikiami skyrelyje naudojami žymėjimai.

2.1 lentelė

Koreliacinėje analizėje naudojami žymėjimai

Žymėjimas	Prasmė
$G = \{G_1, G_2, \dots, G_q\}$	Grupuojantys kintamieji, pagal kurių yra formuojamas

2.1 Ietelės tėsiny

Žymėjimas	Prasmė
	pradinės imties duomenų pjūvis ir taikomi koreliacinės analizės metodai
X, Y	Studentų mokymosi rodikliai, tarp kurių yra įvertinamas koreliacinio ryšio stiprumas (pvz. tarp konkursinio balo ir pajėgumo indekso)
$x_i, y_i, i = 1, 2, 3, \dots, n$	Atitinkamai X ir Y kintamųjų reikšmės
$rx_i, ry_i, i = 1, 2, 3, \dots, n$	X ir Y kintamųjų reikšmių (x_i, y_i) rangai
n	Duomenų pjūvyje esančių studentų skaičius
ρ_S	Spirmeno ranginės koreliacijos koeficientas
$\hat{\rho}_S$	Spirmeno ranginės koreliacijos koeficiento taškinis įvertis
T	Stjudento statistika
\hat{t}_b	Kendalo ranginės koreliacijos koeficiento taškinis įvertis
α	Reikšmingumo lygmuo

Visų pirma yra formuojamas pradinės studijų proceso duomenų imties pjūvis. Panaudojant grupuojančius kintamuosius G yra gaunama charakteringa studentų grupė, kuriai atliekama koreliacinė analizė.

Koreliacinėje analizėje koreliacijos koeficientu yra įvertinamas ryšio stiprumas tarp studentų stojimo pageidavimo numerių, mokyklos metinių pažymių vidurkių, konkursinių balų, stojimo pageidavimų skaičiaus, stojimo prašyme nurodytų studijų sričių skaičiaus, taip pat tarp studentams sudarytų pajėgumo indeksų. Priklausomai nuo gautos koeficiento reikšmės formuluojamos išvados apie ryšio stiprumą. Tarp studentų mokymosi rodiklių gali nebūti ryšio, jis gali būti labai silpnas, silpnas, vidutinio stiprumo arba labai stiprus ryšys.

Jeigu kintamųjų X ir Y reikšmių skirstiniai nėra normalieji, yra naudojamas Spirmeno ranginės koreliacijos koeficientas ρ_S , apibūdinantis ryšio tarp X ir Y stiprumą monotoniškumo prasme. Apskaičiuojant Spirmeno ranginį koreliacijos koeficientą naudojami kintamųjų X ir Y stebėjimų (x_i, y_i), $i = \overline{1, N}$, rangai (rx_i, ry_i). Empirinis Spirmeno koreliacijos koeficientas apskaičiuojamas pagal formulę [9, 10]:

$$\hat{\rho}_S = 1 - \frac{6 \cdot \sum_{i=1}^n (rx_i - ry_i)^2}{n(n^2 - 1)}, -1 < \hat{\rho}_S < 1.$$

Norint nustatyti Spirmeno koreliacijos koeficiento taškinio įverčio $\hat{\rho}_S$ statistinį reikšmingumą ir prie kokio $\hat{\rho}_S$ didumo galima teigti, kad tarp klausimų X ir Y įvertinimų yra statistiškai reikšmingas ryšys, yra tikrinamos hipotezės $H_0 : \hat{\rho}_S = 0, H_a : \hat{\rho}_S \neq 0$.

Hipotezėms H_0 tikrinti naudojama Stjudento statistika [9, 10]:

$$T = \frac{\hat{\rho}_S}{\sqrt{1 - \hat{\rho}_S^2}} \cdot \sqrt{n - 2}, T \sim t(n - 2).$$

Modelyje taip pat yra skaičiuojamas ir Kendalo ranginės koreliacijos koeficientas \hat{t}_b , kuris naudojamas įvertinti ryšio tarp kintamųjų X ir Y stiprumą suderinamumo prasme. Jis apskaičiuojamas pagal formulę [9, 10]:

$$\hat{t}_b = \frac{2S}{n(n - 1)},$$

kur S yra suderintų ir nesuderintų klausimų įvertinimų porų skaičiaus skirtumas.

Hipotezių apie Kendalo ranginės koreliacijos koeficiento reikšmingumą $H_0 : \hat{t}_b = 0, H_a : \hat{t}_b \neq 0$ tikrinimui naudojama Z statistika [9, 10]:

$$Z = \frac{\hat{t}_b}{\sqrt{\frac{2(2n + 5)}{9n(n - 1)}}}.$$

2.2.3. DISPERSINĖ ANALIZĖ

2.2.3.1. VIENFAKTORINĖ DISPERSINĖ ANALIZĖ

2.2 lentelėje pateikiami šiame skyrelyje naudojami žymėjimai.

2.2 lentelė

Vienfaktorinėje dispersinėje analizėje naudojami žymėjimai

Žymėjimas	Prasmė
V	Nepriklausomasis kintamasis (fakultetas, finansavimo tipas, lytis, miesto grupė, miestas/rajonas, pageidavimo numerio grupė, konkursinio balo grupė, studijų kryptis, studijų forma, studijų programa, studijų sritis)

2.2 Ietelės tėsiny

Žymėjimas	Prasmė
I	Faktoriaus V lygmenų skaičius (populiacijų skaičius)
E	Atsitiktinių paklaidų faktorius
ε	Vienfaktorinės dispersinės analizės modelio atsitiktinės paklaidos
Y	Priklausomasis kintamasis (mokyklos metinių pažymių vidurkis, konkursinis balas, studento pajėgumo indeksas, stojimo pageidavimų skaičius, studijų sričių skaičius, konkretaus modulio universitete įvertinimai)
$Y_i, i = 1, 2, \dots, I$	i-tosios populiacijos priklausomas kintamasis
n_i	i-tosios populiacijos imties didumas
n	Jungtinės populiacijos didumas
F	Fišerio statistika
$\bar{Y}_i, i = 1, 2, \dots, I$	i-tosios populiacijos imties vidurkis
$\mu_i, i = 1, 2, \dots, I$	i-tosios populiacijos vidurkis
$\sigma^2_i, i = 1, 2, \dots, I$	i-tosios populiacijos dispersija
\bar{Y}	Priklausomojo kintamojo imties vidurkis
α	Reikšmingumo lygmuo

Kai faktoriaus V lygmenų skaičius $I \geq 3$, atsitiktinės paklaidos ε pasiskirsčiusios pagal normalųjį dėsnį ir populiacijų kintamųjų dispersijos lygios, naudojamas vienfaktorinės dispersinės analizės Fišerio kriterijus. Modelio nepriklausomieji kintamieji yra fakultetas, finansavimo tipas, lytis, miesto grupė, miestas/rajonas, pageidavimo numerio grupė, konkursinio balo grupė, studijų kryptis, studijų forma, studijų programa, studijų sritis. Modelio priklausomieji kintamieji yra mokyklos metinių pažymių vidurkis, konkursinis balas, studento pajėgumo indeksas, stojimo pageidavimų skaičius, studijų sričių skaičius, konkretaus modulio universitete įvertinimai.

Tarkime faktorius V turi I lygmenų ir turime I nepriklausomų populiacijų. Priklausomas kintamasis, matuojamas *i*-toje populiacijoje, yra populiacijos kintamasis. Populiacijų kintamuosius pažymėkime Y_1, Y_2, \dots, Y_I . Iš kiekvienos populiacijos pasirenkama paprastoji atsitiktinė imtis:

$Y_{i1}, Y_{i2}, \dots, Y_{in_i}$; čia $i = 1, 2, \dots, I$ – populiacijos numeris, o n_i – i -tosios imties didumas. Visų populiacijų imčių didumų suma lygi $n = \sum_{i=1}^I n_i$.

Modelis i -tosios imties j -tajam stebėjimui Y_{ij} yra užrašomas taip [9, 10]:

$$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

čia μ_i yra i -tosios populiacijos kintamojo vidurkis; ε_{ij} – nepriklausoma atsitiktinė paklaida, pasiskirsčiusi pagal normalųjį dėsnį; μ – jungtinės populiacijos vidurkis; $\tau_i = \mu_i - \mu$ yra i -tosios populiacijos vidurkio ir bendrojo vidurkio skirtumas (i -tosios populiacijos efektas). Modelis nusako kokią potencialią reikšmę gali įgyti i -tosios imties j -tasis elementas.

Vienfaktorinės dispersinės analizės tikslas – nuspręsti, ar priklausomojo kintamojo, išmatuoto skirtingose populiacijose, vidurkiai skiriasi. Nulinė vienfaktorinės dispersinės analizės hipotezė teigia, kad visi populiacijų kintamųjų vidurkiai yra lygūs:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_I,$$

$$H_a : \text{ bent du } \mu_i \text{ tarpusavyje nelygūs, } i = 1, \dots, I.$$

Hipotezių tikrinimui naudojama Fišerio statistika [9, 10].

Jeigu hipotezė H_0 atmetama, tada taikomas daugkartinio lyginimo metodas. Hipotezių tikrinimui naudojamas Tjukio kriterijus, nes jis yra labai konservatyvus. Naudojant šį kriterijų galima sudaryti homogenines grupes ir, esant didesniai populiacijų skaičiui, šio kriterijaus galia yra didesnė už kitų kriterijų galią. Tjukio kriterijus yra grindžiamas studentizuoto atstumo Q statistika. Tarkime, kad visų populiacijų imčių didumai vienodi $n_1 = n_2 = \dots = n_k = N$. Norint palyginti i -tosios ir j -tosios imčių vidurkius yra skaičiuojama statistika [9, 10]

$$Q(i, j) = \frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{SS_E/N}}$$

Kai faktoriaus V lygmenų skaičius $I \geq 3$, tačiau atsitiktinės paklaidos ε nepasiskirsčiusios pagal normalųjį dėsnį, yra taikomas Kruskalo-Voliso ranginis kriterijus nepriklausomoms imtims. Šis kriterijus naudojamas hipotezei apie trijų ir daugiau imčių skirstinių lygybę tikrinti.

Tarkime, kad Y_1, Y_2, \dots imtys yra $(Y_{11}, Y_{12}, \dots, Y_{1n_1}), (Y_{21}, Y_{22}, \dots, Y_{2n_2}), \dots$. Yra tikrinamos hipotezės:

$$H_0 : \text{ kintamųjų } Y_1, Y_2, \dots \text{ skirstiniai vienodi,}$$

$$H_a : \text{ kintamųjų } Y_1, Y_2, \dots \text{ skirstiniai nėra vienodi.}$$

Hipotezės tikrinamos apskaičiuojant Kruskalo-Voliso statistiką [9, 10].

Kai faktoriaus V lygmenų skaičius $I = 2$ ir atsitiktinės paklaidos ε pasiskirsčiusios pagal normalųjį dėsnį, yra naudojamas Stjudento t kriterijus dviem nepriklausomoms imtims, skirtas palyginti imčių vidurkius. Tarkim turime populiacijų kintamųjų Y_1 ir Y_2 imtis $(Y_{11}, Y_{12}, \dots, Y_{1n_1})$ ir $(Y_{21}, Y_{22}, \dots, Y_{2n_2})$, čia n_1 ir n_2 – atitinkamų imčių didumai. Tikrinamos hipotezės apie vidurkių \bar{Y}_1 ir \bar{Y}_2 lygybę: $H_0 : \mu_1 = \mu_2$, $H_a : \mu_1 \neq \mu_2$. Priklausomai nuo to, ar populiacijų kintamųjų Y_1 ir Y_2 dispersijos lygios, ar nelygios, hipotezių tikrinimui yra skaičiuojamos atitinkamos Stjudento statistikos [9, 10].

Kai faktoriaus V lygmenų skaičius $I = 2$, tačiau atsitiktinės paklaidos ε nepasiskirsčiusios pagal normalųjį dėsnį, yra naudojamas Mano-Vitnio-Vilkoksono rangų sumų kriterijus nepriklausomoms imtims. Jis yra naudojamas tikrinti hipotezes apie kintamųjų skirstinių lygybę.

Tarkime, kad populiacijos kintamųjų Y_1 ir Y_2 imtys yra $(Y_{11}, Y_{12}, \dots, Y_{1n_1})$ ir $(Y_{21}, Y_{22}, \dots, Y_{2n_2})$. Yra tikrinamos hipotezės:

$$H_0 : Y_1 \text{ ir } Y_2 \text{ skirstiniai vienodi,}$$

$$H_a : Y_1 \text{ ir } Y_2 \text{ skirstiniai nėra vienodi.}$$

Priklausomai nuo to, ar imtys mažos ($n_1 \leq 20$, $n_2 \leq 20$), ar didelės ($n_1 > 20$, $n_2 > 20$) hipotezės tikrinamos skaičiuojant atitinkamu atveju naudojamas Mano-Vitnio-Vilkoksono statistikas [9, 10].

2.2.3.2. DVIFAKTORINĖ DISPERSINĖ ANALIZĖ

2.3 lentelėje pateikiami šiame skyrelyje naudojami žymėjimai.

2.3 lentelė

Dvifaktorinėje dispersinėje analizėje naudojami žymėjimai

Žymėjimas	Prasmė
U, W	Nepriklausomieji kintamieji (fakultetas, finansavimo tipas, lytis, miesto grupė, miestas/rajonas, pageidavimo numerio grupė, konkursinio balo grupė, studijų kryptis, studijų forma, studijų programa, studijų sritis)
u	Faktoriaus U lygmenų skaičius (populiacijų skaičius)
w	Faktoriaus W lygmenų skaičius (populiacijų skaičius)
(i,j), $i = 1, \dots, u, j = 1, 2, \dots, w$	Populiacija, kurioje faktorius U yra lygmenyje i, o faktorius W – lygmenyje j
E	Atsitiktinių klaidų faktorius
ε	Dvifaktorinės dispersinės analizės modelio atsitiktinės paklaidos

2.3 Ietelės tėsiny

Žymėjimas	Prasmė
Y	Priklausomasis kintamasis (mokyklos metinių pažymių vidurkis, konkursinis balas, studento pajėgumo indeksas, stojimo pageidavimų skaičius, studijų sričių skaičius, konkretaus modulio universitete įvertinimai)
$Y_{ij}, i = 1, \dots, u, j = 1, 2, \dots, w$	(i, j) populiacijos kintamasis
n	Imties didumas
F_U, F_W, F_{UW}	Fišerio statistikos hipotezių apie atitinkamus faktorius tikrinimui
$\bar{Y}_{i..}, i = 1, 2, \dots, u$	Imčių vidurkis, fiksavus faktoriaus U lygmenį i (vidurkio $\mu_{i..}$ empirinis analogas)
$\bar{Y}_{.j.}, j = 1, 2, \dots, w$	Imčių vidurkis, fiksavus faktoriaus W lygmenį j (vidurkio $\mu_{.j.}$ empirinis analogas)
$\bar{Y}_{ij}, i = 1, 2, \dots, u, j = 1, 2, \dots, w$	(i, j) populiacijos kintamojo vidurkis
\bar{Y}	Priklausomojo kintamojo vidurkis
α	Reikšmingumo lygmuo

Dvifaktorinė dispersinė analizė yra atliekama ir esant bet kokiems faktorių U ir W lygmenims, tačiau tik tuomet, kai modelio atsitiktinės paklaidos ε yra pasiskirsčiusios pagal normalųjį dėsnį. Modelyje yra naudojami Fišerio kriterijai. Faktoriais gali būti fakultetas, finansavimo tipas, lytis, miesto grupė, miestas/rajonas, pageidavimo numerio grupė, konkursinio balo grupė, studijų kryptis, studijų forma, studijų programa, studijų sritis. Modelio priklausomieji kintamieji gali būti mokyklos metinių pažymių vidurkis, konkursinis balas, studento pajėgumo indeksas, stojimo pageidavimų skaičius, studijų sričių skaičius, konkretaus modulio universitete įvertinimai.

Tarkime pagal du nepriklausomus faktorius U ir W turime $u \times w$ nepriklausomų populiacijų. Kiekvienoje populiacijoje matuojamas tas pats priklausomas kintamasis Y . Susiejant šį kintamąjį su konkrečia populiacija (i, j) , gaunamas populiacijos kintamasis Y_{ij} , kurį matuojant sudaroma atsitiktinė imtis $Y_{ij1}, Y_{ij2}, \dots, Y_{ijn}$. Visų imčių didumai lygūs n .

Modelis (i, j) populiacijos k -tajam stebėjimui Y_{ijk} užrašomas taip [9, 10]:

$$Y_{ijk} = \mu + \tau_{i.} + \tau_{.j} + \tau_{ij} + \varepsilon_{ijk},$$

kur ε_{ijk} – nepriklausoma atsitiktinė paklaida, pasiskirsčiusi pagal normalųjį dėsnį; μ – bendrasis visų populiacijų vidurkis (Y vidurkis); $\tau_{i.}$ – faktoriaus U įtaka; $\tau_{.j}$ – faktoriaus W įtaka; τ_{ij} – faktorių U ir W tarpusavio sąveikos įtaka. Dydžiai $\tau_{i.}$, $\tau_{.j}$, ir τ_{ij} apskaičiuojami pagal formules:

$$\tau_{i.} = \mu_{i..} - \mu, \quad \tau_{.j} = \mu_{.j.} - \mu, \quad \tau_{ij} = (\mu_{ij} - \mu) - (\mu_{i..} - \mu) - (\mu_{.j.} - \mu).$$

Dvifaktorinės dispersinės analizės modelyje formuluojamos 3 statistinės hipotezės: apie faktoriaus U įtaką, apie faktoriaus W įtaką ir apie faktorių U ir W tarpusavio sąveiką. Jos užrašomos taip:

$$H_0 : \mu_{1..} = \mu_{2..} = \mu_{3..} = \dots = \mu_{u..},$$

$$H_a : \text{ bent du vidurkiai skiriasi,}$$

čia nulinė hipotezė H_0 reiškia, kad faktoriaus U įtakos nėra;

$$H_0 : \mu_{.1.} = \mu_{.2.} = \mu_{.3.} = \dots = \mu_{.w.},$$

$$H_a : \text{ bent du vidurkiai skiriasi,}$$

čia nulinė hipotezė H_0 reiškia, kad faktoriaus W įtakos nėra;

$$H_0 : \mu_{ij} - \mu_{i..} - \mu_{.j.} + \mu = 0,$$

$$H_a : \mu_{ij} - \mu_{i..} - \mu_{.j.} + \mu \neq 0,$$

čia nulinė hipotezė H_0 reiškia, kad faktorių U ir W tarpusavio įtakos nėra, H_a – kad tarpusavio sąveika yra.

Hipotezių tikrinimui naudojamos Fišerio statistikos [9, 10]:

$$F_U = \frac{\overline{SS_U}}{\overline{SS_E}}, F_W = \frac{\overline{SS_W}}{\overline{SS_E}}, F_{UW} = \frac{\overline{SS_{UW}}}{\overline{SS_E}}.$$

Dvifaktorinės analizės modelyje taip pat yra skaičiuojamas daugkartinio lyginimo Tjukio kriterijus, naudojamas nustatyti, kurie iš skiriamų pagal faktorius U arba W populiacijų kintamieji turi nevienodus vidurkius.

Kriterijų taikant faktoriui U, lyginamos visos pagal faktorių sudarytos populiacijos ir tikrinamos hipotezės $H_0 : \mu_{k..} = \mu_{j..}$, $H_a : \mu_{k..} \neq \mu_{j..}$, kur $k, j = 1, \dots, u, k \neq j$. Skaičiuojama statistika [9, 10]

$$T_U(k, j) = \frac{|\bar{Y}_{k..} - \bar{Y}_{j..}|}{\sqrt{\overline{SS_E} / (nw)}}$$

Analogiškai lyginami populiacijų, besiskiriančių pagal faktorių W , vidurkiai.

2.2.4. LOGISTINĖ REGRESINĖ ANALIZĖ

2.4 lentelėje pateikiami šiame skyrelyje naudojami žymėjimai.

