VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS STATISTIKOS KATEDRA

Donata Vaičiūnaitė (parašas)

LIETUVOS MOKINIŲ RAŠTINGUMO ANALIZĖ, NAUDOJANT TARPTAUTINIO TYRIMO PISA 2009 M. DUOMENIS

Magistro baigiamasis darbas

2012, Vilnius

Darbo vadovas:

Doc., dr. Rimantas Eidukevičius (parašas)

Recenzentas:

Doc., dr. Vytautas Kazakevičius (parašas)

Registracijos Nr.: ....

Darbo gynimo data: ....

TURINYS

Tekste naudojamos santrumpos:

HLM – Hierarcinis tiesinis modeliavimas

IKT – Informacinės ir komunikacinės technologijos.

OECD – Ekonominio bendradarbiavimo ir plėtros organizacija (Organisation For Economic Co-operation and Development).

PISA – Tarptautunis penkiolikmečių tyrimas (Programme For International Student Assessment).

ĮVADAS

Ekonominio bendradarbiavimo ir plėtros organizacijai (OECD) priklausančios šalys ir šalys partnerės, kas trejus metus, vykdo bendrą tarptautinį penkiolikmečių tyrimą PISA (Programme For International Student Assessment). Lietuva šiame tyrime dalyvauja nuo 2006 metų. PISA tiria mokinių gebėjimą pritaikyti savo žinias ir įgūdžius, samprotauti, analizuoti ir spręsti problemas. Taip pat PISA siekia nustatyti ryšį tarp mokymo(si) aplinkų ir mokinių rezultatų, įvertinti mokinių kompiuterinio raštingumo gebėjimus. PISA tyrimu vertinami mokinių skaitymo gebėjimai, matematinis ir gamtamokslis raštingumas bei įtraukiamas kompiuterinis raštingumas. „PISA tyrimo specifinis tikslas yra pateikti patikimą informaciją, kuria remiantis būtų galima pagerinti mokymo programas ir galimybes“ [5]. Švietimo politikams ir sprendimų priėmėjams yra svarbu gauti patikimą informaciją apie Lietuvos penkiolikmečių pasiekimus, todėl yra didelis susidomėjimas PISA tyrimu ir jo rezultatais.

Lietuvos švietimo specialistus domina tarptautinio tyrimo PISA rezultatai, yra atliekamos antrinės analizės. Tačiau norint atsižvelgti į veiksnius, atsirandančius dėl mokinių pasiskirstymo po mokyklas, pavyzdžiui, mokinio socialinė ir ekonominė padėtis gali turėti įtakos pasirenkant lankomos mokyklos tipą, bei dėl duomenų hierarchiškumo, tiesinės regresijos modeliai gali pateikti klaidingą informaciją [7]. Todėl atsirado poreikis sudėtingesnei duomenų analizei, t. y. hierarchinei regresinei analizei.

Šiame darbe analizuojami PISA 2009 metų tyrimo duomenis, nagrinėjama Lietuvos penkiolikmečių mokinių matematinio raštingumo rezultatų priklausomybė nuo mokymo(si) aplinkų veiksnių. Mokinių raštingumą lemiantys veiksniai yra susiję, todėl duomenų analizė turi būti atlikta kompleksiškai. **Magistrinio darbo tikslas** – pritaikius hierarchinę regresiją rasti mokyklos ir namų aplinkų veiksnius įtakojančius Lietuvos penkiolikmečių mokinių matematinį raštingumą. Šiam tikslui pasiekti išsikelėme tokius **uždavinius**:

1. išnagrinėti tarptautinio tyrimo PISA naudojamą metodologiją;
2. išnagrinėti hierarchinių tiesinių modelių teoriją ir taikymą;
3. išnagrinėti programos SAS procedūrą MIXED;
4. atlikti mokinių, jų tėvų ir mokyklų veiksnių, galinčius įtakoti mokinių matematinį raštingumą, pirminę analizę;
5. atlikti hierarchinę regresinę analizę ir nustatyti veiksnius lemiančius penkiolikmečių matematinį raštingumą;
6. sudaryti penkiolikmečių matematinio raštingumo rezultatų prognostinį modelį.

Šį darbą sudaro dvi dalys. Pirmojoje dalyje ,siekiant supažindinti su PISA teikiamų duomenų specifika, aprašoma PISA tyrimo metodologija. Kadangi atliekant duomenų analizę reikia atsižvelgti į specifinę duomenų struktūrą bei, kad mokinių raštingumo rezultatai išreikšti galimomis reikšmėmis. Surašant PISA metodologiją remtasi leidiniu „PISA Data Analysis Manual“ [7]. Taip pat išdėstoma tiesinio hierarchinio modeliavimo teorija, kuria bus remiamasi atliekant duomenų analizę. V. Čekanavičiaus ir G. Murausko knygoje „Statistika ir jos taikymai III“ [2] nagrinėjami hierarchiniai tiesiniai modeliai. Antrojoje dalyje pateikiama pirminė duomenų analizė, norint apžvelgti kai kurių veiksnių įtaką mokinių matematinio raštingumo rezultatams. Remiantis SAS help‘u ir R. C. Littell, G. A. Milliken, W. W. Stroup, R. D. Wolfinger, ir O. Schabenberger knyga „SAS® for Mixed Models“ [9] pristatoma SAS procedūra MIXED. Demonstruojama, kaip naudojant hierarchinius tiesinius metodus kuriamas tinkamiausias modelis penkiolikmečių matematinio raštingumo ir jo aplinkos veiksnių sąsajų radimui bei pateikiama galutinė Lietuvos penkiolikmečių matematinio raštingumo prognozavimo lygtis. Darbo pabaigoje pateikiamos atliktos duomenų analizės išvados.

1. TEORINĖ DALIS

PISA imties išrinkimui naudoja sudėtingą metodologiją, kuri apsunkina duomenų analizę, tiek programoje SAS, tiek kitose statistinėse programose (SPSS, R). Pirma PISA nerenka paprastos atsitiktinės imties iš baigtinio penkiolikmečių sąrašo. Yra naudojama dviejų pakopų imties išrinkimo schema. Dviejų pakopų ėmimas – tai toks tikimybinis ėmimas, kai iš visos populiacijos renkami pirminiai imties elementai[1] (pirmos pakopos imtis), tuomet iš pirmosios pakopos imties elementų išrenkami antriniai imties elementai, t. y. renkama antros pakopos imtis [3]. PISA tyrime naudojant dviejų pakopų ėmimą pirmiausia yra išrenkamos mokyklos, o tada, iš išrinktų mokyklų, išrenkami mokiniai. Mokyklos išrenkamos naudojant sistemį imties išrinkimo būdą (plačiau 1.1 skyrelyje), o mokiniai iš mokyklų išrenkami naudojant paprastą atsitiktinį ėmimą.

Antra tyrimas PISA penkiolikmečių raštingumo balus išreiškia per vadinamąsias *galimas reikšmes*[2]. Kadangi pateikiamos penkios galimos reikšmės, todėl duomenų analizė, su galimomis reikšmėmis, turi būti atliekama penkis kartus, o rezultatai apibendrinami. Plačiau galimų reikšmių skaičiavimą ir jų naudojimą aptarsime 1.3 skyrelyje.

Kadangi PISA imties išrinkimui naudoja dviejų pakopų ėmimą, atsiranda duomenų hierarchiškumas (plačiau aprašytas 1.4 skyrelyje). Duomenų analizė bus atliekama taikant tiesinius hierarchinius modelius[3], dar vadinamais atsitiktinių koeficientų modeliavimu[4]. Tiesinis hierarchinis modeliavimas yra plačiai taikomas edukologijoje, psichologijoje, medicinoje ir kt.

1.1. Imties svoriai

Kiekvienam išrinktam mokiniui ir kiekvienai išrinktai mokyklai yra priskiriamas svoris, kadangi mokiniai, mokyklos turi skirtingas išrinkimo tikimybes.

Paprastoji atsitiktinė imtis – „tai tokia n skirtingų elementų imtis iš N dydžio baigtinės populiacijos, kai bet kuris n skirtingų elementų rinkinys turi vienodą tikimybę būtį išrinktas“ [3].

Paprastos atsitiktinės imties išrinkimo schemų yra daug:

- Paprastosios atsitiktinės imties išrinkimas pagal apibrėžimą,
- Nuoseklus išrinkimas,

---

[1] Primary Sampling Units
[2] Plausible Values
[3] Hierarchical Linear Models
[4] Random Coefficients Modeling

- Naudojant skaičių lentelę,
- Imties išrinkimas, naudojant atsitiktinių skaičių generatorių,
- Imties išrinkimas, taikant atsitiktinį išrykiavimą [3].

Iš N dydžio populiacijos išrenkame n dydžio paprastąją atsitiktinę negrąžintinę imtį, tuomet tikimybė i-am elementui patekti į imtį yra:

$$p_i = \frac{n}{N}. \tag{1.1}$$

Dažniausiai svoris yra atvirkštinis dydis tikimybei būti išrinktam, t. y. elemento svoris yra populiacijos elementų skaičius, kuriam jis atstovauja būdamas išrinktas į imtį [3]. Svoris apskaičiuojamas pagal formulę:

$$w_i = \frac{1}{p_i} = \frac{N}{n}. \tag{1.2}$$

Kadangi kiekvienas stebėjimas turi vienodas išrinkimo tikimybes, todėl kiekvieno stebėjimo svoriai bus vienodi ir išrinktų elementų svorių suma bus lygi populiacijos dydžiui, t.y. N:

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} \frac{N}{n} = N. \tag{1.3}$$

Jeigu elementai turi vienodus svorius, tai svoris neįtakoja populiacijos parametro vertinimo. Pavyzdžiui, norime apskaičiuoti charakteristikos X vidurkį pagal formulę (w_i yra konstanta, todėl ją galime iškelti):

$$\mu_{(X)} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_i \sum_{i=1}^{n} x_i}{w_i \sum_{i=1}^{n} 1} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{1.4}$$

Tačiau paprasta atsitiktinė imtis labai retai naudojama švietimo tyrimuose nes:

- Ji per brangi. Tikėtina, kad išrinkti mokiniai mokysis daugybėje skirtingų mokyklų. Todėl reikės apmokyti daugybė testų administratorių, kompensuoti didžiules keliavimo išlaidas ir pan.;
- Tai nepraktiška. Reikės kontaktuoti su daugybe mokyklų;
- Bus neįmanoma statistiškai susieti mokinio kintamųjų su mokyklos, klasės ar mokytojo kintamaisiais. Iš mokyklos išrinkto vieno ar dviejų mokinių statistiniai ryšiai nebus stabilūs.

Todėl švietimo tyrimuose mokinių imtis renkama dviejų pakopų ėmimu. Pirmiausia išrenkama mokyklų imtis, tada paprastu atsitiktiniu ėmimu iš išrinktų mokyklų išrenkami mokiniai ar klasės. Paprastai PISA iš mokyklos išrenka 35 penkiolikmečius. Jei išrinktoje mokykloje mokosi mažiau nei 35 penkiolikmečiai, tada į imtį įtraukiami visi penkiolikmečiai toje mokykloje. Tokia išrinkimo schema turi įtakos stebėjimo svorio skaičiavimui. Paprasto atsitiktinio ėmimo atveju *i*-ios mokyklos išrinkimo tikimybė apskaičiuojama pagal formulę:

$$p_i = \frac{n_{SC}}{N_{SC}}, \tag{1.5}$$

kur n_{SC} – išrinktų į imtį mokyklų skaičius, N_{SC} – mokyklų skaičius populiacijoje. O *j*-jo mokinio išrinkimo tikimybė:

$$p_j = \frac{n_i}{N_i}, \tag{1.6}$$

kur n_i – mokykloje *i* išrinktų į imtį mokinių skaičius, N_i – mokykloje *i* esančių penkiolikmečių skaičius. Tuomet tikimybė *j*-am mokiniui iš *i*-ios mokyklos būti išrinktam yra lygi:

$$p_{ij} = p_i p_j = \frac{n_{SC} n_i}{N_{SC} N_i}. \tag{1.7}$$

Mokyklos svoris w_i, mokinių svoris w_j ir galutinis mokinio svoris w_{ij}, naudojamas duomenų analizei, apskaičiuojami pagal formules:

$$w_i = \frac{1}{p_i}, \tag{1.8}$$

$$w_j = \frac{1}{p_j}, \tag{1.9}$$

$$w_{ij} = \frac{1}{p_{ij}}. \tag{1.10}$$

Tarptautiniai švietimo tyrimai labiau orientuojasi į mokinių imtį, o ne į mokyklų (mokyklų imtis padeda išrinkti mokinių imtį). Paprastos atsitiktinės imties išrinkimo schema netinkama mokyklų imčiai išrinkti, nes didelės ir mažos mokyklos turės vienodas išrinkimo tikimybes (1.5). Todėl mokyklos renkamos pagal tikimybes proporcingas jų dydžiams[5] (TPD). Šiuo atveju didesnės mokyklos turi didesnę tikimybę būti išrinktos, negu mažos mokyklos, bet mokiniai didelėse mokyklose turi mažesnę tikimybę būti išrinkti, negu mažose mokyklose. Tuomet tikimybė, kad mokykla bus išrinkta, yra lygi mokyklos dydžiui padaugintam iš išrinktų mokyklų skaičiaus ir padalintam iš mokinių skaičiaus visoje populiacijoje:

$$p_i = \frac{N_i n_{SC}}{N}. \tag{1.11}$$

[5] probabilities proportional to their size

Mokinio išrinkimo tikimybės mokykloje ir jo svorio skaičiavimo formulės išlieka tokios pat (1.6 ir 1.9).

Mokyklos išrenkamos naudojant sisteminį imties išrinkimo būdą: pirmiausiai mokyklos išrikiuojamos pagal dydį ir apskaičiuojamas ėmimo žingsnis:

$$Int = \frac{N}{n_{SC}}, \tag{1.12}$$

t.y. mokinių skaičius populiacijoje dalinamas iš mokyklų skaičiaus imtyje. Tuomet atsitiktinai generuojamas skaičius q intervale $[0;1)$ iš tolydaus skirstinio (t. y. $q \sim U[0;1)$). Gautas skaičius q dauginamas iš apskaičiuoto ėmimo žingsnio $q \cdot Int = q_0$ ir mokykla su atitinkamu skaičiu mokinių yra išrenkama. Tuomet prie q_0 pridedamas ėmimo žingsnis Int ir mokykla, kurioje yra atitinkamas mokinių skaičius, išrenkama. Procedūra kartojama tol kol išrenkama norimo dydžio imtis, t. y. $q_0, q_0 + Int, q_0 + 2 \cdot Int, \ldots, q_0 + (n_{SC} - 1) \cdot Int$. Naudojant šią sisteminę procedūrą į imtį pateka ir mažos ir didelės mokyklos.

Tokia išrinkimo schema turėtu užtikrinti, kad visi mokiniai turėtų vienodas išrinkimo tikimybes ir svorius, tačiau taip nėra. Mokinių svorių skirtumus įtakoja keli faktoriai:

- Į imtį įtraukiama daugiau mokinių negu remiantis proporcingu populiacijos ėmimu.
- Neatitikimas tarp realaus penkiolikmečių skaičiaus mokykloje ir mokyklos pateikto skaičiaus sukelia galutinio mokinio svorio kintamumą. Pavyzdžiui, informacija gali būti dviejų metų senumo.
- Kai kurios išrinktos mokyklos ar mokiniai gali atsisakyti bendradarbiauti, todėl reikia koreguoti svorius. Pavyzdžiui, jeigu iš 35 išrinktų mokinių tik 25 sutinka bendradarbiauti, tuomet mokinių svoris dauginamas iš santykio $\frac{35}{25}$. Tokia procedūra taikoma ir mokyklų svoriams koreguoti.

1.2. Rinkčių naudojimas standartinės paklaidos skaičiavimams

Vienas iš skirtumų tarp paprastos atsitiktinės ir dviejų pakopų imčių yra tai, kad dviejų pakopų imties išrinkimo atveju išrinkti mokiniai lankantys tą pačią mokyklą negali būti laikomi nepriklausomais stebėjimais. Taip yra todėl, kad toje pačioje mokykloje besimokantys mokiniai turi bendresnes charakteristikas, t.y. tie patys mokyklos resursai, mokytojai, mokymo planai ir t.t., negu mokiniai iš skirtingų mokyklų. Taip pat didesni skirtumai tarp bendrojo ugdymo ir profesinių, kaimo ir miesto mokyklų, negu tarp dviejų profesinių ar kaimo mokyklų.

Bet kokio populiacijos parametro įvertinio dispersija bus didesnė dviejų pakopų imčiai negu tokio pat didumo paprastai atsitiktinei imčiai. Įvertinio dispersijos padidėjimas atsiranda dėl to, kad mokyklose, pirminiuose imties elementuose, renkama mokinių imtis [3].

Kaip jau minėjome PISA tyrimo populiacijos ir išrinkimo modelio prielaidos:

- mokyklos yra skirtingų dydžių;
- imtis negrąžintinė;
- mokyklos yra atrenkamos sistematiškai, atsižvelgiant į jų dydį;
- į išrinkimo modelį įtraukiami sluoksniniai kintamieji.

Švietimo tyrimui išrinkimo schemos gali būti labai sudėtingos, (gali būti sunku įvertinti netgi imties vidurkį), todėl dispersijai vertinti yra naudojami *rinkčių metodai* [6]. Šie metodai iš esamos imties generuoja naujas rinkčių imtis. Tuomet kiekvienoje rinktyje įvertinama dominanti statistika kuri lyginama su visos imties įvertiniu ir taip gaunamas imties dispersijos įvertinys. Šiuolaikinė skaičiavimo technika leidžia skaičiuoti sudetingų populiacijos įvertinių dispersijas. Sudėtingiems įvertiniams vertinti yra naudojami keli metodai: visrakčio (Jackknife), savirankos (Bootstrap), suderintų kartotinių rinkčių (balanced repeated replications) ir kt. PISA naudoja suderintų kartotinių rinkčių metodą.

Tarkime, kad mokyklų sąrašas yra padalintas į dvi dalis: kaimo ir miesto mokyklas ir mokyklos sluoksniuose yra išrikiuotos didėjimo tvarka. Kiekviename sluoksnyje naudojant sistematinę imties iįrinkimo būdą (atsižvelgiant į mokyklos dydį) išrenkama vienodo dydžio imtis ir panašios mokyklos (panašaus dydžio) iš skirtingų sluoksnių yra suporuojamos ir suskirstomos į pseudo-sluoksnius.

Aprašykime suderintų kartotinių rinkčių metodo algoritmą naudojamą rinkčių išrinkimui: kiekviename pseudo sluoksnyje atsitiktinai pasirienkamas vienas objektas ir jam priskiriamas svors 0 ir padvigubinamas kito pseudo sluoksnyje esančio objekto svoris (2.1 lentelė).

Norint išvengti ilgų skaičiavimų yra naudojamos Adamaro kvadratinės matricos [3]. Tuomet generuojamų rinkčių skaičius yra mažiausias kartotinis 4 skaičius ir yra didesnis arba lygus pseudo–sluoksnių skaičiui.

Naudojant rinkčių metodus parametro θ dispersijai vertinti naudojama formulę:

$$\hat{\boldsymbol{D}}(\theta) = \frac{1}{G}\sum_{i=1}^{G}\left(\theta_{(i)} - \theta\right)^2, \tag{2.18}$$

---

[6] Replication method

kur G – rinkčių skaičius, $\hat{\theta}_{(i)}$ – i-osios rinkties parametro θ įvertinys ir $\hat{\theta}$ – imties parametro θ įvertinys.

Pseudo-sluoksnis	Stebėjimas	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
1	**1**	2	0	0	2	0	0	0	2	2	2	0	2
1	**2**	0	2	2	0	2	2	2	0	0	0	2	0
2	**3**	2	2	0	0	2	0	0	0	2	2	2	0
2	**4**	0	0	2	2	0	2	2	2	0	0	0	2
3	**5**	2	0	2	0	0	2	0	0	0	2	2	2
3	**6**	0	2	0	2	2	0	2	2	2	0	0	0
4	**7**	2	2	0	2	0	0	2	0	0	0	2	2
4	**8**	0	0	2	0	2	2	0	2	2	2	0	0
5	**9**	2	2	2	0	2	0	0	2	0	0	0	2
5	**10**	0	0	0	2	0	2	2	0	2	2	2	0
6	**11**	2	2	2	2	0	2	0	0	2	0	0	0
6	**12**	0	0	0	0	2	0	2	2	0	2	2	2
7	**13**	2	0	2	2	2	0	2	0	0	2	0	0
7	**14**	0	2	0	0	0	2	0	2	2	0	2	2
8	**15**	2	0	0	2	2	2	0	2	0	0	2	0
8	**16**	0	2	2	0	0	0	2	0	2	2	0	2
9	**17**	2	0	0	0	2	2	2	0	2	0	0	2
9	**18**	0	2	2	2	0	0	0	2	0	2	2	0
10	**19**	2	2	0	0	0	2	2	2	0	2	0	0
10	**20**	0	0	2	2	2	0	0	0	2	0	2	2

2.1 lentelė. Pakartojimai naudojant balansinį kartotinį pakartojimų metodą

Naudojant šį rinkčių metodą kiekviena pakartojimų imtis panaudoja tik pusė galimos informacijos. Likusių stebėjimų skaičius gali būti per mažas įvertinti populiacijos statistiką. Fay pasiūlė suderintų kartotinių rinkčių metodo variantą padedantį to išvengti [7]. Užuot dauginus mokyklos svorius iš 0 ir 2, Fay pasiūlė mokyklos svorius dauginti iš mažinimo faktoriaus k, kuris yra arp 0 ir 1, ir iš didinimo faktoriaus $2-k$. PISA naudoja suderintą kartotinių rinkčių metodą su Fay pasiūlytu būdu ir $k = 0.5$ (2.2 pav.).

Naudojant suderintų kartotinių rinkčių metodą su Fay siūlymu imties dispersijos įvertinys apskaičiuojamas:

$$\hat{D}(\hat{\theta}) = \frac{1}{G(1-k)^2}\sum_{i=1}^{G}\left(\hat{\theta}_{(i)} - \hat{\theta}\right)^2. \qquad (2.19)$$

kur G – rinkčių skaičius, k – mažinimo faktorius, $\hat{\theta}_{(i)}$ – i-osios rinkties parametro θ įvertinys ir $\hat{\theta}$ – imties parametro θ įvertinys.

PISA generuoja 80 rinkčių svorių, tada imties dispersijos įvertiniui apskaičiuoti naudojama tokia formulė:

$$\hat{D}(\theta) = \frac{1}{G(1-k)^2}\sum_{i=1}^{G}\left(\theta_{(i)}-\theta\right)^2 = \frac{1}{80(1-0.5)^2}\sum_{i=1}^{G}\left(\theta_{(i)}-\theta\right)^2 = \frac{1}{20}\sum_{i=1}^{G}\left(\theta_{(i)}-\theta\right)^2. \tag{2.20}$$

Pseudo-sluoksnis	Stebėjimas	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
1	1	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5
1	2	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5	0.5
2	3	1.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5
2	4	0.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5
3	5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5
3	6	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5
4	7	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5
4	8	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5
5	9	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5
5	10	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5
6	11	1.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5
6	12	0.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5
7	13	1.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5
7	14	0.5	1.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5
8	15	1.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5
8	16	0.5	1.5	1.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5
9	17	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5
9	18	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5
10	19	1.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5
10	20	0.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5

2.2 lentelė. Fay metodo pakartojimai

1.3 Galimos reikšmės[7]

Kad tyrimas būtų patikimas, galutinį testą turi sudaryti daugybė įvairių užduočių, kurios padengtų plačią sritį. Tačiau kiekvienam atrinktam mokiniui vertinti naudoti visą didžiulį klausimyną yra neišmintinga, nes:

- esant ilgam testavimo laikui mokinio rezultatus pradeda įtakoti nuovargis;

mokyklų direktoriai gali atsisakyti išleisti mokinius labai ilgam testavimo laikui.

Tai gali sumažinti bendradarbiaujančių mokyklų skaičių, todėl mokiniai gauna atlikti tik dalį užduočių. Lentelė 3.1 vaizduoja PISA 2006 testo planą. Užduotys suskirstytos į trylika sluoksniu, kur S1–S7 žymi gamtamokslio sluoksniai, M1–M4 matematikos sluoksniai, R1 ir R2 skaitymo sluoksniai. Iš viso paruošiama 13 sąsiuvinių ir kiekvieną sąsiuvinį sudaro 4 blokai, t.y. kiekvieną sąsiuvinį sudaro 4 sluoksniai. Pavyzdžiui, 3 sąsiuvinį sudaro 3 ir 4 gamtamokslio ir 4 ir 1 matematikos blokai. Esant tokiam planui kiekvienas blokas atsiranda keturis kartus, po vieną kartą kiekvienoje pozicijoje. Taip pat kiekviena blokų pora atsiranda tik vieną kartą.

[7] Plausible values

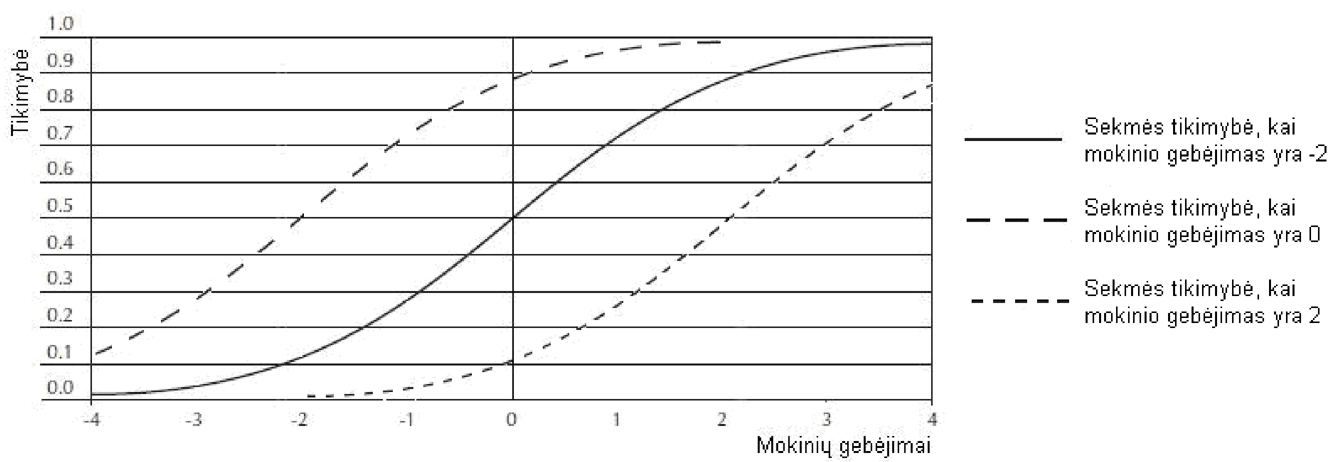
	1 blokas	2 blokas	3 blokas	4 blokas
1 sąsiuvinis	S1	S2	S4	S7
2 sąsiuvinis	S2	S3	M3	R1
3 sąsiuvinis	S3	S4	M4	M1
4 sąsiuvinis	S4	M3	S5	M2
5 sąsiuvinis	S5	S6	S7	S3
6 sąsiuvinis	S6	R2	R1	S4
7 sąsiuvinis	S7	R1	M2	M4
8 sąsiuvinis	M1	M2	S2	S6
9 sąsiuvinis	M2	S1	S3	R2
10 sąsiuvinis	M3	M4	S6	S1
11 sąsiuvinis	M4	S5	R2	S2
12 sąsiuvinis	R1	M1	S1	S5
13 sąsiuvinis	R2	S7	M1	M3

Lentelė 3.1 PISA 2006 testo modelis

1.3.1 Modernioji testų teorija ir Rasch modelis

Klasikinėje testų teorijoje mokinių gebėjimai vertinami pagal jų surinktus už testą taškus. Modernioji testų teorija pateikia tikimybes, kaip mokiniai spręs atitinkamas užduotis. Todėl įmanoma charakterizuoti mokinį nepriklausomai nuo to, kurias užduotis jis sprendė (t. y. nepriklausomai nuo testo) ir charakterizuoti užduotį, nepriklausomai nuo ją išsprendusių mokinių (t. y. nepriklausomai nuo grupės). Taigi modernioji testų teorija remdamasi tam tikromis latentinėmis charakteristikomis[8] apskaičiuoja tikimybę atlikti konkretų testą [6]. Santykis tarp mokinio charakteristikų ir užduoties atlikimo galime pavaizduoti užduoties charakteristikos kreive (3.1 pav.), čia sekmės tikimybė – tikimybė teisingai atsakyti į klausimą. Paprasčiausias moderniosios teorijos modelis, kurį naudoja ir PISA, yra Rasch vieno parametro modelis.

