
STUDENTŲ MATEMATIKOS ŽINIŲ VERTINIMO SUBJEKTYVUMO PROBLEMA: TEORINIAI IR PRAKTINIAI ASPEKTAI

Rima Kriauzienė

Mykolo Romerio universitetas, Lietuva, kriauziene@mruni.eu

Aleksandras Krylovas

Mykolo Romerio universitetas, Lietuva, krylovas@mruni.eu

Natalja Kosareva

Vilniaus Gedimino technikos universitetas, Lietuva, natalja.kosareva@vgtu.lt

Abstraktas

Tikslas. Tyrimo tikslas – ištirti dėstytojų įvertinimų suderinamumą ir pateikti rekomendacijas, kaip būtų galima sumažinti subjektyvumo įtaką vertinant studentų žinias. Pristatyti studentų norminio ir kriterinio vertinimo metodiką, atliktą taikant neparametrinės statistikos metodus.

Metodologija. Tyrimui buvo pasitelkti Mykolo Romerio universiteto Ekonomikos ir finansų valdymo fakulteto pirmo kurso studentų parašyti matematikos kontroliniai darbai. Buvo atrinkti 15 darbų, kuriuos vertino 6 dėstytojai. Atlikta statistinė gautų duomenų analizė. Tyrimo uždaviniai buvo trys: pirmas – dėstytojų įvertinimų suderinamumo nustatymas. Jam tikrinti buvo pasirinkti neparametriniai statistiniai metodai, t. y. buvo apskaičiuotas Kendalo konkordacijos koeficientas W bei Frydmano kriterijaus priklausomoms imtims statistika Q . Patikrinta hipotezė, ar W yra statistiškai reikšmingas.

Antras uždavinys – tiriamųjų studentų norminis vertinimas, t. y. studentų išdėstymas didėjimo tvarka pagal tikrinamų matematikos žinių vertinimo rezultatus. Antro uždavinio sprendimui taikytas Page'o L – testas trendui nustatyti (Page, 1963). Jis leidžia testuojamuosius išdėstyti pažymių vidutinių rangų didėjimo tvarka. Trečias uždavinys – tiriamųjų studentų kriterinis vertinimas, t. y. studentų žinių įvertinimas dešimtbale skale. Tam tikslui pasiekti buvo tikrinama, kurie iš gretimų įvertinimų skiriasi reikšmingai. Pagal tai galima nuspręsti, kurie iš testuojamųjų gaus tuos pačius, o kurie skirtingus įvertinimus. Šiam testuojamųjų kriterinio vertinimo uždaviniui spręsti buvo apskaičiuotas *mažiausias reikšminis skirtumas (least significant difference)* (Sprenst ir Smeeton, 2007; Conover, 1999) tarp bendrojo testo balo suminių stulpelių rangų. Taip pat darbe buvo atlikta atskirų testo užduočių analizė.

Rezultatai. Nustatyta, kad dėstytojų nuomonė vertinant testuojamuosius buvo gana vienoda. Atliktas studentų norminis vertinimas, t. y. pažymiai išdėstyti vidutinių rangų didėjimo tvarka. Tuomet testuojamieji pagal bendrąjį testo balą atsižvelgiant į visų šešių dėstytojų įvertinimus buvo priskirti penkioms skirtingoms grupėms pagal jų tiesinės algebros žinių lygį, t. y. atliktas kriterinis vertinimas. Atlikus konkrečių užduočių analizę, padaryta išvada, jog penktoji užduotis neturėjo įtakos bendram testo balui.

Tyrimo ribotumas. Siūloma metodika yra tik vienas iš galimų žinių vertinimo problemos sprendimo būdų.

Praktinė reikšmė. Metodika, aprašyta šiame darbe, gali būti pritaikyta ne tik matematikos, bet ir kitų gamtos mokslų ar socialinių dalykų žinioms vertinti.

Originalumas / vertingumas. Taikytas Page'o L – testas trendui nustatyti (Page, 1963). Jis nėra įtrauktas į daugelį standartinių statistinių paketų ir gana retai naudojamas statistiniuose tyrimuose.

Raktažodžiai: žinių vertinimas, matematikos dėstymas, nparametriniai statistiniai metodai, mažiausias reikšminis skirtumas, vertinimo subjektyvumas, žinių vertinimo testai.

Tyrimo tipas: tyrimo pristatymas.