2.4 lentelė

Logistinėje regresinėje analizėje naudojami žymėjimai

Žymėjimas	Prasmė
$y_i, i = 1, \dots, n$	Modelio priklausomojo kintamojo Y reikšmės
$X_j, j = 1, \dots, k$	Nepriklausomi kintamieji (konkursinis balas, stojimo pageidavimo numeris, mokyklos metinių pažymių vidurkis, miesto grupė, miestas/rajonas, fakultetas, studijų forma, studento pajėgumo indeksas, finansavimas, lytis)
J	Priklausomojo kintamojo Y lygmenų skaičius
n	Stebėjimų skaičius
k	Modelyje naudojamų nepriklausomų kintamųjų skaičius
ε	Atsitiktinė paklaida
α	Reikšmingumo lygmuo

Kai priklausomas kintamasis Y turi tik du lygmenis ($J = 2$) ir įgyja tik dvi reikšmes - 0 arba 1, naudojame logistinės regresijos modelį, kur Y turi binominį skirstinį. Taikant šį modelį galima spręsti tokį uždavinį: žinant studento mokymosi rezultatus įvertinti tikimybę baigti pirmą kursą;

Taigi modelio priklausomas kintamasis gali būti pirmo kurso baigimo požymis (baigė arba nebaigė), o nepriklausomi kintamieji gali būti konkursinis balas, stojimo pageidavimo numeris, mokyklos metinių pažymių vidurkis, miesto grupė, miestas/rajonas, fakultetas, studijų forma, studento pajėgumo indeksas, finansavimas, lytis. Nepriklausomi kintamieji gali būti vardų, tvarkos, intervalinėje skalėje.

Esant fiksuotoms nepriklausomų kintamųjų reikšmėms $X_1 = x_{1i}, \dots, X_k = x_{ki}$, tiesinės daugialypės regresijos modelis yra toks [6, 10]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon,$$

kur y_i yra atsitiktinis dydis, galintis įgyti 0 arba 1 su tikimybėmis $P(Y_i = 1) = p_i$, $P(Y_i = 0) = 1 - p_i$, o ε yra atsitiktinė paklaida.

Darbe taikomos įvairios ryšio funkcijos. Pavyzdžiui, tikimybės p_i priklausomybės nuo $\vec{x}_i =$

(x_{1i}, \dots, x_{ki}) modelis, gali būti aprašomas tokia lygtimi [6, 10]:

$$p_i = \frac{e^{z(\vec{x}_i)}}{1 + e^{z(\vec{x}_i)}}, \text{ kur } z(\vec{x}_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}.$$

Iš šios lygties išplaukia sekančios lygtys:

$$\frac{p_i}{1 - p_i} = e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}} = e^{z(\vec{x}_i)} \text{ ir } \ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} = z(\vec{x}_i).$$

Lygtis $\ln \frac{p_i}{1 - p_i}$ yra logit tipo ryšio funkcija (angl. link function). Tai yra viena iš keleto galimų ryšio funkcijų. Visos SAS ryšio funkcijos pateiktos 2.5 lentelėje.

2.5 lentelė

Modelyje naudojamos SAS ryšio funkcijos

Ryšio funkcijos pavadinimas	Išraiška
Logit	$g(p) = \ln \frac{p}{1 - p}$
Probit (arba Normit)	$g(p) = \Phi^{-1}(p)$, kur $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x \exp(-z^2/2) dz$ – standartinio normaliojo skirstinio funkcija.
Log-Log	$g(p) = \ln(-\ln(1 - p))$
Apibendrinta logit	$g(p) = \ln \frac{p_{i-1}}{1 - p_j}$, kur $i = \overline{2, J}$

Tarkime, duomenis sudaro stebėjimai $(y_i, x_{1i}, \dots, x_{ki})$, kur $i = \overline{1, n}$, y_i yra 0 arba 1, o x_{1i}, \dots, x_{ki} yra nepriklausomų intervalinių kintamųjų reikšmės. Modelio parametrų $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ įverčiai $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$ parenkami taip, kad logistinės regresijos modelis su turimais duomenimis būtų kuo geriau suderintas. Tam yra taikomas didžiausio tikėtimumo metodas. Parametrų įverčiai yra parenkami taip, kad tikėtimumo funkcija

$$L = \prod_{i: y_i=1} p_i \prod_{i: y_i=0} (1 - p_i)$$

būtų maksimali. Čia $p_i = \frac{e^{z(\vec{x}_i)}}{1 + e^{z(\vec{x}_i)}}$, $z(\vec{x}_i) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \dots + \widehat{\beta}_k x_{ki}$. Tada, gavus parametrų įverčius, yra skaičiuojama tikimybė p_i .

Tikrinant, ar visi į modelį įtraukti nepriklausomi kintamieji prognozavimui svarbūs ir ar modelis gerai tinka duomenims, yra naudojamas χ^2 suderinamumo kriterijus. Tarkime, kad didžiausio tikėtimumo metodu radome parametrų įverčius $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$, juos įstatėme į didžiausio tikėtimumo funkciją L ir radome L maksimumą $L(\widehat{\beta}_0, \widehat{\beta})$. Pasirinkome logistinės regresijos modelį, kur visi $\beta_j = 0$,

t.y. $z_i = \alpha$. Šiuo atveju didžiausio tikėtinumo funkcijos maksimumą pažymėkime $L(\widetilde{\beta}_0, 0)$. χ^2 kriterijus remiasi tuo, kad $L(\widehat{\beta}_0, \widehat{\beta})$ mažai skiriasi nuo $L(\widetilde{\beta}_0, 0)$, kai visi koeficientai $\beta_j = 0$. Kriterijaus statistika skaičiuojama pagal formulę [6, 9, 10]

$$\chi^2 = -2 \ln L(\widetilde{\beta}_0, 0) + 2 \ln L(\widehat{\beta}_0, \widehat{\beta})$$

ir yra tikrinamos tokios hipotezės $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0, H_a : bent\ vienas\ \beta_k \neq 0$.

Hipotezės apie kiekvieno nepriklausomojo kintamojo X_j reikšmingumą prognozavimui yra naudojama Voldo statistika [6, 9, 10]

$$W = \left(\frac{\widehat{\beta}_j}{SE_j} \right)^2, j = \overline{1, k}, \text{ kur } SE_j \text{ yra } \widehat{\beta}_j \text{ standartinio nuokrypio įvertis.}$$

Taikant Voldo kriterijų, tikrinamos hipotezės $H_0 : \beta_j = 0, H_a : \beta_j \neq 0$.

Kaip alternatyva χ^2 kriterijui, tikrinant modelio suderinamumą su duomenimis, taip pat yra naudojamas Hosmerio-Lemešou kriterijus ir Kokso-Snelo pseudodeterminacijos koeficientas [6, 9, 10]. Pirmojo veikimas paremtas tuo, kad visų stebėjimų prognozuojamųjų tikimybių $\widehat{p}_i = \widehat{P}(Y = 1, \overline{x}_i)$ variacinė eilutė padalijama į keletą grupių. Kiekvienos grupės tikimybės sudedamos. Gautos sumos traktuojamos kaip skaičiai, rodantys, kiek kartų kiekvienoje grupėje y_i turėtų įgyti reikšmę 1, jeigu logistinės regresijos modelis būtų suderintas su duomenimis. Taip pat yra sudedamos kiekvienos grupės stebėjimų y_i reikšmės. Gautosios sumos rodo, kiek kartų kiekvienoje grupėje y_i įgijo reikšmę 1. Šių dydžių skirtumą aprašo χ^2 skirstinys su $(k-2)$ laisvės laipsnių, kur k – grupių skaičius. Kai Hosmerio-Lemešou statistikos p-reikšmė $p \geq \alpha$, galime tvirtinti, kad logistinės regresijos modelis su duomenimis yra suderintas.

Kokso-Snelo pseudodeterminacijos koeficientas skaičiuojamas pagal formulę:

$$r_{CS}^2 = 1 - \left(\frac{L(\widehat{\beta}_0, \widehat{\beta})}{L(\widetilde{\beta}_0, 0)} \right)^{2/n}.$$

Kuo koeficiento reikšmė didesnė, tuo logistinės regresijos modelis yra geriau suderintas su duomenimis.

Norint nustatyti, kurie kintamieji modelyje yra pertekliniai ir neneša papildomos informacijos, yra sudaroma klasifikacijos lentelė. Joje yra pateikiamos modelio stebėtos ir prognozuotos reikšmės bei teisingai ir neteisingai klasifikuotų y reikšmių dažniai. Siekiant nustatyti pasirinkto logit modelio tinkamumą duomenims yra sudaromos visos galimos duomenų $Y=0$ ir $Y=1$ poros ir tikrinamas jų suderinamumas (naudojamas Somerso kriterijus). Kuo kriterijaus reikšmė didesnė, tuo modelis labiau tinkamas duomenims.

Kai Y lygmenų skaičius yra $J > 2$, yra naudojama politominė logistinė regresija (angl.

multinomial logistic regression). Taikant šį modelį galima spręsti tokį uždavinį: remiantis mokymosi rezultatais mokykloje nustatyti, ar jis per sesiją gaus puikų konkretaus modulio įvertinimą (10 arba 9), gerą įvertinimą (8 arba 7) ar, patenkinamą įvertinimą (6 arba 5).

Šiuo atveju modelio priklausomasis kintamasis gali būti studento pasiekimai atsiskaitant konkretų modulį. Nepriklausomi kintamieji gali būti konkursinis balas, stojimo pageidavimo numeris, mokyklos metinių pažymių vidurkis, miesto grupė, miestas/rajonas, fakultetas, studijų forma, studento pajėgumo indeksas, finansavimas, lytis. Nepriklausomi kintamieji gali būti vardų, tvarkos, intervalinėje skalėje.

Pažymėkime $p_j, j = \overline{1, J}$ tikimybes, kad priklausomas kintamasis įgis vienos iš J kategorijų reikšmę ir $\sum_{i=1}^J p_i = 1$. Tada sudaromas politominės logistinės regresinės analizės modelis, kai priklausomi kintamieji išmatuoti vardų skalėje. Viena iš priklausomojo kintamojo Y kategorijų yra pasirenkama kaip bazinė ir ja remiantis sudaromas modelis. Kaip bazinę pasirenkame, pavyzdžiui, J -tąją kategoriją.

Tikimybė, kad priklausomas kintamasis Y įgis reikšmę y_j^* ($j = \overline{1, J}$), yra [6, 9, 10]

$$p_j = \frac{e^{z(\vec{x}_j)}}{\sum_{h=1}^J e^{z(\vec{x}_h)}}, \text{ kur } z(\vec{x}_j) = \beta_{0j} + \beta_j x_1 + \beta_j x_2 + \dots + \beta_j x_k, \quad 0$$

$$z(\vec{x}_h) = \beta_{0h} + \beta_h x_1 + \beta_h x_2 + \dots + \beta_h x_k.$$

Iš šios lygties išplaukia sekančios lygtys:

$$\ln \frac{p_{ij}}{p_{ij}} = \beta_{0j} + \beta_{ji} x_{1i} + \dots + \beta_{ji} x_{ki}, \quad \text{kur } j = \overline{1, J-1}, i = \overline{1, n}.$$

Parametrų β reikšmės skiriasi ir priklauso nuo to, kokios Y kategorijos yra lyginamos.

Kai priklausomas kintamasis Y išmatuotas tvarkos skalėje, yra sudaromas politominės logistinės regresijos suminio logit (angl. cumulative logit) modelis. Šis modelis yra paprastesnis ir galingesnis už modelius, taikomus vardų skalės nepriklausomiems kintamiesiems. Šiuo atveju formuojant logitus, yra naudojama priklausomojo kintamojo kategorijų tvarka.

Pažymėkime $p_j, j = \overline{1, J}$ tikimybes, kad priklausomas kintamasis įgis vienos iš J kategorijų reikšmę ir $\sum_{i=1}^J p_i = 1$. Sumines tikimybes pažymėkime $P(j) = p_1 + p_2 + \dots + p_j = \sum_{h=1}^j p_h$, kur $j = \overline{1, J-1}$. Tada turime

$$\ln \frac{P(j)}{1 - P(j)} = \ln \frac{p_1 + p_2 + \dots + p_j}{p_{j+1} + \dots + p_J}, \quad \text{kur } j = \overline{1, J-1}.$$

Modelis yra užrašomas taip [6, 9, 10]:

$$\text{logit}[P(j)] = \ln \frac{P(j)}{1 - P(j)} = \beta_{0j} + \beta x_1 + \beta x_2 + \dots + \beta x_k, \quad \text{kur } j = \overline{1, J-1}$$

ir suminės tikimybės išreiškiamos tokia lygtimi

$$P(j) = \frac{e^{\beta_{0j} + \beta x_1 + \beta x_2 + \dots + \beta x_k}}{1 + e^{\beta_{0j} + \beta x_1 + \beta x_2 + \dots + \beta x_k}}, \quad \text{kur } j = \overline{1, J-1}.$$

Tikimybė, kad priklausomas kintamasis įgis reikšmę y_j^* ($j = \overline{1, J}$), yra gaunama skaičiuojant skirtumą tarp suminių tikimybių:

$$p_j = P(j) - P(j-1), \quad \text{kur } j = \overline{2, J}.$$

2.2.5. KLASTERINĖ ANALIZĖ

2.6 lentelėje pateikiami skyrelyje naudojami žymėjimai.

2.6 lentelė

Klasterinėje analizėje naudojami žymėjimai

Žymėjimas	Prasmė
$G = \{G_1, G_2, \dots, G_q\}$	Grupuojantys kintamieji
n	Klasifikuojamų studijų programų skaičius
m	Požymių, pagal kuriuos yra klasifikuojamos studijų programos, skaičius
$PR = \{X_1, X_2, \dots, X_n\}$	Analizuojamų studijų programų aibė
$(x_{i1}, x_{i2}, \dots, x_{im}), i = 1, 2, 3, \dots, n$	Studijų programos požymių reikšmių vektorius

Panaudojant grupuojančius kintamuosius G yra formuojamas pradinių studijų proceso duomenų pjūvis. Pagal pjūvį suskaičiuojamos studijų programų požymių reikšmės ir gautai grupei taikomi klasterinės analizės metodai.

Klasterinės analizės modelyje studijų programos suskirstomos į klasterius pagal grupėse esančių studentų mokyklinius matematikos, fizikos, informatikos, lietuvių kalbos, anglų kalbos įvertinimų ir metinių pažymių vidurkių aritmetinius vidurkius. Gaunamos grupės, kurių viduje studijų programos pagal vieną ar kelis požymius yra labai panašios, o iš skirtingų grupių labai skiriasi. Studijų programų tarpusavio panašumas gali būti matuojamas naudojant vieną iš keleto skirtingų matų: Euklido, Euklido atstumo kvadrato, Minkovskio, Manheteno, Čebyšovo, Mahalanobio atstumo kvadrato. Šio uždavinio sprendimui buvo pasirinktas Euklido atstumo kvadrato matas, kadangi jis geriausiai tinka skirtumams

tarp studijų programų įvertinti dėl geometrinės prigimties. Euklido atstumo kvadrato mato išraiška yra tokia [9, 10]:

$$\|X_q - X_r\|^2 = \sum_{i=1}^m (x_{qi} - x_{ri})^2, \text{ kur } X_q \in PR \text{ ir } X_r \in PR.$$

Studijų programų klasteriams sudaryti modelyje yra naudojami hierarchinio jungimo klasterinės analizės metodai. Pateikiamas metodų bendrasis algoritmas:

- 1) pradžioje turime n klasterių po 1 studijų programą ir $n \times n$ simetrinę atstumų matricą;
- 2) pagal atstumų matricą nustatomi du klasteriai U ir V , tarp kurių atstumas yra mažiausias;
- 3) klasteriai U ir V sujungiami. Naujas klasteris pavadinamas (UV) ir atstumų matrica pakeičiama taip:
 - a) išbraukiami stulpeliai ir eilutės, atitinkančios klasterius U ir V ;
 - b) pridedama eilutė ir stulpelis su atstumais tarp (UV) ir likusių klasterių.
- 4) 2 ir 3 žingsniai kartojami $n-1$ kartų. Procesas baigiamas, kai visos studijų programos yra viename klasteryje.

Dažniausiai naudojami atstumo $d(U, V)$ tarp dviejų klasterių U ir V matai pateikti 2.7 lentelėje.

2.7 lentelė

Klasterių atstumo matai

Atstumas	Atstumo $d(U, V)$ formulė
Vienetinės jungties	$d(U, V) = \min_{X_q \in U, X_r \in V} d(X_q, X_r),$ <p>X_q – q-toji U studijų programa, X_r – r-toji V studijų programa</p>
Pilnosios jungties	$d(U, V) = \max_{X_q \in U, X_r \in V} d(X_q, X_r)$
Vidutinės jungties	$d(U, V) = \frac{\sum_{X_q \in U} \sum_{X_r \in V} d(X_q, X_r)}{n_U n_V},$ <p>n_U, n_V – atitinkamai klasterių U ir V studijų programų skaičius</p>
Centrų	$d(U, V) = d(\bar{U}, \bar{V}),$ <p>\bar{U}, \bar{V} – klasterius sudarančių studijų programų požymių vektorių vidurkiai</p>
Vordo	$d(U, V) = \ \bar{U} - \bar{V}\ ^2 / (1/n_U + 1/n_V)$

Modelyje klasifikavimo rezultatai iliustruojami klasterizavimo dendogramomis, kurios vaizdžiai parodo susidariusias homogenines studijų programų grupes, leidžiančios nustatyti labiausiai besiskiriančias studijų programas/programų grupes.

2.3. MODELIŲ PROGRAMINĖ REALIZACIJA

Studijų proceso duomenų analizės sistema buvo kuriama panaudojant SAS programų paketą ir jo makroprogramavimo priemones [26], internetinių sistemų kūrimo įrankio SAS AppDev Studio posistemę webAF [7, 8, 20] ir JAVA servlet, JSP, Javascript, HTML internetines technologijas [22-25]. Šios programinės priemonės buvo pasirinktos dėl turimų statistinės analizės metodų gausos, aukšto interaktyvumo lygio, mobilumo bei labai lankstaus bendravimo tarp programos ir vartotojo. Sistemos veikimo schema pateikta 2.4 pav. Ją sudaro 12 etapų:

- 1) sistemos vartotojas analizei paruošia pradinį studijų proceso duomenis: kintamiesiems yra priskiriamos sistemoje naudojamos programinės žymės, visi duomenys yra sujungiami į vieną bendrą bylą;
- 2) pradiniai duomenys yra paverčiami SAS duomenų rinkiniu ir pagal pasirinktus grupuojančius kintamuosius yra suformuojamas duomenų pjūvis;
- 3) vartotojas pasirenka statistinės analizės modelį ir, nurodydamas parametrus, pateikia užklausą sistemai;
- 4) internetinė naršyklė apdoroja užklausą;
- 5) užklausa yra siunčiama į nuotolinį JAVA Apache Tomcat serverį;
- 6) JAVA serveris, naudodamas SAS IOM protokolą, jungiasi prie SAS serverio;
- 7) SAS serveryje yra vykdomos statistinės analizės modelių SAS makrokomandos ir gaunami rezultatai;
- 8) analizės rezultatai naudojant SAS IOM protokolą perduodami JAVA serveriui;
- 9) JAVA serveryje suformuojamas rezultatų atsakas;
- 10) atsakas perduodamas internetinei naršyklei;
- 11) naršyklė rezultatus pateikia HTML arba RTF formatu;
- 12) vartotojas rezultatus interpretuoja ir daro išvadas.

Kiekvienos fazės metu vyksta sąveika tarp JSP internetinio puslapio, apjungiančio JavaScript ir HTML programinį kodą ir SAS sistemos, vykdančios pateiktas užklausas. Sąveika vykdoma naudojant Java Servlet priemones kaip tarpininką tarp JSP puslapio ir SAS sistemos.