3.1 pav. Užduoties charakteristikos kreivė



[8] latent traits

Tarkime turime tokius testo rezultatus, pavaizduotus 3.2 lentelėje (šis pavyzdys paimtas iš Chong Ho Yu [1]), kur eilutėse pateikiami mokinių atsakymai į atitinkamus uždavinius ir preliminarūs mokinių gebėjimai (PMG), čia PMG apskaičiuojamas kaip teisingai atliktų užduočių skaičiaus ir visų užduočių skaičiaus santykis. Stulpeliuose pateikiama informacija apie užduotis ir preliminarūs užduočių sudetingumai (PUS), čia PUS apskaičiuojamas iš vieneto atimant teisingai atliktų užduočių skaičiaus ir visų užduočių skaičiaus santykį. Iš 3.2 lentelės matome, kad į pirmą užduotį teisingai atliko tik vienas mokinys, todėl preliminarus pirmos užduoties sudėtingumas yra 0,8, t.y. 80 % mokinių neteisingai atliko pirmą užduotį, analogiškai 60 % mokinių klaidingai atliko antrą užduotį ir t.t.

	1 užduotis	2 užduotis	3 užduotis	4 užduotis	5 užduotis	PMG
1 mokinys	1	1	1	1	1	1
2 mokinys	0	1	1	1	1	0.8
3 mokinys	0	0	1	1	1	0.6
4 mokinys	0	0	0	1	1	0.4
5 mokinys	0	0	0	0	1	0.2
PUS	0.8	0.6	0.4	0.2	0	

3.2 lentelė. Preliminarūs mokinių gebėjimai ir užduočių sunkumai

Tegul $X_{ij} \in \{0,1\}$ kintamasis nurodantis *i*–jo mokinio ($i = 1,\ldots,m$) atsakymą į *j*–ją užduotį ($j = 1,\ldots,n$), kur 0 – klaidingas, o 1 – teisingas atsakymas. Užduoties sudėtingumas matuojamas parametru δ, tuomet *j*–joios užduoties sudėtingumas žymimas δ_j. Latentinis kintamasis θ_i yra *i*-jo mokinio gebėjimų lygis. Mokinys į užduotį gali atsakyti teisingai arba klaidingai, todėl yra du galimi atsakymai. Turint preliminarią informaciją galime apskaičiuoti tikimybę, kad mokinys su θ_i gebėjimais teisingai atsakys į δ_j sudėtingumo uždavinį. Teisingo atsakymo tikimybei apskaičiuoti naudosime Rasch modelį:

$$p = P(X_{ij} = 1 \mid \theta_i, \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}. \tag{3.1}$$

Klaidingo atsakymo tikimybė , t.y. tikimybė klaidingai atlikti užduotį apskaičiuojama pagal formulę:

$$1 - p = P(X_{ij} = 0 \mid \theta_i, \delta_j) = \frac{1}{1 + \exp(\theta_i - \delta_j)}. \tag{3.2}$$

Lengvai galima parodyti, kad sėkmės ir klaidos tikimybių suma lygi 1:

$$P(X_{ij}=1|\theta_i,\delta_j)+P(X_{ij}=0|\theta_i,\delta_j)=\frac{\exp(\theta_i-\delta_j)}{1+\exp(\theta_i-\delta_j)}+\frac{1}{1+\exp(\theta_i-\delta_j)}=1. \tag{3.3}$$

Pasinaudokime (3.1) formule ir apskaičiuokime teisingo atsakymo tikimybes. Pavyzdžiui, tikimybė, kad pirmas mokinys, kurio preliminarūs gebėjimai yra 1, teisingai išspręs pirmąją užduotį, kurios preliminarus sudėtungumas 0.8, yra:

$$p=\frac{\exp(1-0.8)}{1+\exp(1-0.8)}=0.55.$$

Sėkmės tikimybė lygi 0,5, kai mokinio gebėjimai yra lygūs uždavinių sudėtingumui (3.3 lentelė).

	1 klausimas	2 klausimas	3 klausimas	4 klausimas	5 klausimas	PMG
1 mokinys	0.55	0.60	0.65	0.69	0.73	1
2 mokinys	0.50	0.55	0.60	0.65	0.69	0.8
3 mokinys	0.45	0.50	0.55	0.60	0.65	0.6
4 mokinys	0.40	0.45	0.50	0.55	0.60	0.4
5 mokinys	0.35	0.40	0.45	0.50	0.55	0.2
PKS	0.8	0.6	0.4	0.2	0	

3.3 lentelė. Teisingo atsakymo tikimybės

Norėdami mokinių gebėjimus ir klausimų sudėtingumus pažymėti vientisoje matavimų skalėje, vadinamojeje logit[9] skalėje, skaičiuojamas šansų logaritmas. Mokinių gebėjimai logit skalėje apskaičiuojami pagal formulę:

$$Logit=Log\left(\frac{p}{1-p}\right), \tag{3.4}$$

o užduočių:

$$Logit=Log\left(\frac{1-p}{p}\right), \tag{3.5}$$

kur p skaičiuojamas pagal (3.1) formulę.

Galime parodyti, kad skirtumas $\theta_i-\delta_j$, kur $i=1,\ldots,m$, $j=1,\ldots,n$, parodo i–ojo mokinio tikimybę sėkmingai atsakyti j–ąją užduotį.

$$\log\left(\frac{p}{1-p}\right)=\theta_i-\delta_j \tag{3.6}$$

Įrodymas:

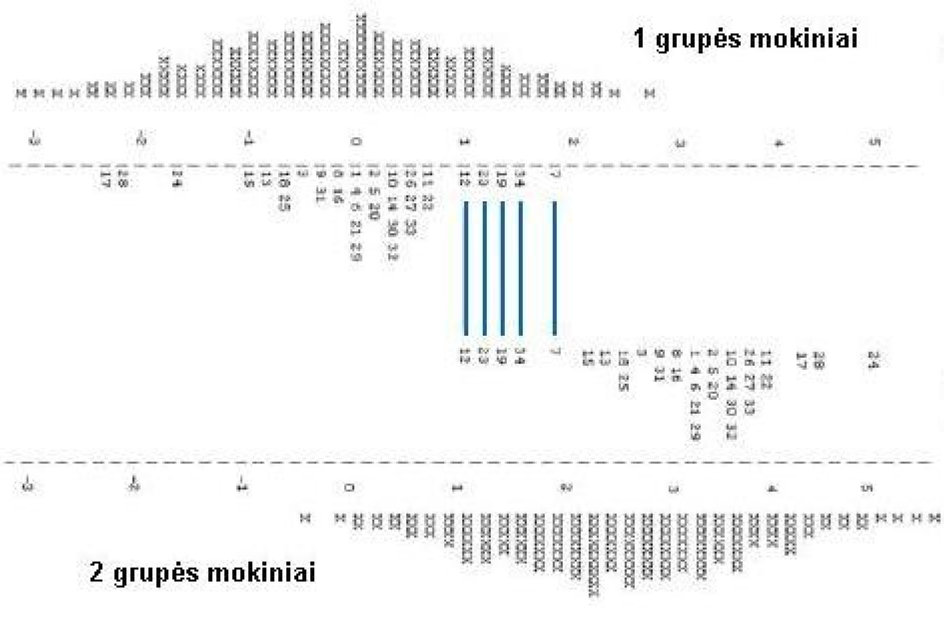
[9] log odds unit

$$\log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p) = \log\left(\frac{\exp(\theta-\delta)}{1+\exp(\theta-\delta)}\right) - \log\left(\frac{1}{1+\exp(\theta-\delta)}\right) =$$

$$= \log(\exp(\theta-\delta)) - \log(1+\exp(\theta-\delta)) - \log(1) + \log(1+\exp(\theta-\delta)) = \log(\exp(\theta-\delta)) = \theta - \delta$$

Kaip buvo minėta skyrelio pradžioje PISA testavimui pateikia trylika sąsiuvinių, sudarytų iš keturių užduočių blokų ir du sąsiuviniai turi po bendrą užduočių bloką. Tuomet naudojantis vienodais užduočių blokais galime susieti skirtingų sąsiuvinių užduotis. Paveikslėlis 3.2 vaizduoja tokį klausimų susiejimą, t.y. 7, 34, 19, 23, ir 12 užduotys buvo bendri pirmai ir antrai mokinių grupėms. Užduočių sunkumai užrašyti virš linijos (žiūrint į pirmos grupės grafiką), užduočių numeriai – po linija, mokinių rezultatų paskirstymas pavaizduotas virš linijos (žymima x).

3.2 pav. Užduočių – atsakymų žemėlapiai ir užduočių susiejimas



Susiejus užduočių sudėtingumus, galime apskaičiuoti mokinių rezultatus naudodami bendrą skalę. Rasch modelyje tikimybė atsakyti į užduotį nepriklauso nuo atsakymų į kitas užduotis, taip pat užduotys laikomos nepriklausomomis. Skaičiavimams naudojamas vieno parametro logistinis modelis (3.1). Naudojamas vieno parametro modelis nes užduoties charakteristikos kreivė (3.1 pav.) priklauso tik nuo klausimo sudėtingumo.

Tarkime turime testą sudarytą iš penkių užduočių, kurių sunkumai: -0,9, -0,4, 0,1, 0,7, 1. Į pirmas tris (lengvas) užduotis mokinys, kurio gebėjimai yra 0, atsakė teisingai, o į likusias (sunkesnias) atsakė klaidingai. Naudodamiesi formulėmis (3.1) ir (3.2) apskaičiuokime tikimybę atlikti tokį testą, kai atsakymai yra (1, 1, 1, 0, 0), čia 0 – klaidingas, o 1 – teisingas atsakymas.

$$P(X_{11} = 1 \mid \theta, \delta_1) = P(X_{11} = 1 \mid 0, -0.9) = \frac{\exp(0 - (-0,9))}{1 + \exp(0 - (-0,9))} = 0.71$$

$$P(X_{14} = 0 \mid \theta, \delta_4) = P(X_{14} = 0 \mid 0, 0.7) = \frac{1}{1 + \exp(0 - 0,7)} = 0.67$$

Kadangi užduotys laikomos nepriklausomomis, tad mokinio, kurio gebėjimai 0, į testą atsakyti (1, 1, 1, 0, 0) tikimybė yra $0.71 * 0.60 * 0.48 * 0.67 * 0.73 = 0.10$. Taigi mokinys, kurio gebėjimai $\theta = 0$ turi 10 šansų iš 100 teisingai atsakyti į 1, 2 ir 3 uždavinius ir klaidingai atsakyti į 4 ir 5 uždavinius.

1.3.2 Galimų reikšmių skaičiavimas

Galimos reikšmės reprezentuoja mokinio gebėjimų apimtį, kurią, atsižvelgant į jo atsakymus į užduotis, jis gali turėti. Galimos reikšmės generuojamos pagal taip vadinamą aposteriorinį mokinio skirstinį[10]. Aprašykime kaip gaunamas šis skirstinys. Tegu $\mathbf{x}$ yra atsakymai į užduotis, θ mokinio gebėjimai, tuomet $f(\mathbf{x}|\theta)$ yra uždavinių atsakymų tikimybė, čia $f(\mathbf{x}|\theta)$ yra vieno parametro Rasch modelis. Darome prielaidą, kad θ yra normaliai pasiskirstęs su vidurkiu 500 ir dispersija 100, t.y. populiacijos modelis $g(\theta) \sim N(500, 100)$. Taigi aposteriorinis skirstinys $h(\theta|\mathbf{x})$ yra:

$$h(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)g(\theta)}{\int f(\mathbf{x}|\theta)g(\theta)d\theta}. \tag{3.7}$$

Kiekvienam mokiniui yra skaičiuojamos penkios galimos reikšmės ir norint gauti dominančią populiacijos statistiką, ši tūrėtų būti skaičiuojama naudojant visas penkias galimas reikšmes. Norėdami įvertinti populiacijos statistiką θ, pirmiausiai skaičiuojame dominančią statistiką $\hat{\theta}_m$, $m = 1, \ldots, M$ visoms galimoms reikšmėms (čia M yra 5) ir gautas reikšmes suvidurkiname:

$$\bar{\theta}_M = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m. \tag{3.8}$$

Statistikos $\hat{\theta}_m$, $m = 1, \ldots, M$ dispersijos įvertinio skaičivimui pritaikomos 80 rinkčių: kiekvienoje rinktyje įvertinama dominanti statistika, kuri lyginama su visos imties įvertiniu ir gaunamas imties dispersijos įvertinys t.y. skaičiavimams naudojama (2.20) formulė

[10] the posterior distribution for a student

$$\hat{\boldsymbol{D}}\theta = \frac{1}{20}\sum_{i=} \left(\hat{\theta} - \hat{\theta}\right)^2,$$

kur $m = ,\ldots,M$, $\hat{\theta}$ –dominančios statistikos įvertinys i–joje rinktyje. Tuomet apskaičiuojamas vidinės dispersijos įvertinys

$$W_M = \frac{1}{M}\sum_{m=1}^{M} \hat{\boldsymbol{D}}\theta \quad , \tag{3.9}$$

ir dispersija tarp galutinio įvertinio ir galimų reikšmių įvertinių

$$\hat{B}_M = \frac{1}{M-1}\sum_{m=1}^{M}\left(\theta - \bar{\theta}_t\right)^2. \tag{3.10}$$

Galimų reikšmių metodu gauto parametro θ įverčio $\hat{\theta}_t$ dispersijai vertinti siūlomas įvertinys:

$$\hat{\boldsymbol{D}}\bar{\theta}_t = V_M + \frac{M+}{M}\hat{B}_M. \tag{3.11}$$

Jei iš skirtingų galimų reikšmių rinkinių gaunami įvertiniai labai skiriasi, tai $\hat{B}_M$ reikšmė yra didelė. Jei įverčiai mažai skiriasi, tai $\hat{B}_M$ mažas ir $\hat{\boldsymbol{D}}\bar{\hat{\theta}}_t$ – mažas [3].

1.4 Tiesiniai hierarchiniai modeliai

Imties išrinkimui naudojant sluoksnines ar lizdines imties išrinkimo metodus, gaunami hierarchiniai duomenys. Pavyzdžiui, mokinys (pirmojo lygmens elementas) yra klasės (antrojo lygmens elemento) sudėtinė dalis, klasė yra mokyklos (trečiojo lygmens elemento) sudėtinė dalis, mokykla yra šalies (ketvirto lygmens elementas) sudėtinė dalis ir pan. Norėdami nagrinėti kaip rezultatai kinta tam tikrose grupėse (pvz., mokinių testų rezultatai skirtingose mokyklose) naudojame hierarchinius modelius. Taikant hierarchinius tiesinius modelius (HLM) yra galimybė nustatyti kontekstinę aplinkos įtaką, t. y. atsižvelgti į antro, trečio lygmenų kintamųjų įtaką. Tolesniuose šio skyriaus poskyriuose, remiantis V. Čekanavičiaus ir G. Murausko knyga *Statistika ir jos taikymai* III ir R. C. Littell, George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger knyga *SAS® for Mixed Models* bus išdėstyta hierarchinių tiesinių modelių teorija.

Hierarchiniai modeliai yra plačiai taikomi įvairiose srityse: medicininiuose tyrimuose, vaistų tyrimuose, sociologiniuose tyrimuose, švietimo tyrimuose ir kt. Pavyzdžiui švietimo tyrimuose į imtį pirmiausiai yra atrenkamos mokyklos, o tada iš atrinktų mokyklų išrenkami mokiniai. Taigi mokiniai, kurie mokosi toje pačioje mokykloje, nėra nepriklausomi ir jų rezultatai koreliuoja. Tokiu atveju duomenų analizei atlikti pasitelkiami hierarchiniai tiesiniai modeliai.

1.4.1 Bendrasis hierarchinis tiesinis modelis

Tarkime turime n normaliai pasiskirsčiusius $Y_1, Y_2, \ldots, Y_n$ kintamuosius ir norime juos paaiškinti vadinamaisiais aiškinamaisiais kintamaisiais $X_1, X_2, \ldots X_k$, $k = 1,2,\ldots,K$ ir kintamasis X gali būti intervalinis arba kategorinis. Tiesinį Gauso modelį galime užrašyti lygtimi:

$$Y_i = \beta_0 + \sum_{j=1}^{k} X_{ij}\beta_j + e_i\,, \tag{4.1}$$

kur $i = 1,2,\ldots,n$, $\beta_1,\ldots,\beta_k$ yra nežinomi fiksuoti parametrai, o paklaidos $e_1,\ldots,e_n$ yra nežinomos, nepriklausomos vienodai pasiskirsčiusios, turinčios normaliųjį skirstinį su vidurkiu 0 ir dispersija σ^2.
Šio modelio matricinė forma:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \vdots & X_{1k} \\ 1 & X_{21} & X_{22} & \vdots & X_{2k} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \vdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}. \tag{4.2}$$

Užrašykime tiesinę Gauso modelio matricinę lygtį paprastesne forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\,, \tag{4.3}$$

kur $\mathbf{Y}$ – normaliai pasiskirstęs kintamųjų $Y_1, Y_2, \ldots, Y_n$ vektorius, $\mathbf{X}$ – aiškinamųjų kintamųjų matrica, $\boldsymbol{\beta}$ – nežinomų fiksuotų parametrų vektorius, $\mathbf{e}$ – paklaidų vektorius turintis daugiamatį normalųjį skirstinį su vidurkiu 0 ir kovariacijų matrica $\boldsymbol{\Sigma}$, čia $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$.

Hierarchiniuosie modeliuose daroma prielaida, kad paklaidos $\mathbf{e}$ yra normaliai pasiskirsčiusios, t. y. $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$. Tuomet hierarchiniuose tiesiniuose modeliuose $\mathbf{Y}$ elementai gali būti priklausomi ir koreliuoti. Individai skiriasi vieni nuo kitų ir teigti, kad visi individai turi tą pačią (fiksuotą) parametro reikšmę yra klaidinga [4]. HLM modelyje parametrai β yra atsitiktiniai, t .y. $\beta = \gamma + u$, kur γ yra fiksuotas parametras, o u yra atsitiktinė paklaida turinti normalųjį skirstinį. HLM modelį gali sudaryti kelių lygmenų lygtys, kadangi šiame darbe analizuosime dviejų lygmenų modelį, todėl pateiksime dviejų lygmenų tiesinio hierarchinio modelio matricinę formą. Užrašykime supaprastintą bendrąjį dviejų lygmenų hierarchinį tiesinį modelį, kur visi β priklauso nuo tų pačių W_{sj} (sudarant modelius, nebūtina į kiekvieną lygtį įtraukti visus W_{sj}). Tarkime turime $j = 1,2,\ldots,J$ grupių. Kiekvienos grupės individų rezultatai priklauso nuo kintamųjų X_k, $k = 1,2,\ldots,K$. HLM modelio pirmojo lygmens priklausomybes:

$$\mathbf{Y}_j = \mathbf{Z}_j \boldsymbol{\beta}_j + \mathbf{e}_j, \tag{4.4}$$

čia $\mathbf{Z}_{ij} = \begin{bmatrix} 1 & X_{1ij} & X_{2ij} & \cdots & X_{Kij} \end{bmatrix}$, $\boldsymbol{\beta}_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{Kj} \end{bmatrix}$, $\mathbf{e}_j = \begin{bmatrix} e_{0j} \\ e_{1j} \\ \vdots \\ e_{Kj} \end{bmatrix}$.

Antrojo lygmens priklausomybės:

$$\boldsymbol{\beta}_j = \mathbf{W}_j \boldsymbol{\gamma} + \mathbf{u}_j, \tag{4.5}$$

čia $\mathbf{W}_j = \begin{bmatrix} 1 & W_{1j} & \cdots & W_{sj} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & W_{1j} & \cdots & W_{tj} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & W_{sj} \end{bmatrix}$, $\boldsymbol{\gamma} = \begin{bmatrix} \gamma_{00} \\ \vdots \\ \gamma_{0s} \\ \gamma_{10} \\ \vdots \\ \gamma_{1s} \\ \vdots \\ \gamma_{Ks} \end{bmatrix}$, $\mathbf{u}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \\ \vdots \\ u_{Kj} \end{bmatrix}$.

Tarkime $\mathbf{X}_j = \begin{bmatrix} \mathbf{Z}_{1j}\mathbf{W}_j \\ \mathbf{Z}_{2j}\mathbf{W}_j \\ \vdots \\ \mathbf{Z}_{n_j j}\mathbf{W}_j \end{bmatrix}$ ir įstatę (4.5) lytį į (4.4) lygtį gauname bendrojo HLM modelio užrašą:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{4.6}$$

čia $j = 1,\ldots,J$, $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_J \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_J \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_J \end{bmatrix}$, $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_J \end{bmatrix}$

ir $\mathbf{X}$ – fiksuoto poveikio kintamieji, $\boldsymbol{\gamma}$ – fiksuoto poveikio parametrai, $\mathbf{Z}$ – atsitiktinio poveikio kintamieji, $\mathbf{u}$ – antrojo lygmens atsitiktinės paklaidos ir $\mathbf{e}$ – pirmojo lygmens atsitiktinės paklaidos.

Bendrojo modelio prielaidos:

1. $e_{ij} \sim N(0,\sigma^2)$ ir e_{ij} tarpusavyje yra nepriklausomi;
2. $u_{kj} \sim N(0,\tau_{kk})$;

3. $Cov(u_{kj}, u_{mj}) = \tau_{km}$;

4. u_{kj} ir u_{mj} su visais *k*, *m* yra nepriklausomi, jei tik $j \neq j'$;

5. u_{ij} ir e_{ij} yra nepriklausomi;

6. kintamasis *Y* yra intervalinis;

7. kintamieji $X_1, X_2, \ldots, X_K$, $W_1, \ldots, W_S$ yra intervaliniai arba pseudokintamieji.

Nagrinėjant HLM modelį reikia įvertinti nežinomus parametrus ir atsitiktinių paklaidų įtaką. Informacija apie paklaidų elgesį ir galimas priklausomybes nusako kovariacijų matricos. Vektoriaus $\mathbf{u}$ kovariacijų matricą pažymėkime $\mathbf{T}$, vektoriaus $\mathbf{e}$ kovariacijų matricą pažymėkime $\mathbf{R}$, o vektoriaus $\mathbf{Y}$ kovariacijų matricą pažymėkime $\mathbf{V}$. Kadangi paklaidos u_{ij} ir e_{ij} yra nepriklausomos normaliai pasiskirsčiusios, tai:

$$E\begin{pmatrix}\mathbf{u}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \text{ ir } Cov\begin{pmatrix}\mathbf{u}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{T} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{pmatrix},$$

kur $\mathbf{T} = \begin{pmatrix}\mathbf{T_1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{T_2} & \mathbf{0} & \cdots & \mathbf{0}\\ \vdots & \vdots & \vdots & & \vdots\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T_J}\end{pmatrix}$, kur $\mathbf{T}_j = \begin{pmatrix}\tau_{00} & \tau_{01} & \cdots & \tau_{0K}\\ \tau_{10} & \tau_{11} & & \tau_{1K}\\ \vdots & \vdots & & \vdots\\ \tau_{K0} & \tau_{K1} & \cdots & \tau_{KK}\end{pmatrix}$

ir $\mathbf{R} = \begin{pmatrix}\mathbf{R_1} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{R_2} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R_J}\end{pmatrix}$, kur $\mathbf{R}_j = \sigma^2\mathbf{I}$, $\mathbf{I}$ – vienetinė matrica.

Vektoriaus $\mathbf{Y}$ kovariacijų matrica lygi

$$\mathbf{V} = \mathbf{ZTZ}' + \mathbf{R}. \tag{4.7}$$

1.4.2. Besąliginis HLM modelis

Besąliginis arba nulinis tiesinis hierarchinio modelio analizė atskleidžia ar duomenims tinkamas hierarchinis modeliavimas. Aprašykime dviejų lygių – individo ir grupės (pvz., mokinio ir mokyklos) – besąlyginį modelį. Tarkime yra *J* grupių, t.y. $j = 1,2,\ldots,J$, o individų *j*-je grupėje yra n_j, t.y. $i = 1,2,\ldots,n_j$. Tada Y_{ij} yra *i*-ojo individo iš *j*-osios grupės rezultatas. Pirmasis (individo) lygmuo užrašomas lygtimi:

$$Y_{ij} = \beta_{0j} + e_{ij}, \quad i = 1,\dots,n_j,\ j = 1,\dots,J, \tag{4.8}$$

antrasis (grupės) lygmuo:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad j = 1,\dots,J. \tag{4.9}$$

Pirmojo lygmens lygtį sudaro atsitiktinis koeficientas β_{0j} ir atsitiktinė paklaida e_{ij}. Antrojo lygmens lygtį sudaro konstanta γ_{00} ir atsitiktinė paklaida u_{0j}. Laikoma, kad atsitiktinės paklaidos e_{ij} ir u_{0j} yra nepriklausomos, normaliai pasiskirsčiusios su dispersijomis σ^2 ir τ_{00}, t.y. $e_{ij} \sim N(0,\sigma^2)$ ir e_{ij} nepriklausomi visiems $i = 1,\dots,n_j$, $j = 1,\dots,J$ bei $u_{0j} \sim N(0,\tau_{00})$ ir u_{0j} nepriklausomi visiems $j = 1,\dots,J$. Atsitiktinės paklaidos e_{ij} ir u_{0j} nekoreliuoja.

Besąlyginį modelį galime užrašyti jungtine lygtimi, įstačius (4.9) išraišką į (4.8) lygtį:

$$Y_{ij} = \gamma_{00} + (u_{0j} + e_{ij}). \tag{4.10}$$

Kadangi paklaidos yra atsitiktiniai dydžiai, todėl vertinamos dispersijos σ^2 ir τ_{00}. Didelės įverčių σ^2 ir τ_{00} reikšmės rodo, kad į modelį reikia įtraukti kintamuosius, paaiškinančius grupių ir individų skirtumus. Besąlyginiame modelyje vertinami trys parametrai: γ_{00} – fiksuoto poveikio parametras, σ^2 ir τ_{00} – pirmojo ir antrojo lygmenų atsitiktinio poveikio parametrai.