Įvadas

Lietuvos aukštosios mokyklos šiuo metu naudoja kriterinę studentų žinių vertinimo sistemą, kai įgytos žinios lyginamos su dalyko programoje numatytą žinių apimtimi ir išreiškiamos atitinkamu procentu. Alternatyvus kriteriniam – norminis vertinimas numato studento mokymosi pasiekimų palyginimą su kitų studijuojančiųjų rezultatais (Girdzijauskas, 1997). Taip vertinamos, pavyzdžiui, vidurinių mokyklų absolventų žinios valstybiniuose brandos egzaminuose. Universitetai (pavyzdžiui, VDU) kartais taiko ir norminio vertinimo principus, tačiau ir teisiniai aktai, ir kompetencijų aprašai, ir dėstytojų patirtis orientuoja į žinių palyginimą su tam tikrais standartais, o tai numato

būtent kriterijais grįstas vertinimas. Tačiau toks vertinimas neretai būna gana subjektyvus (Krylovas et al, 2002; Krylovas, Suboč, 2005; Krylovas et al, 2007; Kriausienė et al, 2010) ir praktikoje dažnai įvairiais būdais koreguojamas (kaupiamojo balo principu, lankomumo įskaitymu ir pan.) (Bulajeva, 2007; Andziulienė, 2004.). Šio darbo tikslas – išnagrinėti kai kuriuos studentų žinių vertinimo teorinius aspektus ir autorių sukauptą praktinę patirtį bei apžvelgti autorių atliktus eksperimentus.

Studentų (moksleivių ir kt.) žinių vertinimo subjektyvumas, t. y. vertinimo rezultato stipri priklausomybė nuo vertintojo, yra gerai žinoma ir nagrinėta literatūroje problema (Gage, Berliner, 1994). Dažniausiai tokiuose tyrimuose kalbama apie psichologinius vertinimo aspektus, vertintojo asmenybės ypatumus ir kt. Vertinimas gali būti atliekamas tiek neformalioju, tiek formalioju būdu. Neformalioju vertinimu gaunamos žinios apie studentą yra tik bendros. O pagrindinis studento vertinimas vyksta formalioju būdu: rašto darbai, kontroliniai darbai, testai. Tačiau vertintojas gali pasinaudoti bendriausiomis žiniomis, kurios gali būti netikslios ir nepatikimos. Pavyzdžiui, vertinamasis, sėdintis auditorijos gale ir nepasižymintis aktyvumu, gali toks būti dėl to, kad yra ramaus būdo, o ne todėl, kad mažiau išmano už savo kolegas. Vertinamasis, sėdintis pirmoje eilėje ir nuolat pritariantis dėstytojui, gali iš tiesų suvokti, tai kas aiškinama, bet gali būti, jog jis yra tik socialiai paklusnus. Gautos neformalaus vertinimo žinios gali būti netikslios dar ir todėl, kad vertinamieji gali perprasti vertintoją, t. y. vertintojas (mūsų atveju dėstytojas) atsižvelgia rašydamas pažymį į tai, kad vertinamasis nuolat pritaria, t. y. pritarimą išreiškia galvos linktelėjimu arba tuo metu, kai vertintojas aiškina, vertinamasis užsirašinėja arba dažnai reiškia savo, kad ir visiškai neteisingą, nuomonę konkrečiu klausimu. Vertinamasis gali kelti ranką ir atėjęs prie lentos nieko nemokėti, bet žinodamas, kad vertintojas padės atlikti užduotį ir atsižvelgs į studento aktyvumą rašydamas vertinimą, tai daro dažnai. Patyrę vertintojai atskiria vertinamuosius ir jų elgesį, jie taip pat nekreipia dėmesio į kolegų nuomonę apie konkrečius vertinamuosius ir neperžiūri esamų įrašų apie juos. Yra dvi pagrindinės neformalaus vertinimo problemos: patikimumas ir validumas. Patikimumas susijęs su tuo, kad nelengva per trumpą laikotarpį įvertinti mokymo efektyvumą ir vertinamųjų išitraukimą į darbą. Vertintojas dažniausiai kontroliuoja grupelę vertinamųjų (4 ar 5), juos seka ir įvertina mokymo efektyvumą. Tačiau toks skaičius vertinamųjų yra per mažas, kad būtų galima spręsti dėl viso efektyvaus grupės darbo. Vertinamieji, kurių vertintojas nepastebėjo, gali mokyti ar nesimokyti. Dėl vertintojų šališkumo neformalaus mokymo validumas irgi yra abejotinas. Dažniausiai vertintojai ieško argumentų, patvirtinančių, kad jie gerai dėsto ir kad studentai informaciją priima gerai. Taip pat validumas yra abejotinas dėl to, kad apie informacijos supratimą sprendžiama iš vertinamųjų elgesio, t. y. ar jie žiūri į dėstytoją, ar užsirašinėja. Tačiau tai darydami vertinamieji gali iš tikrųjų nieko nesuprasti. Knygoje (Gage, Berliner, 1994) pateikiamas pavyzdys, kaip yra vertinamas rašinėlis. Jo vertinimui įtakos gali turėti daug faktorių: patraukli ar nepatraukli autoriaus išvaizda, sugebėjimas tvarkingai rašyti, net vertinamojo lytis. Jau XX a. pradžioje buvo parodyta, kad subjektyvumas veikia pažymius. Buvo patikrintos ir palygintos atliktos užduotys mokinių, kuriuos mokytojai laikė gabiais, ir tų, kuriuos laikė mažiau išmanančiais. Palyginus atliktas užduotys paaiškėjo, kad gabesniais laikomų mokinių darbuose mokytojai