2.5 pav. pateikta visų duomenų analizės sistemos bylų struktūra ir tarpusavio sąveika. JSP puslapiai žymi sistemos vartotojo sąsajos dalį, per kurią sistema yra valdoma, o SAS bylos žymi

statistinės analizės modelius realizuojančias makrokomandas. Bylų tarpusavio sąveika vykdoma ir sistema valdoma vartotojui pasirenkant mygtukus internetinės naršyklės lange. Sukurtų programinių priemonių tekstų apimtis yra 4000 eilučių. Pagrindinių sistemos bylų tekstai ir logistinės regresinės analizės modelio programiniai tekstai pateikti 2 priede. Kitų bylų tekstai įrašyti į kompaktinį diską (CD).

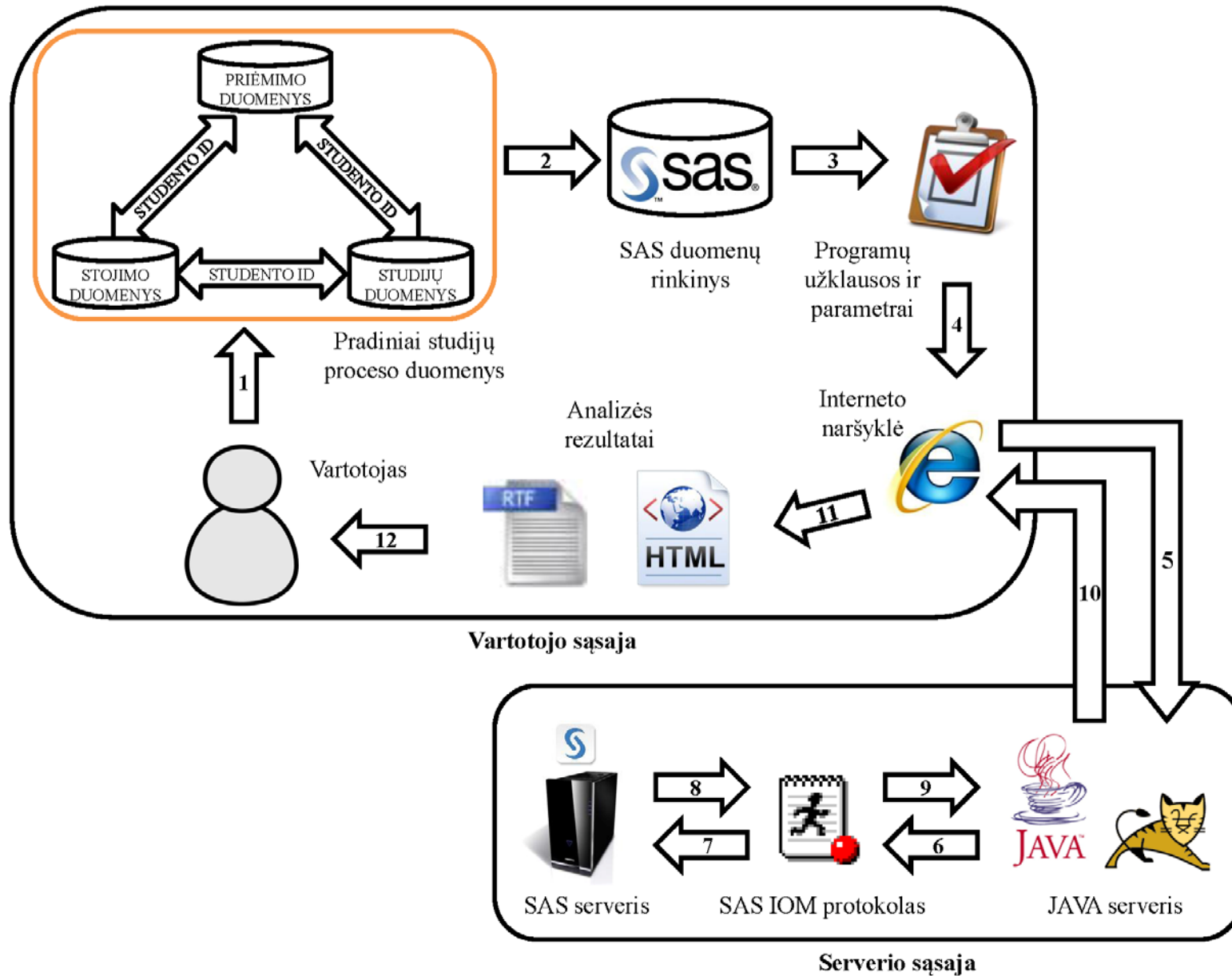
Vartotojas darbą su sistema pradeda iškviesdamas analizės sistemos duomenų įvesties puslapį (2.6 pav.). Yra pasirenkama pradinis duomenų byla, kuri, paspaudus mygtuką „Įkelti duomenis“, yra įkeliami į SAS sistemą ir paverčiama jos duomenų rinkiniu.



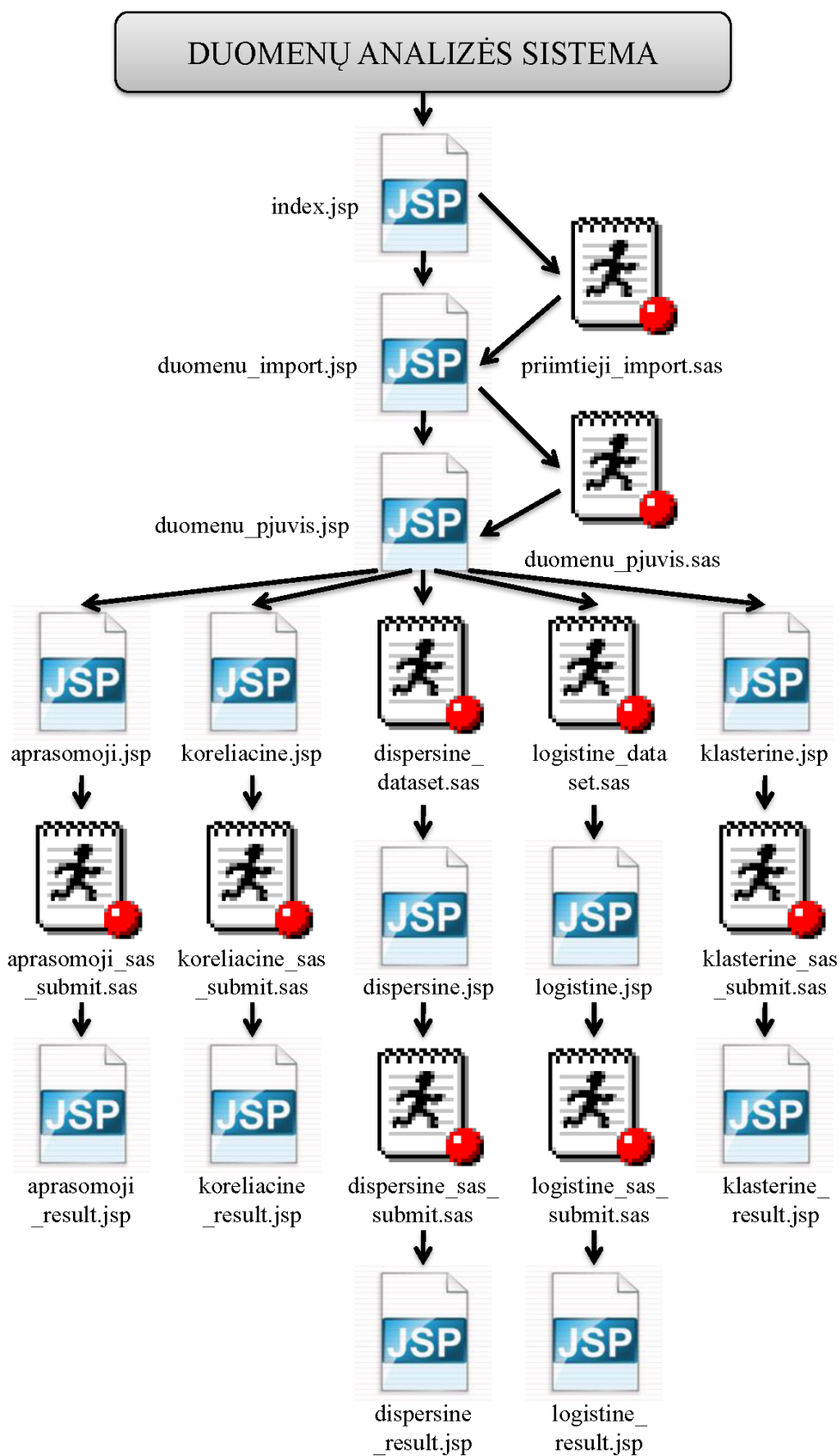
2.6 pav. Pradinis duomenų analizės sistemos langas

Kitame lange vartotojas turi pasirinkti grupuojančius kintamuosius ir suformuoti pradinis duomenų pjūvį (2.7 pav.). Pasirinkus atitinkamą mygtuką SAS sistema tai atlieka ir paruošia duomenis statistinei analizei.

2.7 pav. Duomenų pjūvio formavimas



2.4 pav. Studijų proceso duomenų analizės sistemos veikimo schema



2.5 pav. Duomenų analizės sistemos bylų struktūra ir tarpusavio sąveika

Trečiame lange vartotojui suteikiama galimybė suformuoti naują duomenų pjūvį arba dirbti su jau esančiu SAS sistemoje (2.8 pav.). Pastaruoju atveju jis turi pasirinkti vieną iš kelių statistinių modelių. Paspaudus mygtuką atveriamas modelio sąsaja.

Naudojant aprašomosios statistikos modelį pasirenkamas vienas iš kelių analizuojamų kintamųjų ir grafikų tipų (2.9 pav.).

2.8 pav. Statistinių modelių pasirinkimas

2.9 pav. Aprašomosios statistikos modelio vartotojo sąsaja

Naudojant koreliacinės analizės modelį pasirenkami keli studentų mokymosi rodikliai ir grafinių rezultatų tipas (2.10 pav.).

Naudojant dispersinės analizės modelius, sąsaja leidžia pasirinkti analizės tipą, analizuojamus priklausomus kintamuosius ir faktorius bei grafinių rezultatų tipą (2.11 pav.). Priklausomu kintamuoju

pasirinkus konkretaus studijų modulio įvertinimus, reikia tą modulį nurodyti.

Naudojant logistinės regresinės analizės modelį pasirenkamas analizės tipas, naudojama ryšio funkcija, vienas ar keli analizuojami kintamieji ir grafinių rezultatų tipas (2.12 pav.). Pasirinkus antrąją analizės tipą ir prognozuojant studijų modilio įvertinimo tikimybes, reikia tą modulį nurodyti.

Koreliacinės analizės modelis

Lentelės

Analizuojami kintamieji:

Stojimo pageidavimo numeris
 Konkursinis balas
 Mokyklos metinių pažymių vidurkis
 Stojimo pageidavimų skaičius
 Studijų sričių skaičius
 Studentų pažangumo indeksas

Grafikai

Rezultatai su grafikais
 Bendras visų kintamųjų analizės grafikas
 Be grafikų
 Detalūs visų kintamųjų porų analizės grafikai

Rezultatų išvedimas: HTML ir RTF

RTF failą saugoti į: C:\Koreliacine_analize.rtf

[Įkelti naujus duomenis](#)
[Formuoti naują duomenų pjūvį](#)

2.10 pav. Koreliacinės analizės modelio vartotojo sąsaja

Dispersinės analizės modeliai

Analizės tipas: Vienfaktorinė dispersinė analizė
 Dvifaktorinė dispersinė analizė

Lentelės

Pirmas faktorius: Studijų kryptis, Studijų forma, Studijų programa, Studijų sritis
Antras faktorius: Miesto grupė, Miestas/rajonas, Pageidavimo numerio grupė, Konkursinio balo grupė
Analizuojamas kintamasis: Mokyklos metinių pažymių vidurkis, Konkursinis balas, Studento pažangumo indeksas, Stojimo pageidavimų skaičius
Studijų modulis: Matematika

Grafikai

Rezultatai su grafiku vienfaktorinės analizės normaliuoju atveju
 Be grafikų

Rezultatų išvedimas: HTML ir RTF

RTF failą saugoti į: C:\Dispersine_analize.rtf

[Įkelti naujus duomenis](#)
[Formuoti naują duomenų pjūvį](#)

2.11 pav. Dispersinės analizės modelių vartotojo sąsaja

Naudojant klasterinės analizės modelį pasirenkamas klasterizavimo metodas, analizuojami kintamieji ir grafinių rezultatų tipas (2.13 pav.).

Visose modelių sąsajose pasirenkamas rezultatų išvedimo tipas: tik naršyklės lange (HTML) ar naršyklės lange ir ataskaitos forma atskiroje byloje (HTML ir RTF). Paspaudus mygtuką „Analizuoti“ SAS sistema vykdo sukurtas makrokomandas ir rezultatai pateikiami vartotojui. 3 priede pateiktas logistinės regresinės analizės modelio rezultatų pavyzdys. Kitų modelių rezultatų pavyzdžiai įrašyti į kompaktinį diską (CD).

Logistinės regresinės analizės modelis

Analizės tipas: **Studijų modulis:**

Prognozuojama tikimybė baigti pirmą kursą
 Prognozuojama tikimybė gauti puikų modulio įvertinimą
 Geometrija

Lentelės

Ryšio funkcija: **Analizuojami kintamieji:**

<input type="checkbox"/> Stojimo pageidavimo numeris <input type="checkbox"/> Fakultetas	<input checked="" type="checkbox"/> Konkursinis balas <input type="checkbox"/> Studijų forma	<input type="checkbox"/> Mokyklos metinių pažymių vidurkis <input checked="" type="checkbox"/> Studentų pažangumo indeksas	<input checked="" type="checkbox"/> Miesto grupė <input type="checkbox"/> Finansavimo tipas	<input type="checkbox"/> Miestas/rajonas <input type="checkbox"/> Lytis
---	---	---	--	--

Grafikai

Rezultatai su grafiku pirmu analizės atveju
 Be grafiko

Rezultatų išvedimas: HTML ir RTF
 RTF failą saugoti į: C:\Logistine_regresine_analize.rtf

Analizuoti
Atgal Baigti darbą

[Įkelti naujus duomenis](#)
[Formuoti naują duomenų pjūvį](#)

2.12 pav. Logistinės regresinės analizės modelio vartotojo sąsaja

Klasterinės analizės modelis

Lentelės

Klasterizavimo metodas: **Analizuojami kintamieji:**

Pilnosios jungties Vidutinės jungties Centrų Vordo	<input checked="" type="checkbox"/> Matematikos įvertinimų vidurkis <input type="checkbox"/> Fizikos ir astronomijos įvertinimų vidurkis	<input type="checkbox"/> Lietuvių kalbos įvertinimų vidurkis <input checked="" type="checkbox"/> Informatikos ir informacinių technologijų įvertinimų vidurkis	<input checked="" type="checkbox"/> Anglų kalbos įvertinimų vidurkis <input type="checkbox"/> Mokyklos metinių pažymių vidurkių vidurkis
---	---	---	---

Grafikai

Rezultatai su klasterizavimo dendograma
 Be dendogramos

Rezultatų išvedimas: HTML ir RTF
 RTF failą saugoti į: C:\Klasterine_analize.rtf

Analizuoti
Atgal Baigti darbą

[Įkelti naujus duomenis](#)
[Formuoti naują duomenų pjūvį](#)

2.13 pav. Klasterinės analizės modelio vartotojo sąsaja

2.4. MODELIŲ TAIKYMO PAVYZDŽIAI

Skyrelyje yra pateikiami Kauno technologijos universiteto studijų proceso duomenų analizės pavyzdžiai. Buvo naudojami 2009 ir 2010 metų KTU studijų proceso duomenys. Taikant aprašomosios statistikos, koreliacinės, dispersinės, logistinės regresinės ir klasterinės analizės modelius yra analizuojami šie studentų mokymosi rodiklius apibūdinantys kintamieji:

- studijų sritis, į kurios programą studentas įstojo;
- stojimo prašyme nurodytas pageidavimo numeris;
- pageidavimo grupė;
- stojimo pageidavimų skaičius
- mokyklos metinių pažymių vidurkis;
- stojimo konkursinis balas;
- studento pajėgumo indeksas;
- studijų programa;
- miestas, kuriame baigta vidurinė mokykla ar gimnazija
- mokyklos matematikos įvertinimų vidurkis;
- mokyklos fizikos ir astronomijos įvertinimų vidurkis;
- mokyklos informatikos ir informacinių technologijų įvertinimų vidurkis;
- mokyklos anglų kalbos įvertinimų vidurkis.

Studento pajėgumo indeksas buvo sudarytas darbo metu ir yra skaičiuojamas pagal tokį algoritmą:

- 1) jeigu studentas stojimo prašyme nurodė ne daugiau kaip 4 pageidavimus ir į KTU įstojo pirmuoju pageidavimu, indeksas skaičiuojamas pagal formulę

$$IND = 0,3 \cdot (1,5 \cdot (PAGSK - PAGNR) + 5) + 0,2 \cdot (16 - PAGSK) + 0,1 \cdot (6 - SRITSK) + \\ + 0,2 \cdot METVID + 0,2 \cdot EGZVID;$$

- 2) jeigu studentas stojimo prašyme nurodė daugiau kaip 4, bet mažiau nei 9 pageidavimus ir į KTU neįstojo pirmuoju pageidavimu, indeksas skaičiuojamas pagal formulę

$$IND = 0,3 \cdot (1,5 \cdot (PAGSK - PAGNR) + 1) + 0,2 \cdot (16 - PAGSK) + 0,1 \cdot (6 - SRITSK) + \\ + 0,2 \cdot METVID + 0,2 \cdot EGZVID;$$

- 3) jeigu studentas stojimo prašyme nurodė daugiau kaip 8 pageidavimus ir į KTU neįstojo pirmuoju pageidavimu, indeksas skaičiuojamas pagal formulę

$$IND = 0,3 \cdot (PAGSK - PAGNR + 1) + 0,2 \cdot (16 - PAGSK) + 0,1 \cdot (6 - SRITSK) + \\ + 0,2 \cdot METVID + 0,2 \cdot EGZVID,$$

kur **IND** – studento pajėgumo indeksas; **PAGSK** – stojimo prašyme nurodytų pageidavimų skaičius; **PAGNR** - pageidavimo numeris, kuriuo studentas įstojo į KTU; **SRITSK** - stojimo prašyme nurodytų studijų sričių skaičius; **METVID** – mokyklos metinių pažymių vidurkis; **EGZVID** – normuotas egzaminų įvertinimų vidurkis, įvertinant egzamino lygį.

Taikymuose hipotezių tikrinimui naudojamas reikšmingumo lygmuo $\alpha = 0,05$. Visų analizėje galimų naudoti kintamųjų matrica pateikiama 1 priede.

2.4.1. APRAŠOMOSIOS STATISTIKOS MODELIO TAIKYMAS

2.8 lentelėje pateiktas aprašomosios statistikos modelio taikymo pavyzdys. Analizuojami du atvejai, kai pradiniai duomenys buvo suformuoti skirtingais pjūviais: pagal studijų sritį ir metus. Lyginami studentai, įstoję į KTU fizinius ir technologijų mokslus 2009 ir 2010 metais. Abiem atvejais analizuojamas studento mokyklos metinių pažymių vidurkis.

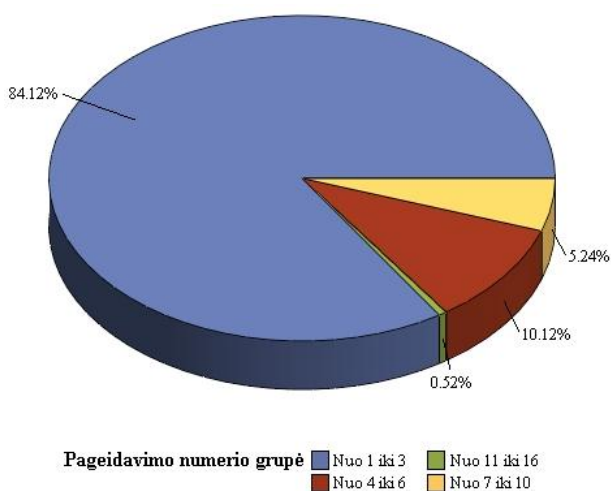
Gavome, kad pažangesni studentai įstojo į fizinių mokslų studijų programas (pažymių vidurkis $\bar{X}_1 = 8,47$ ir $\bar{X}_2 = 8,10$). Tai patvirtina mažiausios pažymių vidurkių reikšmės (fizinių mokslų atveju mažiausia reikšmė buvo 4,7, o technologijų atveju – 4,42) ir dispersijos ($\hat{\sigma}_1^2 = 0,96$ ir $\hat{\sigma}_2^2 = 1,06$).