Analizuojant besąlyginį modelį galime patikrinti tris hipotezes apie parametrų lygybę nuliui:

- Hipotezė ar vidutinis individų rezultatas nėra lygus nuliui:

$$\begin{cases} H_0: \gamma_{00} = 0, \\ H_1: \gamma_{00} \neq 0. \end{cases}$$

- Ar tarp individų rezultatų yra ne grupių nulemtų skirtumų:

$$\begin{cases} H_0: \sigma^2 = 0, \\ H_1: \sigma^2 > 0. \end{cases}$$

- Ar grupių įtaka rezultatams skiriasi:

$$\begin{cases} H_0: \tau_{00} = 0, \\ H_1: \tau_{00} > 0. \end{cases}$$

Norėdami įvertinti, kaip stipriai skiriasi grupių rezultatai, palyginti su rezultatais grupėse, skaičiuojamas *tarpklasinės koreliacijos koeficientas*[11] ICC:

[11] Intraclass Correlation Coefficient

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}. \qquad (4.11)$$

Jeigu $ICC = 0$ (t. y. $\tau_{00} = 0$), tai grupių įtakos nėra ir analizei nereikia naudoti hierarchinių modelių. Kuo didesnis ICC, tuo labiau reikia atsižvelgti į duomenų hierarchinę struktūrą [2]. Pavyzdžiui, jeigu $ICC = 0{,}42$, tuomet 42 % individų rezultatų skirtumų lemia grupės ir duomenų analizei reikia naudoti HLM.

1.4.3 Dviejų lygmenų tiesinių hierarchinių modelių pavyzdžiai

Plačiau panagrinėkime keleta hierarchinių tiesinių modelių pavyzdžių. Tarkime, kad Y_{ij} priklauso tik nuo pirmojo lygmens kintamojo X_{ij}. Tuomet modelio pirmojo lygmens lygtis yra:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}, \quad i = 1, \ldots, n_j, \; j = 1, \ldots, J,$$

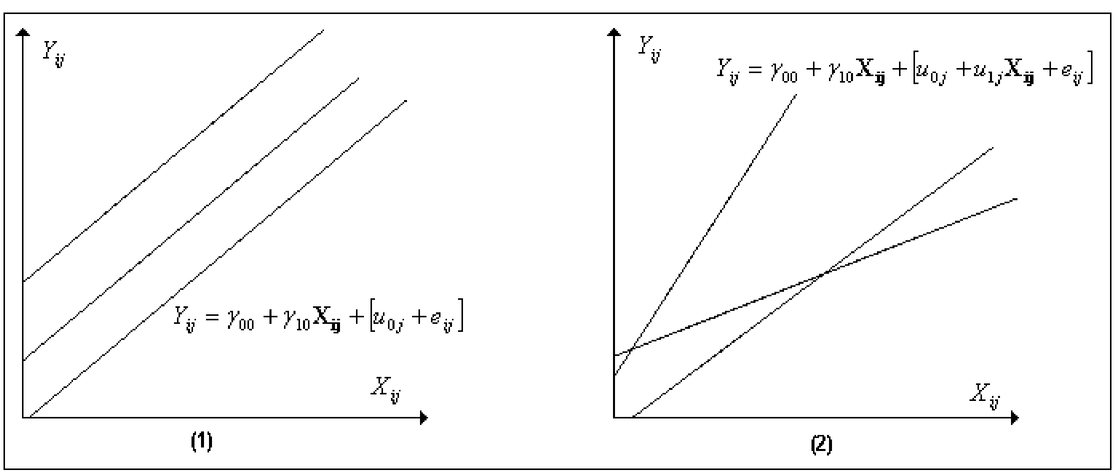
o antrojo lygmens lygtys:

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j}, \\ \beta_{1j} = \gamma_{10} \end{cases}, \quad j = 1, \ldots, J.$$

Toks modelis vadinamas **atsitiktinio postūmio ir pastovaus posvyrio modeliu**, o jo jungtinė lygtis atrodo taip:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + [u_{0j} + e_{ij}].$$

Matome, kad kintamojo X_{ij} įtaka Y_{ij} yra fiksuota, t.y. visoms grupėms Y_{ij} priklausomybė nuo X_{ij} vienoda, grupės skiriasi tik postūmiu (4.1 pav. (1)).



4.1 pav. Y_{ij} priklausomybės nuo X_{ij} grafikai

Tarkime Y_{ij} priklausomybė nuo kintamojo X_{ij} yra atsitiktinė, t.y. vienose grupėse X_{ij} ir Y_{ij} priklausomybės yra stipresnės, kitose ne. Toks modelis vadinams **atsitiktinio postūmio ir atsitiktinio posvyrio modeliu**. Tokio modelio pirmojo lygmens lygtis:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}, \quad i = 1,\dots,n_j,\ j = 1,\dots,J.$$

Antrojo lygmens lygtys:

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j}, \\ \beta_{1j} = \gamma_{10} + u_{1j} \end{cases}, \quad j = 1,\dots,J.$$

Jungtinė lygtis:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \left[u_{0j} + u_{1j} X_{ij} + r_{ij}\right].$$

Šiuo atveju kintamojo X_{ij} įtaka Y_{ij} yra atsitiktinė, t.y. visoms grupėms Y_{ij} priklausomybė nuo X_{ij} skirtinga, grupės skiriasi ir postūmiu, ir posvyriu (4.1 pav. (2)).

Ankstesnių modelių pavyzdžiuose aiškinamieji kintamieji buvo įtraukti tik į pirmąjį lygmenį, užrašykime kaip atrodo **modelis su antrojo lygmens kintamaisiais**. Tarkime, kad kiekvieno individo Y_{ij} priklauso nuo X_{ij}, o grupės įtaka (antrasis lygmuo) – nuo Z_j ir V_j. Tokio modelio pirmojo lygmens lygtis:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}, \quad i = 1,\dots,n_j,\ j = 1,\dots,J$$

Antrojo lygmens lygtis:

$$\begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{02} V_j + u_{0j}, \\ \beta_{1j} = \gamma_{10} + \gamma_{11} V_j + u_{1j} \end{cases}, \quad j = 1,\dots,J$$

Tariame, kad β_{0j} yra atsitiktinis parametras ir priklauso nuo kintamųjų Z_j ir V_j. β_{1j} taip pat atsitiktinis parametras ir priklauso nuo kintamojo V_j. Užrašykime jungtinę lygtį:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{02} V_j + \gamma_{10} X_{ij} + \gamma_{11} X_{ij} V_j + \left[u_{1j} X_{ij} + u_{0j} + r_{ij}\right].$$

Iš jungtinės lygties matome, kad yra kintamųjų X_{ij} ir V_j sąveika, t.y. pirmojo ir antrojo lygmenų kintamųjų sąveika.

1.4.3 Informaciniai indeksai, parametrų įverčiai ir hipotezės

Hierarchiniai modeliai yra sudaromi iteracijos būdu, vis pridedant naujus kintamuosius. Besąlyginį modelį papildžius kintamaisiais, naujasis modelis lyginamas su besąlyginiu modeliu ir tikrinama ar naujasis modelis geriau tinka duomenims. Tada į modelį įtraukiame naujus kintamuosius ir vėl tikriname ar naujasis modelis geriau tinka duomenims negu senasis ir t. t. Dviejų modelių palyginimui naudojami **informaciniai indeksai**. Hierarchiniam tiesinio modelio parametrų įverčiams skaičiuoti naudojamas didžiausio tikėtinumo metodas (ML) arba apriboto didžiausio tikėtinumo metodas (REML). Reiktų paminėti, kad dviejų HLM modelių parametrus galima lyginti tik tada, kai jų įverčiai skaičiuojami naudojant ML metodą [2].

Informacinių indeksų skaičiavimams naudojamos formulės pateiktos 4.1 lentelėje, kur l – didžiausio tikėtinumo funkcijos logaritmo maksimumas. Pirmojo lygmens duomenų ir fiksuoto poveikio parametrų skaičiaus skirtumas lygus n, o d yra atsitiktinio poveikio parametrų skaičius.

Kriterijus	Formulė
Dvigubas didžiausio tikėtinumo funkcijos logaritmas	$-2l$
Akaikės informacinis kriterijus (AIC)	$-2l+2d$
Hurvičiaus–Cajo indeksas (AICC)	$-2l+2dn/(n-d-1)$
Švarco–Bajeso kriterijus (BIC)	$-2l+d\ln n$

4.1 lentelė. Informacinių indeksų formulės

Modelis, kurio informacinių indeksų reikšmės mažesnės, geriau tinka duomenims. Taigi lyginant dviejų modelių informacinius indeksus nusprendžiama, kuris modelis tinkamesnis duomenims.

Jeigu parametro įvertinys σ^2 įgyja didelę reikšmę, tai rodo, kad pirmojo lygmens kintamieji paaiškina ne visą Y elgesį. Analogiškai, jeigu τ_{00} įgyja didelę reikšmę, tai rodo, kad antrojo lygmens kintamieji silpnai paaiškina atitinkamo atsitiktinio koeficiento β elgesį. Lyginat modelius, žiūrima kiek sumažėjo naujojo modelio įverčiai σ^2 ir τ_{00}. Pavyzdžiui, atlikus analizę gavome, kad $\sigma^2 = 86$ ir $\tau_{00} = 44$, įtraukus į modelį naujus pirmojo ir antrojo lygmenų kintamuosius gavome tokius įverčius $\sigma^2 = \ldots1$ ir $\tau_{00} = \ldots4$. Apskaičiuokime santykinius įvertinių pokyčius:

$\frac{86-\ldots5}{86} = 0{,}24$ ir $\frac{54-\ldots4}{44} = 0{,}23$. Taigi galime daryti išvadą, kad į modelį itraukus papildomus kintamuosius, 24 % geriau paaiškinami skirtumai tarp individų ir 23 % geriau paaiškinami skirtumai tarp grupių.

Trumpai aprašykime, kaip skaičiuojami įvertiniai. Bendrasis HLM modelio užrašas pateiktas (4.6) lygtyje. Taikant apibendrintą mažiausių kvadratų metodą apskaičiuojami įvertiniai, fiksuoto poveikio parametrų įvertiniai:

$$\hat{\gamma} = (X' V^{-1} X)^{-1} X' V^{-1} Y, \tag{4.12}$$

atsitiktinio poveikio parametrų įvertiniai:

$$\beta_{\smile} = Z'_{\smile} R_j^{-} Z_j)^{-} Z'_{\smile} R_j^{-} Y_j, \tag{4.13}$$

įstačius $\hat{\gamma}$ ir $\hat{\beta}_{\smile}$ įverčius į lygtį $\beta_{\smile} = V_j\gamma + {}_j$ gauname u_j įvertinį:

$$\hat{u}_j = \hat{3}_j - V_j\hat{\gamma}. \tag{4.14}$$

Pakankamai tiksliai modelio analizei pakanka $\hat{\gamma}$, $\hat{\sigma}$ ir $\hat{\tau}$ įvertinių. Turint įvertinius $\hat{\gamma}$ galime užrašyti prognozės funkciją, pavyzdžiui, turint tokį jungtinį modelį:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + {}_{1j} X_{ij} + {}_{0j} + {}_{ij},$$

i-ojo individo j-ios grupės prognozės funkcija bus:

$$Y = {}_{00} + {}_{01} Z_j + {}_{0} X.$$

Trumpai aprašykime kokios **hipotezės** yra tikrinamos HLM modeliui. V. Čekanavičius ir G. Murauskas [2] didesnį dėmesį skiria vieno parametro hipotezėms, todėl šiame skyrelyje aptarsime tokias (vieno parametro) hipotezes. Jeigu nulinė hipotezė apie parametro lygybę nuliui neatmetama, tai vietoj to parametro įrašę nulį turėtume ištirti redukuotą modelį. Ar redukuotasis modelis geresnis nusprendžiame atsižvelgę į parametrų įvertinius ir informacinių indeksų reikšmes. Jeigu parametrų įverčiai ir informaciniai indeksai sumažėja, tai redukuotasis arba naujasis (jeigu į modelį įtraukiame naujus kinamuosius ir tikriname ar naujasis modelis geresnis) modelis yra geresnis. Redukuoto modelio atveju, jeigu pastarasis modelis geresnis, iš modelio galime pašalinti parametrą vietoj kurio redukuotame modelyje įrašėme nulį. Nagrinėjant HLM modelius yra tikrinamos tokio hipotezės:

- Fiksuoto poveikio parametrui γ_{s} tikrinama hipotezė, ar jis nelygus nuliui:

$$\begin{cases} H_0 : \gamma_{s} = 0, \\ H_1 : \gamma_{s} \neq 0. \end{cases}$$

Nulinė hipotezė atmetama, jei statistikos p reikšmė mažesnė už pasirinktą reikšmingumo lygmenį α, t.y. $p < \alpha$. Dažniausiai $\alpha = 0{,}05$. Jeigu hipotezė neatmetama, tai rodo, kad iš modelio reikėtų pašalinti kintamąjį, kurio daugiklio reikšmingumą tikrinome ir analizuoti redukuotą modelį.

- Taip pat tikrinama hipotezė, ar pirmojo lygmens paklaidų dispersija σ lygi nuliui:

$$\begin{cases} H_0 : \sigma = 0, \\ H_1 : \sigma > 0. \end{cases}$$

Sprendimo priėmimo taisyklė yra analogiška anksčiau aprašytai, t.y. nulinė hipotezė atmetama jei $p < \alpha$, čia α – pasirinktasis reikšmingumo lygmuo. Neatmesta nulinė hipotezė gali reikšti, kad į pirmojo lygmens modelį įtraukti kintamieji gana tiksliai aprašo rezultatą. Arba į modelį įtrauktas kintamasis stipriai koreliuoja su rezultatu, nors nėra susijęs priežastiniais ryšiais.

- Analogiškai tikrinama ir interpretuojama hipotezė, ar antrojo lygmens paklaidų dispersija τ_{k} lygi nuliui:

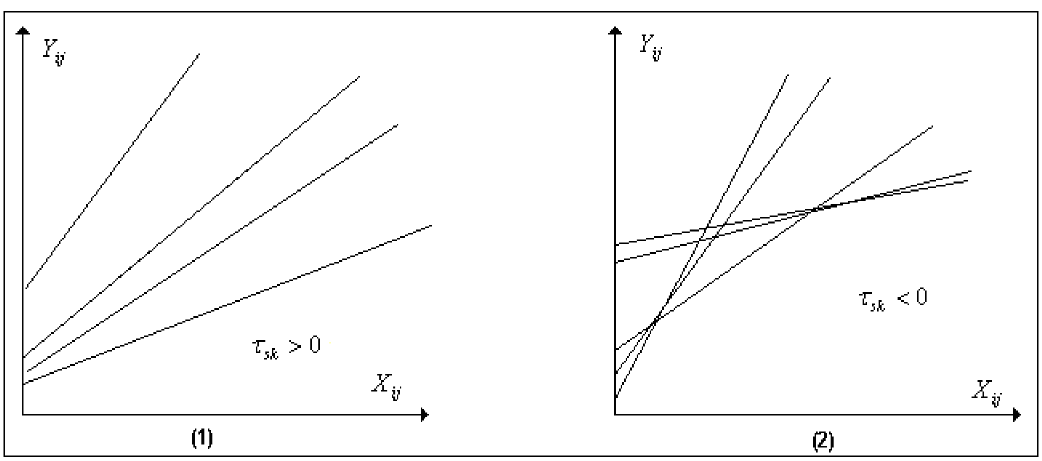
$$\begin{cases} H_0 : \tau_{k} = 0, \\ H_1 : \tau_{k} > 0. \end{cases}$$

Vėlgi nulinė hipotezė atmetama jei $p < \chi$, čia α – pasirinktasis reikšmingumo lygmuo. Neatmesta nulinė hipotezė gali reikšti, kad τ_{k} statistiškai reikšmingai lygus 0, t. y. koeficientas β neatsitiktinis.

- Kadangi antrojo lygmens paklaidos gali koreliuoti, todėl tikrinamos hipotezės dėl jų koreliuotumo, t.y. nulinė hipotezė, kad paklaidos nekoreliuoja. Hipotezę galime užrašyti:

$$\begin{cases} H_0 : \tau_{k} = 0, \\ H_1 : \tau_{k} \neq 0, \end{cases}$$

kur $s \neq :$. Nulinė hipotezė atmetama, t.y. β ir β statistiškai reikšmingai koreliuoja, jei $p < \chi$, čia α yra pasirinktasis reikšmingumo lygmuo. Jeigu patikrinę nulinę hipotezę gauname, kad $\tau_{k} >)$ tai galime manyti, kad grupėse, kuriose rezultatai aukštesni, Y_{ij} priklausomybė nuo X_{ij} yra stipresnė (4.2 pav. (1)), jeigu $\tau_{k} <)$, tai grupėse, kuriose rezultatai aukštesni, Y_{ij} priklausomybė nuo X_{ij} yra silpnesnė (4.2 pav. (2)).



4.2 pav. Koreliuotų parametrų priklausomybių stiprumai

- Suformuluokime hipotezę apie atsitiktinius koeficientus β:

$$\begin{cases} H_0 : \beta_{kj} = 0, \\ H_1 : \beta_{kj} \neq 0. \end{cases}$$

Nulinė hipotezė atmetama, jei $p < \alpha$, čia α – pasirinktasis reikšmingumo lygmuo. Jeigu nulinė hipotezė neatmesta, t.y. $\beta_{kj} = 0$, tai galime įtarti, kad Y ir pirmojo lygmens k-tasis kintamasis (X_k) nesusiję ir reiktų nagrinėti modelį, kurio pirmojo lygmens lygtyje nebūtų kintamojo X_k.

- Hipotezė, kuomet tikrinama ar parametrų γ_k tiesinių darinių vektorius lygus nuliniam vektoriui, vadinama vektorine hipoteze. Vektorinių hipotezių prireikia, kai norime įsitikinti, kad visose antrojo lygmens lygtyse kuris nors kintamasis nereikalingas. Pavyzdžiui, turime tokią antrojo lygmens lygtį:

$$\begin{cases} \beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + \gamma_{12} V_j + u_{0j}, \\ \beta_{2j} = \gamma_{20} + \gamma_{21} V_j + u_{1j} \end{cases} \quad j = 1,\dots,J,$$

ir norime patikrinti hipotezę ar β_{1j} ir β_{2j} nepriklauso nuo V_j, $j = 1,\dots,J$. Tuomet formuluojame hipotezę:

$$\begin{cases} H_0 : \gamma_{12} = \gamma_{21} = 0, \\ H_1 : kita. \end{cases}$$

Nulinė hipotezė atmetama, nusprendžiama , kad bent vienas iš parametrų γ_{11} arba γ_{21} statistiškai reikšmingai skiriasi nuo 0, jei $p < \alpha$. Tačiau nulinės hipotezės atmetimas nereiškia, kad reikia visiškai iš modelio pašalinti kintamąjį V_j (redukuojame modelį ir tikriname ar pašalinus V_j modelis pagerėja ar ne).

- Dviejų modelių palyginimui formuluojame tokią hipotezę:

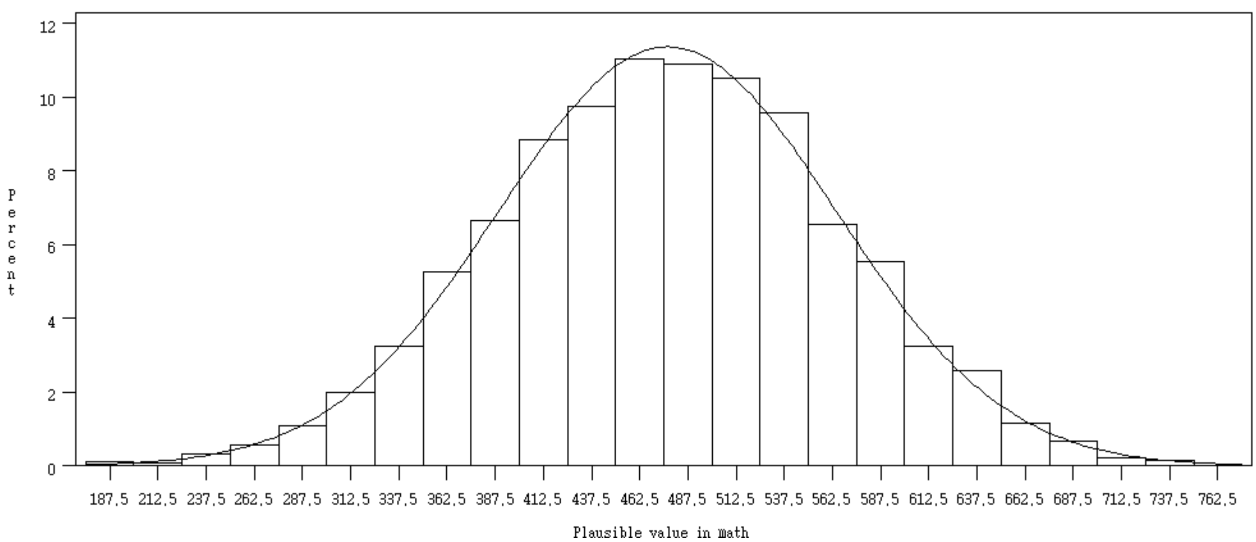
$$\begin{cases} H_0 : \text{modeliai nesiskiria}, \\ H_1 : \text{modeliai skiriasi}. \end{cases}$$

Tarkime pirmojo modelio didžiausio tikėtinumo funkcijos logaritmo maksimumą pažymėkime l_1, antrojo modelio – l_2. Šiuo atveju kriterijaus statistika yra $(-2l_2) - (-2l_1)$. Nulinė hipotezė atmetama, jei ši reikšmė didesnė nei χ^2 skirstinio su *k* laisvės laipsnių α lygmens kritinė reikšmė. Čia *k* pirmo ir antro modelių vertinamų parametrų skaičiaus skirtumas, α – pasirinktasis reikšmingumo lygmuo. Neatmesta nulinė hipotezė rodo, kad antrasis modelis nėra geresnis už pirmąjį. Liginimo su besąlyginiu modeliu testas vadinamas *nulinio modelio tikėtinumo santykio* *testu* (*Null Model Likelihood Ratio Test*). Šis testas yra realizuotas SAS programoje. Taigi nulinė hipotezė atmetama, t.y. nagrinėjamas modelis statistiškai skiriasi nuo nulinio modelio, jei $p < \chi$.

2. PRAKTINĖ DALIS

2.1. Duomenys

PISA tiriamoji grupė yra 15 metų mokiniai. 2009 metų tyrime dalyvavo 196 Lietuvos mokyklos, iš kurių buvo atrinkta 4528 mokiniai. PISA tyrimą sudaro testų sąsiuviniai (2009 m. tyrime jų buvo 13), kur kiekvienas mokinys gauna po vieną sąsiuvinį, ir kontekstiniai klausimynai, kuriuose mokiniai pateikiama informaciją apie save, savo įpročius, pomėgius, mokymosi būdus, namų aplinką ir pan. Taip pat kompiuterinio raštingumo, tėvų klausimynai, klausimynas apie mokymąsi. Mokyklų dalyvaujančių tyrime direktoriai taip pat gauna klausimyną apie mokyklos žmogiškuosius ir materialinius išteklius, finansavimą, personalą ir pan. Šiame darbe analizuojami PISA 2009 metų tyrimo Lietuvos penkiolikmečių mokinių duomenys. PISA tiria tris sritis: gamtamokslį, matematinį raštingumą ir skaitymo gebėjimus. Šiame darbe duomenų analizei pasitelkę hierarchinius modelius, bandysime rasti mokinių matematikos raštingumą įtakojančius veiksnius. PISA tyrimų užduočių pavyzdžius, mokinių, tėvų bei mokyklos klausimynus, duomenų bazes galima rasti internete adresu: www.pisa.oecd.org.



2.1 pav. Matematinio raštingumo histograma

Analizuodami veiksnius įtakojančius penkiolikmečių mokinių matematinį raštingumą naudosime tiesinius hierarchinius modelius. Prieš pradedant sudetingesnę duomenų analizę atlikime pradinę duomenų analizę, t. y. pasinaudokime aprašomosios statistikos metodais ir panagrinėkime duomenų savybes. 2.1 paveiksle pavaizduota mokinių matematinio raštingumo histograma atskleidžia, kad mokinių matematinio raštingumo skalės taškai pasiskirstę normaliai ir apytiksliai kinta nuo 175 iki 775 balų. Lietuvos penkiolikmečių matematinio raštingumo vidurkis yra 476,6 skalės taškų, o OECD vidurkis 496 skalės taškų [5], taigi Lietuvos penkiolikmečiai yra žemiau vidurkio.

Panagrinėkime kokie veiksniai gali įtakoti matematinio raštingumo rezultatus (kaip mokinio lytis, kompiuterio namuose turėjimas ir pan.). Vidutinis merginų matematinis raštingumas yra 479,75 skalės taškų, o vaikinų – 473,54 taškų, šis skirtumas nėra didelis. Net 93,3 % penkiolikmečių namuose turi kompiuterį ir tokių mokinių matematinio raštingumo vidukis siekia 482,64 balų. Kompiuterio namuose neturinčių mokinių yra apie 6,7 % ir jų matematinio raštingumo vidurkis yra 407,83 balų, o tai yra beveik 80 balų mažiau nei turinčių kompiuterius mokinių. 2.1 lentelėje pateikti rezultatai parodo, kad matematinio raštingumo rezultatai skiriasi ne tik tarp mokinių turinčių namuose kompiuterius ir neturinčių, bet rezultatai skiriasi ir tarp vietovės kurioje mokykla įsikūrusi bei mokyklos tipo. Skliausteliuose nurodytas procentinis mokinių pasiskirstymas.

Apie 4 % penkiolikmečių dažniausiai namuose kalba kita kalba, o ne lietuvių. Tokių mokinių vidutinis matematinis raštingumas yra 455,33 balų ir yra 23,39 balu žemesnis negu mokinių, kurie namuose dažniausiai kalba lietuviškai.

	Mokyklos įsikūrimo vieta				**Mokyklos tipas**			
	Kaimas	**Miestelis**	**Miestas**	**Vilnius**	**Kita**	**Pagrindinė**	**Vidurinė**	**Gimnazija**
Namuose turi kompiuterį	454,23 (22,5%)	484,81 (35,4%)	494,52 (22,8%)	505,31 (15,2%)	365,06 (1,4%)	439,81 (14,4%)	473,14 (40,8%)	514,61 (36,6%)
Namuose neturi kompiuterio	400,75 (3,2%)	406,82 (2,3%)	407,47 (0,7%)	455,07 (0,6%)	328,37 (0,4%)	392,19 (2,3%)	415,92 (2,5%)	437,02 (1,6%)

2.1 lentelė. Penkiolokamečių matematinio raštingumo balai ir pasiskirstymas pagal mokinių kompiuterio namuose (ne)turėjimą ir mokyklos vietovę ir tipą.