buvo praleidę daugiau klaidų. Mokytojai tikėjosi daugiau klaidų rasti, jų nuomone, mažiau išmanančių mokinių sąsiuvinuose ir todėl jų daugiau rado. Iš tiesų mokytojų buvo susidaryta klaidinga nuomonė. Taigi vertinimui įtakos gali turėti daug veiksnių, tačiau net ir norint būti objektyviam, vertintojui gana sunku įvertinti objektyviai. Vertinimo problema atsiranda, kai vertinamasis padaro klaidą. Jos įvertinimas ir yra subjektyvus sprendimas. Ta klaida gali atsirasti dėl nežinojimo arba atidumo stokos. Tai gali būti ir įgūdžių stoka, kuri atsiranda dėl per mažo pratybų skaičiaus, pavyzdžiui, matematikoje nepakanka mokėti arba tiksliai žinoti apibrėžimus, lemas, teoremas, reikia mokėti jas pritaikyti ir praktikoje, t. y. konkrečioms uždaviniais spręsti. Tam reikia įgūdžių, kurie atsiranda nagrinėjant pavyzdžius ir daug sprendžiant. Visuomenėje pasigirsta įvairių nuomonių, vienos iš jų prieštarauja šiems teiginiams. Teigiama, jog matematiką studentas gebas išmokyti savarankiškai, nelankydamas paskaitų.

Ši tema gana aktuali, nes daug kalbama apie nuotolines studijas, mokymąsi savarankiškai namuose arba bibliotekoje, todėl buvo atliktas tyrimas, kurio tikslas – ištirti studentų gebėjimą skaityti ir suprasti matematinę tekstą.

Gebėjimo skaityti ir suprasti matematinę tekstą tyrimas: eiga, rezultatai, išvados

Tyrimo dalyvavo Mykolo Romerio universiteto pirmo kurso studentai, kurie rudens semestre klausė tiesinės algebros ir matematinės analizės kursą. Tyrimo tikslas – nustatyti gebėjimą skaityti ir suprasti matematinę tekstą. Tema apie determinantus parinkta dėl to, kad pirmakursiams studentams buvo nežinoma ir negirdėta, antra vertus – nesudėtinga ir įkandama mokantis savarankiškai. Atsitiktinai pasirinktos trys pirmo kurso studentų grupės. Grupėse studentų skaičius buvo panašus: 29, 26, 30. Vienos grupės studentai turėjo susipažinti su šia tema savarankiškai namuose, antros grupės studentai paskaitos pradžioje skaitė savarankiškai apie 20 min., o vėliau rašė testą, su trečios grupės studentais buvo dirbama paskaitos metu, t. y. dėstytojas aiškino, kas yra determinantas, jo savybės, ir tik likus 20 min. iki paskaitos pabaigos studentai rašė testą.

Testo užduotys buvo sudarytos iš trijų dalių (žr. 1 pav.). Pirmoje užduotyje reikėjo apskaičiuoti tris determinantus: vieną antros eilės ir du trečios eilės, viename iš jų, pritaikius determinanto savybę, iš karto buvo galima parašyti atsakymą. Antroje užduotyje taip pat reikėjo apskaičiuoti du trečios eilės determinantus, tačiau jeigu studentas gerai suprato savybes, tai pakako apskaičiuoti vieną, o prie transponuotos matricos determinanto parašyti tą patį atsakymą. Trečioje dalyje reikėjo įrašyti trūkstantus žodžius. Šia užduotimi buvo tikrinamas determinanto savybių mokėjimas. Testo variantai buvo keturi, tačiau užduotys buvo tokios pat, tik sąlygos sukeistos vietomis.

Teisingai atlikta užduotis buvo vertinama vienu tašku, o neteisingai atlikta – 0 taškų. Už pirmą užduotį buvo galima surinkti 3 taškus, nes reikėjo apskaičiuoti tris determinantus, už antrą užduotį – 2, o už trečią – 5 taškus. Atlikus skaičiavimus paaiškėjo, jog grupės, kuri turėjo savarankiškai pasiruošti namie, rezultatas buvo blogiausias iš visų testuojamųjų, t. y. vidutiniškai 20 proc. teisingai atliktų užduočių, tik vienas tes-

1. Apskaičiuokite šiuos determinantus:

a) $\begin{vmatrix} 3 & 1 & 2 \\ 4 & 3 & 2 \\ -1 & 0 & 3 \end{vmatrix} =$

b) $\begin{vmatrix} 2 & 3 \\ 4 & 1 \end{vmatrix} =$

c) $\begin{vmatrix} 6 & 4 & 3 \\ -4 & 0 & -2 \\ 2 & 5 & 1 \end{vmatrix} =$

2. Apskaičiuokite $\det A$ ir $\det A^T$, kai $A = \begin{pmatrix} 2 & 4 & -1 \\ 7 & 3 & 2 \\ 3 & 1 & -2 \end{pmatrix}$.

3. Įrašykite trūkstamus žodžius:

a) Sukeitus dvi determinanto eilutes vietomis,

_____.

b) Jeigu determinantas turi du vienodus stulpelius, tai jis lygus

_____.

c) Jeigu determinanto eilutės yra proporcingos, tai toks determinantas

lygus _____.

d) Jeigu kurio nors determinanto stulpelio elementai turi bendrą

daugiklį, tai jį galima

_____.

_____.

e) Jeigu prie vienos determinanto eilutės elementų pridėsime kitos

eilutės elementus, padaugintus iš bet kurio (to paties) skaičiaus,

nelygaus nuliui, determinanto reikšmė _____.