2.14 ir 2.15 pav. atitinkamai kiekvienam atvejui pateiktos studentų stojimo pageidavimo grupės pasiskirstymo diagramos. Matome, kad didžiausia dalis studentų ir į fizinius, ir į technologijų mokslus įstojo pageidavimo numeriais nuo 1 iki 3 (84,12% ir 84,21%). Lyginant tarpusavyje matome, kad į šią grupę pateko didesnė technologijų mokslų studentų dalis.

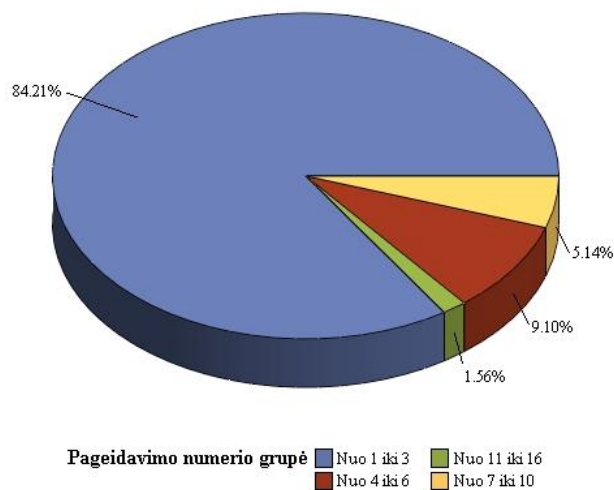
2.8 lentelė

Aprašomosios statistikos modelio taikymas

Aprašomosios statistikos modelio taikymo rezultatai			
		Analizuojamas kintamasis: mokyklos metinių pažymių vidurkis	
Tyrimo rezultatai	Tyrimo numeris	1	2
		Pjūvis Studijų sritis: Fiziniai mokslai; Metai: 2009-2010	Pjūvis Studijų sritis: Technologijos mokslai; Metai: 2009-2010
	Vidurkis \bar{X}	8,47	8,10
	Standartinis nuokrypis $\hat{\sigma}$	0,98	1,03
	Dispersija $\hat{\sigma}^2$	0,96	1,06
	Mažiausia reikšmė	4,7	4,42
	Didžiausia reikšmė	10	10
	Studentų skaičius n	573	2116



2.14 pav. Fizinių mokslų studentų pageidavimo grupių pasiskirstymas



2.15 pav. Technologijos mokslų studentų pageidavimo grupių pasiskirstymas

2.4.2. RYŠIŲ TARP STUDENTŲ MOKYMOŠI RODIKLIŲ TYRIMAS

Sudarysime koreliacinės analizės modelius su 4 kintamaisiais (2.9 lentelė) trims skirtingiems duomenų pjūviams: studijų sritis – fiziniai mokslai, priėmimo metai – 2009-2010; studijų sritis – technologijų mokslai, priėmimo metai – 2009-2010; studijų sritis – socialiniai mokslai, priėmimo metai – 2009-2010. Analizuojami kintamieji yra studento stojimo pageidavimų skaičius, konkursinis balas, mokyklos metinių pažymių vidurkis ir studento pajėgumo indeksas.

Nagrinėjant Spirmeno koreliacijos atvejį gavome, kad visais atvejais labai silpna priklausomybė monotoniškumo prasme yra tarp studentų nurodytų pageidavimų skaičiaus ir studentų pajėgumo indekso (visi $|\hat{\rho}_S| < 0,3$), o tarp mokyklos metinių pažymių vidurkio ir pajėgumo indekso yra stebima vidutinio stiprumo priklausomybė (visi $0,45 < \hat{\rho}_S < 0,7$).

Pirmuoju atveju tarp stojimo pageidavimų skaičiaus ir mokyklos metinių pažymių vidurkio statistiškai reikšmingos priklausomybės nėra ir hipotezė apie koreliacijos koeficiento lygybę 0 priimama ($p_1 = 0,1775$). Tarp tų pačių kintamųjų antruoju ir trečiuoju atvejais yra stebima labai silpna priklausomybė monotoniškumo prasme (abi $|\hat{\rho}_S| < 0,3$).

Galime teigti, kad fizinių, technologijų ir socialinių mokslų atvejais tarp studentų pajėgumo sėkmingai studijuoti ir studijų pasirinkimo yra silpnas ryšys, o priklausomybė tarp studentų pajėgumo ir pažangumo yra vidutinio stiprumo. Tarp studijų pasirinkimo ir pažangumo statistinis ryšys yra tik socialinių ir technologijų mokslų atvejais.

Nagrinėjant Kendalo koreliacijos atvejį gavome, kad visais atvejais tarp studentų nurodytų pageidavimų skaičiaus ir pajėgumo indekso yra labai silpna priklausomybė suderinamumo prasme

2.9 lentelė

Ryšių tarp KTU studentų mokymosi rodiklių tyrimas

Koreliacinės analizės rezultatai															
Analizuojami kintamieji:										x ₁ : stojimo pageidavimų skaičius					
										x ₂ : konkursinis balas					
										x ₃ : mokyklos metinių pažymių vidurkis					
										x ₄ : studento pajėgumo indeksas					
Tyrimo numeris	1					2					3				
	Pjūvis Studijų sritis: Fiziniai mokslai; Metai: 2009-2010					Pjūvis Studijų sritis: Technologijų mokslai; Metai: 2009-2010					Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009-2010				
Tyrimo rezultatai															
		x ₁	x ₂	x ₃	x ₄		x ₁	x ₂	x ₃	x ₄		x ₁	x ₂	x ₃	x ₄
Spirmeno koreliacijos koeficientai $\hat{\rho}_5$ ir p-reikšmės	x ₁	1,00	-0,17 0,0001	-0,06 0,1775	-0,20 0,0001	x ₁	1,00	0,01 0,5965	0,05 0,0300	-0,06 0,0037	x ₁	1,00	0,18 0,0001	0,18 0,0001	-0,15 0,0001
	x ₂	-0,17 0,0001	1,00	0,63 0,0001	0,64 0,0001	x ₂	0,01 0,5965	1,00	0,65 0,0001	0,64 0,0001	x ₂	0,18 0,0001	1,00	0,77 0,0001	0,46 0,0001
	x ₃	-0,06 0,1775	0,63 0,0001	1,00	0,51 0,0001	x ₃	0,05 0,0300	0,65 0,0001	1,00	0,53 0,0001	x ₃	0,18 0,0001	0,77 0,0001	1,00	0,49 0,0001
	x ₄	-0,20 0,0001	0,64 0,0001	0,51 0,0001	1,00	x ₄	-0,06 0,0037	0,64 0,0001	0,53 0,0001	1,00	x ₄	-0,15 0,0001	0,46 0,0001	0,49 0,0001	1,00
Kendalo $\hat{\tau}_b$ koreliacijos koeficientai ir p-reikšmės	x ₁	1,00	-0,12 0,0001	-0,04 0,1846	-0,14 0,0001	x ₁	1,00	0,01 0,7364	0,03 0,0300	-0,04 0,0037	x ₁	1,00	0,12 0,0001	0,12 0,0001	-0,11 0,0001
	x ₂	-0,12 0,0001	1,00	0,46 0,0001	0,46 0,0001	x ₂	0,01 0,7364	1,00	0,47 0,0001	0,46 0,0001	x ₂	0,12 0,0001	1,00	0,58 0,0001	0,32 0,0001
	x ₃	-0,04 0,1846	0,46 0,0001	1,00	0,36 0,0001	x ₃	0,03 0,0300	0,47 0,0001	1,00	0,38 0,0001	x ₃	0,12 0,0001	0,58 0,0001	1,00	0,34 0,0001
	x ₄	-0,14 0,0001	0,46 0,0001	0,36 0,0001	1,00	x ₄	-0,04 0,0037	0,46 0,0001	0,38 0,0001	1,00	x ₄	-0,11 0,0001	0,32 0,0001	0,34 0,0001	1,00
Stebėjimų skaičius	573					2116					1511				

(visi $|\hat{t}_b| < 0,3$), o tarp mokyklos metinių pažymių vidurkio ir pajėgumo indekso yra stebima silpna priklausomybė (visi $|\hat{t}_b| < 0,5$).

Pirmuoju atveju tarp stojimo pageidavimų skaičiaus ir mokyklos metinių pažymių vidurkio statistiškai reikšmingos priklausomybės nėra ir hipotezė apie koreliacijos koeficiento lygybę 0 priimama ($p_1 = 0,1846$), tuo tarpu tarp tų pačių kintamųjų antruoju ir trečiuoju atvejais yra stebima labai silpna priklausomybė suderinamumo prasme (abi $|\hat{t}_b| < 0,3$).

Apibendrinus abiejų koreliacijų rezultatus, galime daryti išvadą, kad ir monotoniškumo, ir suderinamumo prasme priklausomybė tarp studijų pasirinkimo ir pajėgumo bei tarp studijų pasirinkimo ir pažangumo yra labai silpna, o tarp studentų pažangumo ir pajėgumo stipresnis ryšys yra stebimas monotoniškumo pasme.

2.4.3. STOJANČIŪJŲ METINIŲ PAŽYMIŲ VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ SRITIES IR PAGEIDAVIMO GRUPĖS TYRIMAS

2.10 lentelėje pateikti konkursinio balo vidurkio priklausomybės nuo studijų srities ir pageidavimo grupės tyrimo rezultatai. Analizuojami du skirtingi pjūviai: miestas – Kaunas, priėmimo metai – 2009-2010; miestas – Panevėžys, priėmimo metai – 2009-2010.

Gavome, kad studentų, baigusių Kauno vidurines mokyklas, metinių pažymių vidurkis priklauso tik nuo studijų srities ($p_1 \leq 0,05$). Statistiškai reikšmingos faktorių tarpusavio sąveikos nepastebėta ($p_1 > 0,05$). Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime teigti, kad metinių pažymių vidurkio priklausomybė nuo studijų srities nėra stipri, nes kai kurie pasikliautinieji intervalai persidengia.

Panevėžiečių studentų atveju gavome, kad konkursinio balo vidurkis taip pat priklauso tik nuo studijų srities ($p_2 \leq 0,05$), o statistiškai reikšmingos faktorių tarpusavio sąveikos nėra ($p_2 > 0,05$). Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime spręsti, kad metinių pažymių vidurkio priklausomybė nuo studijų srities taip pat nėra stipri ir yra silpnesnė nei kauniečių atveju.

Lyginant abu atvejus matome, kad kauniečių atveju metinių pažymių vidurkiai yra didesni, dispersijos mažesnės ir pasikliautinieji intervalai siauresni. Galime daryti išvadą, kad abiem atvejais studijų sritis metinių pažymių vidurkiui nėra stiprus faktorius, o Kauno vidurines mokyklas ar gimnazijas baigę studentai yra pažangesni.

2.4.4. STOJANČIŲJŲ KONKURSINIO BALO VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ PROGRAMOS IR PAGEIDAVIMO GRUPĖS TYRIMAS

2.11 lentelėje pateikti konkursinio balo vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimo rezultatai. Analizuojami du skirtingi pjūviai: studijų sritis – socialiniai mokslai, miestas – Kaunas, priėmimo metai – 2009-2010; studijų sritis – socialiniai mokslai, miestas – visi, išskyrus Kauną, priėmimo metai – 2009-2010.

Gavome, kad studentų, įstojusių į socialinius mokslus ir baigusių Kauno vidurines mokyklas, konkursinio balo vidurkis priklauso tik nuo studijų programos ($p_1 \leq 0,05$). Statistiškai reikšmingos faktorių tarpusavio sąveikos nepastebėta ($p_1 > 0,05$). Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime teigti, kad konkursinio balo vidurkis nuo studijų programos priklauso silpnai, nes kai kurie pasikliautinieji intervalai persidengia.

Ne Kauno mokyklas baigusių studentų atveju gavome, kad konkursinio balo vidurkis taip pat priklauso tik nuo studijų programos ($p_2 \leq 0,05$), o statistiškai reikšmingos faktorių tarpusavio sąveikos nėra ($p_2 > 0,05$). Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime spręsti, kad konkursinio balo vidurkio priklausomybė nuo studijų programos nėra stipri, tačiau stipresnė nei kauniečių atveju.

Lyginant abu atvejus matome, kad ir pirmuoju, ir antruoju atveju konkursinio balo vidurkiai yra panašūs, tačiau antruoju atveju vidurkių dispersijos mažesnės ir pasikliautinieji intervalai siauresni. Galime daryti išvadą, kad abiem atvejais studento pasirinkta studijų programa konkursinio balo vidurkiui nėra stiprus faktorius, tačiau antruoju atveju priklausomybė yra stipresnė – ne Kauno mokyklas baigusių studentų konkursinio balo vidurkiui studijų programa turi didesnę įtaką.

2.4.5. STUDENTŲ PAJĖGUMO INDEKSO VIDURKIO PRIKLAUSOMYBĖS NUO STUDIJŲ PROGRAMOS IR PAGEIDAVIMO GRUPĖS TYRIMAS

2.12 lentelėje pateikti studentų pajėgumo indekso vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimo rezultatai. Analizuojami du skirtingi pjūviai: studijų sritis – fiziniai mokslai, priėmimo metai – 2009-2010; studijų sritis – humanitariniai mokslai, priėmimo metai – 2009-2010.

Gavome, kad studentų, įstojusių į fizinius mokslus, pajėgumo indekso vidurkis priklauso ir nuo studijų programos, ir nuo pageidavimo grupės ($p \leq 0,05$). Statistiškai reikšmingos faktorių tarpusavio sąveikos nepastebėta ($p_1 > 0,05$). Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime teigti, kad nuo pageidavimo grupės indekso vidurkis priklauso žymiai stipriau.

2.10 lentelė

Stojančiųjų metinių pažymių vidurkio priklausomybės nuo studijų srities ir pageidavimo grupės tyrimas

Dvifaktorinės dispersinės analizės rezultatai							
Pirmas faktorius: Studijų sritis			Antras faktorius: Pageidavimo grupė			Priklausomas kintamasis: Metinių pažymių vidurkis	
Tyrimo numeris Tyrimo rezultatai		1			2		
		Pjūvis Miestas: Kaunas; Metai: 2009-2010			Pjūvis Miestas: Panevėžys; Metai: 2009-2010		
H ₀ : m _{1**} = m _{2**} = ... = m _{u**} , p-reikšmė		0,0001			0,0121		
H ₀ : m _{*1*} = m _{*2*} = ... = m _{*w*} , p-reikšmė		0,1379			0,5027		
H ₀ : Faktorių tarpusavio sąveikos nėra		0,9490			0,2633		
Naudojamas kriterijus		Fišerio			Fišerio		
R ² koeficientas		0,032			0,094		
Normalumo hipotezės tikrinimo p-reikšmė		0,2776			0,7952		
Pirmo faktoriaus lygmenų skaičius		5			4		
Antro faktoriaus lygmenų skaičius		4			4		
Stebėjimų skaičius		1281			245		
Pirmas faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautiniai intervalai	Fiziniai mokslai	Menas	Technologijos mokslai	Fiziniai mokslai	Technologijos mokslai	Socialiniai mokslai
		8,59	8,41	8,24	8,27	7,44	7,19
		(8,39; 8,79)	(8,08; 8,74)	(8,12; 8,35)	(7,52; 9,02)	(7,16; 7,71)	(6,90; 7,34)
		Socialiniai mokslai	Humanitariniai mokslai		Humanitariniai mokslai		
		8,09	8,08		7,96		
	(7,96; 8,22)	(7,65; 8,51)		(6,12; 9,80)			
	Tjukio daugkartinio lyginimo rezultatai	Fiziniai mokslai – Technologijos mokslai			Fiziniai mokslai - Technologijos mokslai		
Fiziniai mokslai – Socialiniai mokslai			Fiziniai mokslai - Socialiniai mokslai				
Fiziniai mokslai – Humanitariniai mokslai							
Antras faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautiniai intervalai	Nuo 11 iki 16	Nuo 1 iki 3	Nuo 7 iki 10	Nuo 11 iki 16	Nuo 1 iki 3	Nuo 7 iki 10
		8,42	8,29	8,09	7,56	7,32	6,81
		(7,82; 9,03)	(8,20; 8,37)	(7,85; 8,33)	(6,40; 8,72)	(7,13; 7,50)	(6,05; 7,56)
		Nuo 4 iki 6			Nuo 4 iki 6		
	8,06			7,35			
(7,86; 8,25)			(6,85; 7,84)				
Tjukio daugkartinio lyginimo rezultatai	Statistiškai reikšmingai besiskiriančių porų nėra			Statistiškai reikšmingai besiskiriančių porų nėra			

2.11 lentelė

Stojančiųjų konkursinio balo vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimas

Dvifaktorinės dispersinės analizės rezultatai							
Pirmas faktorius: Studijų programa		Antras faktorius: Pageidavimo grupė			Priklausomas kintamasis: Konkursinis balas		
Tyrimo numeris		1			2		
		Pjūvis Studijų sritis: Socialiniai mokslai; Miestas: Kaunas; Metai: 2009-2010			Pjūvis Studijų sritis: Socialiniai mokslai; Miestas: Visi, išskyrus Kauną; Metai: 2009-2010		
Tyrimo rezultatai							
$H_0: m_{1**} = m_{2**} = \dots = m_{u**}$, p-reikšmė		0,0088			0,0001		
$H_0: m_{*1*} = m_{*2*} = \dots = m_{*w*}$, p-reikšmė		0,7270			0,8640		
H_0 : Faktorių tarpusavio sąveikos nėra		0,9487			0,0534		
Naudojamas kriterijus		Fišerio			Fišerio		
R^2 koeficientas		0,186			0,276		
Normalumo hipotezės tikrinimo p-reikšmė		0,1819			0,0873		
Pirmo faktoriaus lygmenų skaičius		10			12		
Antro faktoriaus lygmenų skaičius		4			4		
Stebėjimų skaičius		453			1068		
Pirmas faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautiniai intervalai	Viešojo sektoriaus ekonomika	Finansai	Lyderystė	Viešojo sektoriaus ekonomika	Finansai	Lyderystė
		15,75 (7,57; 23,93)	15,65 (12,84; 18,45)	15,65 (11,28; 20,02)	13,82 (11,05; 16,6)	15,82 (14,44; 17,21)	13,25 (10,24; 16,27)
		Socialinė pedagogika	Ekonomika	Viešasis administravimas	Socialinė pedagogika	Ekonomika	Viešasis administravimas
		15,11 (11,02; 19,20)	14,63 (13,73; 15,52)	11,99 (10,68; 13,30)	13,35 (11,79; 14,91)	15,45 (14,88; 16,02)	13,33 (12,57; 14,09)
		Verslo administravimas	Vadyba	Socialinė edukologija	Verslo administravimas	Vadyba	Socialinė edukologija
		11,73 (10,27; 13,20)	11,44 (10,13; 12,75)	11,07 (7,58; 14,58)	12,8 (11,74; 13,86)	12,01 (11,05; 12,98)	12,26 (10,3; 14,22)
		Sociologija			Sociologija	Verslo administravimas (Panevėžyje)	Vadyba (Panevėžyje)
		9,35 (6,94; 11,76)			10,52 (8,98; 12,07)	10,48 (9,6; 11,36)	10,03 (8,85; 11,2)