Mokinių klausimyne buvo klausimas „kiek knygų yra jūsų namuose“, pažiūrėkime, kaip namuose esančių knygų skaičius įtakoja mokinių pasiekimus (2.2 lentelė). Trečdalio mokinių namuose yra nuo 26 iki 100 knygų. Mokinių, kurių namuose knygų yra daugiau nei 100, matematinio raštingumo vidurkis didesnis nei 500 balų. Galime įtarti, kad labiau išsilavinę ir labiau pasiturinčių šeimų tėvai perka ir namuose turi daugiau knygų negu mažiau išsilavinę tėvai. Panagrinėkime tėvų išsilavinimo lygį ir mokinių matematinio raštingumo rezultatus.

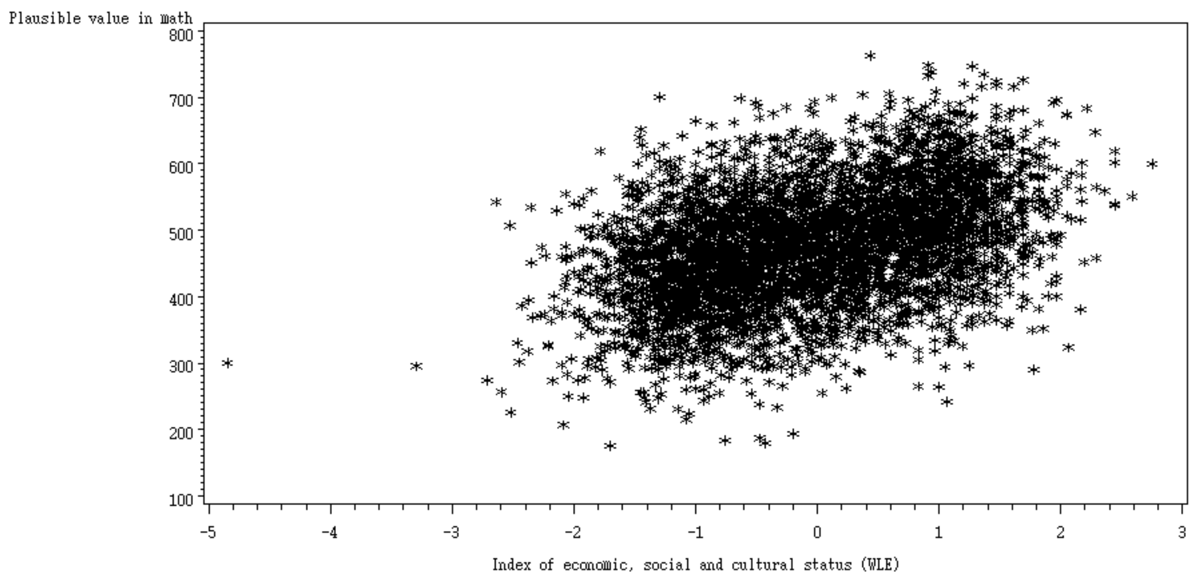
0-10 knygų	**11-25 knygų**	**26-100 knygų**	**101-200 knygų**	**201-500 knygų**	**> 500 knygų**
425,01 (14,7 %)	444,88 (20,3 %)	485,06 (33,4 %)	505,68 (15,8 %)	525,85 (10 %)	520,81 (5,8 %)

2.2 lentelė. Matematinio raštingumo vidurkiai ir mokinių pasiskirstymas (skliausteliuose) pagal namuose esančių knygų skaičių.

Iš 2.3 lentelės matome, kad beveik pusė tėvų (48,9 %) yra kvalifikuoti tarnautojai. Esant didesniam kompiuterių skaičiui namuose – mokinių matematinio raštingumo rezultatai aukštesni. Taip pat pastebima tendencija, kad matematinis raštingumas didesnis tų mokinių, kurių tėvai yra kvalifikuoti tarnautojai ir namuose yra trys ir daugiau kompiutrių, negu mokinių kurių, tėvai yra nekvalifikuoti darbininkai ir namuose neturi kompiuterio.

		Tėvų darbo tipas ir kvalifikacija			
Kompiuterių skaičius		**Kvalifikuoti tarnautojai**	**Nekvalifikuoti tarnautojai**	**Kvalifikuoti darbininkai**	**Nekvalifikuoti darbininkai**
	0	467,71 (0,9 %)	396,73 (1,2 %)	406,03 (2 %)	394,45 (1,2 %)
	1	504,26 (28,4 %)	464,85 (17,2 %)	449,78 (12,8 %)	449,77 (6,4 %)
	2	516,50 (14,4 %)	475,07 (4,5 %)	458,39 (2,9 %)	450,40 (1,6 %)
	> 2	531.74 (5,2 %)	468,98 (0,7 %)	472,01 (0,5 %)	403,09 (0,1 %)

2.3 lentelė. Matematinio raštingumo vidurkiai ir mokinų pasiskirstymas (skliausteliuose) pagal aukščiausią tėvų darbo pobūdį.



2.2 pav. Mokinių matematinio raštingumo ir socialinio ir ekonominio statuso sklaidos diagrama.

Penkiolikmečių matematinį raštingumą veikia įvairūs veiksniai: tėvų išslavinimo lygis, kompiuterio ir knygų namuose turėjimas ir pan., visa tai galima pavadinti socialiniais ir ekonominiais veiksniais. PISA pagal mokinių klausimyno atsakymus kiekvienam mokiniui skaičiuoja socialinio ir ekonominio statuso indeksa (ESCS). Šis indeksas apskaičiuojamas atsižvelgiant į tėvų darbo lygį, tėvų išsilavinimo lygį, namuose esančią nuosavybę (kompiuterio, savo kambario, knygų, televizoriaus ir t. t. turėjimas). 2.2 paveiksle matome mokinių matematinio raštingumo ir socialinio ir ekonominio statuso sklaidos diagramą. Galime pastebėti tendenciją, kad esnt didesniam socialiniam ir ekonominia indeksui mokinio matematinio raštingumo balas didesnis.

	Matematinis raštingumas	Mėgimas skaityti	Mokyklos užduotims atlikti naudoja IKT	Drausminis klasės „klimatas“	IKT turėjimas namuose
Matematinis raštingumas	1,0000	0,2658 (<,0001)	-0,0165 (0,2693)	0,1765 (<,0001)	0,1252 (<,0001)
Mėgimas skaityti	0,2658 (<,0001)	1,0000	0,0397 (0,0085)	0,1378 (<,0001)	-0,0348 (0,0205)
Mokyklos užduotims atlikti naudoja IKT	-0,0165 (0,2693)	0,0397 (0,0085)	1,0000	0,0309 (0,0390)	0,3817 (<,0001)
Drausminis klasės „klimatas“	0,1765 (<,0001)	0,1378 (<,0001)	0,0309 (0,0390)	1,0000	0,0319 (0,0328)
IKT turėjimas namuose	0,1252 (<,0001)	-0,0348 (0,0205)	0,3817 (<,0001)	0,0319 (0,0328)	1,0000

2.4. lentelė. Koreliacijos koeficientai ir p-reikšmės (skliaustuose)

Penkiolikmečių matemetinis raštingumas ir mokinių socialinis ir ekonominis statusas vidutiniškai koreliuoja (koreliacijos koeficientas 0,42). Taigi ankstesnis pastebėjimas, kad didėjant mokinio socialiniam ir ekonominiam statusui didėja jo matematinio raštingumo rezultatas yra teisingas. Taip pat penkiolikmečių matematinis raštingumas statistiškai reikšmingai (su reikšmingumo lygmeniu 0,05) silpnai koreliuoja su mėgimo skaityti, drausminiu klasės „klimatu“ ir IKT turėjimu namuose indeksais (2.4 lentelė). Galime įtarti, kad mokiniai, kurie mėgsta skaityti, mėgta ir kitokią lavinančią veiklą ir jiems geriau sekasi matematika. Matome, kad silpnai koreliuoja IKT turėjimo namuose indeksas ir mokyklos užduotims atlikti naudojamos IKT indeksas (koreliacijos koeficientas $0{,}382$). Galima buvo įtarti, kad penkiolikmetis, kuris namuose turi IKT, atlikdamas užduotis mokyklai jomis naudosis. Statistiškai reikšmingai nekoreliuoja (p-reikšmė $0{,}269$) mokinių matematinis raštingumas ir mokyklos užduotims atlikti naudojamos IKT indeksas.

2.2. SAS procedūra PROC MIXED

Hierarchinių duomenų analizė SAS programoje atliekama naudojant PROC MIXED procedūrą. Procedūros MIXED sintaksė:

```
PROC MIXED <pasirinktys>;
      CLASS <kategoriniai kintamieji>;
      MODEL <Priklausomas kintamasis> = <Fiksuotieji kintamieji> /
<pasirinktys>;
      RANDOM INTERCEPT <Atsitiktiniai kintamieji> /
            SUBJECT = <grupavimo kintamieji> <pasirinktys>;
      WEIGHT <kintamasis>;
      ODS OUTPUT <Lentelė > = <Rezultatas>;
RUN;
```

Sakinyje **PROC MIXED** pasirinktimi METHOD nurodoma kaip bus vertinami parametrai, pagal nutilėjimą naudojamas apriboto didžiausio tikėtinumo (REML) metodas. Atlikdami analizę įvertiniams sakaičiuoti naudosime mažiausių kvadratų (ML) metodas, kuris sintaksėje nurodomas įrašius METHOD=ML. Pasirinktimi COVTEST prašoma pateikti kovariacijų įvertinių standartines paklaidas, Voldo z–testo reikšmes ir p–reikšmes. Sakinyje CLASS nurodomi kategoriniai kintamieji.

Sakinio MODEL kairėje lygybės pusėje nurodomas priklausomas kintamasis, o dešinėje pusėje – fiksuotieji modelio kintamieji. Pasirinktis SOLUTION atspausdina fiksuotų kintamųjų įvertinius. Nurodžius pasirinktį CHISQ tikrinama hipotezė apie fiksuoto poveikio parametrų lygybę nuliui, o pasirinktis COVB apskaičiuoja fiksuoto poveikio parametrų standartines paklaidas. Pasirinktis OUTP=<Rezultatas> į lentelę įrašo prognozės rezultatus.

Sakinyje RANDOM reikia nurodyti laisvąjį narį, nurodoma įrašius INTERCEPT, bei išvardijami visi atsitiktiniai kintamieji. Pasirinktimi SUBJECT nurodomi grupavimo kintamieji, o SOLUTION atspausdina atsitiktinių kintamųjų įvertinius. Pasirinktis TYPE nurodo kovariacijų matricos pavidalą, pavyzdžiui, nuodžius UN įvertinamos visos kintamųjų kovariacijos ir dispersijos (2.1).

Sakinyje WEIGHT nurodomas kintamasis, kurio reikšmės yra stebėjimų svoriai. Sakinyje ODS OUTPUT nurodoma kuriuos procedūros MIXED rezultatus įrašyti į lenteles. Pavyzdžiui, pasirinktis SOLUTIONF=<Lentelė> į lentelę įrašoma fiksuotų parametrų įvertiniai, SOLUTIONR=<Lentelė> į lentelę įrašoma atsitiktinių parametrų įvertiniai, COVPARMS=<Lentelė> į lentelę įrašoma kovariacijos parametrų įvertiniai.

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{\ 1} & \tau_{2n} & \cdots & \tau_{\ n} \end{bmatrix} \tag{2.1}$$

2.3 Analizė taikant hierarchinę regresiją

2.3.1 Besąlyginio modelio analizė

Kaip buvo minėta 1.1 skyrelyje atliekant duomenų analizę naudojami mokinių, mokyklos svoriai. Atliekant hierarchinę analizę svoriai turi būti normalizuoti, t.y. svorių suma turi būti lygi mokinių skaičiui duomenų rinkinyje, o ne populiacijai. Mokinių svorių normalizavimui atlikti naudota SAS sintaksė bei visos analizei naudotos sintaksės pateiktos 1 Priede.

NAME	_LABEL_	N	MIN	MAX	MEAN	STD	miss
ESCS	ESCS	4481	-4,8453	2,75	-0,036	0,966	1,03799
ST04Q01	ST04Q01	4528	1,0000	2,00	1,505	0,500	0,00000
ICTHOME	ICTHOME	4493	-4,1929	1,41	-0,472	0,901	0,77297
ST22Q01	ST22Q01	4487	1,0000	6,00	3,035	1,368	0,90548
ST31Q02	ST31Q02	4441	1,0000	2,00	1,828	0,377	1,92138
ST31Q06	ST31Q06	4448	1,0000	2,00	1,850	0,358	1,76678
SCHSIZE	SCHSIZE	4500	23,0000	1556,00	629,686	313,107	0,61837
SC11Q02	SC11Q02	4485	1,0000	4,00	1,186	0,519	0,94965
STUDBEHA	STUDBEHA	4500	-3,4173	2,36	0,188	0,719	0,61837
TEACBEHA	TEACBEHA	4500	-3,0014	2,12	0,590	0,685	0,61837
ABGROUP	ABGROUP	4449	1,0000	3,00	1,924	0,616	1,74470
IRATCOMP	IRATCOMP	4362	0,0400	2,50	0,476	0,291	3,66608
SC05Q01	SC05Q01	4500	1,0000	3,00	1,649	0,796	0,61837
CULTPOSS	CULTPOSS	4469	-1,5394	0,97	0,005	0,905	1,30300
HOMSCH	HOMSCH	4471	-1,9204	3,04	0,059	0,949	1,25883
WEALTH	WEALTH	4510	-3,0661	3,12	-0,228	0,891	0,39753
ENTUSE	ENTUSE	4478	-3,0980	2,99	0,191	1,004	1,10424
STUDREL	STUDREL	4495	-2,8994	2,45	0,140	1,095	0,72880
DISCLIMA	DISCLIMA	4498	-2,8091	1,84	0,296	0,950	0,66254
PARINVOL	PARINVOL	4438	-1,5050	3,88	-0,066	0,994	1,98763

2.3 pav. Informacija apie kintamuosius

Duomenų rinkinyje pasitaikė neatsakymų į klausimus, t.y. kai kurie mokiniai, mokyklos ar tėvai neatsakė į kai kuriuos klausimus. Kadangi neatsakymų dažnis žemas, mažiau nei 2 % (2.3 pav.), todėl praleistos reikšmės vietoje įrašėme suvidurkintą reikšmę. Jeigu mokinio klausimyne pasitaikė praleista reikšmė, tai vietoje praleistos reikšmės įrašėme mokyklos, kurioje tas mokinys mokosi, vidurkį, o vietoje praleistos mokyklos klausimyno reikšmės – šalies vidurkį. Taip buvo pasielgta norint duomenų analizei turėti didžiausią galimų atvejų (mokinių) skaičių, t. y. turėti pilną duomenų rinkinį. Nors dėl praleistų reikšmių įrašymo atsiranda įvertinių paslinktumas, tačiau kadangi neatsakymų dažnis mažas, paslinktumas nežymus [7].

Norėdami išsiaiškinti ar duomenų analizei yra tinkami hierarchiniai modeliai, pirmiausiai analizuosime tiesinį hierarchinį besąlyginį modelį. Grupavimo kintamasis SCHOOLID (mokyklos kodas) nurodo kuriai mokyklai priklauso mokinys. Mus dominantis mokinio matematinis raštingumas – MAT1. Atlikdami analizę naudosime pirmąją matematinio raštingumo galimą reikšmę, o sudarant galutinį modelį, regresijos koeficientus apskaičiuosime visoms penkioms galimoms reikšmėms ir suvidurkine gausime galutinius regresijos koeficientus.

Kaip buvo aprašyta 1.4.2 skyrelyje, besąlyginio modelio pirmajį lygmenį sudaro atsitiktinis koeficientas β_{0j} ir atsitiktinė paklaida e_{ij}, o antrajį lygmenį – konstanta γ_{00} ir atsitiktinė paklaida u_{0j}. Tuomet pirmojo (mokinio) lygmens lygtis:

$$MAT1_{ij} = \beta_{0j} + e_{ij}, \quad i = 1,\ldots,n_j,\ j = 1,\ldots,196.$$

Antrojo (mokyklos) lygmens lygtis:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad j = 1,\ldots,196.$$

Žemiau pateiktas programos SAS procedūros MIXED sintaksė besąlyginiam modeliui:

```
PROC MIXED DATA=bendras COVTEST METHOD=ML;
      CLASS schoolid;
      MODEL mat1= /SOLUTION CHISQ;
      RANDOM intercept / SUBJECT = schoolid TYPE=UN SOLUTION;
      WEIGHT msvoris;
RUN;
```

Covariance Parameter Estimates

atsitiktinio poveikio parametrų įvertiniai ir reikšmingumai

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
$\hat{\tau}_{00}$	UN(1,1)	SCHOOLID	2434.41	271.69	8.96	<.0001
$\hat{\sigma}^2$	Residual		5280.91	113.45	46.55	<.0001

Fit Statistics

-2 Log Likelihood	52189.3
AIC (smaller is better)	52195.3
AICC (smaller is better)	52195.3
BIC (smaller is better)	52205.2

informacinių indeksų reikšmės

lyginimo su besąlyginiu modeliu testas

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	1263.20	<.0001

Solution for Fixed Effects

fiksuoto poveikio parametro įvertis ir reikšmingumas

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
$\hat{\gamma}_{00}$	Intercept	470.37	3.7127	195	126.69	<.0001

2.4 pav. Procedūros MIXED rezultatas taikant besąlyginį modelį

Rezultatai pateikiami 2.4 paveiksle. Atsitiktinių paklaidų dispersijų įverčiai: individualių mokinių skirtumus apibūdinantis dispersijos įvertinys $\sigma^2 = 5280{,}91$ ir skirtumus tarp mokyklų apibūdinantis dispersijos įvertinys $\tau_{00} = 2434{,}41$. Hipotezės $H_0 : \sigma^2 = 0$ ir $H_0 : \tau_{00} = 0$ atmetamos su reikšmingumo lygmeniu 0,05. Taigi tiek mokinių, tiek mokyklų rezultatų skirtumus lemia kažkokie mums nežinomi kintamieji (tolesnis žingsnis išsiaiškinti kokie tai kintamieji). Galime apskaičiuoti ICC koeficientą ir pasižiūrėti kaip stipriai skiriasi rezultatai tarp mokyklų, palyginti su rezultatais tarp mokinių:

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}} = \frac{2434{,}41}{5280{,}91 + 2434{,}41} = 0{,}32 .$$

Galime teigti, kad 32 % mokinių matematinio raštingumo balų skirtumų lemia mokyklos. Kadangi ICC koeficientas yra didelis, todėl galime daryti išvadą, kad reikia atsižvelgti į hierarchinę duomenų struktūrą.

Fiksuoto poveikio parametro įvertinys, nusakantis visų mokyklų matematinio raštingumo balų vidurkį, yra $\gamma_{00} = 470{,}37$. Toje pačioje eilutėje pateikta statistinės hipotezės $H_0 : \gamma_{00} = 0$, $H_1 : \gamma_{00} \neq 0$ p-reikšmė. Pagal nutilėjimą reikšmingumo lygmuo $\alpha = 0{,}05$. Kadangi p-reikšmė mažesnė už 0,05, galime daryti išvadą, kad $\gamma_{00} \neq 0$, t.y. nulinė hipotezė atmetama. Tačiau, kaip buvo minėta 1.4.2 skyrelyje, ši išvada nėra vertinga, kadangi ji parodo, kad vidutinis mokinių matematinis raštingumas nėra lygūs nuliui.

2.3.2 Modelio su pirmojo lygmens kintamaisiais analizė

Besąlyginio modelio analizė atskleidė, kad duomenys yra hierarchinės struktūros ir duomenų anlizei turėtume naudoti hierarchinius tiesinius modelius. Pirmiausiai į modelio pirmajį (mokinio) lygmenį įtraukime kintamuosius, galinčius paaiškinti skirtumus tarp penkiolikmečių matematinio raštingumo balų, tai:

- ESCS – mokinio socialinis ir ekonominis statusas;
- LYTIS – mokinio lytis;
- pseudo kintamasis N_KALBA – kalba, kuria mokinys dažniausiai kalba namuose: 1 – lietuvių kalba, 0 – kita kalba;
- ICTHOME – mokinio namuose esančių IKT indeksas, ar mokinio namuose yra kompiuteris, nešiojamasis kompiuteris, spausdintuvas, interneto ryšys, mobilusis telefonas, mp3 grotuvas, internetinis ryšys ir pan.;
- pseudo–kintamasis KNYGOS – turimų knygų skaičius namuose: 0 – nuo 0 iki 25 knygų, 1 – daugiau kaip 25 knygos;

Nors 2.1 skyrelyje atlikdami paprastą duomenų analizę gavome, kad merginų ir vaikinų matematinio raštingumo balai skiriasi tik 7 balai, tačiau kintamąjį lytis įtraukime į modelį norėdami patikrinti ar tikrai mokinio lytis yra statistiškai nereikšminga. Šio modelio (1M[12]) pirmojo lygmens lygtis:

$$MAT1_{ij} = \beta_{0j} + \beta_{1j}(ESCS)_{ij} + \beta_{2j}(LYTIS)_{ij} + \beta_{3j}(N\_KALBA)_{ij} + \beta_{4j}(ICTHOME)_{ij} + \\ + \beta_{5j}(KNYGOS)_{ij} + e_{ij}, \quad i = 1,\dots,n_j, \; j = 1,\dots,196,$$

antrojo lygmens lygtys:

[12] Numeruosime modelius, kad išvengtume painiavos lyginant skirtingus modelius

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j}, \\ \beta_{1j} = \gamma_{10} + u_{1j}, \\ \beta_{2j} = \gamma_{20} + u_{2j}, \\ \beta_{3j} = \gamma_{30} + u_{3j}, \\ \beta_{4j} = \gamma_{40} + u_{4j}, \\ \beta_{5j} = \gamma_{50} + u_{5j}, \end{cases} \qquad j = 1,\ldots,196.$$

Galime užrašyti jungtinę lygtį, kuri atrodo taip:

$$MAT1_{ij} = \gamma_{00} + \gamma_{10}(ESCS)_{ij} + \gamma_{20}(LYTIS)_{ij} + \gamma_{30}(N\_KALBA)_{ij} + \gamma_{40}(ICTHOME)_{ij} +$$
$$+ \gamma_{50}(KNYGOS)_{ij} + [u_{0j} + u_{1j}(ESCS)_{ij} + u_{2j}(LYTIS)_{ij} + u_{3j}(N\_KALBA)_{ij} + u_{4j}(ICTHOME)_{ij} +$$
$$+ u_{5j}(KNYGOS)_{ij} + r_{ij}].$$

Tariame, kad visi kintamieji yra atsitiktiniai, t. y. turi paklaidas u_j. Jungtinėje lygtyje atsitiktiniai kintamieji yra laužtinuose skliaustuose, o SAS programos MIXED procedūros sintaksėje fiksuoti kintamieji įrašomi MODEL, o atsitiktiniai kintamieji RANDOM sakiniuose. Žemiau pateiktas pirmojo modelio (1M) SAS sintaksė, o 2.5 paveiksle programos rezultatai:

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs lytis n_kalba icthome knygos/ SOLUTION chisq;
      RANDOM intercept escs lytis n_kalba icthome knygos / SUBJECT =
schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

Hipotezė $H_0: \gamma_{20} = 0$ neatmetama, t. y. kintamasis LYTIS statistiškai reikšmingai nesiskiria nuo nulio. Taigi mokinio lytis statistiškai reikšmingai neturi įtakos jo matematiniam raštingumui. Iš pirmos lentelės (*Covariante Parameter Estimate*) pastebime, kad hipotezės $H_0: \tau_{00} = 0$ ir $H_0: \sigma^2 = 0$ atmetamos su reikšmingumo lygmeniu 0,05. Taigi mokinių matematinis raštingumas statistiškai reikšmingai skiriasi dėl mums nežinomų mokyklų kintamųjų (įvertinys τ_{00} didelis) ir mokinių individualių savybių (įvertinys σ^2 didelis).

Atmesta hipotezė $H_0: \tau_{44} = 0$ parodo, kad įvertinys τ_{44} statistiškai reikšmingai skiriasi nuo nulio, t. y. koeficientas β_{4j} atsitiktinis. Taigi galime įtarti, kad matematinio raštingumo priklausomybės nuo IKT turėjimo namuose indekso (kintamojo ICTHOME) yra skirtingos visose mokyklose, t.y. vienose mokyklose šios priklausomybės stipresnės, kitose silpnesnės. Kitoms τ_{i}, kur $i = \,,2,3,5$, nulinės hipotezės neatmetamos, taigi galime įtarti, kad koeficientai β β ,β ,β neatsitiktiniai.

Redukuokime modelį pašalindami kintamąjį LYTIS ir tardami, kad koeficientai $\beta = 0, \beta = 0, \beta = 0, \beta = 0$. Redukuoto modelio jungtinė lygtis atrodo taip:

$$MAT1_{ij} = \gamma_{00} + \gamma_{0}(ESCS)_{ij} + \gamma_{0}(N\_KALBA)_{ij} + \gamma_{0}(ICTHOME)_{ij} + \gamma_{0}(KNYGOS)_{ij} +$$
$$+ u_{0j} + u_{3j}(ICTHOME)_{ij} + r_{ij}].$$

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
$\hat{\tau}_{00}$	UN(1,1)	SCHOOLID	927,23	467,85	1,98	0,0237
	UN(2,1)	SCHOOLID	17,8306	106,95	0,17	0,8676
$\hat{\tau}_{11}$	UN(2,2)	SCHOOLID	3,5068	38,8616	0,09	0,4640
	UN(3,1)	SCHOOLID	-94,9365	173,59	-0,55	0,5844
	UN(3,2)	SCHOOLID	-59,8864	41,5783	-1,44	0,1498
$\hat{\tau}_{22}$	UN(3,3)	SCHOOLID	127,11	100,53	1,26	0,1031
	UN(4,1)	SCHOOLID	161,59	309,68	0,52	0,6018
	UN(4,2)	SCHOOLID	-16,4269	102,07	-0,16	0,8721
	UN(4,3)	SCHOOLID	-197,95	162,54	-1,22	0,2233
$\hat{\tau}_{33}$	UN(4,4)	SCHOOLID	232,88	314,68	0,74	0,2296
	UN(5,1)	SCHOOLID	-58,8274	121,71	-0,48	0,6288
	UN(5,2)	SCHOOLID	-8,1569	32,7824	-0,25	0,8035
	UN(5,3)	SCHOOLID	-17,1191	50,2286	-0,34	0,7332
	UN(5,4)	SCHOOLID	-35,1817	111,20	-0,32	0,7517
$\hat{\tau}_{44}$	UN(5,5)	SCHOOLID	128,95	49,4865	2,61	0,0046
	UN(6,1)	SCHOOLID	108,33	178,71	0,61	0,5444
	UN(6,2)	SCHOOLID	74,0372	52,0227	1,42	0,1547
	UN(6,3)	SCHOOLID	-43,1020	85,0795	-0,51	0,6124
	UN(6,4)	SCHOOLID	-101,22	158,62	-0,64	0,5234
	UN(6,5)	SCHOOLID	-92,6199	58,0600	-1,60	0,1107
$\hat{\tau}_{55}$	UN(6,6)	SCHOOLID	123,78	127,96	0,97	0,1667
$\hat{\sigma}^2$	Residual		4546,31	107,19	42,41	<,0001

Solution for Fixed Effects

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
$\hat{\gamma}_{00}$	Intercept	428,12	5,1373	195	83,34	<,0001
	ESCS	21,3358	1,3889	195	15,36	<,0001
$\hat{\gamma}_{20}$	lytis	-2,2242	2,2460	193	-0,99	0,3233
	n_kalba	28,3299	4,7971	84	5,91	<,0001
	ICTHOME	-6,2980	1,5426	194	-4,08	<,0001
	knygos	30,2881	2,5815	191	11,73	<,0001

2.5 pav. Procedūros MIXED rezultatas taikant 1M modelį

Iš RANDOM sakinio pašaliname kintamuosius ESCS, LYTIS, N_KALBA ir KNYGOS, o kintamąjį ICTHOME paliekame. Redukuoto modelio (1RM) SAS sintaksė:

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs n_kalba icthome knygos/ SOLUTION chisq;
      RANDOM intercept icthome/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

Redukuoto modelio rezultatai pateikti 2.6 pav. Palyginti su pilnu modeliu informaciniai indeksai sumažėjo (pavyzdžiui, pilname modelyje AIC=51611,2, o redukuotame modelyje AIC=51599,7), taigi galime teigti, kad redukuotasis modelis geresnis nei pilnasis.