I pav. Testo pavyzdys

tuojamasis atliko teisingai 6 užduotis iš 10 galimų, o savybes pastebėjo ir pritaikė tik 4 testuojamieji iš 29. Testuojamųjų grupės, kurios dalyviai savarankiškai skaitė pratybų metu, rezultatas buvo geresnis – teisingai atlikta 47,69 proc. užduočių, savybes pritaikė 11 iš 26 testuojamųjų. Geriausias rezultatas buvo grupės, kurioje dėstytojas aiškino šią

temą, o vėliau likus 20 min. iki paskaitos pabaigos buvo atliekamos užduotys – atlikta vidutiniškai 63,33 proc. užduočių, savybes pritaikė 19 studentų iš 30, vienas studentas atliko visas užduotis teisingai ir surinko maksimalų taškų skaičių. Taigi galime daryti išvadą, jog savarankiškai studentai sunkiai gali išmokti naują medžiagą, o ypač pritaikyti įgytas žinias praktikoje, kadangi net dirbant kartu su dėstytoju dar atsiranda spragų. Vadinasi, igūdžių stoka gali atsirasti dėl per mažo pratybų skaičiaus, todėl vertinant studentų darbus būna sunku įvertinti klaidą, nes nežinia, ar ši klaida yra atsiradusi dėl vertinamojo nežinojimo, ar dėl igūdžių stokos.

Taigi objektyvus studentų žinių vertinimas yra nelengva užduotis, net ir vertinant tokių lyg ir lengvai išmatuojamą dalyką kaip matematikos žinios. Įvertinti sunku ir tada, kai vertinamą darbą sudaro kelios dalys, pavyzdžiui, keli kontroliniai ir rašto darbai, net ir taikant įvairias skatinimo priemones.

Mes apsiribosime tik didaktiniais matematikos žinių vertinimo aspektais, kai pakankamai gerai išryškintas nuomonių skirtumas ir tyrimui taikytini statistikos metodai.

Straipsnyje (Krylovas, Raulynaitis, 2005) aprašomas eksperimentas, kuriam buvo pasirinktas vienas VGTU Statybos fakulteto srautas (4 grupės). Šio eksperimento tikslas buvo aprašyti studentų darbo vertinimo semestro metu priemonių kompleksą. Aukštųjų mokyklų studentų žinių vertinimas reglamentuotas instrukcijomis, kurios numato, kad egzamino pažymys atitinka modulyje aprašytų žinių lygį: pažymys 5 (silpnai) rašomas už 50–55 proc. žinių ir t. t. Tam tikrą egzamino pažymio dalį studentas semestro metu gauna už semestro metu atliktus laboratorinius, kontrolinius, namų darbus ir kt. Šio straipsnio autoriai dėstė šiam srautui, vienas iš jų skaitė paskaitas, o kitas vedė pratybas. Žinių vertinimo taisyklės buvo apibrėžtos, studentai semestro pradžioje buvo su jomis supažindinti. Studentai semestro metu rašė du testus, atlikinėjo namų darbus, kurių įvertinimas sudarė dalį egzamino pažymio. Taip pat buvo vertinamas darbas pratybų metu. Visa tai tik palengvina egzamino laikymą.

Bet autorių patirtis parodė, kad net ir taikant tokią rimtą skatinimo priemonę kaip galimybė nelaikyti egzamino sesijos metu, tik maža dalis studentų nuolat dirba semestro metu, o nemaža jų dalis nesugeba patenkinamai įsiminti pagrindinę dalyko dalį net po kelių egzamino laikymų. Griežtai apibrėžtas vertinimo taisyklės, įvertinant kiekvieną semestro metu padarytą darbą ar užduotį, galima taikyti tik visą semestrą dirbantiems studentams. O studentų, kurie neįveikė tam tikrų reikalavimų (neišlaikiusių kolokviuumų, neatlikusių namų darbų ar kt.), žinioms įvertinti tų pačių reikalavimų taikyti nebegalima, belieka vertinti tik pagrindines dalyko žinias. Tuomet įvertinimas gali turėti norminio vertinimo elementus, kai skaičiuojami visų laikiusiųjų rezultatai ir išlaikymo kriterijus nustatomas atsižvelgiant į minimalų žinių lygį ir į neišlaikiusiųjų skaičių.

Dar viename straipsnyje (Krylovas et al, 2002) iškelta ta pati studentų žinių vertinimo problema. Šio straipsnio tyrimo objektas buvo skirtingų dėstytojų nuomonės apie tą patį klaidingai išspręstą uždavinį.