2.11 Ietelės tėsiny

	Tjukio daugkartinio lyginimo rezultatai	Finansai - verslo administravimas			Finansai - Socialinė pedagogika		
		Finansai - viešasis administravimas			Finansai - Viešasis administravimas		
		Finansai - vadyba			Finansai - Verslo administravimas		
		Finansai - sociologija			Finansai - Socialinė edukologija		
		Lyderystė - sociologija			Finansai - Vadyba		
		Socialinė pedagogika - sociologija			Finansai - Sociologija		
		Ekonomika - viešasis administravimas			Finansai - Verslo administravimas (Panevėžyje)		
		Ekonomika - verslo administravimas			Finansai - Vadyba (Panevėžyje)		
		Ekonomika - vadyba			Ekonomika - Socialinė pedagogika		
		Ekonomika - sociologija			Ekonomika - Viešasis administravimas		
					Ekonomika - Verslo administravimas		
					Ekonomika - Socialinė edukologija		
					Ekonomika - Vadyba		
					Ekonomika - Sociologija		
					Ekonomika - Verslo administravimas (Panevėžyje)		
					Ekonomika - Vadyba (Panevėžyje)		
					Viešojo sektoriaus ekonomika - Verslo administravimas (Panevėžyje)		
					Viešojo sektoriaus ekonomika - Vadyba (Panevėžyje)		
					Socialinė pedagogika - Sociologija		
					Socialinė pedagogika - Verslo administravimas (Panevėžyje)		
					Socialinė pedagogika - Vadyba (Panevėžyje)		
					Viešasis administravimas - Sociologija		
					Viešasis administravimas - Verslo administravimas (Panevėžyje)		
					Viešasis administravimas - Vadyba (Panevėžyje)		
					Verslo administravimas - Sociologija		
					Verslo administravimas - Verslo administravimas (Panevėžyje)		
					Verslo administravimas - Vadyba (Panevėžyje)		
			Vadyba - Verslo administravimas (Panevėžyje)				
			Vadyba - Vadyba (Panevėžyje)				
Antras faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautinieji intervalai	Nuo 1 iki 3	Nuo 11 iki 16	Nuo 4 iki 6	Nuo 1 iki 3	Nuo 11 iki 16	Nuo 4 iki 6
		13,25	13,19	12,41	13,32	12,85	13,04
		(12,64; 13,87)	(10,10; 16,30)	(11,36; 13,50)	(12,98; 13,66)	(11,23; 14,47)	(12,45; 13,64)
		Nuo 7 iki 10			Nuo 7 iki 10		
		12,35			12,91		
			(12,2; 13,62)				
	Tjukio daugkartinio lyginimo rezultatai	Statistiškai reikšmingai besiskiriančių porų nėra			Statistiškai reikšmingai besiskiriančių porų nėra		

2.12 lentelė

Studentų pajėgumo indekso vidurkio priklausomybės nuo studijų programos ir pageidavimo grupės tyrimas

Dvifaktorinės dispersinės analizės rezultatai

Pirmas faktorius: Studijų programa		Antras faktorius: Pageidavimo grupė			Priklausomas kintamasis: Studento pajėgumo indeksas		
Tyrimo rezultatai		1			2		
Tyrimo numeris		Pjūvis Studijų sritis: Fiziniai mokslai; Metai: 2009-2010			Pjūvis Studijų sritis: Humanitariniai mokslai; Metai: 2009-2010		
$H_0: m_{1**} = m_{2**} = \dots = m_{0**}$, p-reikšmė		0,0001			0,6292		
$H_0: m_{*1*} = m_{*2*} = \dots = m_{*w*}$, p-reikšmė		0,0001			0,0001		
H_0 : Faktorių tarpusavio sąveikos nėra		0,2803			0,4661		
Naudojamas kriterijus		Fišerio			Fišerio		
R^2 koeficientas		0,505			0,646		
Normalumo hipotezės tikrinimo p-reikšmė		0,3719			0,1284		
Pirmo faktoriaus lygmenų skaičius		5			5		
Antro faktoriaus lygmenų skaičius		4			4		
Stebėjimų skaičius		573			117		
Pirmas faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautiniai intervalai	Informacinės sistemos	Taikomoji fizika	Informatika	Technikos kalbos vertimas ir redagavimas (vokiečių kalba)	Technikos kalbos vertimas ir redagavimas (anglų kalba)	Kompiuterinė lingvistika
		8,13	7,95	7,88	7,78	7,20	7,00
		(7,71; 8,55)	(7,67; 8,23)	(7,77; 7,99)	(5,90; 9,67)	(6,98; 7,41)	(6,65; 7,35)
		Taikomoji matematika	Taikomoji chemija		Technikos kalbos vertimas ir redagavimas (rusų kalba)	Technikos kalbos vertimas ir redagavimas (prancūzų kalba)	
		7,77	7,40		6,71	6,65	
		(7,62; 7,90)	(7,20; 7,60)		(5,94; 7,48)	(5,56; 7,74)	
Tjukio daugkartinio lyginimo rezultatai	Taikomoji chemija – taikomoji fizika			Statistiškai besiskiriančių porų nėra			
	Taikomoji chemija – informacinės sistemos						
Antras faktorius	Grupių vidurkių taškiniai įverčiai ir pasikliautiniai intervalai	Nuo 1 iki 3	Nuo 4 iki 6	Nuo 7 iki 10	Nuo 1 iki 3	Nuo 4 iki 6	Nuo 7 iki 10
		8,07	6,69	5,78	7,77	6,44	5,52
		(7,99; 8,16)	(6,46; 6,93)	(5,45; 6,10)	(7,55; 7,98)	(6,09; 6,77)	(5,03; 6,00)
		Nuo 11 iki 16			Nuo 11 iki 16		
		3,80			4,57		
	(2,77; 4,84)			(3,28; 5,86)			
	Tjukio daugkartinio lyginimo rezultatai	Nuo 1 iki 3 - nuo 4 iki 6		Nuo 1 iki 3 - nuo 7 iki 10		Nuo 1 iki 3 - nuo 4 iki 6	
		Nuo 1 iki 3 - nuo 7 iki 10		Nuo 1 iki 3 - nuo 11 iki 16		Nuo 1 iki 3 - nuo 7 iki 10	
		Nuo 1 iki 3 - nuo 11 iki 16		Nuo 4 iki 6 - nuo 7 iki 10		Nuo 1 iki 3 - nuo 11 iki 16	
		Nuo 4 iki 6 - nuo 7 iki 10		Nuo 4 iki 6 - nuo 11 iki 16		Nuo 4 iki 6 - nuo 7 iki 10	
Nuo 4 iki 6 - nuo 11 iki 16		Nuo 7 iki 10 - nuo 11 iki 16		Nuo 4 iki 6 - nuo 11 iki 16			

Humanitarinių mokslų atveju gavome, kad pajėgumo indekso vidurkis priklauso tik nuo pageidavimo grupės ($p_2 = 0,0001$), o statistiškai reikšmingos faktorių tarpusavio sąveikos taip pat nepastebėta. Pagal daugkartinio lyginimo rezultatus ir vidurkių pasikliautuosius intervalus galime spręsti, kad indekso vidurkio priklausomybė nuo pageidavimo grupės yra stipri.

Lyginant abu atvejus matome, kad fizinių mokslų atveju pajėgumo indekso vidurkiai didesni, dispersijos mažesnės, o pasikliautinieji intervalai siauresni. Humanitarinių mokslų atveju vidurkiai mažesni, dispersijos didesnės, o pasikliautinieji intervalai platesni. Galime daryti išvadą, kad abiem atvejais pajėgumo indekso vidurkis labai stipriai priklauso nuo studento noro ir motyvacijos, o fizinius mokslus pasirinkę studentai yra pajėgesni toliau tęsti studijas ir sėkmingai jas baigti.

2.4.6. PIRMO KURSO BAIGIMO TIKIMYBĖS PROGNOZAVIMAS

Sudarysime logistinės regresinės analizės modelius su 5 nepriklausomais kintamaisiais (2.13 lentelė) trim skirtingiems duomenų pjūviams: studijų sritis – fiziniai mokslai, priėmimo metai – 2009; studijų sritis – technologijų mokslai, priėmimo metai – 2009; studijų sritis – socialiniai mokslai, priėmimo metai – 2009. Visais atvejais priklausomas kintamasis yra dvireikšmis pirmo kurso baigimo požymis (baigė arba nebaigė). Modeliuose naudojama logit ryšio funkcija ir prognozuojama tikimybė baigti pirmą kursą. Nepriklausomi kintamieji yra konkursinis balas, mokyklos metinių pažymių vidurkis, miesto grupė ir studento pajėgumo indeksas.

Gavome, kad visais atvejais modelis yra suderintas su duomenimis (visos $p > 0,05$). Bendras teisingai klasifikuotų studentų procentas visur yra aukštas (90,7%, 87,1% ir 87,5%), tačiau tiksliausiai prognozuojama fizinius mokslus studijuojančių studentų tikimybė baigti pirmą kursą.

Visais atvejais hipotezė apie visų koeficientų prie nepriklausomų kintamųjų lygybę 0 atmetama (visos $p \leq 0,05$). Hipotezės apie kiekvieno atskiro koeficiento lygybę 0 pirmu atveju atmetos tik konkursiniam balui ($p_1 \leq 0,05$). Visi kiti kintamieji yra statistiškai nereikšmingi (visos $p > 0,05$). Antru atveju hipotezės atmetos konkursiniam balui ir metinių pažymių vidurkiui (abi $p \leq 0,05$), o trečiuoju – konkursiniam balui, metinių pažymių vidurkiui ir studentų pajėgumo indeksui.

Modelyje naudojant tik statistiškai reikšmingus nepriklausomus kintamuosius (2.14 lentelė) gavome, kad modelis yra suderintas su duomenimis tik antru ir trečiu atvejais (abi $p > 0,05$). Pirmu atveju hipotezė apie duomenų ir modelio suderinamumą atmetama ($p_1 \leq 0,05$). Bendras teisingai klasifikuotų studentų procentas visais atvejais yra aukštas (91,1%, 87,4% ir 87,5%), tačiau tiksliausiai prognozuojama fizinius mokslus studijuojančių studentų tikimybė baigti pirmą kursą.

Lyginant 2.13 ir 2.14 lenteles matome, kad, iš modelių pašalinus nereikšmingus kintamuosius, jų tikslumas, nors ir nežymiai, bet padidėjo. Tačiau pirmu atveju modelis tapo nesuderintas su duomenimis ir jo prognozavimui taikyti nebegalima. Iš likusių dviejų atvejų tiksliau tikimybė baigti

2.13 lentelė

Pirmo kurso baigimo tikimybės prognozavimas

Logistinės regresinės analizės rezultatai							
Priklausomas kintamasis:		Ryšio funkcija:		Nepriklausomi kintamieji			
pirmo kurso baigimas		logit				x_1 : Konkursinis balas	
						x_2 : Mokyklos metinių pažymių vidurkis	
						x_{3a} : Miesto grupė (Apskričių centrai)	
						x_{3b} : Miesto grupė (Kiti miestai)	
						x_4 : Studento pajėgumo indeksas	
Tyrimo rezultatai	Tyrimo numeris	1		2		3	
		Pjūvis Studijų sritis: Fiziniai mokslai; Metai: 2009		Pjūvis Studijų sritis: Technologijų mokslai; Metai: 2009		Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009	
H_0 : modelis gerai suderintas su duomenimis		0,2531		0,4362		0,5545	
Bendras teisingai klasifikuotų studentų procentas		90,7		87,1		87,5	
Teisingai klasifikuotų pirmą kursą baigusiu studentų procentas		98		98,2		99,4	
Teisingai klasifikuotų pirmo kurso nebaigusiu studentų procentas		27,6		14,2		1,8	
$H_0: \beta_1 = \beta_2 = \beta_{3a} = \beta_{3b} = \beta_4 = 0$		0,0001		0,0001		0,0001	
$H_0: \beta_1 = 0$		0,0001		0,0001		0,0001	
$H_0: \beta_2 = 0$		0,0584		0,0013		0,0006	
$H_0: \beta_{3a} = 0$		0,1375		0,3022		0,7493	
$H_0: \beta_{3b} = 0$		0,9347		0,8457		0,1717	
$H_0: \beta_4 = 0$		0,1251		0,3971		0,0361	
Kokso-Snelo determinacijos koeficientas r^2_{CS}		0,1962		0,1281		0,0921	
Stebėti dažniai	Baigė pirmą kursą	252		1019		805	
	Nebaigė pirmo kurso	29		155		112	
Parametras β_0		-4,5281		-4,6890		-1,6235	
Parametras $\beta_i, i = 1, 2, 3, \dots$		β_1	β_2	β_1	β_2	β_1	β_2
Taškinis įvertis		0,4972	0,6136	0,2542	0,3658	0,1568	0,4634
Galimybių santykis		1,644	1,847	1,289	1,442	1,170	1,589
Pasikliautinis intervalas		(1,30; 2,07)	(0,98; 3,49)	(1,19; 1,40)	(1,15; 1,80)	(1,09; 1,26)	(1,22; 2,07)
		β_{3a}	β_{3b}	β_{3a}	β_{3b}	β_{3a}	β_{3b}
		-0,5308	-0,0284	0,1347	-0,0261	-0,0467	0,2332
		0,588	0,556	1,144	0,974	0,954	1,263
		(0,08; 1,45)	(0,13; 2,33)	(0,76; 2,13)	(0,64; 1,84)	(0,64; 2,07)	(0,79; 2,94)
		β_4		β_4		β_4	
		-0,5940		0,0894		-0,2418	
		0,552		1,093		0,785	
		(0,26; 1,18)		(0,89; 1,35)		(0,63; 0,99)	

2.14 lentelė

Pirmo kurso baigimo tikimybės prognozavimas (be nereikšmingų kintamųjų)

Logistinės regresinės analizės rezultatai					
Priklausomas kintamasis:		Ryšio funkcija:		Nepriklausomi kintamieji	
pirmo kurso baigimas		logit		x_1 : Konkursinis balas x_2 : Mokyklos metinių pažymių vidurkis x_{3a} : Miesto grupė (Apskričių centrai) x_{3b} : Miesto grupė (Kiti miestai) x_4 : Studento pajėgumo indeksas	
Tyrimo rezultatai	Tyrimo numeris	1	2	3	
		Pjūvis Studijų sritis: Fiziniai mokslai; Metai: 2009	Pjūvis Studijų sritis: Technologijos mokslai; Metai: 2009	Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009	
H_0 : modelis gerai suderintas su duomenimis		0,0057	0,9343	0,4714	
Bendras teisingai klasifikuotų studentų procentas		91,1	87,4	87,5	
Teisingai klasifikuotų pirmą kursą baigusiu studentų procentas		98	98,3	99,5	
Teisingai klasifikuotų pirmo kurso nebaigusiu studentų procentas		31	15,5	0,9	
$H_0: \beta_1 = \beta_2 = \beta_{3a} = \beta_{3b} = \beta_4 = 0$		0,0001	0,0001	0,0001	
$H_0: \beta_1 = 0$		0,0001	0,0001	0,0001	
$H_0: \beta_2 = 0$		-	0,0007	0,0011	
$H_0: \beta_{3a} = 0$		-	-	-	
$H_0: \beta_{3b} = 0$		-	-	-	
$H_0: \beta_4 = 0$		-	-	0,0433	
Kokso-Snelo determinacijos koeficientas r^2_{CS}		0,1771	0,1268	0,0902	
Stebėti dažniai	Baigė pirmą kursą	252	1019	805	
	Nebaigė pirmo kurso	29	155	112	
Parametras β_0		-4,9067	-4,2157	-1,7805	
Parametras $\beta_i, i = 1, 2, 3, \dots$		β_1	β_1 β_2	β_1 β_2	β_1 β_2
Taškinis įvertis		0,5380	0,2621 0,3797	0,1505 0,4844	
Galimybių santykis		1,713	1,30 1,462	1,162 1,163	
Pasikliautinis intervalas		(1,44; 2,04)	(1,20; 1,40) (1,17; 1,80)	(1,08; 1,25) (1,25; 2,11)	
				β_4	
				-0,2299	
				0,795	
				(0,64; 0,99)	

pirmą kursą yra prognozuojama socialinių mokslų studentams (87,5%), tačiau ji yra mažesnė, nei 2.13 lentelėje pateiktu fizinių mokslų atveju (90,7%). Taigi optimaliausias yra modelis, pritaikytas fizinių mokslų studentams, neišbraukus nereikšmingų kintamųjų. Jis užrašomas taip:

$$\ln \frac{p}{1-p} = -4,5281 + 0,4972 \cdot \textit{Konkursinis balas} + 0,6136 \cdot \textit{Metinių pažymių vidurkis} - \\ -0,5308 \cdot \textit{Miesto grupė (apskričių centrai)} - 0,0284 \cdot \textit{Miesto grupė (kiti miestai)} - \\ -0,5940 \cdot \textit{Studento pajėgumo indeksas}.$$

Apibendrinant rezultatus galime daryti išvadą, kad, norint pagerinti modelių tikslumą, reiktų patobulinti konkursinio balo ir studentų pajėgumo indekso struktūrą ir skaičiavimo tvarką.

2.4.7. MODULIO „MATEMATIKA 1“ ĮVERTINIMO PROGNOZAVIMAS

Sudarysime logistinės regresinės analizės modelius su 5 nepriklausomais kintamaisiais (2.15 lentelė) dviem skirtingiems duomenų pjūviams: studijų sritis – socialiniai mokslai, priėmimo metai – 2009, studento gimtasis miestas – Kaunas; studijų sritis – socialiniai mokslai, priėmimo metai – 2009, gimtasis miestas – visi, išskyrus Kauną. Šiuo atveju analizuojami tik tokie studentai, kurie už studijų modulį atsiskaitė (įvertinimas ≥ 5). Abiem atvejais priklausomas kintamasis yra trireikšmis modulio įvertinimas: įvertinimas puikus (9 arba 10), geras (7 arba 8) arba patenkinamas (5 arba 6). Modeliuose naudojama logit ryšio funkcija ir prognozuojamas puikus modulio įvertinimas. Nepriklausomi kintamieji yra konkursinis balas, mokyklos metinių pažymių vidurkis, stojimo į pasirinktą programą pageidavimo numeris ir studento pajėgumo indeksas.

Gavome, kad abiem atvejais logistinės regresijos modelis yra gerai suderintas su duomenimis (pirmu atveju Score kriterijaus p-reikšmė $p_1 = 0,878$, o Somerso kriterijaus reikšmė lygi 0,808). Tikslus prognozavimas yra pirmu atveju, kadangi tuomet Somerso kriterijaus reikšmė yra didesnė. Taigi puikus modulio įvertinimas geriau yra prognozuojamas socialinius mokslus studijuojantiems kauniečiams.

Hipotezė apie visų koeficientų prie nepriklausomų kintamųjų lygybę 0 abiem atvejais atmesta (abi $p \leq 0,05$). Hipotezė apie kiekvieno atskiro koeficiento lygybę 0 pirmu atveju atmesta tik konkursiniam balui ($p_1 \leq 0,05$). Visi kiti kintamieji yra statistiškai nereikšmingi (visos $p > 0,05$). Antru atveju hipotezės atmestos konkursiniam balui, metinių pažymių vidurkiui ir stojimo pageidavimo numeriui (visos $p \leq 0,05$).

Modeliuose naudojant tik statistiškai reikšmingus nepriklausomus kintamuosius (2.16 lentelė) gavome, kad abiem atvejais logistinės regresijos modelis yra suderintas su duomenimis, o tiksliau prognozuojama kauniečių atveju.