Pirmojo lygmens atsitiktinio parametro įvertis $\sigma^2 = 4634{,}12$ ir šis įvertis statistiškai reikšmingai didesnis už nulį, t.y p-reikšmė mažesnė už 0,05. Besąlyginiame modelyje pirmojo lygmens atsitiktinio parametro įvertis buvo $\sigma^2 = 5280{,}91$, galime apskaičiuokime santykinį įverčio pokytį:

$$\frac{5280{,}91 - 4634{,}12}{5280{,}91} = 0{,}122,$$

Taigi į modelį įtraukus papildomus kintamuosius 12,2 % geriau paaiškinami skirtumai tarp mokinių. Atrojo lygmens atsitiktinio parametro įvertis $\tau_{00} = 1299{,}32$, palyginti su besąlyginiu modeliu ($\tau_{00} = 2434{,}41$), taip pat sumažėjo. Apskaičiavę santykinį įverčio pokytį:

$$\frac{2434{,}41 - 1299{,}32}{2434{,}41} = 0{,}466,$$

Matome, kad šis modelis (1RM) 46,6 % geriau paaiškina matematinio raštingumo balų skirtumus tarp mokyklų.

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
$\hat{\tau}_{00}$	UN(1,1)	SCHOOLID	1229,32	156,28	7,87	<,0001
$\hat{\tau}_{03}$	UN(2,1)	SCHOOLID	-140,68	56,8864	-2,47	0,0134
$\hat{\tau}_{33}$	UN(2,2)	SCHOOLID	118,19	42,1910	2,80	0,0025
$\hat{\sigma}^2$	Residual		4634,12	101,89	45,48	<,0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	427,30	5,2619	195	81,21	<,0001
ESCS	21,4862	1,3886	4133	15,47	<,0001
n_kalba	28,6492	4,7177	4133	6,07	<,0001
ICTHOME	-6,1137	1,5179	195	-4,03	<,0001
knygos	30,0482	2,4494	4133	12,27	<,0001

2.6 pav. Procedūros MIXED rezultatas taikant redukuotą (1RM) modelį

Nulinė hipotezė $H_0 : \tau_{33} = 0$ lieka atmesta, t.y. matematinio raštingumo priklausomybės nuo IKT turėjimo namuose indekso (ICTHOME) visose mokyklose yra skirtingos. Atmesta hipotezė

$H_0: \tau_{03} = 0$, parodo, kad β_{0j} ir β_{3j} statistiškai reikšmingai koreliuoja, t. y. mokyklos matematinio raštingumo vidurkis koreliuoja su IKT turėjimu namuose (ICTHOME) indeksu. Kadangi įvertinys $\tau_{03} = -40{,}68$ yra neigiamas, tai mokykloje kurioje matematinio raštingumo vidurkis didesnis, mokinių rezultatų priklausomybė nuo ICTHOME indekso yra silpnesnė ir atvirkščiai (4.2 pav. (2)).

Antroje lentelėje (2.6 pav.) pateikti fiksuotų parametrų įverčiai (*Solution for Fixed Effects*). Šiuos įverčius įstatę į jungtinę lygtį gautume:

$$MAT1_{ij} = 427{,}30 + 21{,}49(ESCS)_{ij} + 28{,}65(N\_KALBA)_{ij} - 5{,}11(ICTHOME)_{ij} + 30{,}05(KNYGOS)_{ij}.$$

Iš lygties matome, kad mokinio socialinio ir ekonominio (ESCS) statuso padidėjimas vienetu prie mokinio matematinio raštingumo rezultato prideda 21,49 balo. Taip pat mokinio, kuris dažniausiai namuose kalba lietuvių kalba arba kurio namuose yra daugiau nei 26 knygos, matematinis raštingumas, atitinkamai 28,62 arba 30,05 balo, aukštesnis.

Papildykime modelį į pirmąjį lygmenį įtraukdami šiuos pseudo–kintamuosius ir tarkime, kad jie atsitiktiniai:

MOKMAT: 1 – mokinys papildomai mokosi matematikos, 0 – ne;

PAGMAT: 1 – mokinys lanko pataisomojo pobūdžio matematikos užsiemimus, 0 – jei ne;

GERMAT: 1 – mokinys lanko sustiprintą matematiką, 0 – jeigu ne;

- NAGLENT: 1 – mokinys dažnai arba labai dažnai nagrinėja lenteles: 0 – retai arba niekada.

Naujojo modelio (2M) jungtinė lygtis:

$$\begin{aligned} MAT1_{ij} &= \gamma_{00} + \gamma_{10}(ESCS)_{ij} + \gamma_{20}(N\_KALBA)_{ij} + \gamma_{30}(ICTHOME)_{ij} + \gamma_{40}(KNYGOS)_{ij} \\ &+ \gamma_{50}(MOKMAT)_{ij} + \gamma_{60}(GERMAT)_{ij} + \gamma_{70}(PAGMAT)_{ij} + \gamma_{80}(NAGLENT)_{ij} + u_{0j} + u_{3j}(ICTHOME)_{ij} + \\ &+ u_{5j}(MOKMAT)_{ij} + u_{6j}(GERMAT)_{ij} + u_{7j}(PAGMAT)_{ij} + u_{8j}(NAGLENT)_{ij} + r_{ij}]. \end{aligned}$$

SAS programos MIXED procedūros sintaksė naujajam (2M) modeliui:

```
PROC MIXED DATA=bendras covtest METHOD=ML;
      CLASS schoolid;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent /
SOLUTION chisq;
      RANDOM intercept icthome mokmat germat pagmat naglent/ SUBJECT = schoolid
TYPE=UN;
      WEIGHT msvoris;
RUN;
```

Rezultatai pateikti 2.7 paveiksle. Palyginti su ankstesniu modeliu sumažėjo visi informaciniai indeksai (2 Priedas. 2.4). Pirmojo lygmens atsitiktinis parametro įvertinys $\sigma^2 = 4146{,}97$ sumažėjo

10,5 %, antrojo lygmens atsitiktinis parametro įvertinys $\tau_{10} = 298{,}19$ nežymiai sumažėjo. Taigi galime daryti išvadą, kad naujasis modelis (2M) yra geresnis, palyginti su ankstesniuoju (1RM).

Hipotezės $H_0 : \tau_{33} = 0$ ir $H_0 : \tau_{66} = 0$ atmetamos, t.y. mokinio matematinio raštingumo priklausomybės nuo mokinio turėjimo namuose IKT indekso (ICTHOME) ir papildomų matematikos pamokų lankymo (GERMAT) yra skirtingos mokyklose. Atmesta hipotezė $H_0 : \tau_{13} = 0$ atskleidžia, kad β_{j} ir β_{j} statistiškai reikšmingai koreliuoja, taip pat atmetama hipotezė $H_0 : \tau_{8} = 0$ parodo, kad β_{j} ir β_{j} statistiškai reikšmingai koreliuoja.

Matome, kad yra neatmetamos nulinės hipotezės $H_0 : \tau_{5} = 0$, $H_0 : \tau_{7} = 0$ ir $H_0 : \tau_{8} = 0$, todėl galime įtarti, kad β_{j}, β_{j} ir β_{j} yra neatsitiktiniai. Taip pat matome, kad visi naujai į modelį įtraukti kintamųjų iverčiai γ_{0}, γ_{0}, γ_{0} ir γ_{0} yra statistiškai reikšmingi (su reikšmingumo lygmeniu 0,05).

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
	UN(1,1)	SCHOOLID	1298,19	215,82	6,02	<,0001
	UN(2,1)	SCHOOLID	-133,99	64,6877	-2,07	0,0383
$\hat{\tau}_{33}$	UN(2,2)	SCHOOLID	99,7663	38,6394	2,58	0,0049
	UN(3,1)	SCHOOLID	-43,8243	132,71	-0,33	0,7412
	UN(3,2)	SCHOOLID	36,5666	50,6122	0,72	0,4700
$\hat{\tau}_{55}$	UN(3,3)	SCHOOLID	197,66	142,67	1,39	0,0830
	UN(4,1)	SCHOOLID	-45,7663	167,96	-0,27	0,7852
	UN(4,2)	SCHOOLID	5,9377	69,5099	0,09	0,9319
	UN(4,3)	SCHOOLID	124,35	130,84	0,95	0,3419
$\hat{\tau}_{66}$	UN(4,4)	SCHOOLID	603,60	215,48	2,80	0,0025
	UN(5,1)	SCHOOLID	-79,3945	138,86	-0,57	0,5675
	UN(5,2)	SCHOOLID	-6,8562	62,0500	-0,11	0,9120
	UN(5,3)	SCHOOLID	112,92	102,80	1,10	0,2720
	UN(5,4)	SCHOOLID	80,3506	155,26	0,52	0,6048
$\hat{\tau}_{77}$	UN(5,5)	SCHOOLID	2,66E-30	.	.	.
	UN(6,1)	SCHOOLID	84,7931	117,94	0,72	0,4722
	UN(6,2)	SCHOOLID	-117,43	50,2275	-2,34	0,0194
	UN(6,3)	SCHOOLID	-10,7399	95,2906	-0,11	0,9103
	UN(6,4)	SCHOOLID	140,79	124,43	1,13	0,2578
	UN(6,5)	SCHOOLID	122,24	105,03	1,16	0,2445
$\hat{\tau}_{88}$	UN(6,6)	SCHOOLID	127,37	122,06	1,04	0,1484
	Residual		4146,97	96,3678	43,03	<,0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	415,53	5,4188	195	76,68	<,0001
ESCS	20,6792	1,3317	3427	15,53	<,0001
n_kalba	25,8459	4,5770	3427	5,65	<,0001
ICTHOME	-5,4912	1,4325	195	-3,83	0,0002
knygos	27,4390	2,3433	3427	11,71	<,0001
mokmat	32,4947	2,6231	185	12,39	<,0001
germat	9,0385	3,5173	169	2,57	0,0110
pagmat	-25,4948	2,9561	163	-8,62	<,0001
naglent	8,3034	2,6073	185	3,18	0,0017

2.7 pav. Procedūros MIXED rezultatas taikant 2M modelį

Redukuokime modelį tardami, kad $\beta_{j} = 0$, $\beta_{j} = 0$ ir $\beta_{j} = 0$. Redukuotasis modelis (2.8 pav.) yra šiek tiek geresnis modelis nei pilnasis modelis (informaciniai indeksai sumažėjo (2 Priedas 2.5 pav)), nors dispersijos įverčiai τ_{00} ir σ šiek tiek padidėjo. Darome išvadą, kad redukuotasis modelis (2RM) yra geresnis nei pilnasis modelis (2M).

Į jungtinę lygtį įrašę regresijos koeficientus galime užrašyti lygtį, kurią naudotume prognozuojant mokinių matematinį raštingumą:

$$MAT1_{ij} = 413{,}58 + 20{,}51(ESCS)_{ij} + 26{,}32(N\_KALBA)_{ij} - 5{,}39(ICTHOME)_{ij} + 27{,}76(KNYGOS)_{ij} +$$
$$+ 33{,}33(MOKMAT)_{ij} + 8{,}99(GERMAT)_{ij} - 25{,}63(PAGMAT)_{ij} + 9{,}16(NAGLENT)_{ij}.$$

Iš lygties matome, kad mokinio, kuris papildomai mokosi matematikos (MOKMAT), matematinis raštingumas vidutiniškai 33,33 balo aukštesnis, negu mokinio, kuris papildomai nesimoko matematikos. Mokinio, kuris lanko sustiprintas matematikos pamokas ar nagrinėja tekste esančias lenteles, matematinis raštingumas atitinkamai 8,99 arba 9,16 balo aukštesnis. Tačiau penkiolikmečio, kuris lanko ištaisomojo pobūdžio matematikos užsiemimus, matematinis raštingumas vidutiniškai 25,63 balo žemesnis. Galime įtarti, kad klaidų ištaisymas, o ne leidimas pačiam mokiniui surasti klaidą ir ją ištaisyti, neigiamai veikia mokinių gebėjimus, mąstymą.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	1391.46	175.16	7.94	<.0001
UN(2,1)	SCHOOLID	-127.24	57.7534	-2.20	0.0276
UN(2,2)	SCHOOLID	105.31	39.0335	2.70	0.0035
UN(3,1)	SCHOOLID	39.3658	144.78	0.27	0.7857
UN(3,2)	SCHOOLID	-8.2398	67.8194	-0.12	0.9033
UN(3,3)	SCHOOLID	572.44	199.69	2.87	0.0021
Residual		4183.15	93.7162	44.64	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	413.58	5.4526	195	75.85	<.0001
ESCS	20.5119	1.3352	3957	15.36	<.0001
n_kalba	26.3198	4.5955	3957	5.73	<.0001
ICTHOME	-5.3860	1.4482	195	-3.72	0.0003
knygos	27.7649	2.3481	3957	11.82	<.0001
mokmat	33.3331	2.4022	3957	13.88	<.0001
germat	8.9851	3.5068	172	2.56	0.0113
pagmat	-25.6295	2.9846	3957	-8.59	<.0001
naglent	9.1627	2.4683	3957	3.71	0.0002

2.8 pav. 2RM modelis

Rezultatai parodo, kad nulinė hipotezė $H_0 : \sigma = 0$ atmetama su reikšmingumo lygmeniu 0,05, t.y. mokinių matematinis raštingumas skiriasi dėl individualių mokinių charakteristikų. Taip pat atmetama nulinė hipotezė $H_0 : \tau_{00} = 0$ ($p < 0{,}05$), t.y. mokinių matematinis raštingumas mokyklose statistiškai skiriasi dėl mokyklų charakteristikų.

2.3.3 Modelio su antrojo lygmens kintamaisiais analizė

Kadangi nulinė hipotezė $H_0 : \tau_{00} = 0$ atmetama ($p < 0{,}05$), t.y. mokinių matematinis raštingumas mokyklose statistiškai skiriasi dėl mokyklų charakteristikų, todėl į modelį įtraukime antrojo, mokyklos, lygmens kintamuosius:

- M_ESCS – mokyklų socialinis ir ekonominis statusas, jis apskaičiuojamas kiekvienoje mokykloje suvidurkinant tos mokyklos mokinių socialinį ir ekonominį statusą (ESCS):

  $M\_ESCS = \frac{1}{n_j}\sum_{i=1}^{n_j} ESCS_i$, kur n_j – mokinių skaičius *j*-ojoje mokykloje;
- kategorinis kintamasis TIPAS – mokyklos tipas: 1 – pagrindinė mokykla, 2 – vidurinė mokykla, 3 – gimnazija, 0 – kito tipo mokykla;
- pseudo kintamasis TRMOK: 1 – mokykloje trūksta matematikos mokytojų, 0 – ne;
- STUDBEHA – mokinių elgesio mokykloje indeksas. Į indekso skaičiavimą įtraukiama mokinių pamokų praleidinėjimas, mokytojų negerbimas, ar mokiniai naudoja alkoholį, narkotikus ir pan.;
- TEACBEHA – mokytojų elgesio mokykloje indeksas. Indeksas apskaičiuojamas atsižvelginat į mokinių ir mokytojų santykius, mokytojų griežtumą su mokiniais ir pan.;
- IRATCOMP - kompiuetrių ir mokinių skaičiaus santykis mokykloje;
- Kategorinis kintamasis SC04Q01 – vietovė kurioje įsikūrusi mokykla: 1 – kaimas, 2 – mažas miestelis, 3 – miestelis, 4 – miestas.

SAS programos MIXED procedūros sintaksė 3M modeliui:

```
PROC MIXED DATA=bendras covtest METHOD=ML;
      CLASS schoolid TIPAS sc04q01;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent
M_ESCS TIPAS TRMOK STUDBEHA TEACBEHA IRATCOMP sc04q01/ SOLUTION chisq;
      RANDOM intercept icthome germat/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

Modelio (3M) pirmojo lygmens lygtis:

$$MAT1_{ij} = \beta_{0j} + \beta_{1j}(ESCS)_{ij} + \beta_{2j}(N\_KALBA)_{ij} + \beta_{3j}(ICTHOME)_{ij} + \beta_{4j}(KNYGOS)_{ij} +$$
$$+ \beta_{5j}(MOKMAT)_{ij} + \beta_{6j}(GERMAT)_{ij} + \beta_{7j}(PAGMAT)_{ij} + \beta_{8j}(NAGLENT)_{ij} + r_{ij}.$$

antrojo lygmens lygtis (tik koeficientai β_{0j}, β_{3j} ir β_{6j} atsitiktiniai):

$$\beta_{j} = \gamma_{01}(M\_ESCS)_j + \gamma_{02}(TIPAS)_j + \gamma_{03}(TRMOK)_j + \gamma_{04}(STUDBEHA)_j + \gamma_{05}(TEACBEHA)_j +$$
$$+ \gamma_{06}(IRATCOMP)_j + \gamma_{07}(SC04Q01)_j + u_{0j},$$
$$\beta_{j} = \gamma_{0},$$
$$\beta_{j} = \gamma_{20},$$
$$\beta_{j} = \gamma_{0} + u_{3j},$$
$$\beta_{j} = \gamma_{40},$$
$$\beta_{j} = \gamma_{0},$$
$$\beta_{j} = \gamma_{60} + u_{6j},$$
$$\beta_{j} = \gamma_{0},$$
$$\beta_{j} = \gamma_{0}.$$

Aptarkime gautus rezultatus, kurie pateikti 2.9 paveiksle. Visi informaciniai indeksai (2 Priedas. 2.5) sumažėjo. Į modelį įtraukus mokyklos (antrojo) lygmens kintamuosius antrojo lygmens atsitiktinio parametro τ_{00} įvertinys sumažėjo net 52 %. Pirmojo lygmens atsitiktinio poveikio parametro σ įvertinys nežymaiai sumažėjo. Taigi palyginti su ankstesniu modeliu naujasis modelis yra geresnis ir 52 procentais geriau paaiškina skirtumus tarp mokyklų.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	668,41	100,09	6,68	<,0001
UN(2,1)	SCHOOLID	22,7959	47,0921	0,48	0,6283
UN(2,2)	SCHOOLID	106,28	38,3896	2,77	0,0028
UN(3,1)	SCHOOLID	108,82	113,08	0,96	0,3359
UN(3,2)	SCHOOLID	-5,3058	67,8547	-0,08	0,9377
UN(3,3)	SCHOOLID	586,27	202,23	2,90	0,0019
Residual		4180,16	93,8002	44,56	<,0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3931	192,48	192,48	<,0001	<,0001
n_kalba	1	3931	34,84	34,84	<,0001	<,0001
ICTHOME	1	194	13,87	13,87	0,0002	0,0003
knygos	1	3931	126,39	126,39	<,0001	<,0001
mokmat	1	3931	189,30	189,30	<,0001	<,0001
germat	1	171	6,19	6,19	0,0129	0,0138
pagmat	1	3931	76,05	76,05	<,0001	<,0001
naglent	1	3931	13,71	13,71	0,0002	0,0002
m_escs	1	3931	31,76	31,76	<,0001	<,0001
tipas	3	3931	23,34	7,78	<,0001	<,0001
trmok	1	3931	2,00	2,00	0,1578	0,1579
STUDBEHA	1	3931	7,39	7,39	0,0066	0,0066
TEACBEHA	1	3931	1,47	1,47	0,2247	0,2248
IRATCOMP	1	3931	0,32	0,32	0,5687	0,5688
SC04Q01	3	3931	1,66	0,55	0,6449	0,6450

2.9 pav. Procedūros MIXED rezultatas taikant 3M modelį

Kintamieji TRMOK, TEACBEHA, IRATCOMP ir mokyklos bendruomenė (SC04Q01) yra statistiškai nereikšmingi (2 Priedas 2.6), todėl redukuokime modelį iš jo pašalindami šiuos kintamuosius. Redukuotame modelyje (pilni rezultatai 2 Priedas 2.7 pav.) informaciniai indeksai sumažėjo. Pirmojo ir antrojo lygmenų atsitiktiniai parametrai šiek tiek padidėjo. Taigi pašalinus kintamuosius TRMOK, TEACBEHA, IRATCOMP ir kategorinį kintamąjį SC04Q01 gauname geresnį modelį.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	684,81	101,97	6,72	<,0001
UN(2,1)	SCHOOLID	24,5612	47,4726	0,52	0,6049
UN(2,2)	SCHOOLID	108,33	38,6364	2,80	0,0025
UN(3,1)	SCHOOLID	111,27	112,11	0,99	0,3210
UN(3,2)	SCHOOLID	-5,2766	68,1267	-0,08	0,9383
UN(3,3)	SCHOOLID	589,60	202,88	2,91	0,0018
Residual		4181,14	93,8115	44,57	<,0001

Solution for Fixed Effects

Effect	tipas	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		431,16	6,0500	189	71,27	<,0001
ESCS		18,8118	1,3567	3932	13,87	<,0001
n_kalba		26,5716	4,3610	3932	6,09	<,0001
ICTHOME		-5,4529	1,4571	194	-3,74	0,0002
knygos		26,5038	2,3506	3932	11,28	<,0001
mokmat		32,6004	2,3722	3932	13,74	<,0001
germat		8,8413	3,5338	171	2,50	0,0133
pagmat		-25,9912	2,9944	3932	-8,68	<,0001
naglent		9,0454	2,4707	3932	3,66	0,0003
m_escs		35,8325	5,4788	3932	6,54	<,0001
tipas	0	-73,5212	15,7570	3932	-4,67	<,0001
tipas	1	-18,8869	7,5396	3932	-2,51	0,0123
tipas	2	-20,1658	5,3582	3932	-3,76	0,0002
tipas	3	0	.	.	.	.
STUDBEHA		8,5936	3,1168	3932	2,76	0,0059

2.10 pav. Procedūros MIXED rezultatas taikant 3RM modelį

Aptarkime redukuoto modelio rezultatus, pateiktus 2.10 paveiksle. Mokyklos lygmens fiksuoto parametro γ_{01} įvertinys $\gamma_{01} = 5,83$, taigi matome, kad mokyklos socialiniui ir ekonominiui statusui padidėjus vienetu mokinio matematinis raštingumas vidutiniškai padidėja 35,83 balo. Tuo tarpu mokinio, kuris lanko kitą mokyklą, o ne pagrindinę, vidurinę ar gimnaziją (TIPAS = 0), matematinis raštingumas vidutiniškai sumažėja net 73,5 balo. Mokinio, kuris lanko vidurinę mokyklą, matematinis raštingumas vidutiniškai 20,17 balų žemesnis negu mokinio kuris lanko gimnaziją. Šio modelio prognozavimui naudojama lygtis atrodo taip:

$$MAT1_{ij} = 431{,}16 + 18{,}81(ESCS)_{ij} + 26{,}57(N\_KALBA)_{ij} - 5{,}45(ICTHOME)_{ij} + 26{,}5(KNYGOS)_{ij} +$$
$$+ 32{,}6(MOKMAT)_{ij} + 8{,}84(GERMAT)_{ij} - 25{,}99(PAGMAT)_{ij} + 9{,}05(NAGLENT)_{ij} + 35{,}83(M\_ESCS)_{j} +$$
$$+ \gamma_{02}(TIPAS)_{j} + 8{,}59(STUDBEHA)_{j}.$$

Lygtyje vietoj γ_{02} įrašomas atitinkamas regresijos koeficientas, pavyzdžiui jei kategorinis kintamasis TIPAS=2, t.y. mokinys mokosi vdurinėje mokykoje, tai regresijos koeficientas $\gamma_{02} = -0{,}17$.

Vis dar atmetamos nulinės hipotezės $H_0 : \tau_{03} = 0$ ir $H_0 : \tau_{06} = 0$, t.y. mokinio matematinio raštingumo priklausomybės nuo mokinio turėjimo namuose IKT indekso (ICTHOME) ir papildomų matematikos pamokų lankymo (GERMAT) mokyklose yra skirtingos. Norėdami paaiškinti ir sumažinti didelius dispersijos įverčius į β ir β lygtis įveskime mokyklos kintamuosius, kurie galėtų turėti įtakos. Taip pat modelį papildykime kai kurių kintamųjų sąveikomis ir naujais kintamaisiais:

- pseudo kintamasis TKOMP: 1 – mokinys namuose turi kompiuterį, 0 – ne;
- HOMSCH – ruošiant užduotis mokyklai naudojamasi esančiomis namuose IKT indeksas (naudojamasi internetu, mokyklos internetine svetaine ir pan.);
- WEALTH – mokinio šeimos turtingumo indeksas. Skaičiuojant indeksą atsižvelgiama ar mokinys turi savo kambarį, namuose yra indaplovė, televizorius, automobilis, vonios kambarys ir pan.;
- ENTUSE – IKT naudojimo pramogoms indeksas (žaidžia kompiuterinius žaidimus, siunčiasi muziką, filmus per internetą ir pan.);
- LIBUSE – naudojimosi bibliotekos paslaugomis indeksas;
- READRES – mokinių skaitymo šaltiniai namuose (elektroninis paštas, laikraščiai, žurnalai ir pan.);
- JOYREAD – mėgimo skaityti indeksas (ar mėgsta skaityti, nuomonė apie skaitymą ir pan.);
- STUDREL – mokytojų ir mokinių klasėje santykių indeksas (ar sutaria su mokytojais, reikiant pagalbos jos sulaukia, mokytojai su mokiniais elgiasi teisingai ir pan.);
- DISCLIMA – „klimato“ klasėje indeksas (mokiniai neklauso mokytojų, klasėje triukšmaujama, mokiniai pamokai prasidėjus ilgai nepradeda dirbti ir pan.);
- PARINVOL – tėvų įsitrukimo į vaiko lankomos mokyklos veiklą indeksas (tėvai padeda mokyklos bibliotekoje, organizuojant užklasines veiklas, su mokytojais aptaria vaiko pažangą ir elgesį, tėvai yra tėvų taryboje ir pan.);
- PCGIRLS – mergaičių ir berniukų santykis mokykloje.

Naujo (4M) modelio SAS programos MIXED procedūros sintaksė:

```
PROC MIXED DATA=bendras covtest METHOD=ML;
      CLASS schoolid TIPAS;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent
tkomp homsch wealth entuse libuse readres joyread studrel disclima
icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq ;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

Į senąjį modelį (3RM) vis įtraukiant kelių kintamųjų grupeles ir atsižvelgiant į p-reikšmes, informacinius indeksus, kintamųjų reikšmingumą buvo sprendžiama kokie kintamieji bus įtraukiami į modelį. Tačiau to nedemonstruosime pažingsniui, į galutinį modelį įtrauksime tik statistiškai reikšmingus kintamuosius. Sudarinėjant modelį kintamieji kaip, darželio lankymas (DARŽELIS), mokinys namuose turi mokymosi vietą (MVIETA), tėvų darbo tipas ir kvalifikacija (HSECATEG), mokyklos švietimo ištekliai (SCMATEDU), mokiniai dažnai gauna standartizuotus testus (STEST), mokiniai dažnai gauna mokytojų sudarytus testus (MTEST), arti yra kita mokykla (KITMOK) ir mokykla yra privati ar valstybinė (SC02Q01) buvo statistiškai nereikšmingi. Modelio, į kurį įtraukti statistiškai nereikšmingi kintamieji (4TM), programos SAS sintaksė ir rezultatas pateikti 1 Priede 1.21 ir 2 Priede 2.10 pav.