Tyrimo metu dėstytojams buvo pateiktos egzamino užduotys ir šių uždavinių sprendimai su padarytomis klaidomis. Dėstytojai gavo anketą, kurioje jie turėjo skalėje nuo 1 iki 100 įvertinti studento klaidingai išspręstą uždavinį. Apklausoje dalyvavo 22 VGTU,

VPU ir VU dėstytojai, dėstantys aukštosios matematikos kursą. Tik dviejų iš jų pedagoginis stažas buvo trumpesnis kaip 5 metai, o 16 pedagoginis stažas viršijo 10 metų.

Atlikus tyrimą ir apdorojus rezultatus, pastebėtas gerai žinomas faktas, jog skirtingi dėstytojai tų pačių studentų žinias vertina skirtingai. Straipsnio autoriai šiai problemai sušvelninti siūlo griežčiau apibrėžti vertinimo kriterijus, išskiriant tipines klaidas. Kyla klausimas, ar tos klaidos liudija tam tikrų žinių, gebėjimų trūkumą, ar jos gali būti atsitiktinės. Tačiau atsakymas juk irgi gali būti labai subjektyvus.

Kitame panašios tematikos straipsnyje (Krylovas et al, 2007) atliktas tyrimas, kuriame dalyvavo du dėstytojai, kurie tikrino tuos pačius studentų atliktus darbus vertindami nuo 0 iki 1 taško, tačiau intervalo skaidymo žingsnis nebuvo reglamentuojamas. Atlikus tyrimą pasitvirtino prieš tai atlikto tyrimo išvados: net pakankamai detaliam suderinę vertinimo kriterijus, tuos pačius darbus dėstytojai įvertina skirtingai. Tuomet reiktų priimti tokias studentų žinių vertinimo sistemas kaip studentų per semestrą atliktų darbų kaupiamojo balo skaičiavimas (Krylovas, Raulynaitis, 2005), kontroliniai darbai testų pavidalu.

Naudojant testus galima tikrinti labai didelius studentų srautus (Krylovas, Raulynaitis, 2005; Krylovas, Suboč, 2006) ir gauti objektyvius studentų žinių įvertinimus. Atsakymai į atskirus testo klausimus gali suteikti informaciją apie testuojamųjų tam tikrų temų žinias bei įgūdžius, taip pat gauta informacija yra naudinga sudarant testus, atitinkančius apibrėžtus tikrinimo tikslus (Krylovas, Raulynaitis, 2005). Kito straipsnio (Krylovas, Suboč, 2005) autoriai studentų žinioms įvertinti pateikė uždarojo tipo (t. y. su pasiūlytais atsakymais) klausimų testus. Skirtingi testo variantai buvo išlygiagretinti: turėjo skirtingas formules bei kintamųjų žymėjimus. Studentai pildė nustatyto pavidalo atsakymų korteles ir tikrinant buvo skaičiuojami teisingi ir klaidingi atsakymai. Dėl gana didelio klausimų skaičiaus įvertinimas galėjo turėti sistemine teigiama paklaidą, jei studentas bandytų atspėti atsakymą. Todėl studentai buvo įspėti, kad, esant neteisingam atsakymui, bendras surinktų balų skaičius bus mažinamas. Atlikus tyrimą paaiškėjo, jog ši taktika pasiteisino. Studentai nebandė spėlioti atsakymų į testo klausimus ir paliko šiuos klausimus nepažymėtus, o tie, kurie spėliojo, gavo vienu balu mažiau už kiekvieną neteisingai atspėtą atsakymą.

Testai galėtų būti vienas iš studento žinių vertinimo subjektyvumo problemos sprendimo būdų, tačiau skirtingų testo variantų klausimai turi būti lygiaverčiai. Straipsnyje (Krylovas et al, 2007) nagrinėjamas testų su išlygiagretintais klausimų variantais lygiavertiškumas sunkumo prasme. Aprašomas tyrimas, kai nagrinėjami du diskrečiosios matematikos testo klausimai ir kiekvienas iš klausimų turėjo dešimt skirtingų variantų. Buvo lyginamas šių variantų sunkumas. Ištyrus pirmąjį klausimą, tyrėjai padarė išvadą, jog visi variantai yra lygiaverčiai sunkumo prasme.

Kitame analizuotame klausime iš 10 variantų tik antras ir trečias, atlikus tyrimo statistinę analizę, buvo pripažinti skirtingo sunkumo. Tačiau autoriai šį faktą aiškina tuo, kad dažnai studentai klaidingai mano, kad formulės gylis (Krylovas et al, 2007) lygus kintamųjų (raidžių arba simbolių) skaičiui. Šiai prielaidai tikrinti autoriai ateityje ketina kurti naujus testų klausimus ir atlikti tyrimą. Tyrėjai pataria vengti tokių uždarojo

tipo testų klausimų, kai neteisingi samprotavimai veda prie teisingo atsakymo. Straipsnyje autoriai pateikia formulės tikrinimo reikalavimus. Uždarąjo tipo testai, kuriuose išlygiagretinti variantai yra lygiaverčiai sunkumo prasme, galėtų išspręsti vertinimo subjektyvumo problemą.