2.15 lentelė

Modulio „Matematika 1“ įvertinimo prognozavimas

Politominės tvarkos skalės logistinės regresinės analizės rezultatai					
Priklausomas kintamasis: studijų modulio įvertinimas (3 kategorijų skalė)	Studijų modulis: „Matematika 1“	Ryšio funkcija: cumulative logit	Nepriklausomi kintamieji		x_1 : Konkursinis balas
					x_2 : Mokyklos metinių pažymių vidurkis
					x_3 : Stojimo pageidavimo numeris
					x_4 : Studento pajėgumo indeksas
Tyrimo numeris Tyrimo rezultatai		1		2	
		Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009; Miestas: Kaunas		Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009; Miestas: Visi, išskyrus Kauną	
H_0: modelis gerai suderintas su duomenimis (Score kriterijaus p-reikšmė ir Somerso kriterijaus reikšmė)		0,878; 0,808		0,659; 0,781	
$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$		0,0001		0,0001	
$H_0: \beta_1 = 0$		0,0001		0,0001	
$H_0: \beta_2 = 0$		0,4504		0,0301	
$H_0: \beta_3 = 0$		0,8956		0,04	
$H_0: \beta_4 = 0$		0,5126		0,5372	
Kokso-Snelo determinacijos koeficientas r^2_{CS}		0,5063		0,4902	
Stebėjimų dažnis	Patenkinamai		73	142	
	Gerai		25	47	
	Puikiai		16	37	
Parametras β_0		12,9616		10,3748	
Parametras $\beta_i, i = 1, 2, 3, \dots$		β_1	β_2	β_1	β_2
Taškinis įvertis		-0,8101	-0,3952	-0,4603	-0,6948
Galimybių santykis		0,445	0,674	0,631	0,499
Pasikliautinis intervalas		(0,29; 0,68)	(0,24; 1,88)	(0,52; 0,77)	(0,27; 0,94)
		β_3	β_4	β_3	β_4
		-0,03	0,3867	0,3147	0,2440
		0,970	1,472	1,370	1,276
		(0,62; 1,52)	(0,46; 4,68)	(1,01; 1,85)	(0,59; 2,77)

2.16 lentelė

Modulio „Matematika 1“ įvertinimo prognozavimas (be nereikšmingų kintamųjų)

Politominės tvarkos skalės logistinės regresinės analizės rezultatai						
Priklausomas kintamasis: studijų modulio įvertinimas (3 kategorijų skalė)	Studijų modulis: „Matematika 1“	Ryšio funkcija: cumulative logit	Nepriklausomi kintamieji			x_1 : Konkursinis balas
						x_2 : Mokyklos metinių pažymių vidurkis
						x_3 : Stojimo pageidavimo numeris
						x_4 : Studento pajėgumo indeksas
Tyrimo numeris		1		2		
Tyrimo rezultatai		Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009; Miestas: Kaunas		Pjūvis Studijų sritis: Socialiniai mokslai; Metai: 2009; Miestas: Visi, išskyrus Kauną		
H ₀ : modelis gerai suderintas su duomenimis (Score kriterijaus p-reikšmė ir Somerso kriterijaus reikšmė)		0,489; 0,797		0,451; 0,774		
H ₀ : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$		0,0001		0,0001		
H ₀ : $\beta_1 = 0$		0,0001		0,0001		
H ₀ : $\beta_2 = 0$		-		0,0492		
H ₀ : $\beta_3 = 0$		-		0,0314		
H ₀ : $\beta_4 = 0$		-		-		
Kokso-Snelo determinacijos koeficientas r^2_{CS}		0,4987		0,4796		
Stebėjimų dažnis	Patenkinamai	73		142		
	Gerai	25		47		
	Puikiai	16		37		
Parametras β_0		12,0484		11,5608		
Parametras $\beta_i, i = 1, 2, 3, \dots$		β_1		β_1 β_2		
Taškinis įvertis		-0,7339		-0,4926 -0,4758		
Galimybių santykis		0,480		0,611 0,621		
Pasikliautinis intervalas		(0,38; 0,61)		(0,53; 0,71) (0,39; 1,00)		
				β_3		
				0,1467		
				1,158		
				(1,01; 1,32)		

Lyginant 2.15 ir 2.16 lenteles matome, kad iš modelių pašalinus nereikšmingus kintamuosius, jų tikslumas sumažėjo. Taigi optimaliausias yra modelis, pritaikytas socialinius mokslus studijuojantiems kauniečiams, neišbraukus nereikšmingų kintamųjų. Jis užrašomas taip:

$$\ln \frac{p}{1-p} = 12,9616 - 0,8101 \cdot \textit{Konkursinis balas} - 0,3952 \cdot \textit{Metinių pažymių vidurkis} - \\ - 0,03 \cdot \textit{Stojimo pageidavimo numeris} + 0,3867 \cdot \textit{Studento pajėgumo indeksas}.$$

Apibendrinant rezultatus galime daryti išvadą, kad, norint pagerinti modelių tikslumą, reiktų patobulinti konkursinio balo struktūrą ir skaičiavimo tvarką.

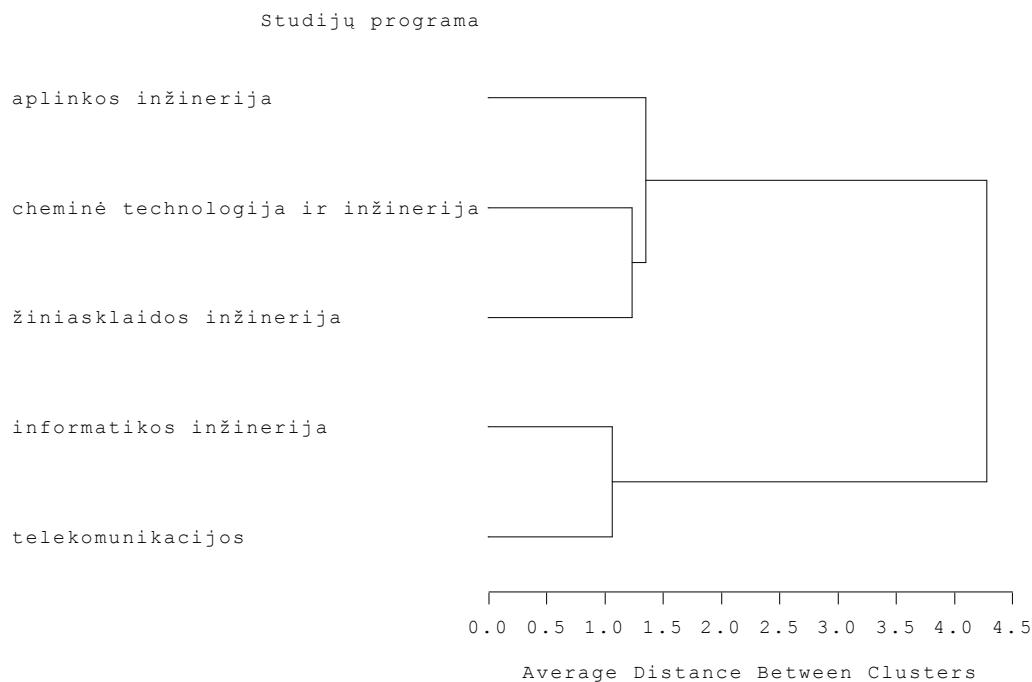
2.4.8. STUDIJŲ PROGRAMŲ KLASIFIKAVIMAS

Sudarysime klasterinės analizės modelius su 3 klasifikuojančiais kintamaisiais (2.17 lentelė) dviem skirtingais pjūviais: studijų sritis – technologijų mokslai, priėmimo metai – 2009-2010, studento gimtasis miestas – Kaunas; studijų sritis – technologijų mokslai, priėmimo metai – 2009-2010; gimtasis miestas – visi, išskyrus Kauną. Abiem atvejais naudojamas vidutinės jungties klasterizavimo algoritmas ir klasifikuojamos studijų programos pagal studentų metinius mokyklos matematikos, fizikos ir astronomijos bei informatikos ir informacinių technologijų įvertinimų vidurkius.

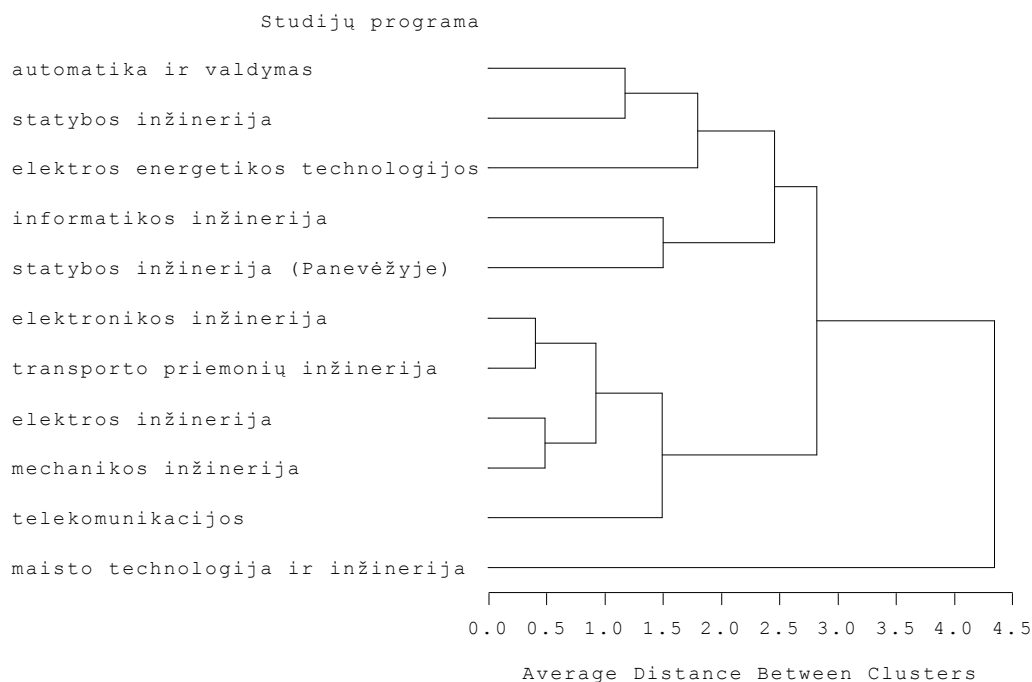
Pirmuoju atveju gavome, kad gerai atsiskyrė dvi studijų programų grupės. Į pirmąją grupę pateko informatikos inžinerijos ir telekomunikacijų studijų programos, o į antrąją – aplinkos inžinerijos, cheminės technologijos ir žiniasklaidos inžinerijos studijų programos. Tai patvirtina 2.16 pav. dendograma. Galime daryti išvadą, kad tarp informatikos inžinerijos ir telekomunikacijų studentų skirtumų nėra ir, sudarant bendrą universiteto matematikos, fizikos, informatikos dalykų tvarkaraštį, juos galima jungti į vieną srautą.

Antruoju atveju gavome, kad pagal pasirinktus požymius gerai atsiskyrė tik maisto technologijos ir inžinerijos studijų programa ir ši klasifikacija nėra informatyvi. Tai patvirtina 2.17 pav. dendograma.

Lyginant abu atvejus tarpusavyje matome, kad geriau klasifikuojamos Kauno vidurines mokyklas ar gimnazijas baigusių studentų pasirinktos studijų programos.



2.16 pav. Technologijos mokslų studijų programų klasifikacija pagal Kauno mokyklas baigusių studentų 2009-2010 m. metinius mokyklos matematikos, fizikos ir informatikos įvertinimus



2.17 pav. Technologijos mokslų studijų programų klasifikacija pagal ne Kauno mokyklas baigusių studentų 2009-2010 m. metinius mokyklos matematikos, fizikos ir informatikos įvertinimus

2.17 lentelė

KTU studijų programų klasifikavimas

Klasterizavimo metodas: Vidutinės jungties		Požymiai, pagal kuriuos klasifikuojama		x_1 : matematikos įvertinimų vidurkis		
				x_2 : fizikos ir astronomijos įvertinimų vidurkis		
				x_3 : informatikos ir informacinių technologijų įvertinimų vidurkis		
Tyrimo numeris	Tyrimo rezultatai	Klasterio numeris	Sujungti klasteriai		Klasterių standartinis nuokrypis $\hat{\sigma}$	Euklido atstumas tarp klasterių d
1	Pjūvis Studijų sritis: technologijų mokslai; Miestas: Kaunas; Metai: 2009-2010	4	informatikos inžinerija	telekomunikacijos	0,43	1,06
		3	cheminė technologija ir inžinerija	žiniasklaidos inžinerija	0,50	1,23
		2	aplinkos inžinerija	KLASTERIS 3	0,54	1,35
		1	KLASTERIS 2	KLASTERIS 4	1,39	4,28
2	Pjūvis Studijų sritis: technologijų mokslai; Miestas: Visi, išskyrus Kauną; Metai: 2009-2010	10	elektronikos inžinerija	transporto priemonių inžinerija	0,17	0,40
		9	elektros inžinerija	mechanikos inžinerija	0,20	0,49
		8	KLASTERIS 10	KLASTERIS 9	0,33	0,92
		7	automatika ir valdymas	statybos inžinerija	0,48	1,17
		6	KLASTERIS 8	telekomunikacijos	0,46	1,49
		5	informatikos inžinerija	statybos inžinerija (Panevėžyje)	0,61	1,50
		4	KLASTERIS 7	elektros energetikos technologijos	0,66	1,80
		3	KLASTERIS 4	KLASTERIS 5	0,88	2,45
		2	KLASTERIS 3	KLASTERIS 6	0,98	2,81
1	KLASTERIS 2	maisto technologija ir inžinerija	1,16	4,35		

2.5. REKOMENDACIJOS DĖL TOLIMESNIO SUKURTŲ MODELIŲ IR PROGRAMINIŲ PRIEMONIŲ TOBULINIMO

Darbe sukurti duomenų analizės modeliai ir programinės priemonės suteikia galimybę greitai, įvairiais pjūviais atlikti studijų proceso duomenų analizę ir gauti vertingus rezultatus. Siekiant analizės procesą padaryti lankstesniu ir dar patogesniu vartotojui, neturinčiam pakankamų statistikos žinių, rekomenduojama statistinius modelius ir juos realizuojančias programines priemones toliau tobulinti.

Statistiniai modeliai turėtų būti tobulinami į juos įtraukiant daugiau grafikos elementų visiems analizės atvejams. Pavyzdžiui, dvifaktorišinėje dispersinėje analizėje turėtų būti pateikiami grafikai, atspindintys faktorių įtaką analizuojamiems mokymosi rodikliams bei pačių faktorių tarpusavio sąveiką, logistinės regresijos atveju turėtų būti grafiškai pateikiamas pritaikyto modelio galimybių santykis ir prognozuojamos tikimybės. Ateityje sistemą rekomenduojama papildyti naujais modeliais (pvz. diskriminantinės ir faktorinės analizės modeliais).

Programinės priemonės turėtų būti tobulinamos sukuriant universalesnę vartotojo sąsają, analizės procesui suteikiant lankstumo ir jį pritaikant nepatyrusiam vartotojui, neturinčiam specialaus statistinio pasirengimo ir darbo su statistiniais paketais patirties. Lankstumas galėtų būti realizuotas įgalinant programą dalinai interpretuoti gaunamus rezultatus ir pateikti ataskaitas su atitinkamomis išvadomis. Tokiu būdu programines priemones galėtų naudoti didesnis vartotojų ratas, kadangi jiems nereiktų ypatingų žinių analizuojant pateikiamus rezultatus. Vartotojo sąsaja galėtų būti tobulinama įdiegiant daugiau sistemos valdymo galimybių ir apsaugų nuo neteisingos duomenų ir parametrų įvesties.

IŠVADOS

1. Išanalizavus literatūros šalinius paaiškėjo, kad studijų proceso duomenų analizei Lietuvoje ir užsienio šalyse dažniausiai yra naudojami tik pavieniai aprašomosios statistikos, išlikimo analizės, logistinės regresinės analizės ir neparametriniai tikimybių branduolio tankio įvertinimo metodai, todėl reikia kurti duomenų analizės modelių sistemas ir patogiai jas realizuojančias programines priemones. Populiariausios duomenų analizei naudojamos programinės priemonės yra SAS, SPSS ir PASW Statistics.
2. Atlikus sistemos SAS objektinio programavimo galimybių analizę nustatyta, kad joje yra visos duomenų analizės sistemos kūrimui reikalingos priemonės.
3. Studijų proceso duomenų analizei sudaryti koreliacinės, dispersinės, logistinės regresinės ir klasterinės analizės modeliai, leidžiantys analizuoti studentų mokymosi rodiklių tarpusavio ryšius, mokymosi rodiklių vidurkių priklausomybę nuo įvairių faktorių, prognozuoti sėkmingų studijų tikimybes ir klasifikuoti studijų programas.
4. Panaudojus SAS programinę įrangą, jos makroprogramavimo kalbą, SAS webAF posistemę ir Java technologijas sukurtos studijų proceso duomenų analizės programinės priemonės, veikiančios interneto aplinkoje realaus laiko režime, ženkliai sutrumpinančios analizės laiką ir rezultatus pateikiančios vartotojui patogiu formatu (HTML ir MS Word RTF). Sistemos struktūra yra lanksti ir jos galimybes galima nesunkiai išplėsti pridedant naujų modelių, vizualizacijos priemonių, praplečiant vartotojo sąsajos galimybes. Tokio pobūdžio sistema yra pirmoji Lietuvoje.
5. Sukurti modeliai ir programinės priemonės panaudotos realių KTU studijų proceso duomenų analizei. Taikymas parodė, kad programinės priemonės sėkmingai sprendžia darbe suformuluotus uždavinius.
6. Rekomenduojama programines priemones tobulinti papildant naujais statistiniais modeliais, grafinėmis priemonėmis ir sukuriant universalesnę vartotojo sąsają.

LITERATŪRA

1. Leonavičienė T. Studijų proceso analizė taikant išlikimo analizės metodus / *Pedagogy Studies (Pedagogika)* iš www.ceeol.com, 2009, 93, p. 98-10.
2. Saldauskienė V., Virkutis V. Matematikos dalyko studijų proceso tobulinimo aspektai Vilniaus kolegijoje / *Vilniaus kolegija*.
3. Samusenko P., Radavičius M. Palyginamoji pažangumo analizė: panelinės logistinės regresijos modelis / *Lietuvos matematikos rinkinys. LMD darbai*, 2010, 51 t., 1-14.
4. Murauskas G., Radavičius M. Studentų pasiekimų analizė naudojant ranginę logistinę regresiją / *Lietuvos matematikos rinkinys. LMD darbai*, 2010, 51 t., 11-14.
5. Figini S., Giudici P. Statistical models for e-learning data / *Stat Methods Appl.*, 2009, 18, p. 293-304.
6. Stokes, Maura E., Davis, Charles S., Koch Gary G. Categorical Data Analysis Using the SAS System / *Cary, NC: SAS Institute Inc.*, 1995, 499 p.
7. Pratter, F. Web Development with SAS® by Example / *Second Edition. Cary, NC: SAS Institute Inc.*, 2006, 341 p.
8. Bohnenkamp W., Iverson J. SAS® Graphics for Java: Examples Using SAS® AppDev Studio™ and the Output Delivery System. / *Cary, NC: SAS Institute Inc.*, 2007, 346 p.
9. Čekanavičius V, Murauskas G. Statistika ir jos taikymai I. / *Vilnius*, 2006, 238 p.
10. Čekanavičius V, Murauskas G. Statistika ir jos taikymai II. / *Vilnius*, 2008, 268 p.
11. Kernel Density Estimation. <http://www.technion.ac.il/docs/sas/stat/chap13/sect13.htm>, 2010 01 21.
12. SAS/STAT® User's Guide, Version 8 / *SAS Institute Inc., Cary*, 1999.
13. The LOGISTIC procedure – Overview.
<http://www.tau.ac.il/cc/pages/docs/sas8/stat/chap39/sect1.htm>, 2010 01 21.
14. Karp A. H. Getting started with PROC LOGISTIC / *Sierra Information Services, Inc. Sonoma, California USA*, paper 248-26.
15. The REG procedure – Overview. <http://www.technion.ac.il/docs/sas/stat/chap55/sect1.htm>, 2010 01 21.
16. Overview of SAS PROC REG. http://www.ats.ucla.edu/stat/sas/library/SASReg_mf.htm, 2011 05 17.
17. Survival Analysis Procedures. <http://www.technion.ac.il/docs/sas/stat/chap10/sect3.htm>, 2010 01 21.
18. SAS Survival Analysis PROC LIFETEST. <http://www.stattutorials.com/SAS/TUTORIAL-PROC-LIFETEST.htm>, 2011 05 17.