Į modelį įtraukdami sąveikas tikėjomės tiksliau nustatyti, kas lemia skirtumus tarp mokyklų (2.11 pav.). Hipotezės $H_0 : \tau_{3} = 0$ ir $H_0 : \tau_{6} = 0$ atmetamos, su reikšmingumo lygmeniu 0,05, taigi mokyklų skirtumus lemia nežinomi veiksniai ir parametrai β_{j} ir β_{j} yra atsitiktiniai. Į modelį įtraukėme naują kintamąjį: ruošiant užduotis mokyklai naudojamasi esančiomis namuose IKT indeksą (HOMSCH) ir tariame, kad jis atsitiktinis. Nulinė hipotezė $H_0 : \tau_{0,10} = 0$ atmetama, taigi parametras β yra atsitiktinis.

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
	UN(1,1)	SCHOOLID	580.47	87.0814	6.67	<.0001
	UN(2,1)	SCHOOLID	-1.3531	38.8593	-0.03	0.9722
	UN(2,2)	SCHOOLID	56.8756	32.7352	1.74	0.0412
	UN(3,1)	SCHOOLID	138.40	92.9272	1.49	0.1364
	UN(3,2)	SCHOOLID	-3.5170	54.2044	-0.06	0.9483
	UN(3,3)	SCHOOLID	401.34	164.55	2.44	0.0074
	UN(4,1)	SCHOOLID	12.9027	36.5530	0.35	0.7241
	UN(4,2)	SCHOOLID	-32.2595	21.6759	-1.49	0.1367
	UN(4,3)	SCHOOLID	25.7124	52.2315	0.49	0.6225
$\hat{\tau}_{10,10}$	UN(4,4)	SCHOOLID	55.0498	29.1264	1.89	0.0294
	Residual		3666.11	83.4199	43.95	<.0001

2.11 pav. Procedūros MIXED rezultatas taikant 4M modelį

Į modelį įtrauktų naujų kintamųjų fiksuoto poveikio parametrų γ įvertiniai statistiškai reikšmingai nelygūs nuliui (2.12 pav.). Mokinių, kurie namuose turi kompiuterį, vidutiniškai 17,9 balo aukštesni matematikos raštingumo rezultatai. Mokinio, kurio mėgimo skaityti indeksui padidėjus vienetu, matematinio raštingumo rezultatas vidutiniškai padidėja 14,4 balo, o mokinio IKT naudojimo pramogai indeksui padidėjus vienetu, jo matematinis raštingumo rezultatas vidutiniškai padidėja 7,15 balo, nors galima buvo tikėtis atvirkštinės įtakos, t. y., kad mokinio matematinis raštingumas sumažės padidėjus IKT naudojimo pramogoms indeksui. Tačiau mokinio mokyklos užduotims atlikti naudojimosi IKT (HOMSCH) arba turto (WEALTH) indeksui padidėjus vienetu mokinio matematinis raštingumas vidutiniškai sumažėja atitinkamai 11,51 arba 4,58 balo. Galime įtarti, kad mokiniai, kurie ruošdami užduotis mokyklai naudojasi namuose esančiomis IKT, naudojasi palengvinančiomis darbą programomis (kaip Excel programa) ir tai nelavina mokinių matematinio raštingumo.

Matome, kad yra svarbūs santykiai tarp mokinių ir mokytojų bei mokinių darbas klasėje, pavyzdžiui, jeigu klasė nedrausminga, neklauso mokytojo, tai galima įtarti, kad tokį elgesį perims visi ar bent dauguma tos klasės mokinių t. y. mokyklos ir klasės „klimatas“ įtakoja mokiniu matematinį raštingumą. Mokinių ir mokytojų santykių klasėje indeksui padidėjus vienetu, t. y. esant geresniems santykiams, mokinių matematinis raštingumas vidutiniškai padidėja 4,18 balo. O klasės „klimato“ indeksui padidėjus vienetu mokinių matematinis raštingumas vidutiniškai padidėja 7,09 balo.

Solution for Fixed Effects

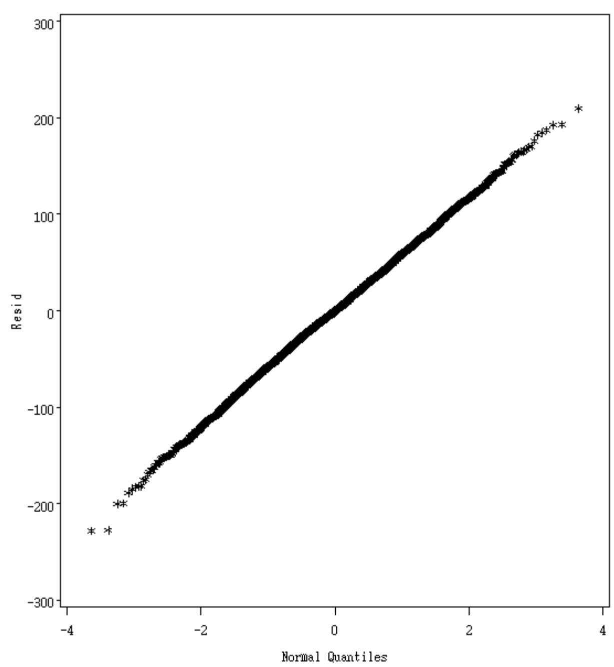
Effect	tipas	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		372,38	23,0865	189	16,13	<,0001
ESCS		17,4848	1,4779	3743	11,83	<,0001
n_kalba		17,4579	4,1336	3743	4,22	<,0001
ICTHOME		-6,4821	1,5495	193	-4,18	<,0001
knygos		20,4317	2,2353	3743	9,14	<,0001
mokmat		28,0289	2,2322	3743	12,56	<,0001
germat		10,9468	3,2029	170	3,42	0,0008
pagmat		-22,6916	2,8035	3743	-8,09	<,0001
naglent		7,5850	2,3150	3743	3,28	0,0011
tkomp		17,9103	4,6249	3743	3,87	0,0001
HOMSCH		-11,5062	1,3727	194	-8,38	<,0001
WEALTH		-4,5840	1,5424	3743	-2,97	0,0030
ENTUSE		7,1536	1,2653	3743	5,65	<,0001
LIBUSE		-16,0726	1,2787	3743	-12,57	<,0001
READRES		25,3215	8,6483	3743	2,93	0,0034
JOYREAD		14,4165	1,4485	3743	9,95	<,0001
STUDREL		4,1842	0,9104	3743	4,60	<,0001
DISCLIMA		7,0880	1,0529	3743	6,73	<,0001
ICTHOME*m_escs		-7,1521	2,3718	3743	-3,02	0,0026
germat*m_escs		-15,3559	6,4037	3743	-2,40	0,0165
ESCS*m_escs		6,3423	2,3503	3743	2,70	0,0070
READRES*PCGIRLS		-0,3390	0,1692	3743	-2,00	0,0452
JOYREAD*tipas	0	11,1179	9,7649	3743	1,14	0,2550
JOYREAD*tipas	1	-7,7740	2,6474	3743	-2,94	0,0033
JOYREAD*tipas	2	-1,8316	1,9775	3743	-0,93	0,3544
JOYREAD*tipas	3	0	.	.	.	.
m_escs		22,8675	5,3305	3743	4,29	<,0001
tipas	0	-50,3997	17,2123	3743	-2,93	0,0034
tipas	1	-14,4071	7,4728	3743	-1,93	0,0539
tipas	2	-15,6715	5,3269	3743	-2,94	0,0033
tipas	3	0	.	.	.	.
STUDBEHA		7,1686	2,9603	3743	2,42	0,0155
PARINVOL		-3,4575	0,9721	3743	-3,56	0,0004
PCGIRLS		0,9239	0,4104	3743	2,25	0,0244

2.12 pav. Procedūros MIXED rezultatas taikant 4M modelį

Modelį papildėme mokyklos lygmens kintamaisiais, kurie statistiškai reikšmingai nelygūs nuliui esant 0,05 reikšmingumo lygmeniui. Matome, kad tėvų įsitraukimo į mokyklos veiklą indeksui padidėjus vienetu, mokinio matematinis raštingumas vidutiniškai sumažėja 3,46 balo. O merginų ir vaikinų santykiui mokykloje padidėjus vienetu, tos mokyklos mokinių matematinis raštingumas vidutiniškai 0,92 balo didesnis.

Mokinio IKT turėjimas namuose ir mokyklos socialinio ir ekonominio statuso, papildomo matematikos mokymosi ir mokyklos socialinio ir ekonominio statuso, socialinio ir ekonominio statuso ir mokyklos socialinio ir ekonominio statuso, skaitymo šaltinių indekso ir merginų ir vaikinų santykiui mokykloje, mėgimo skaityti indekso ir mokyklos tipo sąveikos yra statistiškai reikšmingos su reikšmingumo lygmeniu 0,05 (2 Priedas 2.9). Pavyzdžiui, matome, kad yra mokinio socialinio ir ekonominio statuso ir mokyklos socialinio ir ekonominio statuso priklausomybė

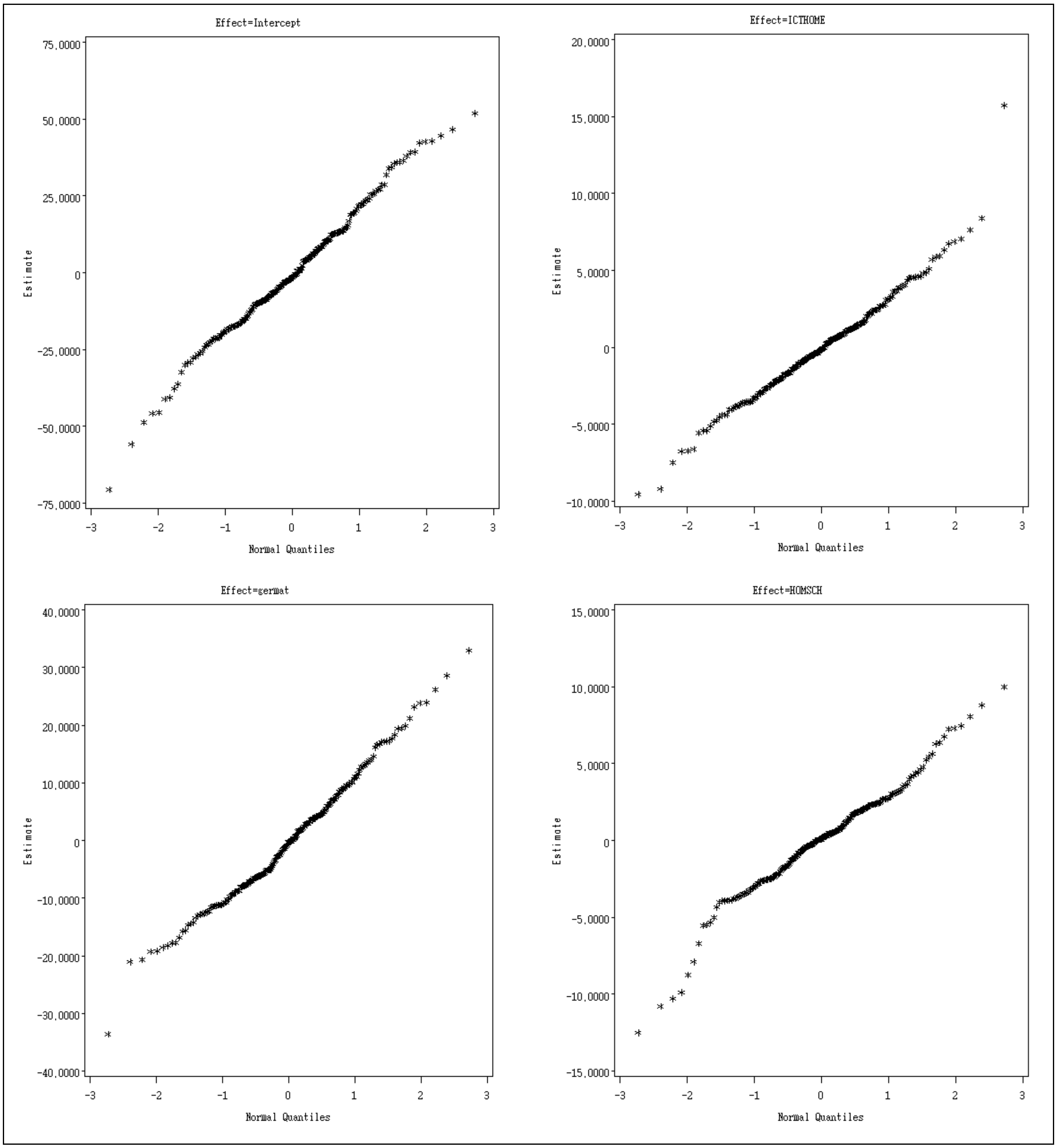
Nors hipotezės $H_0 : \sigma = 0$ ir $H_0 : \tau_{00} = 0$, $H_0 : \tau_{3} = 0$, $H_0 : \tau_{66} = 0$, $H_0 : \tau_{0,10} = 0$ lieka atmetamos, tačiau tarus, kad $\tau_{3} = 0$, $\tau_{66} = 0$ ir $\tau_{0,10} = 0$ gautas redukuotasis modelis prastesnis, nes padidėja informaciniai indeksai (pvz., AIC, AICC žiūrėkite 2 Priedas 2.9 pav.) bei parametrų σ ir τ_{00} įvertinių dispersijos. Taigi paskutinį (pilną 4M) modelį laikysime geriausiu hierarchiniu modeliu su atitinkamais kintamaisiais.



2.13 pav. *e* paklaidų kvantilių grafikas

Patikrinkime ar tenkinamos hierarchinio modelio prielaidos, t. y. atlikime modelio diagnostiką. Patikrinkime dvi prielaidas:

- Pirmojo lygmens kintamųjų paklaidos e_{ij} yra normaliai pasiskirsčiusios ir turi tą pačią dispersiją σ^2 (t. y. $e_{ij} \sim N(0,\sigma^2)$, kur $i = 1,2,\ldots,n_j$, $j = 1,2,\ldots,J$);
- Antrojo lygmens paklaidos u_{kj} normaliai pasiskirsčiusios (t. y. $u_{kj} \sim N(0,\tau_k^2)$, kur $k = 1,2,\ldots,K$, $j = 1,2,\ldots,J$).



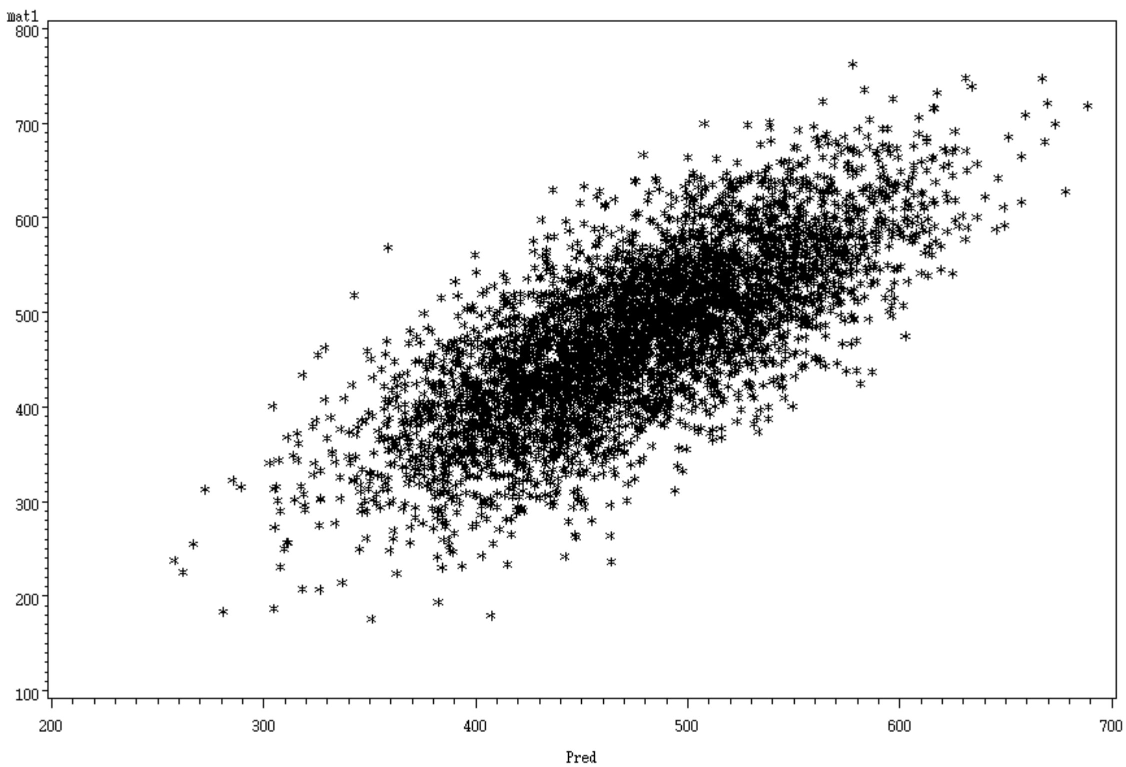
4.14 pav. *u* paklaidų kvantilių grafikai

Jeigu šios prielaidos nėra tenkinamos, tai galime įtarti, kad modelio parametrų įvertiniai nepatikimi ir netgi pats modelis netinkamas. Paprasčiausias būdas vizualiai įsitikinti ar paklaidos normaliai pasiskirsčiusios yra nubraižyti kvantilių grafikus (QQ plot).

Braižant kvantilių grafikus x ašyje atidedami standartinio normaliojo skirstinio $\frac{1}{n}, \frac{2}{n}, \ldots, 1$ kvantiliai, kur n yra imties dydis. y ašyje atidedami atitinkami tiriamojo dydžio empiriniai kvantiliai. Kuo labiau tiriamasis dydis panašus į normalųjį, tuo gauti taškai yra arčiau tiesės [2]. Nubraižykime kvantilių grafikus e_{ij} ir u_{kj} paklaidų homogeniškumui tikrinti.

Gauti grafikai (4.13 ir 4.14 pav.) atskleidžia, kad abiejų lygmenų paklaidas galime laikyti normaliai pasiskirsčiusiomis.

Su procedūra MIXED galime prognozuoti mokinių matematinį raštingumą pagal mūsų sudarytą modelį, SAS sintaksėje MODEL sakinyje nurodę pasirinktį OUTP=<Lentelė> programa įrašys prognozės rezultatus į lentelę. Galime nubraižyti matematinio raštingumo rezultatų ir prognozuojamų matematinio raštingumo rezultatų sklaidos diagramą (2.15 pav.). Matome, kad „debesėlis“ gana siauras, žinome, kad kuo siauresnis „debesėlis“, tuo tikslesnė prognozė.



2.15 pav. Realių ir prognozuojamų matematinio raštingumo rezultatų sklaidos diagrama

Skyriuje 1.3 buvo aprašyta PISA taikoma metodologija, kuomet mokiniams yra skaičiuojama ne viena reikšmė nusakanti jų matematinį raštingumą, o penkios galimos reikšmės. Analizę atlikome su viena galima reikšme, tačiau į galutinių regresijos koeficientų skaičivimus turime įtraukti visas penkias galimas reikšmes. Visoms penkioms galimoms reikšmėms yra pritaikomas tas pats geriausias mūsų sudarytas modelis. Gautos regresijos koeficientų reikšmės suvidurkinamos ir gaunamos galutinės regresinių koeficientų reikšmės (skaičiavimai atliekami su SAS programa ir SAS sintaksė pateikiama 1 Priede 1.26 ir 1.27).

Galutinė prognozės lygtis atrodo taip:

$$
\begin{aligned}
MAT_{ij} = {} & 65{,}94 + 7{,}51(ESCS)_{ij} + 6{,}51(N\_KALBA)_{ij} - {,}46(ICTHOME)_{ij} + 0{,}74(KNYGOS)_{ij} \\
& + 8{,}04(MOKMAT)_{ij} + 0{,}78(GERMAT)_{ij} - 2{,}02(PAGMAT)_{ij} + {,}44(NAGLENT)_{ij} + 5{,}95(TKOMP)_{ij} \\
& - 2{,}03(HOMSCH)_{ij} - {,}5(WEALTH)_{ij} + {,}26(ENTUSE)_{ij} - 5{,}13(LIBUSE)_{ij} + 5{,}32(READRES)_{ij} \\
& + 3{,}09(JOYREAD)_{ij} + {,}67(STUDREL)_{ij} + {,}94(DISCLIMA)_{ij} - {,}5(ICTHOME)_{ij} * (M\_ESCS)_{j} \\
& - 5{,}59(GERMAT)_{ij} * (M\_ESCS)_{j} + {,}39(ESCS)_{ij} * (M\_ESCS)_{j} - {,}34(READRES)_{ij} * (PCGIRLS)_{j} \\
& + \gamma_{5,2}(JOYREAD)_{ij} * (TIPAS)_{j} + 3{,}58(M\_ESCS)_{j} + \gamma_{2}(TIPAS)_{j} + {,}23(STUDBEHA)_{j} \\
& - {,}28(PARINVOL)_{j} + {,}1(PCGIRLS)_{j},
\end{aligned}
$$

kur parametras γ_{2} įgyja reikšmes, kai TIPAS=0 $\gamma_{2} = -3{,}87$, TIPAS=1 $\gamma_{2} = -3{,}56$, TIPAS=2 $\gamma_{2} = -4{,}35$, TIPAS=3 $\gamma_{2} =$ ir parametras $\gamma_{5,2}$ įgyja reikšmes, kai TIPAS=0 $\gamma_{5,2} = {,}25$, TIPAS=1 $\gamma_{5,2} = -{,}15$, TIPAS=2 $\gamma_{5,2} = {,}45$, TIPAS=3 $\gamma_{5,2} =$.

Regresijos koeficientų įvetiniai šiek tiek skiriasi palyginus su galutinio modelio koeficientų įvertiniais (2.12 pav.), kurie gauti skaičiavimams naudojant tik pirmąją matematinio raštingumo galimą reikšmę.

IŠVADOS

Šiame darbe analizuodami PISA 2009 m. tyrimo duomenis, bandėme rasti mokinio, jo tėvų, namų ir mokyklos veiksnius įtakojančius penkiolikmečių matematinį raštingumą. Analizė buvo atliekama taikant hierarchinį tiesinį modelį. Atlikus duomenų analizę galime daryti tokias išvadas:

- Besąlyginio hierarchinio tiesinio modelio analizė atskleidė, kad 32 % Lietuvos penkiolikmečių matematinio raštingumo rezultatų skirtumus lemia mokyklos;
- Kadangi atlikus modelio diagnostiką yra tenkinamos abi modelio prielaidos (pirmojo ir antrojo lygmenų paklaidas galime laikyti normaliai pasiskirsčiusiomis) ir redukuotasis modelis nėra geresnis už pilnąjį, todėl galutinis modelis yra geriausias su atitinkamais kintamaisiais;
- Mokinio ir mokyklos socialinis ir ekonominis statusas statistiškai reikšmingai (su reikšmingumo lygmeniu 0,05) turi įtakos matematiniam raštingumui, t. y. esant didesniam mokinio ir/ar mokyklos socialiniam ir ekonominiam statusui, penkiolikamečių matematinio raštingumo balai aukštesni;
- Matematinio raštingumo balai aukštesni tų penkiolikamečių, kurie papildomai mokosi matematikos ir/ar lanko sustiprintą matematiką. Tačiau mokinių, lankančių pataisomojo pobūdžio matematikos užsiėmimus (kuomet mokytojas ištaiso padarytas klaidas), matematinio raštingumo balai žemesni;
- Penkiolikamečių matematiniams raštingumo balams statistiškai reikšmingai turi įtakos lankomos mokyklos tipas, tačiau vietovė kurioje mokykla įsikūrusi (miestas, miestelis ar kaimas) ir mokyklos dydis neįtakoja matematinio raštingumo.
- Mokinio lytis statistiškai reikšmingai neturi įtakos matematiniam raštingumui;
- Penkiolikmečių tėvų darbo tipas ir jų kvalifikacija statistiškai reikšmingai įtakos neturi matematiniam raštingumui;
- Mokinių elgesys mokykloje įtakoja jų matematinį raštingumą (didėjant mokinių elgesio indeksui, didėja jų matematinio raštingumo balai), tačiau mokytojų elgesys neturi įtakos mokinių matematiniam raštingumui;
- Mėgimas skaityti ir namuose turėjimas įvairių skaitymo šaltinių (knygos, žurnalai, internetas ir pan.) statistiškai reikšmingai įtakoja matematinį raštingumą, t. y. mėgatančių skaityti penkiolikamečių matematinio raštingumo balai aukštesni.

SANTRAUKA

Tarptautinis švietimo tyrimas PISA (*Programme for International Student Assessment*) tiria penkiolikmečių mokinių skaitymo gebėjimus, matematinį ir gamtamokslį raštingumą. Taip pat tyrimo metu renka informaciją apie mokinį, jo pomegius, informacinius ir komunikacinių technologijų išmanymą ir naudojimą, namų aplinką, mokyklą kurioje jis mokosi ir pan. Informacija surinkta tyrimo metu yra vertinga švietimo tyrėjams, sprendimų priėmėjams, švietimiečiams ir pan.

Šio darbo tikslas – taikant hierarchinė regresiją rasti mokinių matematinio raštingumo rezultatų prikausomybę nuo mokymo(si) aplinkos veiksnių. Kadangi mokiniai yra pirmojo lygmens objektai, o mokyklos – antrojo lygmens, todėl analizuojant duomenis reikia atsižvelgti į duomenų hierarchiškumą. Besąlyginio hierarchinio modelio analizė atskleidė, kad 32 % Lietuvos penkiolikmečių matematinio raštingumo rezultatų skirtumų lemia mokyklos. Taikant hierarchinį tiesinį modeliavimą, sudarytas geriausias hierarchinis tiesinis modelis su atitinkamais kintamaisiais bei užrašyta lygtis, kuri gali būti naudojama Lietuvos penkiolikmečių matematinio raštingumo rezultatų prognozavimui.

SUMMARY

International education survey PISA (*Programme for International Student Assessment*) examine 15-yers-old reading, mathematical and science literacy. A lot of information about student‘s, theirs hobbies, knowledge and use of information and comunication technology, home and school environment and etc. is collected during PISA survey. The collected information is valuable source of information for researchers, policy makers, educators and ect.