Šiame darbe aprašomas tyrimas, kurio tikslas – ištirti dėstytojų įvertinimų suderinamumą ir pateikti rekomendacijas, kaip sumažinti subjektyvumo faktoriaus įtaką vertinant studentų žinias. Tyrimui pasitelkti Mykolo Romerio universiteto Ekonomikos ir finansų valdymo fakulteto pirmo kurso studentų 2009 m. rudens semestre parašyti matematikos kontroliniai darbai. Atrinktus 15 darbų vertino 6 dėstytojai, kurie nedėstė

1. Bankas moka 9 proc. metinių palūkanų, kurias perskaičiuoja kas 4 mėnesiai. Klientas įneša 4000 Lt ir kas keturis mėnesius sąskaitą papildo vienoda suma. Po dvejų metų sąskaitoje buvo 6963,58 Lt.

Kiek litų palūkanų gavo klientas?

4 taškai

2. Apskaičiuokite determinantą:

$$\begin{vmatrix} 3 & -2 & -2 & 3 \\ 2 & 2 & 1 & -2 \\ 4 & 3 & 3 & 4 \\ -2 & 1 & 2 & 3 \end{vmatrix}$$

4 taškai

3. Rasti: $AB^{-1}C^T$.

$$A = \begin{pmatrix} 2 & 3 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -5 & 2 & 3 \end{pmatrix}.$$

4 taškai

4. Išspręskite sistemą Gauso metodu.

$$\begin{cases} x_1 - 3x_2 + 4x_3 - x_4 = 1, \\ 7x_1 + 3x_2 - 5x_3 + 5x_4 = 10, \\ 2x_1 + 2x_2 - 3x_3 + 2x_4 = 3. \end{cases}$$

5 taškai

5. Išspręskite grafiškai šį tiesinio programavimo uždavinį: $\min(x - 2y)$, kai

$$\begin{cases} x + 2y \leq 4, \\ x + y \geq 1, \\ x \geq 0, \\ y \geq \frac{1}{2}. \end{cases}$$

3 taškai

šiems studentams, ir gauti statistiniai duomenys teikia informaciją apie atlikto kontrolinio darbo gautų pažymių pagrįstumą ir dėstytojų įvertinimų suderinamumą.

Tyrimo eiga

Tikrintą kontrolinį darbą sudarė penki tiesinės algebros uždaviniai. Tyrimui buvo atrinkti 15 studentų darbų, kuriems buvo keliamas tik vienas reikalavimas, kad darbas būtų pakankamai tvarkingas ir įskaitomas. Buvo padarytos šių darbų kopijos ir dėstytojai nepriklausomai tikrino kiekvieno iš studentų darbą. Kiekvienai užduočiai atsižvelgiant į jos sunkumą buvo nustatytas maksimalus taškų skaičius, suteikiamas teisingai atlikus šią užduotį (žr. 2 pav.).

Pirmoji užduotis – sudėtinių palūkanų uždavinys. Antroje užduotyje reikėjo apskaičiuoti ketvirtos eilės determinantą, o trečiojoje – atlikti veiksmus su matricomis. Už kiekvieną iš šių pirmųjų trijų užduočių galima buvo surinkti po 4 taškus. Ketvirtoje užduotyje reikėjo išspręsti tiesinę lygčių sistemą Gauso metodu. Teisingai atlikęs užduotį studentas galėjo gauti 5 taškus. Paskutinės penktosios užduoties sąlyga buvo tokia: „Išspręskite grafiškai tiesinio programavimo uždavinį“. Už šią užduotį studentas galėjo surinkti 3 taškus. Taigi maksimalus taškų skaičius už visas penkias užduotis – 20 taškų.

Tyrimo uždaviniai:

- 1) Dėstytojų įvertinimų suderinamumo nustatymas.
- 2) Tiriamųjų studentų norminis vertinimas, t. y. studentų išdėstymas didėjimo tvarka pagal tikrinamų matematikos žinių vertinimo rezultatus.
- 3) Tiriamųjų studentų kriterinis vertinimas, t. y. studentų žinių įvertinimas dešimtbale skale.

Tyrimo metodika ir rezultatai

Ekspertų vertinimų suderinamumui tikrinti neparametriniai statistiniai metodai buvo pasirinkti pirmaisia dėl to, kad jie yra tinkami, kai duomenų nėra daug ir kai kintamųjų skirstinys skiriasi nuo normaliojo skirstinio. Buvo apskaičiuotas Kendalo konkordacijos koeficientas W bei Frydmano kriterijaus priklausomoms imtims statistika Q (Bagdonavičius ir Kruopis, 2007; Conover, 1999). Patikrinta hipotezė, ar W statistiškai reikšmingai skiriasi nuo nulio. Laikome, kad 6-ių ekspertų vertinimai yra stebėjimų blokai, o 15-os testuojamųjų testo rezultatai yra kintamieji. Rezultatai, apskaičiuoti visų penkių užduočių ir bendrojo testo balo įvertinimams, pateikti 1 lentelėje.