19. SAS AppDev Studio 3.0 Developer's Site – Java Technologies for Clients.
<http://support.sas.com/rnd/appdev/V30/tech/clientjava.htm>, 2010 01 21.
20. SAS AppDev Studio 3.0 Developer's Site – webAF Software.
<http://support.sas.com/rnd/appdev/V30/webAF/index.htm>, 2010 01 21.
21. SAS AppDev Studio 3.0 Developer's Site – webEIS Software.
<http://support.sas.com/rnd/appdev/V30/webEIS/index.htm>, 2010 01 21.
22. Java Servlet – Wikipedia. http://lt.wikipedia.org/wiki/Java_Servlet, 2010 01 21.
23. Java ServerPages – Wikipedia. http://lt.wikipedia.org/wiki/JavaServer_Pages, 2010 01 21.
24. JavaScript – Wikipedia. <http://en.wikipedia.org/wiki/JavaScript>, 2011 05 16.
25. HTML – Wikipedia. <http://en.wikipedia.org/wiki/HTML>, 2011 05 16.
26. The SAS system. <http://www.sas.com>, 2011 05 16.
27. Anderson, Carolyn J. Lecture notes. Introduction to Generalized Linear Models (or GLM for short). Department of Educational Psychology, University of Illinois.
28. Anderson, Carolyn J. Lecture notes. Logistic Regression for Dichotomous Responses. Department of Educational Psychology, University of Illinois.
29. Anderson, Carolyn J. Lecture notes. Multiple Logistic Regression for Dichotomous Responses. Department of Educational Psychology, University of Illinois.
30. Anderson, Carolyn J. Lecture notes. Logit Models for Multicategory Responses. Department of Educational Psychology, University of Illinois.
31. Pabarčius L., Janilionis V. Studentų apklausos rezultatų dispersinės analizės modeliai / *Aštuntosios studentų taikomosios matematikos konferencijos medžiaga*, Kaunas, 2010, p. 82-83.
32. Pabarčius L., Janilionis V. Studijų proceso dispersinės analizės modeliai / *Devintosios studentų taikomosios matematikos konferencijos medžiaga*, Kaunas, 2011, p. 47-48.

1 PRIEDAS. STUDIJŲ PROCESO DUOMENŲ MATRICA

1.1 priedo lentelėje pateikiama visų duomenų analizės sistemos naudojamų kintamųjų matrica.

1.1 priedo lentelė

Duomenų analizėje naudojamų kintamųjų matrica

Kintamojo pavadinimas	Prasmė
Metai	Studento priėmimo į universitetą metai
Lytis	Studento lytis
Miestas	Studento gimtasis miestas
Miesto grupė	Trijų reikšmių kintamasis: 1 – gimtasis miestas priklauso apskrities centrui, 2- savivaldybių centrui, 3 – kitiems miestams
Miestas/rajonas	Dvireikšmis kintamasis: 1 – kintamojo „Miestas“ reikšmė yra apskrities arba savivaldybės centras, 2 – miestelis rajone
Universitetas	Universitetas, į kurį buvo priimtas studentas
Pageidavimo numeris	Stojimo pageidavimo numeris, kuriuo studentas įstojo į KTU
Pageidavimo numerio grupė	1 – pageidavimo numeris nuo 1 iki 3, 2 – nuo 4 iki 6, 3 – nuo 7 iki 10, 4 – nuo 11 iki 16
Studijų sritis	Pasirinkta studijų sritis
Studijų kryptis	Pasirinkta studijų kryptis
Fakultetas	Pasirinktas fakultetas
Studijų programa	Pasirinkta studijų programa
Studijų forma	1 – nuolatinė, 2 - išštinė
Finansavimas	1 – valstybės finansuojama, 2 – valstybės nefinansuojama
Konkursinis balas	Studento konkursinis balas
Konkursinio balo grupė	1 – [0; 6), 2 – [6; 12), 3 – [12; 18), 4 – [18; 23]
Metinių pažymių vidurkis	Studento mokyklos metinių pažymių vidurkis
Pirmo kurso baigimo požymis	1 – studentas primą kursą baigė, 2 – nebaigė
Stojimo pageidavimų skaičius	Studento stojimo prašyme nurodytų pageidavimų skaičius
Studijų sričių skaičius	Studento stojimo prašyme nurodytų studijų sričių skaičius
Studento pajėgumo indeksas	Šio darbo metu sudarytas studento pajėgumo indeksas
Modulio įvertinimas	KTU studijų modulio įvertinimas
Modulio įvertinimo tipas	0 – įvertinimas patenkinamas (5 arba 6), 1 – įvertinimas geras (7 arba 8), 2 – įvertinimas puikus (9 arba 10)
Matematika	Studento mokyklinis matematikos metinis pažymys
Anglų kalba	Studento mokyklinis anglų kalbos pažymys
Fizika ir astronomija	Studento mokyklinis fizikos ir astronomijos metinis pažymys
Informatika ir informacinės technologijos	Studento mokyklinis informatikos ir informacinių technologijų metinis pažymys
Lietuvių kalba	Studento mokyklinis lietuvių kalbos metinis pažymys

2 PRIEDAS. MODELIŲ IR VARTOTOJO SĄSAJOS PROGRAMŲ TEKSTAI

Šiame priede pateikiami analizės sistemos bylų ir statistinius modelius realizuojančių programų tekstai. Dėl didelės apimties yra parodyti tik pagrindinių sistemos bylų ir logistinės regresinės analizės modelio programų tekstai. Kiti programų tekstai įrašyti kompaktiniame diske (CD).

JSP bylų programiniai tekstai

Byla index.jsp

```
<%@ taglib uri="http://www.sas.com/taglib/sas" prefix="sas" %>
<%@ taglib uri="http://www.sas.com/taglib/sasads" prefix="sasads" %>
<%@ page import="com.sas.sasserver.datasetinfo.DataSetInfoInterface" %>
<%@ page import="com.sas.collection.OrderedCollection" %>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<%@ page contentType="text/html; charset=windows-1257"%>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</title>
</head>
<style>
form {display: inline;}
</style>
<body bgcolor="CEDEF4">
<sasads:Connection id="connection1" serverArchitecture="IOM" scope="session">
<div align="center">
<h1><b>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</b></h1><br />
<h2>Duomenų įkėlimas</h2><br />
<sasads:Form id="duomenų_import" action="duomenų_import.jsp" >
<sasads:TextEntry id="import_name" prolog="Duomenų failas: " size="50"
text="C:\\Users\\Laurynas\\Laurynas Pabarcius (studijos)\\1_MAGISTRINIS
DARBAS\\MAGISTRINIO DARBO SAS\\DATA_ALL.xls"/><br /><br />
<sasads:PushButton id="duomenų_import_submit" text="Įkelti duomenis" style="font-
size:12pt;"/>
</sasads:Form>
<sasads:Form id="end_button" action="javascript:window.close();">
<sasads:PushButton id="end" text="Baigti darbą" style="font-size:12pt;"/>
</sasads:Form>
</div>
</sasads:Connection>
<br />
<hr align=center noshade size=8>
<br /><br />
<div align="center"></div>
</body>
</html>
```

Byla duomenų_import.jsp

```
<%@ taglib uri="http://www.sas.com/taglib/sas" prefix="sas" %>
<%@ taglib uri="http://www.sas.com/taglib/sasads" prefix="sasads" %>
<%@ page import="com.sas.sasserver.datasetinfo.DataSetInfoInterface" %>
<%@ page import="com.sas.collection.OrderedCollection" %>
```

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<%@ page language="java" contentType="text/html; charset=windows-1257"
pageEncoding="Cp1257"%>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</title>
</head>
<style>
form {display: inline;}
</style>
<body bgcolor="CEDEF4">
<sasads:Connection id="NAMUCIAI_localhost" serverArchitecture="IOM"
commandPrompt="" port="5310" persistedName="NAMUCIAI_localhost"
initialStatement="" command="" appletCodebaseRelative="true" accessMethod="IOM"
scope="session"/>
<sasads:Form id="duomenų_pjuvis" action="duomenų_pjuvis.jsp" >
<sasads:Submit id="duomenų_import" connection="NAMUCIAI_localhost"
display="LASTOUTPUT" scope="page">
    <% include file="Priimtieji_import.sas" %>
</sasads:Submit>
<% // Java scriptlet to add choicebox values to page context
DataSetInfoInterface dsinfo =
(DataSetInfoInterface) com.sas.servlet.util.Util.newInstance
(NAMUCIAI_localhost.getClassFactory(),NAMUCIAI_localhost,DataSetInfoInterface.class);
dsinfo.setDataSet("SASUSER.PRIIMTIEJI");
// display unique values
int index_univ = dsinfo.getVariableIndex("UNIVERSITETAS");
int index_fakult = dsinfo.getVariableIndex("FAKULTETAS");
int index_stud_srit = dsinfo.getVariableIndex("STUDIJU_SRITIS");
int index_stud_krypt = dsinfo.getVariableIndex("STUDIJU_KRYPTIS");
int index_stud_progr = dsinfo.getVariableIndex("STUDIJU_PROGRAMA");
int index_lytis = dsinfo.getVariableIndex("LYTIS");
int index_metai = dsinfo.getVariableIndex("METAI");
int index_miestas = dsinfo.getVariableIndex("MIESTAS");

com.sas.collection.OrderedCollection univ_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_univ),1);
univ_col.add("Visi");
pageContext.setAttribute("values_univ",univ_col);
com.sas.collection.OrderedCollection fakult_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_fakult),1);
fakult_col.add("Visi");
pageContext.setAttribute("values_fakult",fakult_col);
com.sas.collection.OrderedCollection stud_srit_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_srit),1);
stud_srit_col.add("Visi");
pageContext.setAttribute("values_stud_srit",stud_srit_col);
com.sas.collection.OrderedCollection stud_krypt_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_krypt),1);
stud_krypt_col.add("Visi");
pageContext.setAttribute("values_stud_krypt",stud_krypt_col);
com.sas.collection.OrderedCollection stud_progr_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_progr),1);
stud_progr_col.add("Visi");
pageContext.setAttribute("values_stud_progr",stud_progr_col);
com.sas.collection.OrderedCollection lytis_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_lytis),1);
lytis_col.add("Visi");
pageContext.setAttribute("values_lytis",lytis_col);
com.sas.collection.OrderedCollection metai_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_metai),1);

```

```

metai_col.add("Visi");
pageContext.setAttribute("values_metai",metai_col);
com.sas.collection.OrderedCollection miestas_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_miestas),1);
miestas_col.add("Visi");
pageContext.setAttribute("values_miestas",miestas_col);
%>
<div align="center">
<h1><b>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</b></h1><br />
<h2>Duomenų pjūvio formavimas</h2><br />
</div>

<div align="center">
<table border="0" cellpadding="2" cellspacing="2" width=1060>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Studijų sritis:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_srit" model="values_stud_srit" selectedItem="Visi"></sasads:Choicebox>
</td>
    <td width=187 align="Right" valign="Middle">&nbsp; Metai:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="metai" model="values_metai" selectedItem="Visi"></sasads:Choicebox> </td>
  </tr>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Studijų kryptis:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_krypt" model="values_stud_krypt" selectedItem="Visi"></sasads:Choicebox>
</td>
    <td width=300 align="Right" valign="Middle">&nbsp; Studijų programa:
    <td width=230 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_progr" model="values_stud_progr" selectedItem="Visi"></sasads:Choicebox>
</td>
  </tr>
  <tr>
    <td width=300 align="Right" valign="Middle">&nbsp; Universitetas:
    <td width=230 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="univ" model="values_univ"></sasads:Choicebox> </td>
    <td width=187 align="Right" valign="Middle">&nbsp; Miestas:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="miestas" model="values_miestas" selectedItem="Kaunas"></sasads:Choicebox>
</td>
  </tr>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Fakultetas:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="fakult" model="values_fakult" selectedItem="Visi"></sasads:Choicebox> </td>
    <td width=187 align="Right" valign="Middle">&nbsp; Lytis:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="lytis" model="values_lytis" selectedItem="Visi"></sasads:Choicebox> </td>
  </tr>
</table>
<br /></br>
<sasads:PushButton id="duomenų_pjuvis_submit" text="Formuoti duomenų pjūvį"
style="font-size:12pt;"/>
</sasads:Form>
<sasads:Form id="back_button" action="index.jsp">
<sasads:PushButton id="back" text="Atgal" style="font-size:12pt;"/>
</sasads:Form>
<sasads:Form id="end_button" action="javascript:window.close();">
<sasads:PushButton id="end" text="Baigti darba" style="font-size:12pt;"/>
</div>
</sasads:Form>

```

```

</sasads:Connection>
<div align = "left"><font size="3" color="red">Duomenys ikelti!</font></div>
<hr align=center noshade size=8>
</body>
</html>

```

Byla duomenu_pjuvis.jsp

```

<%@ taglib uri="http://www.sas.com/taglib/sas" prefix="sas" %>
<%@ taglib uri="http://www.sas.com/taglib/sasads" prefix="sasads" %>
<%@ page import="com.sas.sasserver.datasetinfo.DataSetInfoInterface" %>
<%@ page import="com.sas.collection.OrderedCollection" %>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<%@ page contentType="text/html; charset=windows-1257"%>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</title>
</head>
<style>
form {display: inline;}
</style>
<body bgcolor="#CEDEF4">
<sasads:Connection id="NAMUCIAI_localhost" serverArchitecture="IOM"
commandPrompt="" port="5310" persistedName="NAMUCIAI-localhost"
initialStatement="" command="" appletCodebaseRelative="true" accessMethod="IOM"
scope="session"/>
<sasads:Form id="duom_pjuv" action="duomenu_pjuvis.jsp">
<sasads:Submit id="priimtieji_pjuvis" connection="NAMUCIAI_localhost"
display="LASTOUTPUT" scope="session">
    <%@ include file="duomenu_pjuvis.sas" %>
</sasads:Submit>
<%
// Java scriptlet to add choicebox values to page context
DataSetInfoInterface dsinfo =
(DataSetInfoInterface) com.sas.servlet.util.Util.newInstance
(NAMUCIAI_localhost.getClassFactory(),NAMUCIAI_localhost,DataSetInfoInterface.class);
dsinfo.setDataSet("SASUSER.PRIIMTIEJI");
// display unique values
int index_univ = dsinfo.getVariableIndex("UNIVERSITETAS");
int index_fakult = dsinfo.getVariableIndex("FAKULTETAS");
int index_stud_srit = dsinfo.getVariableIndex("STUDIJU_SRITIS");
int index_stud_krypt = dsinfo.getVariableIndex("STUDIJU_KRYPTIS");
int index_stud_progr = dsinfo.getVariableIndex("STUDIJU_PROGRAMA");
int index_lytis = dsinfo.getVariableIndex("LYTIS");
int index_metai = dsinfo.getVariableIndex("METAIS");
int index_miestas = dsinfo.getVariableIndex("MIESTAS");

com.sas.collection.OrderedCollection univ_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_univ),1);
univ_col.add("Visi");
pageContext.setAttribute("values_univ",univ_col);
com.sas.collection.OrderedCollection fakult_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_fakult),1);
fakult_col.add("Visi");
pageContext.setAttribute("values_fakult",fakult_col);
com.sas.collection.OrderedCollection stud_srit_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_srit),1);
stud_srit_col.add("Visi");
pageContext.setAttribute("values_stud_srit",stud_srit_col);

```



```

com.sas.collection.OrderedCollection stud_krypt_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_krypt),1);
stud_krypt_col.add("Visi");
pageContext.setAttribute("values_stud_krypt",stud_krypt_col);
com.sas.collection.OrderedCollection stud_progr_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_stud_progr),1);
stud_progr_col.add("Visi");
pageContext.setAttribute("values_stud_progr",stud_progr_col);
com.sas.collection.OrderedCollection lytis_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_lytis),1);
lytis_col.add("Visi");
pageContext.setAttribute("values_lytis",lytis_col);
com.sas.collection.OrderedCollection metai_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_metai),1);
metai_col.add("Visi");
pageContext.setAttribute("values_metai",metai_col);
com.sas.collection.OrderedCollection miestas_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_miestas),1);
miestas_col.add("Visi");
pageContext.setAttribute("values_miestas",miestas_col);
%>
<div align="center">
<h1><b>Kauno Technologijos Universiteto studijų proceso duomenų analizės
sistema</b></h1>
<h2>Duomenų pjūvio formavimas</h2><br />
</div>
<div align="center">
<table border="0" cellpadding="2" cellspacing="2" width=1060>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Studijų sritis:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_srit" model="values_stud_srit"></sasads:Choicebox> </td>
    <td width=187 align="Right" valign="Middle">&nbsp; Metai:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="metai" model="values_metai"></sasads:Choicebox> </td>
  </tr>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Studijų kryptis:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_krypt" model="values_stud_krypt"
selectedItem="Matematika"></sasads:Choicebox> </td>
    <td width=300 align="Right" valign="Middle">&nbsp; Studijų programa:
    <td width=230 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="stud_progr" model="values_stud_progr" selectedItem="taikomoji
matematika"></sasads:Choicebox> </td>
  </tr>
  <tr>
    <td width=300 align="Right" valign="Middle">&nbsp; Universitetas:
    <td width=230 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="univ" model="values_univ"></sasads:Choicebox> </td>
    <td width=187 align="Right" valign="Middle">&nbsp; Miestas:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="miestas" model="values_miestas" selectedItem="Kaunas"></sasads:Choicebox>
</td>
  </tr>
  <tr>
    <td width=50 align="Right" valign="Middle">&nbsp; Fakultetas:
    <td width=231 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="fakult" model="values_fakult" selectedItem="Fundamentaliųjų mokslų
fakultetas"></sasads:Choicebox> </td>
    <td width=187 align="Right" valign="Middle">&nbsp; Lytis:
    <td width=150 align="Left" valign="Middle">&nbsp; <sasads:Choicebox
id="lytis" model="values_lytis" selectedItem="Vyras"></sasads:Choicebox> </td>
  </tr>

```

```

    </table>
<br /></br>
<sasads:PushButton id="duomenu_pjuvis_submit" text="Formuoti duomenų pjūvį"
style="font-size:12pt;"/>
</sasads:Form>
<sasads:Form id="back_button" action="duomenu_import.jsp">
<sasads:PushButton id="back" text="Atgal" style="font-size:12pt;"/>
</sasads:Form>
<sasads:Form id="end_button" action="javascript:window.close();">
<sasads:PushButton id="end" text="Baigti darba" style="font-size:12pt;"/>
</sasads:Form>
</div>
</sasads:Connection>
<%
DataSetInfoInterface dsinfo_pjuvis1 =
(DataSetInfoInterface) com.sas.servlet.util.Util.newInstance
(NAMUCIAI_localhost.getClassFactory(),NAMUCIAI_localhost,DataSetInfoInterface.class);
dsinfo_pjuvis1.setDataSet("priimtieji_pjuvis");

int stud_index1 = dsinfo_pjuvis1.getVariableIndex("ASM_KODAS");
com.sas.collection.OrderedCollection studentail = new
OrderedCollection(dsinfo_pjuvis1.getVariableUniqueValues(stud_index1));
int skaicius1 = studentail.count();
%>
<div id='if-part1' style='visibility: hidden;'><div align = "left"><font size="3"
color="red">Duomenų pjūvis suformotas!</font></div></div>
<div id='else-part1' style='visibility: hidden;'><div align = "left"><font size="3"
color="red">Pasirinkite kitus grupuojančius kintamuosius!</font></div></div>
<script>
var nodel;
if(<%=skaicius1%> > 8) {
    nodel = document.getElementById('if-part1');
}
else {
    nodel = document.getElementById('else-part1');
}
nodel.style.visibility = 'visible';
</script>

<hr align=center noshade size=8>

<div id='if-part' style='visibility: hidden;'>
<div align="center">
<h2>Statistinių duomenų analizės modelių pasirinkimas</h2>

    <table border="0" cellpadding="1" cellspacing="1" width=600>
        <tr>
            <td width=200 align="Center" valign="Top">&nbsp;
                <sasads:Form id="aprasomoji" action="aprasomoji.jsp" >
                    <sasads:PushButton id="aprasomoji_submit" text="Aprašomoji statistika"
style="font-size:13pt;"/>
                </sasads:Form>
            </td>
            <td width=200 rowspan="2" align="Center" valign="Middle">&nbsp;
                <sasads:Form id="dispersine" action="dispersine.jsp" >
                    <sasads:PushButton id="dispersine_submit" text="Dispersinė analizė"
style="font-size:13pt;"/>
                </sasads:Form>
            </td>
            <td width=200 align="Center" valign="Top">&nbsp;
                <sasads:Form id="logistine" action="logistine.jsp" >
                    <sasads:PushButton id="logistine_submit" text="Logistinė regresinė
analizė" style="font-size:13pt;"/>

```