The main purpose of this work, “Analysis of Lithuanian Students Literacy Using International Survey PISA 2009 Data“, is using hierarchical regression to find the students' mathematical literacy dependency of training and learning environment factors. The students are first-level objects and schools - at the second level, so data analysis should take into account the data hierarchy. Unconditional hierarchical model analysis revealed that 32 % of Lithuanian 15-years-old students' mathematical literacy outcomes determined by the school. Best hierarchical linear model with relevant variables was done using hierarchical linear modelling. Also a prognostic equation of Lithuaninan 15-years-old students matematic literacy was created.

LITERATŪRA IR ŠALTINIAI

1. Chong Ho Yu. *A Simple Guide to the Item Responce Theory (IRT) and Rasch Modeling*. 2010. www.creative-wisdom.com/computer/sas/IRT.pdf
2. Čekanavičius, V. ir Murauskas, G. *Statistika ir jos taikymai*. III. Vilnius: TEV, 2009.
3. Krapavickaitė, D. ir Plikusas, A. *Imčių teorijos pagrindai*. Vilnius: Technika, 2005.
4. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. *SAS® for Mixed Models, Second Edition*. Cary, NC: SAS Institute Inc. 2006.
5. Nacionalinis egzaminų centras. *Tarptautinio penkiolikmečių tyrimo OECD PISA 2009 rezultatai*, 2010. http://www.nec.lt/failai/1810_PISA_Rezultatai.pdf
6. Nacionalinis egzaminų centras. *Teorinė metodinė medžiaga pagrindinio ugdymo pasiekimų patikrinimo (PUPP) ir brandos egzaminų (BE) užduočių rengėjams*. 2011. http://www.nec.lt/failai/2204_NEC_teorine_metodine_medziaga_spaudai_A4+5mm_crops.pdf
7. OECD. *PISA Data Analysis Manual*, OECD, Paris, 2009.
8. SAS help
9. Singer, J. D. *Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models*. Journal of Educational and Behavioral Statistics Winter 1998, Vol. 24, No. 4, 323-355 p.
10. Wu, M. *The Role of Plausible Values in Large-scale Surveys*, Studies in Educational Evaluation, Volume 31, 2005, 114–128 p. http://www.sciencedirect.com/science/article/pii/S0191491X05000209

PRIEDAI

1. Priedas. Programos SAS kodai

1.1 Tvarkome mokinių duomenų rinkinį

```
options fmtsearch=(pisa.stud);
data mokiniai;
set pisa.lt_stud;
w_fstr0=w_fstuwt;
mat1=pv1math;
mat2=pv2math;
mat3=pv3math;
mat4=pv4math;
mat5=pv5math;
keep schoolid stidstd mat1-mat5 read1-read5 w_fstr0-w_fstr80 st05q01 st01q01
ST04Q01 famstruc fisced misced hisced fsecateg msecateg hsecateg hisei immig iscedo
langn lmins mmins smins progn escs st17q01 st20q01 st20q02 st20q03 st20q04 st20q06
st20q07 st22q01 st26q07 st28q02 st29q02 st31q02 st31q06 st32q02 ic04q06 ic06q03
ic06q08 rfs1q03 rfs1q07 ATTCOMP ATSCHL CSTRAT CULTPOSS DISCLIMA DIVREAD ELAB ENTUSE
HEDRES HIGHCONF HOMEPOS HOMSCH ICTHOME ICTRES ICTSCH JOYREAD LIBUSE MEMOR ONLNREAD
STIMREAD STRSTRAT STUDREL USESCH WEALTH RFSINTRP RFSNCONT RFSTRLIT RFSFUMAT;
run;
proc sort data=mokiniai;
by stidstd;
run;
```

1.2 Tvarkome mokyklos duomenų rinkinį

```
options fmtsearch=(pisa.sch);
data mokykla;
set pisa.lt_sch;
keep schoolid ABGROUP COMPWEB IRATCOMP PCGIRLS PROPCERT PROPQUAL SCHSIZE SCHTYPE
SELSCH STRATIO EXCURACT LDRSHP RESPCURR RESPRES SCMATEDU STUDBEHA TCHPARTI TCSHORT
TEACBEHA W_FSCHWT STRATUM sc04q01 SCHOOLID SC02Q01 SC05Q01 SC11Q02 SC15Q01 SC15Q02
SC15Q05 SC16Q05 SC18Q01 SC23Q01 SC26Q01 SC26Q08 SC26Q10 SC26Q11;
run;
proc sort data=mokykla;
by schoolid;
run;
```

1.3 Tvarkome tėvų duomenų rinkinį

```
options fmtsearch=(pisa.par);
data tevai;
set pisa.lt_par;
keep stidstd schoolid PQMISCED PQFISCED PQHISCED CURSUPP MOTREAD PARINVOL PQSCHOOL
PRESUPP READRES STRATUM;
run;
proc sort data=tevai;
by stidstd;
run;
```

1.4 Sujungiame tris duomenų rinkinius į vieną bendrą

```
data mokt;
merge mokiniai tevai;
by stidstd;
```

```
run;
proc sort data=mokt;
by schoolid;
run;
data bendras;
merge mokt mokykla;
by schoolid;
run;
```

1.5 Braižome matematinio raštingumo balų histogramą

```
proc univariate data=bendras; var pv1math;
histogram mat1/normal(noprint color=black) cbarline=black caxis=black cframe=white;
run;
```

1.6 Skaičiuojamas Lietuvos matematinio raštingumo vidurkis su penkiomis galimomis reikšmėmis

```
/*Skaičiuojame vidurkius visoms penkioms galimoms reikšmėms*/
proc means data=bendras;
var mat1 mat2 mat3 mat4 mat5;
weight w_fstuwt;
output out=vidm;
run;
proc transpose data=vidm out=vidm;
id  stat_; var mat1 mat2 mat3 mat4 mat5;
run;
/*Skaičiuojame galutinį vidurkį*/
proc means data=vidm;
var mean;
run;
```

1.7 Sukuriame naujus kintamuosius - mokyklos TIPAS ir VIETA

```
data bendras;
set bendras;
tipas=.;
if (stratum in (44001,44005,44009,44013)) then tipas=1;
if (stratum in (44002,44006,44010,44014)) then tipas=2;
if (stratum in (44003,44007,44011,44015)) then tipas=3;
if (stratum in (44004,44008,44012,44016)) then tipas=0;
vieta=.;
if (stratum in (44001,44002,44003,44004)) then vieta=3;
if (stratum in (44005,44006,44007,44008)) then vieta=2;
if (stratum in (44009,44010,44011,44012)) then vieta=1;
if (stratum in (44013,44014,44015,44016)) then vieta=0;
run;
```

1.8 Apskaičiuojame matematinio raštingumo vidurkį ir pasiskirstymą pagal atitinkamą kintamąjį

```
/*Apskaičiuojame matematinio raštingumo vidurkį pagal mokinio lytį*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;
class st04q01 ;weight w_fstuwt;output out=vidm; run;
proc freq data=bendras;tables st04q01;run;
/*Apskaiciuojame matematinio rastingumo vidurkius ir pasiskirstyma pagal tai ar
mokinys namuose turi kompiuteri ar ne*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class st20q04 ;weight
w_fstuwt; run;
proc freq data=bendras;tables st20q04;run;
```

```
/*Apskaiciuojame matematinio rastingumo vidurkius ir pasiskirstyma pagal
kompiuterio turejima ir mokyklos vietove bei tipa*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class st20q04 vieta;weight
w_fstuwt; run;
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class st20q04 tipas;weight
w_fstuwt; run;
proc freq data=bendras;tables st20q04*vieta;run;
proc freq data=bendras;tables st20q04*tipas;run;
/*Apskaiciuojame matematinio rastingumo vidurkius ir pasiskirstyma pagal tai kiek
namuose yra knygu*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class st22q01 ;weight
w_fstuwt; run;
proc freq data=bendras;tables st22q01;run;
/*Apskaiciuojame matematinio rastingumo vidurkius ir pasiskirstyma pagal tai, kokia
kalba mokinys dazniausiai kalba namuose*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class st19q01 ;weight
w_fstuwt; run;
proc freq data=bendras;tables st19q01;run;
/*Apskaiciuojame matematinio rastingumo vidurkius ir pasiskirstyma pagal tevu darbo
tipa ir kompiuteriu skaiciu namuose*/
proc means data=bendras;var mat1 mat2 mat3 mat4 mat5;class hsecateg st21q03;weight
w_fstuwt; run;
proc freq data=bendras;tables hsecateg*st21q03;run;
```

1.9 Braizoma sklaidos diagrama matematiniam rastingumui ir sicialiniam ir ekonominiam statusui

```
proc gplot data=bendras;
   plot mat1*escs; symbol value=star;
run;quit;
```

1.10 Skaiciuojami koreliacijos koeficientai

```
proc corr data=bendras;
var mat1 escs;run;
proc corr data=bendras;
var mat1 joyread homsch disclima icthome;run;
```

1.11 Skaičiuojame ir įrašome praleistas reikšmes

```
/*Skaičiuojame mokinių klausimyno vidurkius*/
proc means data=bendras noprint;
var st17q01 st20q01 st20q02 st20q03 st20q04 st20q06 st20q07 st22q01 st26q07 st28q02
st29q02 st31q02 st31q06 st32q02
ic04q06 ic06q03 ic06q08 wealth studrel disclima PARINVOL cultposs homsch wealth
entuse joyread libuse readres;
by schoolid;
weight w_fstr0;
output out=vidurkiai mean(st17q01 st20q01 st20q02 st20q03 st20q04 st20q06 st20q07
st22q01 st26q07 st28q02 st29q02
st31q02 st31q06 st32q02 ic04q06 ic06q03 ic06q08 wealth studrel disclima PARINVOL
cultposs homsch wealth
entuse joyread libuse readres)=m_st17q01 m_st20q01 m_st20q02 m_st20q03 m_st20q04
m_st20q06 m_st20q07 m_st22q01 m_st26q07
m_st28q02 m_st29q02 m_st31q02 m_st31q06 m_st32q02 m_ic04q06 m_ic06q03 m_ic06q08
m_wealth m_studrel m_disclima m_PARINVOL
m_cultposs m_homsch m_wealth m_entuse m_joyread m_libuse m_readres;
run;
data bendras;
```

```
merge bendras vidurkiai;
by schoolid;
run;
/*Skaičiuojame mokyklos klausimyno praleistas reikšmes*/
proc means data=bendras noprint;
var SC02Q01 SC05Q01 SC11Q02 SC15Q01 SC15Q02 SC15Q05 SC16Q05 SC18Q01 SC23Q01 SC26Q01
SC26Q08 SC26Q10 SC26Q11;
weight W_FSCHWT;
output out=vidurkiai1 mean(SC02Q01 SC05Q01 SC11Q02 SC15Q01 SC15Q02 SC15Q05 SC16Q05
SC18Q01 SC23Q01 SC26Q01 SC26Q08 SC26Q10
SC26Q11)=m_SC02Q01 m_SC05Q01 m_SC11Q02 m_SC15Q01 m_SC15Q02 m_SC15Q05 m_SC16Q05
m_SC18Q01 m_SC23Q01 m_SC26Q01 m_SC26Q08
m_SC26Q10 m_SC26Q11;
run;
data bendras;
merge bendras vidurkiai1;
by schoolid;
run;

/*Praleistų reikšmių įrašymas*/
data bendras;
set bendras;
if st17q01=. then st17q01=ROUND(m_st17q01);
if st20q01=. then st20q01=ROUND(m_st20q01);
if st20q02=. then st20q02=ROUND(m_st20q02);
if st20q03=. then st20q03=ROUND(m_st20q03);
if st20q04=. then st20q04=ROUND(m_st20q04);
if st20q06=. then st20q06=ROUND(m_st20q06);
if st20q07=. then st20q07=ROUND(m_st20q07);
if st22q01=. then st22q01=ROUND(m_st22q01);
if st26q07=. then st26q07=ROUND(m_st26q07);
if st28q02=. then st28q02=ROUND(m_st28q02);
if st29q02=. then st29q02=ROUND(m_st29q02);
if st31q02=. then st31q02=ROUND(m_st31q02);
if st31q06=. then st31q06=ROUND(m_st31q06);
if st32q02=. then st32q02=ROUND(m_st32q02);
if ic04q06=. then ic04q06=ROUND(m_ic04q06);
if ic06q03=. then ic06q03=ROUND(m_ic06q03);
if ic06q08=. then ic06q08=ROUND(m_ic06q08);
if joyread=. then joyread=m_joyread;
if libuse=. then libuse=m_libuse;
if readres=. then readres=m_readres;
if studrel=. then studrel=m_studrel;
if disclima=. then disclima=m_disclima;
if parinvol=. then parinvol=m_parinvol;
if wealth=. then wealth=m_wealth;
if cultposs=. then cultposs=m_cultposs;
if homsch=. then homsch=m_homsch;
if entuse=. then entuse=m_entuse;
if hsecateg=. then hsecateg=ROUND(hsecateg);
if immig=. then immig=ROUND(immig);
if SC02Q01=. then SC02Q01=1;
if SC05Q01=. then SC05Q01=2;
if SC11Q02=. then SC11Q02=1;
if SC15Q01=. then SC15Q01=2;
if SC15Q02=. then SC15Q02=4;
if SC15Q05=. then SC15Q05=4;
if SC16Q05=. then SC16Q05=1;
if SC18Q01=. then SC18Q01=2;
```

```
if SC23Q01=. then SC23Q01=1;
if SC26Q01=. then SC26Q01=4;
if SC26Q08=. then SC26Q08=3;
if SC26Q10=. then SC26Q10=3;
if SC26Q11=. then SC26Q11=3;
if PCGIRLS=. then PCGIRLS=49.3067;
if SCHSIZE=. then SCHSIZE=499.8549;
if STUDBEHA=. then STUDBEHA=0.1515;
if TEACBEHA=. then TEACBEHA=0.6267;
run;
```

1.12 Pseudo kintamųjų įrašymas

```
data bendras;
set bendras;
lytis=.;/*Mokinio lytis: 1 - mergina, 0 - vaikinas*/
if (st04q01 in (1)) then lytis=1;
if (st04q01 in (2)) then lytis=0;
mvieta=.;/*Ar turi mokymosi vietą: 1 - taip, 0 - ne*/
if (st20q03 in (1)) then mvieta=1;
if (st20q03 in (2)) then mvieta=0;
tkomp=.;/*Ar turi kompiuterį:1 - taip, 0 - ne*/
if (st20q04 in (1)) then tkomp=1;
if (st20q04 in (2)) then tkomp=0;
germat=.;/*Mokosi sustiprintos matematikos:1 - taip, 0 - ne*/
if (st31q02 in (1)) then germat=1;
if (st31q02 in (2)) then germat=0;
pagmat=.;/*Lanko pataisomojo pobūdžio matematikos užsiėmimus:1 - taip, 0 - ne*/
if (st31q06 in (1)) then pagmat=1;
if (st31q06 in (2)) then pagmat=0;
mokmat=.;/*Papildomai mokosi matematikos:1 - taip, 0 - ne*/
if (st32q02 in (1)) then mokmat=1;
if (st32q02 in (2)) then mokmat=0;
nagdiag=.;/*Nagrinėja diagramas:1 - taip, 0 - ne*/
if (rfs1q03 in (1)) then nagdiag=1;
if (rfs1q03 in (2)) then nagdiag=0;
naglent=.;/*Nagrinėja lenteles:1 - taip, 0 - ne*/
if (rfs1q07 in (1)) then naglent=1;
if (rfs1q07 in (2)) then naglent=0;
darzelis=.;/*Lankė darželį:1 - taip, 0 - ne*/
if (st05q01 in (1)) then darzelis=0;
if (st05q01 in (2,3)) then darzelis=1;
knygos=.;/*Namuose yra daugiau nei 25 knygos:1 - taip,0 - ne*/
if (knygsk in (1,2)) then knygos=0;
if (knygsk in (3,4,5,6)) then knygos=1;
run;
```

1.13 Svorių normalizavimas

```
data temp1; set bendras;
keep stidstd w_fstr0; run;
proc sort data=temp1;
by stidstd; run;
proc univariate data=temp1;
var w_fstr0; output out=temp2 sum=suma N=viso;run;
data temp3;
set temp1; msvoris=(w_fstr0*4528)/40530.1785; run;
data bendras;
merge bendras temp3; by stidstd; run;
```

DUOMENŲ ANALIZĖ

1.14 Nulinis modelis (0M)

```
proc mixed data=bendras COVTEST METHOD=ML NOINFO NOCLPRINT NOITPRINT;
      class schoolid;
      model mat1= /solution;
      random intercept / subject = schoolid type=un;
      weight msvoris;
run;
```

1.15 Pirmojo modelio(1M) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs lytis n_kalba icthome knygos / SOLUTION chisq;
      RANDOM intercept escs lytis n_kalba icthome knygos / SUBJECT = schoolid
TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.16 Redukuoto modelio (1RM) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs n_kalba icthome knygos/ SOLUTION chisq;
      RANDOM intercept icthome / SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.17 Antrojo modelio (2M) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent/
SOLUTION chisq;
      RANDOM intercept icthome mokmat germat pagmat naglent/ SUBJECT = schoolid
TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.18 Redukuoto antrojo modelio (2RM) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent/
SOLUTION chisq;
      RANDOM intercept icthome germat/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.19 Trečiojo modelio (3M) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS sc04q01;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent M_ESCS
TIPAS TRMOK STUDBEHA TEACBEHA
IRATCOMP sc04q01/ SOLUTION chisq;
```

```
      RANDOM intercept icthome germat/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.20 Redukuoto trečiojo modelio (3RM) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent M_ESCS
TIPAS STUDBEHA / SOLUTION chisq;
      RANDOM intercept icthome germat/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.21 Modelio (4TM) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS hsecateg SC02Q01;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent
darzelis mvieta hsecateg M_ESCS TIPAS STUDBEHA
SCHSIZE ABGROUP SCMATEDU STEST MTEST KITMOK SC02Q01/ SOLUTION chisq;
      RANDOM intercept icthome germat/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.22 Galutinio modelio (4M) SAS sintaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq OUTP=prog1;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid SOLUTION TYPE=UN;
      WEIGHT msvoris;
      ods output solutionf=reg1 solutionr=ats1;
RUN;
```

1.23 Redukuoto modelio (4RM) SAS sntaksė

```
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat1 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq ;
      RANDOM intercept/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
RUN;
```

1.24 Braižoma matematinio raštingumo balų ir prognozių sklaidos diagrama

```
proc gplot data=prog1;
   plot mat1*pred; symbol value=star;
run;quit;
```

1.25 Modeliu diagnostika - paklaidų homogeniškumo tikrinimas braižant kvantilių grafikus

```
proc capability data=prog1 noprint;
      qqplot resid / normal square vaxis=axis1;axis1 label=(a=90 r=0);
run;
proc sort data=ats1; by effect;run;
proc capability data=ats1 noprint;
      qqplot estimate / normal square vaxis=axis1;axis1 label=(a=90 r=0);
        by effect;
run;
```

1.26 Skaičiuojami regresijos koeficientai kitoms galimoms reikšmėms

```
/*Skaičiuojame su antra galima reikšme*/
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat2 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq OUTP=prog2;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
      ods output solutionf=reg2;
RUN;
/*Skaičiuojame su trečia galima reikšme*/
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat3 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq OUTP=prog3;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
      ods output solutionf=reg3;
RUN;
/*Skaičiuojame su ketvirta galima reikšme*/
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat4 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq OUTP=prog4;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
      ods output solutionf=reg4;
RUN;
/*Skaičiuojame su penkta galima reikšme*/
PROC MIXED DATA=bendras covtest METHOD=ML noinfo NOCLPRINT NOITPRINT;
      CLASS schoolid TIPAS;
      MODEL mat5 = escs n_kalba icthome knygos mokmat germat pagmat naglent tkomp
homsch wealth entuse libuse readres
joyread studrel disclima icthome*M_ESCS germat*M_ESCS escs*M_ESCS readres*PCGIRLS
joyread*TIPAS
M_ESCS TIPAS STUDBEHA PARINVOL PCGIRLS / SOLUTION chisq OUTP=prog5;
      RANDOM intercept icthome germat homsch/ SUBJECT = schoolid TYPE=UN;
      WEIGHT msvoris;
```

```
      ods output solutionf=reg5;
RUN;
```

1.27 Skaičiuojami galutiniai regresijos koeficientai

```
data reg11; set reg1;pred1=estimate; keep effect tipas est1;run;
data reg21; set reg2;est2=estimate; keep effect tipas est2;run;
data reg31; set reg3;est3=estimate; keep effect tipas est3;run;
data reg41; set reg4;est4=estimate; keep effect tipas est4;run;
data reg51; set reg5;est5=estimate; keep effect tipas est5;run;
proc sort data=reg11;by effect tipas;run;
proc sort data=reg21;by effect tipas;run;
proc sort data=reg31;by effect tipas;run;
proc sort data=reg41;by effect tipas;run;
proc sort data=reg51;by effect tipas;run;