1 lentelė. Konkordacijos koeficiento W ir Frydmano kriterijaus statistikos Q reikšmės kiekvienai iš penkių užduočių ir bendrojo balo įvertinimams

| | 1 užd. | 2 užd. | 3 užd. | 4 užd. | 5 užd. | Bendrasis balas |
|---|--------|--------|--------|--------|--------|-----------------|
| W | 0,957 | 0,83 | 0,908 | 0,898 | 0,617 | 0,893 |
| Q | 80,388 | 69,687 | 76,263 | 75,459 | 51,852 | 74,97 |

Kadangi visų penkių užduočių ir bendrojo testo balo šešių ekspertų konkordacijos koeficientai reikšmingai skiriasi nuo nulio (p reikšmės visais atvejais lygios 0,000), galime teigti, kad ekspertų nuomonė apie testuojamuosius yra gana vienoda. Kiek mažesnė konkordacijos koeficiento W reikšmė vertinant penktąją užduotį.

Visais atvejais Q reikšmės viršija χ^2 kriterijaus su $k-1=14$ laisvės laipsnių $\alpha=0,05$ lygmens kritinę reikšmę $\chi^2_{0,05}(14)=23,685$, todėl galime teigti, kad nulinė hipotezė H_0 : ekspertų nuomonė vertinant testuojamuosius nėra vienoda (arba kad ekspertai neatskiria testuojamųjų), yra atmetama.

Norėdami gauti tikslesnes išvadas, galime taikyti Page'o L – testą (Page, 1963) trendui nustatyti. Page'as 1963 m. sudarė statistikos L tikslų kritinių reikšmių lenteles mažoms k ir n reikšmėms, tačiau, jei tose lentelėse nėra atitinkamų n arba k reikšmių, iš L reikšmės apskaičiuojama χ^2 statistikos reikšmė, kuri turi apytiksliai χ^2 skirstinį su vienu laisvės laipsniu. Aproximaciją galima taikyti tik tam tikromis sąlygomis – kai yra bent 20 stebėjimų ir bet koks kintamųjų skaičius arba bent 9 kintamieji ir bet koks stebėjimų skaičius. Aptariamuoju atveju šios sąlygos tenkinamos, nes kintamųjų skaičius 15 yra didesnis už 9. Page'o L testas neįtrauktas į daugelį standartinių statistinių paketų. Tikrinsime hipotezę apie pažymių vidutinių rangų lygybę su alternatyva, kad testuojamieji pagal savo žinių lygį gali būti surikiuoti į eilę pažymių vidutinių rangų didėjimo tvarka:

$$\begin{cases} H_0 : m_1 = m_2 = \dots = m_{15}, \\ H_1 : m_1 < m_2 < \dots < m_{15}, \end{cases}$$

čia m_1, m_2, \dots, m_{15} – testuojamųjų vidutiniai rangai, išdėstyti nemažėjimo tvarka. Jei H_0 būtų atmetama, galėtume teigti, kad visi 6 ekspertai sutinka, kad testuojamieji pagal žinių lygį gali būti išdėstyti pažymių rangų didėjimo tvarka. Pradinė duomenų matrica turi būti išdėstyta taip, kad testuojamieji būtų stulpeliuose, o ekspertų vertinimai – eilutėse.

Apskaičiuosime kiekvieno stebėjimo duomenų rangus (nuo 1 iki 15), po to suskaičiuosime stulpelių rangų sumas ir stulpelius išdėstysime suminių rangų nemažėjimo tvarka. Šitai išdėstytiems stulpeliams priskirsime numerius nuo 1 iki 15. Page'o L sta-

tistika skaičiuojama taip: $L = \sum_{i=1}^{15} x_i R_i$, čia x_i – i -tojo stulpelio numeris, R_i – i -tojo stulpelio rangų suma. Tuomet statistika

$$X^2 = \frac{(12L - 3kn(n+1))^2}{kn^2(n^2-1)(n+1)}$$

turi apytiksliai χ^2 skirstinį su 1 laisvės laipsniu. Kritinė reikšmė, kai $\alpha=0,05$ lygi $\chi_{0,05}^2(1) = 3,841$. Patikrinus hipotezę bendrajam testo balui gavome, kad $L=7311$, $X^2=71,59$. Kadangi $\chi^2 > \chi_{0,05}^2(1)$, hipotezė H_0 atmetama. Galime teigti, kad visi 6 dėstytojai sutinka, kad studentai pagal bendrojo balo įvertinimą galėtų būti išdėstyti suminių rangų didėjimo tvarka taip:

$$m_{14} < m_{10} < m_2 < m_4 < m_{12} < m_9 < m_{15} < m_7 < m_{11} < \\ m_{13} < m_3 < m_8 < m_5 < m_6 < m_1.$$

Taigi Page'o L testas leidžia atlikti norminį testuojamųjų vertinimą.