```

        </sasads:Form>
    </td>
</tr>
<tr>
    <td width=200 align="Center" valign="Top">&nbsp;
        <sasads:Form id="koreliacine" action="koreliacine.jsp" >
            <sasads:PushButton id="koreliacine_submit" text="Koreliacinė analizė"
style="font-size:13pt;" />
        </sasads:Form>
    </td>
    <td width=200 align="Center" valign="Top">&nbsp;
        <sasads:Form id="klasterine" action="klasterine.jsp" >
            <sasads:PushButton id="klasterine_submit" text="Klasterinė analizė"
style="font-size:13pt;" />
        </sasads:Form>
    </td>
</tr>
</table>
<br />
<hr align=center noshade size=8>
</div>
</div>
<%
DataSetInfoInterface dsinfo_pjuvis =
(DataSetInfoInterface) com.sas.servlet.util.Util.newInstance
(NAMUCIAI_localhost.getClassFactory(),NAMUCIAI_localhost,DataSetInfoInterface.class);
dsinfo_pjuvis.setDataSet("priimtieji_pjuvis");

int stud_index = dsinfo_pjuvis.getVariableIndex("ASM_KODAS");
com.sas.collection.OrderedCollection studentai = new
OrderedCollection(dsinfo_pjuvis.getVariableUniqueValues(stud_index));
int skaicius = studentai.count();
%>
<script language="Javascript">
<!--
var node;
if(<%=skaicius%> < 8)
{
    alert ("Duomenų pjūvyje yra mažiau nei 8 studentai! Pasirinkite kitokius
grupuojančius kintamuosius.")
}
else
{
    node = document.getElementById('if-part');
    node.style.visibility = 'visible';
}
//-->
</script>

</body>
</html>

```

Byla logistine.jsp

```

<%@ taglib uri="http://www.sas.com/taglib/sas" prefix="sas" %>
<%@ taglib uri="http://www.sas.com/taglib/sasads" prefix="sasads" %>
<%@ page import="com.sas.sasserver.datasetinfo.DataSetInfoInterface" %>
<%@ page import="com.sas.collection.OrderedCollection" %>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<%@ page contentType="text/html; charset=windows-1257"%>
<html xmlns="http://www.w3.org/1999/xhtml">

```

```

<head>
<title>Logistinės regresinės analizės modelis</title>
</head>
<style>
form {display: inline;}
</style>
<body bgcolor="#CEDEF4">
<div align="center">
<h1><b>Logistinės regresinės analizės modelis</b></h1>
</div>

<sasads:Connection id="NAMUCIAI_localhost" serverArchitecture="IOM"
commandPrompt="" port="5310" persistedName="NAMUCIAI-localhost"
initialStatement="" command="" appletCodebaseRelative="true" accessMethod="IOM"
scope="session"/>
<sasads:Submit id="log_data" connection="NAMUCIAI_localhost" display="LASTOUTPUT"
scope="page">
    <%@ include file="Logistine_dataset.sas" %>
</sasads:Submit>

<%
DataSetInfoInterface dsinfo =
(DataSetInfoInterface)com.sas.servlet.util.Util.newInstance
(NAMUCIAI_localhost.getClassFactory(),NAMUCIAI_localhost,DataSetInfoInterface.class);
dsinfo.setDataSet("PRIIMTIEJI_PJUVIS_LOG");
int index_modul = dsinfo.getVariableIndex("PAVADIN");

com.sas.collection.OrderedCollection modul_col = new
OrderedCollection(dsinfo.getVariableUniqueValues(index_modul));
pageContext.setAttribute("values_modul",modul_col);

pageContext.setAttribute("log_tipas", new OrderedCollection
("DVIREIKSME,TRIUREIKSME"));
pageContext.setAttribute("log_tipas_lab", new OrderedCollection
("Prognozuojama tikimybė baigti pirma kursa, Prognozuojama tikimybė gauti puiku
modulio įvertinimą"));
%>

<sasads:Form id="logistine" action="logistine_result.jsp">

<table border="0" cellpadding="0" cellspacing="0" align="Center" height=34
width=650>
    <tr>
        <td align="Center" valign="Middle" height=32><b>Analizės
tipas:</b></td>
        <td align="Center" valign="Middle" height=32><b>Studiju
modulis:</b></td>
    </tr>
    <tr>
        <td align="Left" valign="Middle" height=32><sasads:Radio
id="log_choice" model="log_tipas" descriptionModel="log_tipas_lab"
selectedIndex="0"/></td>
        <td align="Center" valign="Middle" height=32><sasads:Choicebox
id="log_modul" model="values_modul"/></td>
    </tr>
</table>

<div align = "center">
<h2>Lentelės</h2>
</div>

<%
pageContext.setAttribute("log_gr", new OrderedCollection

```

```

("GRAFIKA,NONE"));
pageContext.setAttribute("log_gr_lab", new OrderedCollection
("Rezultatai su grafiku pirmu analizės atveju, Be grafiko"));

pageContext.setAttribute("link_tip", new OrderedCollection
("CLOGLOG,GLOGIT,LOGIT,PROBIT"));
pageContext.setAttribute("link_tip_lab", new OrderedCollection
("Log-log, Apibendrinta logit, Logit, Probit"));

pageContext.setAttribute("log_saug", new OrderedCollection("HTML,HTML_RTF"));
pageContext.setAttribute("log_saug_lab", new OrderedCollection("Tik HTML,HTML ir
RTF"));
%>

<div align="center">

<table border="0" cellpadding="2" cellspacing="2" align="Center" width=1200>
  <tr>
    <td align="Center" valign="Center"><b>Ryšio funkcija:</b></td>
    <td align="Center" valign="Center" colspan="5"><b>Analizuojami
kintamieji:</b></td>
  </tr>
  <tr>
    <td align="Center" valign="Top" rowspan="2">&nbsp; <sasads:ListBox
id="link_tipas" model="link_tip" descriptionModel="link_tip_lab"
selectedIndexes="2"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_pag_nr" text="Stojimo pageidavimo numeris" value="PAGEIDAVIMO_NUMERIS"
/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_konk_balas" text="Konkursinis balas" value="KONK_BALAS"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_met_vid" text="Mokyklos metinių pažymių vidurkis"
value="METINIU_VIDURKIS"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_mst_grp" text="Miesto grupė" value="MIESTO_GRUPE"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_mst_raj" text="Miestas/rajonas" value="MIESTO_TIPAS"/></td>
  </tr>
  <tr>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox id="var_fak"
text="Fakultetas" value="Fakultetas"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_stud_frm" text="Studijų forma" value="STUDIJU_FORMA"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_paz_ind" text="Studentų pažangumo indeksas"
value="PAZANGUMO_INDEKSAS"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_fin_tipas" text="Finansavimo tipas" value="FINANSAVIMAS"/></td>
    <td align="Left" valign="Center">&nbsp; <sasads:Checkbox
id="var_lytis" text="Lytis" value="LYTIS"/></td>
  </tr>
</table>
</div>

<div align = "center">
<h2>Grafikai</h2>

<table border="0" cellpadding="0" cellspacing="1" align="Center">
  <tr>
    <td width=300 align="Left" valign="Middle"><sasads:Radio
id="log_grafika" model="log_gr" descriptionModel="log_gr_lab"
selectedIndex="0"/></td>

```



```

initialStatement="" command="" appletCodebaseRelative="true" accessMethod="IOM"
scope="session"/>
<%
com.sas.servlet.util.SocketListener socket = new
com.sas.servlet.util.SocketListener();
int port = socket.setup();
socket.start();
String localhost = (java.net.InetAddress.getLocalHost()).getHostAddress();
%>
<sasads:Submit id="logist_sas_submit" connection="NAMUCIAI_localhost"
display="LASTOUTPUT" scope="page">
  <%@ include file="log_sas_submit.sas" %>
</sasads:Submit>
</div>
<%
socket.write(out);
socket.close();
%>

</body>
</html>

```

SAS bylą programiniai tekstai

logistine_dataset.sas

```

proc sql;
create table priimtieji_pjuvis_log as
select * from priimtieji_pjuvis,sasuser.rezultatai_ktu
where priimtieji_pjuvis.ASM_KODAS = rezultatai_ktu.ASM_KODAS;
quit;

```

logistine_sas_submit.sas

```

filename sock SOCKET '<%=localhost%>:<%=port%>';
filename sock_rtf SOCKET '<%=localhost%>:<%=port%>';
filename webapp "D:\AppDevStudio\webAF\Projects\Duomenų_analizes_sistema\webapp";
options nodate nonumber noovp center pagesize=20;
goptions reset=all device=activex;
ods listing close;
%macro ods_html(class_varlist,varlist,log_gr,link_tipas,an_tipas,modul);
ods html body=sock rs=none style=analysis gpath=webapp;
ods noptitle;

data priimtieji_pjuvis_2009;
set priimtieji_pjuvis;
where META1 = 2009;
run;

data priimtieji_pjuvis_log_mod;
set priimtieji_pjuvis_log;
where PAVADIN = &modul;
run;

%if &an_tipas = "DVIREIKSME" %then
%do;
  %if &log_gr = "GRAFIKA" %then
  %do;
    ods graphics on;
    ods exclude FitStatistics LackFitPartition NObs;
    title "<h2>Logistinės regresinės analizės modelio rezultatai</h2>";
    proc logistic data=priimtieji_pjuvis_2009 desc;

```

```

        class &class_varlist;
        model BAIG_POZ=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas outroc=rocl selection=s;
        run;
        ods graphics off;
    %end;
    %else
    %do;
        ods graphics on;
        ods exclude FitStatistics LackFitPartition NObs;
        title "<h2>Logistinės regresinės analizės modelio rezultatai</h2>";
        proc logistic data=priimtieji_pjuvis_2009 desc;
        class &class_varlist;
        model BAIG_POZ=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas;
        run;
        ods graphics off;
    %end;
%end;
%else
%do;
    %if &log_gr = "GRAFIKA" %then
    %do;
        ods graphics on;
        ods exclude FitStatistics LackFitPartition NObs;
        title "<h2>Logistinės regresinės analizės modelio rezultatai</h2>";
        proc logistic data=priimtieji_pjuvis_log_mod;
        class &class_varlist;
        model PAZ_TIPAS=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas outroc=rocl selection=s;
        run;
        ods graphics off;
    %end;
    %else
    %do;
        ods graphics on;
        ods exclude FitStatistics LackFitPartition NObs;
        title "<h2>Logistinės regresinės analizės modelio rezultatai</h2>";
        proc logistic data=priimtieji_pjuvis_log_mod;
        class &class_varlist;
        model PAZ_TIPAS=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas;
        run;
        ods graphics off;
    %end;
%end;

ods html close;

%mend ods_html;

%macro ods_rtf(class_varlist,varlist,log_gr,link_tipas,an_tipas,vardas,modul);
ods rtf body=sock_rtf rs=none file=&vardas;
ods noprt;

data priimtieji_pjuvis_2009;
set priimtieji_pjuvis;
where METAI = 2009;
run;

data priimtieji_pjuvis_log_mod;
set priimtieji_pjuvis_log;
where PAVADIN = &modul;
run;

```



```

%if &an_tipas = "DVIREIKSME" %then
%do;
  %if &log_gr = "GRAFIKA" %then
  %do;
    ods graphics on;
    ods exclude FitStatistics LackFitPartition NObs;
    title "Logistinės regresinės analizės modelio rezultatai";
    proc logistic data=priimtieji_pjuvis_2009 desc;
    class &class_varlist;
    model BAIG_POZ=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas outroc=rocl selection=s;
    run;
    ods graphics off;
  %end;
  %else
  %do;
    ods graphics on;
    ods exclude FitStatistics LackFitPartition NObs;
    title "Logistinės regresinės analizės modelio rezultatai";
    proc logistic data=priimtieji_pjuvis_2009 desc;
    class &class_varlist;
    model BAIG_POZ=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas;
    run;
    ods graphics off;
  %end;
%end;
%else
%do;
  %if &log_gr = "GRAFIKA" %then
  %do;
    ods graphics on;
    ods exclude FitStatistics LackFitPartition NObs;
    title "Logistinės regresinės analizės modelio rezultatai";
    proc logistic data=priimtieji_pjuvis_log_mod;
    class &class_varlist;
    model PAZ_TIPAS=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas outroc=rocl selection=s;
    run;
    ods graphics off;
  %end;
  %else
  %do;
    ods graphics on;
    ods exclude FitStatistics LackFitPartition NObs;
    title "Logistinės regresinės analizės modelio rezultatai";
    proc logistic data=priimtieji_pjuvis_log_mod;
    class &class_varlist;
    model PAZ_TIPAS=&varlist / expb rsquare ctable pprob=0.5 nodummyprint
lackfit link=&link_tipas;
    run;
    ods graphics off;
  %end;
%end;

ods rtf close;

%mend ods_rtf;

%macro
logistine(pag_nr, konk_balas, met_vid, mst_gr, mst_tipas, fak, stud_frm, paz_ind, finansav
imas, lytis, log_gr, link_tipas, an_tipas, mark, vardas, modul);

```

```

%let class_var_tmp = &mst_gr &mst_tipas &fak &stud_frm &finansavimas &lytis;
%let class_varlist = %sysfunc(tranwrd(&class_var_tmp,%quote(null),%quote()));
%let var_tmp = &pag_nr &konk_balas &met_vid &mst_gr &mst_tipas &fak &stud_frm
&paz_ind &finansavimas &lytis;
%let varlist = %sysfunc(tranwrd(&var_tmp,%quote(null),%quote()));

%if &mark = "HTML" %then
%do;
%ods_html(&class_varlist,&varlist,&log_gr,&link_tipas,&an_tipas,&modul);
%end;
%else
%do;
%ods_html(&class_varlist,&varlist,&log_gr,&link_tipas,&an_tipas,&modul);
%ods_rtf(&class_varlist,&varlist,&log_gr,&link_tipas,&an_tipas,&vardas,&modul);
%end;
%mend logistine;
%logistine(<%= request.getParameter("var_pag_nr") %>,
          <%= request.getParameter("var_konk_balas") %>,
          <%= request.getParameter("var_met_vid") %>,
          <%= request.getParameter("var_mst_grp") %>,
          <%= request.getParameter("var_mst_raj") %>,
          <%= request.getParameter("var_fak") %>,
          <%= request.getParameter("var_stud_frm") %>,
          <%= request.getParameter("var_paz_ind") %>,
          <%= request.getParameter("var_fin_tipas") %>,
          <%= request.getParameter("var_lytis") %>,
          "<%= request.getParameter("log_grafika") %>",<%=
request.getParameter("link_tipas") %>,"<%= request.getParameter("log_choice") %>",<%=
request.getParameter("logist_saug") %>",<%=
request.getParameter("logist_saug_name") %>",<%=
request.getParameter("log_modul") %>");

```

3 PRIEDAS. MODELIŲ TAIKYMO REZULTATAI

Šiame priede pateikiami darbo su duomenų analizės sistema rezultatų pavyzdžiai. Dėl didelės apimties parodyti tik logistinės regresinės analizės modelio taikymo rezultatai. Kitų modelių rezultatų pavyzdžiai įrašyti kompaktiniame diske (CD).

Logistinės regresinės analizės modelio rezultatai

Model Information		
Data Set	WORK.PRIIMTIEJI_PJUVIS_2009	
Response Variable	BAIG_POZ	Pirmo kurso baigimo požymis
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Response Profile		
Ordered Value	BAIG_POZ	Total Frequency
1	1	252
2	0	29

Probability modeled is BAIG_POZ=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

R-Square	0.1962	Max-rescaled R-Square	0.4044
----------	--------	-----------------------	--------

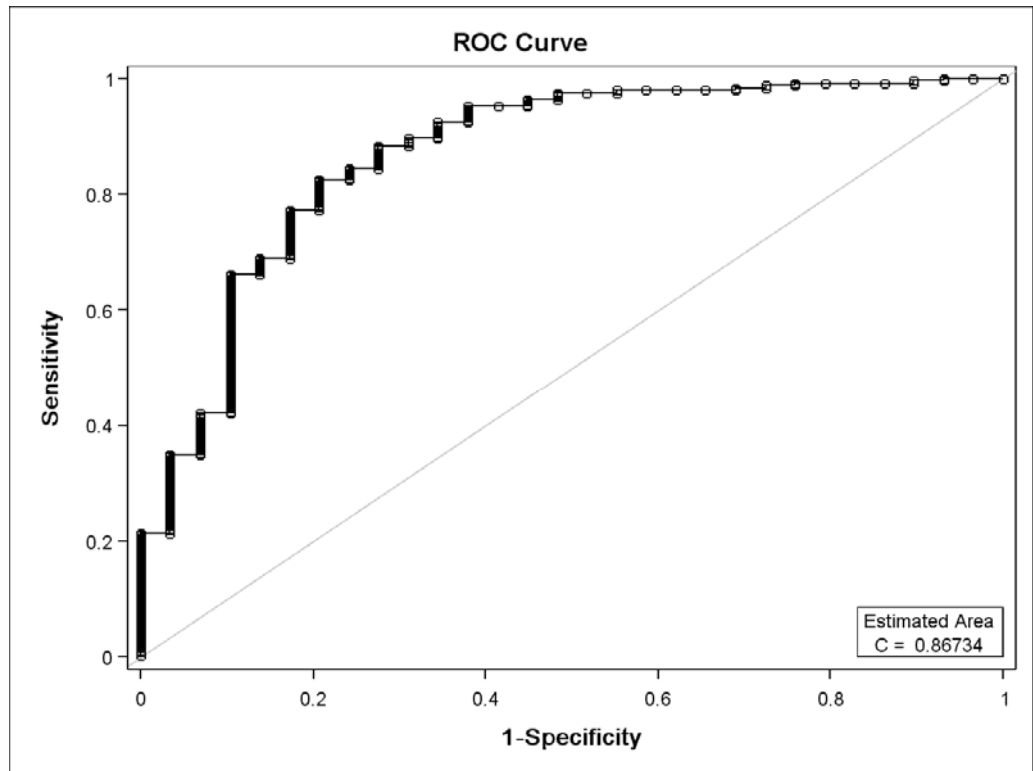
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	61.3840	5	<.0001
Score	64.4676	5	<.0001
Wald	38.5214	5	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
KONK_BALAS	1	17.5703	<.0001
METINIU_VIDURKIS	1	3.5833	0.0584
MIESTO_GRUPE	2	2.2977	0.3170
PAZANGUMO_INDEKSAS	1	2.3520	0.1251

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	-4.5281	2.4105	3.5286	0.0603	0.011
KONK_BALAS		1	0.4972	0.1186	17.5703	<.0001	1.644
METINIU_VIDURKIS		1	0.6136	0.3242	3.5833	0.0584	1.847
MIESTO_GRUPE	Apskričių centrai	1	-0.5308	0.3574	2.2059	0.1375	0.588
MIESTO_GRUPE	Kiti miestai	1	-0.0284	0.3459	0.0067	0.9347	0.972
PAZANGUMO_INDEKSAS		1	-0.5940	0.3873	2.3520	0.1251	0.552

Odds Ratio Estimates			
Effect		Point Estimate	95% Wald Confidence Limits
KONK_BALAS		1.644	1.303 2.074
METINIU_VIDURKIS		1.847	0.979 3.487
MIESTO_GRUPE	Apskričių centrai vs Savivaldybių centrai	0.336	0.078 1.454
MIESTO_GRUPE	Kiti miestai vs Savivaldybių centrai	0.556	0.133 2.327
PAZANGUMO_INDEKSAS		0.552	0.258 1.180

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.5	Somers' D	0.735
Percent Discordant	13.0	Gamma	0.738
Percent Tied	0.5	Tau-a	0.136
Pairs	7308	c	0.867



Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.1731	8	0.2531

Classification Table									
Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.500	247	8	21	5	90.7	98.0	27.6	7.8	38.5