data reg;
merge reg11 reg21 reg31 reg41 reg51;
by effect tipas;
galut=(est1+est2+est3+est4+est5)/5;
run;
```

2. Priedas Programos SAS rezultatai

2.1 Besąlyginio modelio (0M) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	2434.41	271.69	8.96	<.0001
Residual		5280.91	113.45	46.55	<.0001

Fit Statistics

-2 Log Likelihood	52189.3
AIC (smaller is better)	52195.3
AICC (smaller is better)	52195.3
BIC (smaller is better)	52205.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	1263.20	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	470.37	3.7127	195	126.69	<.0001

2.2 Pirmojo modelio (1M) programos SAS rezultatas:

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	927,23	467,85	1,98	0,0237
UN(2,1)	SCHOOLID	17,8306	106,95	0,17	0,8676
UN(2,2)	SCHOOLID	3,5068	38,8616	0,09	0,4640
UN(3,1)	SCHOOLID	-94,9365	173,59	-0,55	0,5844
UN(3,2)	SCHOOLID	-59,8864	41,5783	-1,44	0,1498
UN(3,3)	SCHOOLID	127,11	100,53	1,26	0,1031
UN(4,1)	SCHOOLID	161,59	309,68	0,52	0,6018
UN(4,2)	SCHOOLID	-16,4269	102,07	-0,16	0,8721
UN(4,3)	SCHOOLID	-197,95	162,54	-1,22	0,2233
UN(4,4)	SCHOOLID	232,88	314,68	0,74	0,2296
UN(5,1)	SCHOOLID	-58,8274	121,71	-0,48	0,6288
UN(5,2)	SCHOOLID	-8,1569	32,7824	-0,25	0,8035
UN(5,3)	SCHOOLID	-17,1191	50,2286	-0,34	0,7332
UN(5,4)	SCHOOLID	-35,1817	111,20	-0,32	0,7517
UN(5,5)	SCHOOLID	128,95	49,4865	2,61	0,0046
UN(6,1)	SCHOOLID	108,33	178,71	0,61	0,5444
UN(6,2)	SCHOOLID	74,0372	52,0227	1,42	0,1547
UN(6,3)	SCHOOLID	-43,1020	85,0795	-0,51	0,6124
UN(6,4)	SCHOOLID	-101,22	158,62	-0,64	0,5234
UN(6,5)	SCHOOLID	-92,6199	58,0600	-1,60	0,1107
UN(6,6)	SCHOOLID	123,78	127,96	0,97	0,1667
Residual		4546,31	107,19	42,41	<,0001

Fit Statistics

-2 Log Likelihood	51555,2
AIC (smaller is better)	51611,2
AICC (smaller is better)	51611,6
BIC (smaller is better)	51703,0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
21	767,55	<,0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	428,12	5,1373	195	83,34	<,0001
ESCS	21,3358	1,3889	195	15,36	<,0001
lytis	-2,2242	2,2460	193	-0,99	0,3233
n_kalba	28,3299	4,7971	84	5,91	<,0001
ICTHOME	-6,2980	1,5426	194	-4,08	<,0001
knygos	30,2881	2,5815	191	11,73	<,0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	195	235,98	235,98	<,0001	<,0001
lytis	1	193	0,98	0,98	0,3220	0,3233
n_kalba	1	84	34,88	34,88	<,0001	<,0001
ICTHOME	1	194	16,67	16,67	<,0001	<,0001
knygos	1	191	137,66	137,66	<,0001	<,0001

2.3 Redukuoto modelio (1RM) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	1229,32	156,28	7,87	<,0001
UN(2,1)	SCHOOLID	-140,68	56,8864	-2,47	0,0134
UN(2,2)	SCHOOLID	118,19	42,1910	2,80	0,0025
Residual		4634,12	101,89	45,48	<,0001

Fit Statistics

-2 Log Likelihood	51581,7
AIC (smaller is better)	51599,7
AICC (smaller is better)	51599,7
BIC (smaller is better)	51629,2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	744,39	<,0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	427,30	5,2619	195	81,21	<,0001
ESCS	21,4862	1,3886	4133	15,47	<,0001
n_kalba	28,6492	4,7177	4133	6,07	<,0001
ICTHOME	-6,1137	1,5179	195	-4,03	<,0001
knygos	30,0482	2,4494	4133	12,27	<,0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	4133	239,43	239,43	<,0001	<,0001
n_kalba	1	4133	36,88	36,88	<,0001	<,0001
ICTHOME	1	195	16,22	16,22	<,0001	<,0001
knygos	1	4133	150,50	150,50	<,0001	<,0001

2.4 Antrojo modelio (2M) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	1298.19	215.82	6.02	<.0001
UN(2,1)	SCHOOLID	-133.99	64.6877	-2.07	0.0383
UN(2,2)	SCHOOLID	99.7663	38.6394	2.58	0.0049
UN(3,1)	SCHOOLID	-43.8243	132.71	-0.33	0.7412
UN(3,2)	SCHOOLID	36.5666	50.6122	0.72	0.4700
UN(3,3)	SCHOOLID	197.66	142.67	1.39	0.0830
UN(4,1)	SCHOOLID	-45.7663	167.96	-0.27	0.7852
UN(4,2)	SCHOOLID	5.9377	69.5099	0.09	0.9319
UN(4,3)	SCHOOLID	124.35	130.84	0.95	0.3419
UN(4,4)	SCHOOLID	603.60	215.48	2.80	0.0025
UN(5,1)	SCHOOLID	-79.3945	138.86	-0.57	0.5675
UN(5,2)	SCHOOLID	-6.8562	62.0500	-0.11	0.9120
UN(5,3)	SCHOOLID	112.92	102.80	1.10	0.2720
UN(5,4)	SCHOOLID	80.3506	155.26	0.52	0.6048
UN(5,5)	SCHOOLID	2.66E-30	.	.	.
UN(6,1)	SCHOOLID	84.7931	117.94	0.72	0.4722
UN(6,2)	SCHOOLID	-117.43	50.2275	-2.34	0.0194
UN(6,3)	SCHOOLID	-10.7399	95.2906	-0.11	0.9103
UN(6,4)	SCHOOLID	140.79	124.43	1.13	0.2578
UN(6,5)	SCHOOLID	122.24	105.03	1.16	0.2445
UN(6,6)	SCHOOLID	127.37	122.06	1.04	0.1484
Residual		4146.97	96.3678	43.03	<.0001

Fit Statistics

-2 Log Likelihood	51201.6
AIC (smaller is better)	51261.6
AICC (smaller is better)	51262.0
BIC (smaller is better)	51359.9

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
20	913.97	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	415.53	5.4188	195	76.68	<.0001
ESCS	20.6792	1.3317	3427	15.53	<.0001
n_kalba	25.8459	4.5770	3427	5.65	<.0001
ICTHOME	-5.4912	1.4325	195	-3.83	0.0002
knygos	27.4390	2.3433	3427	11.71	<.0001
mokmat	32.4947	2.6231	185	12.39	<.0001
germat	9.0385	3.5173	169	2.57	0.0110
pagmat	-25.4948	2.9561	163	-8.62	<.0001
naglent	8.3034	2.6073	185	3.18	0.0017

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3427	241.14	241.14	<.0001	<.0001
n_kalba	1	3427	31.89	31.89	<.0001	<.0001
ICTHOME	1	195	14.70	14.70	0.0001	0.0002
knygos	1	3427	137.11	137.11	<.0001	<.0001
mokmat	1	185	153.46	153.46	<.0001	<.0001
germat	1	169	6.60	6.60	0.0102	0.0110
pagmat	1	163	74.38	74.38	<.0001	<.0001
naglent	1	185	10.14	10.14	0.0014	0.0017

2.5 Redukuoto modelio (2RM) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	1391.46	175.16	7.94	<.0001
UN(2,1)	SCHOOLID	-127.24	57.7534	-2.20	0.0276
UN(2,2)	SCHOOLID	105.31	39.0335	2.70	0.0035
UN(3,1)	SCHOOLID	39.3658	144.78	0.27	0.7857
UN(3,2)	SCHOOLID	-8.2398	67.8194	-0.12	0.9033
UN(3,3)	SCHOOLID	572.44	199.69	2.87	0.0021
Residual		4183.15	93.7162	44.64	<.0001

Fit Statistics

-2 Log Likelihood	51217.0
AIC (smaller is better)	51249.0
AICC (smaller is better)	51249.1
BIC (smaller is better)	51301.5

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	898.55	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	413.58	5.4526	195	75.85	<.0001
ESCS	20.5119	1.3352	3957	15.36	<.0001
n_kalba	26.3198	4.5955	3957	5.73	<.0001
ICTHOME	-5.3860	1.4482	195	-3.72	0.0003
knygos	27.7649	2.3481	3957	11.82	<.0001
mokmat	33.3331	2.4022	3957	13.88	<.0001
germat	8.9851	3.5068	172	2.56	0.0113
pagmat	-25.6295	2.9846	3957	-8.59	<.0001
naglent	9.1627	2.4683	3957	3.71	0.0002

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3957	236.02	236.02	<.0001	<.0001
n_kalba	1	3957	32.80	32.80	<.0001	<.0001
ICTHOME	1	195	13.83	13.83	0.0002	0.0003
knygos	1	3957	139.82	139.82	<.0001	<.0001
mokmat	1	3957	192.55	192.55	<.0001	<.0001
germat	1	172	6.56	6.56	0.0104	0.0113
pagmat	1	3957	73.74	73.74	<.0001	<.0001
naglent	1	3957	13.78	13.78	0.0002	0.0002

2.6 Trečiojo modelio (3M) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	668,41	100,09	6,68	<,0001
UN(2,1)	SCHOOLID	22,7959	47,0921	0,48	0,6283
UN(2,2)	SCHOOLID	106,28	38,3896	2,77	0,0028
UN(3,1)	SCHOOLID	108,82	113,08	0,96	0,3359
UN(3,2)	SCHOOLID	-5,3058	67,8547	-0,08	0,9377
UN(3,3)	SCHOOLID	586,27	202,23	2,90	0,0019
Residual		4180,16	93,8002	44,56	<,0001

Fit Statistics

-2 Log Likelihood	50777,7
AIC (smaller is better)	50831,7
AICC (smaller is better)	50832,0
BIC (smaller is better)	50920,0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	426,06	<,0001

Solution for Fixed Effects

Effect	tipas	SC04Q01	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			469,37	28,7910	184	16,30	<,0001
ESCS			18,8184	1,3564	3931	13,87	<,0001
n_kalba			25,8690	4,3829	3931	5,90	<,0001
ICTHOME			-5,4152	1,4538	194	-3,72	0,0003
knygos			26,4286	2,3508	3931	11,24	<,0001
mokmat			32,6376	2,3722	3931	13,76	<,0001
germat			8,7814	3,5306	171	2,49	0,0138
pagmat			-26,1184	2,9950	3931	-8,72	<,0001
naglent			9,1473	2,4708	3931	3,70	0,0002
m_escs			38,5798	6,8454	3931	5,64	<,0001
tipas	0		-67,7649	16,7224	3931	-4,05	<,0001
tipas	1		-19,6912	8,0016	3931	-2,46	0,0139
tipas	2		-20,4421	5,4941	3931	-3,72	0,0002
tipas	3		0	.	.	.	.
trmok			-39,2386	27,7793	3931	-1,41	0,1579
STUDBEHA			10,6008	3,8998	3931	2,72	0,0066
TEACBEHA			-4,8177	3,9684	3931	-1,21	0,2248
IRATCOMP			4,7866	8,3991	3931	0,57	0,5688
SC04Q01		1	4,7716	7,6158	3931	0,63	0,5310
SC04Q01		2	-1,7473	6,8432	3931	-0,26	0,7985
SC04Q01		3	6,0739	6,5801	3931	0,92	0,3560
SC04Q01		4	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3931	192,48	192,48	<,0001	<,0001
n_kalba	1	3931	34,84	34,84	<,0001	<,0001
ICTHOME	1	194	13,87	13,87	0,0002	0,0003
knygos	1	3931	126,39	126,39	<,0001	<,0001
mokmat	1	3931	189,30	189,30	<,0001	<,0001
germat	1	171	6,19	6,19	0,0129	0,0138
pagmat	1	3931	76,05	76,05	<,0001	<,0001
naglent	1	3931	13,71	13,71	0,0002	0,0002
m_escs	1	3931	31,76	31,76	<,0001	<,0001
tipas	3	3931	23,34	7,78	<,0001	<,0001
trmok	1	3931	2,00	2,00	0,1578	0,1579
STUDBEHA	1	3931	7,39	7,39	0,0066	0,0066
TEACBEHA	1	3931	1,47	1,47	0,2247	0,2248
IRATCOMP	1	3931	0,32	0,32	0,5687	0,5688
SC04Q01	3	3931	1,66	0,55	0,6449	0,6450

2.7 Redukuoto modelio (3RM) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	684,81	101,97	6,72	<,0001
UN(2,1)	SCHOOLID	24,5612	47,4726	0,52	0,6049
UN(2,2)	SCHOOLID	108,33	38,6364	2,80	0,0025
UN(3,1)	SCHOOLID	111,27	112,11	0,99	0,3210
UN(3,2)	SCHOOLID	-5,2766	68,1267	-0,08	0,9383
UN(3,3)	SCHOOLID	589,60	202,88	2,91	0,0018
Residual		4181,14	93,8115	44,57	<,0001

Fit Statistics

-2 Log Likelihood	50783,3
AIC (smaller is better)	50825,3
AICC (smaller is better)	50825,5
BIC (smaller is better)	50894,0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	435,15	<,0001

Solution for Fixed Effects

Effect	tipas	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		431,16	6,0500	189	71,27	<,0001
ESCS		18,8118	1,3567	3932	13,87	<,0001
n_kalba		26,5716	4,3610	3932	6,09	<,0001
ICTHOME		-5,4529	1,4571	194	-3,74	0,0002
knygos		26,5038	2,3506	3932	11,28	<,0001
mokmat		32,6004	2,3722	3932	13,74	<,0001
germat		8,8413	3,5338	171	2,50	0,0133
pagmat		-25,9912	2,9944	3932	-8,68	<,0001
naglent		9,0454	2,4707	3932	3,66	0,0003
m_escs		35,8325	5,4788	3932	6,54	<,0001
tipas	0	-73,5212	15,7570	3932	-4,67	<,0001
tipas	1	-18,8869	7,5396	3932	-2,51	0,0123
tipas	2	-20,1658	5,3582	3932	-3,76	0,0002
tipas	3	0	.	.	.	.
STUDBEHA		8,5936	3,1168	3932	2,76	0,0059

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3932	192,27	192,27	<,0001	<,0001
n_kalba	1	3932	37,12	37,12	<,0001	<,0001
ICTHOME	1	194	14,00	14,00	0,0002	0,0002
knygos	1	3932	127,14	127,14	<,0001	<,0001
mokmat	1	3932	188,86	188,86	<,0001	<,0001
germat	1	171	6,26	6,26	0,0124	0,0133
pagmat	1	3932	75,34	75,34	<,0001	<,0001
naglent	1	3932	13,40	13,40	0,0003	0,0003
m_escs	1	3932	42,77	42,77	<,0001	<,0001
tipas	3	3932	28,44	9,48	<,0001	<,0001
STUDBEHA	1	3932	7,60	7,60	0,0058	0,0059

2.8 Galutinio modelio (4M) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	580,47	87,0814	6,67	<,0001
UN(2,1)	SCHOOLID	-1,3531	38,8593	-0,03	0,9722
UN(2,2)	SCHOOLID	56,8756	32,7352	1,74	0,0412
UN(3,1)	SCHOOLID	138,40	92,9272	1,49	0,1364
UN(3,2)	SCHOOLID	-3,5170	54,2044	-0,06	0,9483
UN(3,3)	SCHOOLID	401,34	164,55	2,44	0,0074
UN(4,1)	SCHOOLID	12,9027	36,5530	0,35	0,7241
UN(4,2)	SCHOOLID	-32,2595	21,6759	-1,49	0,1367
UN(4,3)	SCHOOLID	25,7124	52,2315	0,49	0,6225
UN(4,4)	SCHOOLID	55,0498	29,1264	1,89	0,0294
Residual		3666,11	83,4199	43,95	<,0001

Fit Statistics

-2 Log Likelihood	50418,3
AIC (smaller is better)	50504,3
AICC (smaller is better)	50505,1
BIC (smaller is better)	50645,0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
10	422,02	<,0001

Solution for Fixed Effects

Effect	tipas	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		372,38	23,0865	189	16,13	<,0001
ESCS		17,4848	1,4779	3743	11,83	<,0001
n_kalba		17,4579	4,1336	3743	4,22	<,0001
ICTHOME		-6,4821	1,5495	193	-4,18	<,0001
knygos		20,4317	2,2353	3743	9,14	<,0001
mokmat		28,0289	2,2322	3743	12,56	<,0001
germat		10,9468	3,2029	170	3,42	0,0008
pagmat		-22,6916	2,8035	3743	-8,09	<,0001
naglent		7,5850	2,3150	3743	3,28	0,0011
tkomp		17,9103	4,6249	3743	3,87	0,0001
HOMSCH		-11,5062	1,3727	194	-8,38	<,0001
WEALTH		-4,5840	1,5424	3743	-2,97	0,0030
ENTUSE		7,1536	1,2653	3743	5,65	<,0001
LIBUSE		-16,0726	1,2787	3743	-12,57	<,0001
READRES		25,3215	8,6483	3743	2,93	0,0034
JOYREAD		14,4165	1,4485	3743	9,95	<,0001
STUDREL		4,1842	0,9104	3743	4,60	<,0001
DISCLIMA		7,0880	1,0529	3743	6,73	<,0001
ICTHOME*m_escs		-7,1521	2,3718	3743	-3,02	0,0026
germat*m_escs		-15,3559	6,4037	3743	-2,40	0,0165
ESCS*m_escs		6,3423	2,3503	3743	2,70	0,0070
READRES*PCGIRLS		-0,3390	0,1692	3743	-2,00	0,0452
JOYREAD*tipas	0	11,1179	9,7649	3743	1,14	0,2550
JOYREAD*tipas	1	-7,7740	2,6474	3743	-2,94	0,0033
JOYREAD*tipas	2	-1,8316	1,9775	3743	-0,93	0,3544
JOYREAD*tipas	3	0	.	.	.	.
m_escs		22,8675	5,3305	3743	4,29	<,0001
tipas	0	-50,3997	17,2123	3743	-2,93	0,0034
tipas	1	-14,4071	7,4728	3743	-1,93	0,0539
tipas	2	-15,6715	5,3269	3743	-2,94	0,0033
tipas	3	0	.	.	.	.
STUDBEHA		7,1686	2,9603	3743	2,42	0,0155
PARINVOL		-3,4575	0,9721	3743	-3,56	0,0004
PCGIRLS		0,9239	0,4104	3743	2,25	0,0244

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3743	139,97	139,97	<,0001	<,0001
n_kalba	1	3743	17,84	17,84	<,0001	<,0001
ICTHOME	1	193	17,50	17,50	<,0001	<,0001
knygos	1	3743	83,55	83,55	<,0001	<,0001
mokmat	1	3743	157,66	157,66	<,0001	<,0001
germat	1	170	11,68	11,68	0,0006	0,0008
pagmat	1	3743	65,51	65,51	<,0001	<,0001
naglent	1	3743	10,74	10,74	0,0011	0,0011
tkomp	1	3743	15,00	15,00	0,0001	0,0001
HOMSCH	1	194	70,26	70,26	<,0001	<,0001
WEALTH	1	3743	8,83	8,83	0,0030	0,0030
ENTUSE	1	3743	31,96	31,96	<,0001	<,0001
LIBUSE	1	3743	158,00	158,00	<,0001	<,0001
READRES	1	3743	8,57	8,57	0,0034	0,0034
JOYREAD	1	3743	33,28	33,28	<,0001	<,0001
STUDREL	1	3743	21,12	21,12	<,0001	<,0001
DISCLIMA	1	3743	45,32	45,32	<,0001	<,0001
ICTHOME*m_escs	1	3743	9,09	9,09	0,0026	0,0026
germat*m_escs	1	3743	5,75	5,75	0,0165	0,0165
ESCS*m_escs	1	3743	7,28	7,28	0,0070	0,0070
READRES*PCGIRLS	1	3743	4,01	4,01	0,0451	0,0452
JOYREAD*tipas	3	3743	10,48	3,49	0,0149	0,0150
m_escs	1	3743	18,40	18,40	<,0001	<,0001
tipas	3	3743	12,72	4,24	0,0053	0,0053
STUDBEHA	1	3743	5,86	5,86	0,0155	0,0155
PARINVOL	1	3743	12,65	12,65	0,0004	0,0004
PCGIRLS	1	3743	5,07	5,07	0,0244	0,0244

2.9 Redukuoto modelio (4RM) programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	661,09	86,2418	7,67	<,0001
Residual		3789,49	81,5075	46,49	<,0001

Fit Statistics

-2 Log Likelihood	50446,0
AIC (smaller is better)	50514,0
AICC (smaller is better)	50514,6
BIC (smaller is better)	50625,3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	394,25	<,0001

Solution for Fixed Effects

Effect	tipas	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		376,87	23,5792	189	15,98	<,0001
ESCS		17,7650	1,4852	4300	11,96	<,0001
n_kalba		16,1754	4,1443	4300	3,90	<,0001
ICTHOME		-6,3187	1,4406	4300	-4,39	<,0001
knygos		19,8896	2,2436	4300	8,87	<,0001
mokmat		27,7161	2,2465	4300	12,34	<,0001
germat		12,3717	2,7290	4300	4,53	<,0001
pagmat		-23,5175	2,8040	4300	-8,39	<,0001
naglent		7,6784	2,3289	4300	3,30	0,0010
tkomp		17,7396	4,6126	4300	3,85	0,0001
HOMSCH		-11,5443	1,2505	4300	-9,23	<,0001
WEALTH		-4,6697	1,5499	4300	-3,01	0,0026
ENTUSE		7,2784	1,2638	4300	5,76	<,0001
LIBUSE		-16,3073	1,2805	4300	-12,74	<,0001
READRES		25,2872	8,6119	4300	2,94	0,0033
JOYREAD		14,4977	1,4545	4300	9,97	<,0001
STUDREL		4,1673	0,9129	4300	4,56	<,0001
DISCLIMA		7,2594	1,0596	4300	6,85	<,0001
ICTHOME*m_escs		-7,3737	2,2013	4300	-3,35	0,0008
germat*m_escs		-14,8102	5,3701	4300	-2,76	0,0058
ESCS*m_escs		6,3727	2,3649	4300	2,69	0,0071
READRES*PCGIRLS		-0,3403	0,1685	4300	-2,02	0,0435
JOYREAD*tipas	0	10,1725	9,7926	4300	1,04	0,2990
JOYREAD*tipas	1	-7,7665	2,6525	4300	-2,93	0,0034
JOYREAD*tipas	2	-1,9385	1,9836	4300	-0,98	0,3285
JOYREAD*tipas	3	0	.	.	.	.
m_escs		22,6681	5,5084	4300	4,12	<,0001
tipas	0	-54,1650	17,6040	4300	-3,08	0,0021
tipas	1	-13,9902	7,6091	4300	-1,84	0,0660
tipas	2	-16,2887	5,4451	4300	-2,99	0,0028
tipas	3	0	.	.	.	.
STUDBEHA		8,1803	3,0234	4300	2,71	0,0068
PARINVOL		-3,6459	0,9759	4300	-3,74	0,0002
PCGIRLS		0,8705	0,4198	4300	2,07	0,0382

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	4300	143,07	143,07	<,0001	<,0001
n_kalba	1	4300	15,23	15,23	<,0001	<,0001
ICTHOME	1	4300	19,24	19,24	<,0001	<,0001
knygos	1	4300	78,59	78,59	<,0001	<,0001
mokmat	1	4300	152,21	152,21	<,0001	<,0001
germat	1	4300	20,55	20,55	<,0001	<,0001
pagmat	1	4300	70,35	70,35	<,0001	<,0001
naglent	1	4300	10,87	10,87	0,0010	0,0010
tkomp	1	4300	14,79	14,79	0,0001	0,0001
HOMSCH	1	4300	85,22	85,22	<,0001	<,0001
WEALTH	1	4300	9,08	9,08	0,0026	0,0026
ENTUSE	1	4300	33,17	33,17	<,0001	<,0001
LIBUSE	1	4300	162,19	162,19	<,0001	<,0001
READRES	1	4300	8,62	8,62	0,0033	0,0033
JOYREAD	1	4300	32,29	32,29	<,0001	<,0001
STUDREL	1	4300	20,84	20,84	<,0001	<,0001
DISCLIMA	1	4300	46,94	46,94	<,0001	<,0001
ICTHOME*m_escs	1	4300	11,22	11,22	0,0008	0,0008
germat*m_escs	1	4300	7,61	7,61	0,0058	0,0058
ESCS*m_escs	1	4300	7,26	7,26	0,0070	0,0071
READRES*PCGIRLS	1	4300	4,08	4,08	0,0435	0,0435
JOYREAD*tipas	3	4300	10,17	3,39	0,0172	0,0173
m_escs	1	4300	16,93	16,93	<,0001	<,0001
tipas	3	4300	13,77	4,59	0,0032	0,0033
STUDBEHA	1	4300	7,32	7,32	0,0068	0,0068
PARINVOL	1	4300	13,96	13,96	0,0002	0,0002
PCGIRLS	1	4300	4,30	4,30	0,0381	0,0382

2.10 Modelio 4TM programos SAS rezultatas:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOLID	666,97	99,9186	6,68	<,0001
UN(2,1)	SCHOOLID	27,4657	47,0806	0,58	0,5596
UN(2,2)	SCHOOLID	109,62	38,6165	2,84	0,0023
UN(3,1)	SCHOOLID	128,55	109,82	1,17	0,2418
UN(3,2)	SCHOOLID	-6,6970	67,8302	-0,10	0,9214
UN(3,3)	SCHOOLID	579,21	201,47	2,87	0,0020
Residual		4171,53	93,3056	44,71	<,0001

Fit Statistics

-2 Log Likelihood	51086,7
AIC (smaller is better)	51152,7
AICC (smaller is better)	51153,2
BIC (smaller is better)	51260,8

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	428,62	<,0001

Solution for Fixed Effects

Effect	tipas	HSECATEG	SC02Q01	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				479,33	59,2008	187	8,10	<,0001
ESCS				18,0631	1,8536	3948	9,74	<,0001
n_kalba				25,6904	4,3668	3948	5,88	<,0001
ICTHOME				-5,2803	1,4688	195	-3,59	0,0004
knygos				26,8851	2,3467	3948	11,46	<,0001
mokmat				32,7998	2,3693	3948	13,84	<,0001
germat				9,1414	3,5125	172	2,60	0,0101
pagmat				-26,0439	2,9770	3948	-8,75	<,0001
naglent				9,2577	2,4657	3948	3,75	0,0002
darzelis				-3,1441	2,3608	3948	-1,33	0,1830
mvieta				2,3803	3,6478	3948	0,65	0,5141
HSECATEG		1		-1,4059	4,6759	3948	-0,30	0,7637
HSECATEG		2		-7,6560	3,9953	3948	-1,92	0,0554
HSECATEG		3		-2,2829	4,0790	3948	-0,56	0,5757
HSECATEG		4		0	.	.	.	.
m_escs				34,7803	5,9275	3948	5,87	<,0001
tipas	0			-75,0609	15,6664	3948	-4,79	<,0001
tipas	1			-20,5592	7,5651	3948	-2,72	0,0066
tipas	2			-21,3120	5,3937	3948	-3,95	<,0001
tipas	3			0	.	.	.	.
STUDBEHA				7,7655	3,2471	3948	2,39	0,0168
SCHSIZE				0,004800	0,008096	3948	0,59	0,5533
ABGROUP				-4,7897	3,7133	3948	-1,29	0,1972
SCMATEDU				-1,8198	3,4321	3948	-0,53	0,5960
stest				-0,7463	11,3209	3948	-0,07	0,9474
mtest				-5,1294	18,5892	3948	-0,28	0,7826
kitmok				4,8199	11,7390	3948	0,41	0,6814
SC02Q01			1	-37,3688	54,1242	3948	-0,69	0,4900
SC02Q01			2	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
ESCS	1	3948	94,96	94,96	<,0001	<,0001
n_kalba	1	3948	34,61	34,61	<,0001	<,0001
ICTHOME	1	195	12,92	12,92	0,0003	0,0004
knygos	1	3948	131,26	131,26	<,0001	<,0001
mokmat	1	3948	191,64	191,64	<,0001	<,0001
germat	1	172	6,77	6,77	0,0093	0,0101
pagmat	1	3948	76,53	76,53	<,0001	<,0001
naglent	1	3948	14,10	14,10	0,0002	0,0002
darzelis	1	3948	1,77	1,77	0,1829	0,1830
mvieta	1	3948	0,43	0,43	0,5141	0,5141
HSECATEG	3	3948	7,17	2,39	0,0667	0,0669
m_escs	1	3948	34,43	34,43	<,0001	<,0001
tipas	3	3948	30,63	10,21	<,0001	<,0001
STUDBEHA	1	3948	5,72	5,72	0,0168	0,0168
SCHSIZE	1	3948	0,35	0,35	0,5533	0,5533
ABGROUP	1	3948	1,66	1,66	0,1971	0,1972
SCMATEDU	1	3948	0,28	0,28	0,5960	0,5960
stest	1	3948	0,00	0,00	0,9474	0,9474
mtest	1	3948	0,08	0,08	0,7826	0,7826
kitmok	1	3948	0,17	0,17	0,6814	0,6814
SC02Q01	1	3948	0,48	0,48	0,4899	0,4900

2.11 Galutinių regresijos koeficientų skaičiavimo rezultatai:

	Effect	tipas	est1	est2	est3	est4	est5	galut
1	DISCLIMA	_	7.0880423271	6.6467968051	6.9282223838	7.1413871534	6.8997270217	6.9408351382
2	ENTUSE	_	7.1535931646	7.4034037255	7.0355302124	7.1990825425	7.5159526569	7.2615124604
3	ESCS	_	17.484812082	18.049312491	17.327405312	17.039924892	17.649758563	17.510242668
4	ESCS*m_escs	_	6.3422774154	5.0176868277	6.5868565834	4.8960819426	4.1131757458	5.391215703
5	HOMSCH	_	-11.50619736	-12.38408372	-11.22857632	-12.03117294	-13.00567516	-12.0311411
6	ICTHOME	_	-6.48211667	-4.74290067	-6.382945494	-5.033696026	-4.663321443	-5.460996061
7	ICTHOME*m_escs	_	-7.152110132	-4.696555652	-6.918761999	-6.8826335	-6.835942345	-6.497200726
8	Intercept	_	372.38410071	355.97388061	363.1709985	365.92384413	372.23948572	365.93846193
9	JOYREAD	_	14.416514811	11.888534468	13.18206504	12.800635516	13.160371155	13.089624198
10	JOYREAD*tipas	0	11.117889701	5.5727435187	1.9915519365	0.4236045293	12.164000114	6.2539579599
11	JOYREAD*tipas	1	-7.774013402	-4.91113655	-5.113967066	-5.867907959	-7.092482597	-6.151901515
12	JOYREAD*tipas	2	-1.83158835	2.8153233519	-0.181239179	0.2780499571	1.152628333	0.4466348226
13	JOYREAD*tipas	3	0	0	0	0	0	0
14	LIBUSE	_	-16.07261313	-14.60369334	-14.97311461	-15.55032937	-14.45002976	-15.12995604
15	PARINVOL	_	-3.457537306	-3.020800167	-3.148398954	-3.53786834	-3.254322054	-3.283785364
16	PCGIRLS	_	0.9238877163	1.2690698021	1.1755855497	1.1388647764	0.9734865689	1.0961788827
17	READRES	_	25.32151231	29.318999291	23.455165667	20.606100422	27.898520299	25.320059598
18	READRES*PCGIRLS	_	-0.339020977	-0.424032632	-0.298481961	-0.256344998	-0.383192712	-0.340214656
19	STUDBEHA	_	7.1685796723	7.3650668504	6.7911234855	6.5169538921	8.3131216294	7.2309691059
20	STUDREL	_	4.1842387073	4.0008792787	3.3190733339	3.6972649885	3.1483270309	3.6699566679
21	WEALTH	_	-4.584007879	-4.476082583	-3.53156396	-4.225523658	-5.695467316	-4.502529079
22	germat	_	10.946779457	11.860972976	10.428426934	9.61409316	11.04834333	10.779723171
23	germat*m_escs	_	-15.35589354	-14.89470571	-12.42355826	-14.22376621	-21.04485676	-15.5885561
24	knygos	_	20.431658903	21.691108111	21.602633063	21.33540594	18.628792014	20.737919606
25	m_escs	_	22.867467666	23.281081044	22.52170729	22.526849797	26.712675362	23.581956232
26	mokmat	_	28.028910961	29.417227236	26.567545134	27.206055435	28.974213746	28.038790502
27	n_kalba	_	17.457937426	17.859958738	15.92252873	14.851894384	16.444672607	16.507398377
28	naglent	_	7.5850115181	8.388618253	8.3232139502	8.8395786764	9.0523951015	8.4377634998
29	pagmat	_	-22.69163696	-22.8941375	-21.17356178	-21.52900517	-21.78770747	-22.01520978
30	tipas	0	-50.3997076	-56.83940932	-52.4890373	-56.79607532	-52.83201887	-53.87124968
31	tipas	1	-14.40708236	-15.79743966	-12.08259235	-14.19341377	-11.30277999	-13.55666163
32	tipas	2	-15.67150421	-13.2097088	-14.87372809	-14.77034418	-13.24618466	-14.35429399
33	tipas	3	0	0	0	0	0	0
34	tkomp	_	17.910346866	15.25744042	14.380680051	16.190150838	16.0246479	15.952653215