Toliau patikrinsime, kurie iš gretimų įvertinimų skiriasi reikšmingai. Pagal tai galėsime spręsti, kurie iš testuojamųjų gaus tuos pačius įvertinimus dešimtbalėje vertinimų skalėje. Apskaičiuosime *mažiausią reikšminį skirtumą (least significant difference)* (Sprent, Smeeton, 2007; Conover, 1999) tarp bendrojo testo balo suminių stulpelių rangų. Jei dviejų gretimų stulpelių suminių rangų skirtumas yra didesnis už šį skaičių, tokių testuojamųjų įvertinimai, visų ekspertų nuomone, skiriasi reikšmingai, todėl jiems turėtų būti rašomi skirtingi pažymiai. i -ojo ir j -ojo studento įvertinimai skiriasi reikšmingai, jei

$$|R_i - R_j| > t \cdot \sqrt{\frac{2k \cdot \left(\sum_{j=i+1}^n r_{ij}^2 - \sum_{i=1}^n R_i^2 \right)}{(k-1)(n-1)}},$$

čia R_i, R_j yra i -ojo ir j -ojo stulpelio rangų sumos, k – stebėjimų skaičius ($k=6$), n – kintamųjų skaičius ($n=15$), r_{ij} – atskiro stebėjimo rangas, t – Studento kriterijaus kritinė reikšmė $t=t_{0,05}(k-1)(n-1)$. Mūsų atveju $t=t_{0,05}(70)=1,667$ ir apskaičiuotas mažiausias reikšminis skirtumas bendrojo testo balo atveju yra 9,1587.

$$1,667 \cdot \sqrt{\frac{2 \cdot 6 \cdot (7398 - 43331,5)}{5 \cdot 14}} = 9,1587$$

Studentai, kurių pažymių suminiai rangų skirtumai neviršija 9,1587, turėtų gauti vienodus pažymius, todėl jie priskiriami prie tos pačios grupės. Mūsų testuojamieji pagal bendrąjį testo balą atsižvelgiant į visų šešių dėstytojų įvertinimus buvo priskirti prie penkių skirtingų grupių pagal jų tiesinės algebros žinių lygį tokia tvarka:

$$14,10,2,4 < 12 < 9,15,7,11 < 13,3,8 < 5,6,1. \quad (2)$$

Užduočių tyrimas

Šiame skyrelyje yra tiriamos antrojo testo užduotys (žr. 2 pav.). Tyrime dalyvavo šeši nepriklausomi dėstytojai ir vienas dėstytojas, kuris dėstė šiems studentams.

1 užduotis. Už šią užduotį studentas galėjo gauti iki keturių taškų. Įrašysime į lentelę skirtumus, kurie gauti, kai iš studento taškų, gautų jo tikrajam dėstytojui vertinant šią užduotį, atimti taškai, gauti vertinant kitam eksperimente dalyvavusiam dėstytojui.

2 lentelė. Pirmos užduoties įverčių skirtumai

| | | | | | | | |
|-----------|----|------|----|-----|---|-----|---|
| Skirtumas | -1 | -0,5 | 0 | 0,5 | 1 | 1,5 | 2 |
| Dažnis | 4 | 2 | 74 | 0 | 3 | 0 | 7 |

Kaip matyti iš lentelės, šią užduotį dėstytojai, lyginant su tikruoju dėstytoju, vertino beveik vienodai. Tikimybė gauti kitą pažymį yra tik 0,18. O taip galėjo atsitikti dėl to, kad ši uždavinį daugelis, t. y. 10 studentų iš 15, išsprendė teisingai. Tokių uždavinių kontroliniame darbe reikėtų vengti, nes šis uždavinys nesuteikia pakankamai informacijos apie studentų žinias ar spragas. Jis yra per lengvas šiems studentams.

2 užduotis. Šios užduoties maksimalus galimų surinkti taškų skaičius yra 4. Vėlgi pateikiame lentelę su skirtumais, kurie gauti, kai iš studento taškų, gautų jo tikrajam dėstytojui vertinant šią užduotį, atimti taškai, gauti vertinant jau kitam eksperimente dalyvavusiam dėstytojui.

3 lentelė. Antros užduoties įverčių skirtumai

| | | | | | | | | | | | |
|-----------|----|------|----|-----|----|-----|---|-----|---|-----|---|
| Skirtumas | -1 | -0,5 | 0 | 0,5 | 1 | 1,5 | 2 | 2,5 | 3 | 3,5 | 4 |
| Dažnis | 2 | 6 | 54 | 4 | 11 | 1 | 2 | 2 | 6 | 0 | 2 |

Iš pateiktos dažnių lentelės matome, jog tikimybė gauti kitą įvertinimą už šią užduotį yra 0,4.

3 užduotis. Už užduotį buvo galima surinkti 4 taškus. Pateikiame skirtumų dažnių lentelę:

4 lentelė. Trečios užduoties įverčių skirtumai

| | | | | | | | | |
|-----------|------|----|------|----|-----|---|-----|---|
| Skirtumas | -1,5 | -1 | -0,5 | 0 | 0,5 | 1 | 1,5 | 2 |
| Dažnis | 2 | 12 | 11 | 51 | 5 | 6 | 0 | 3 |

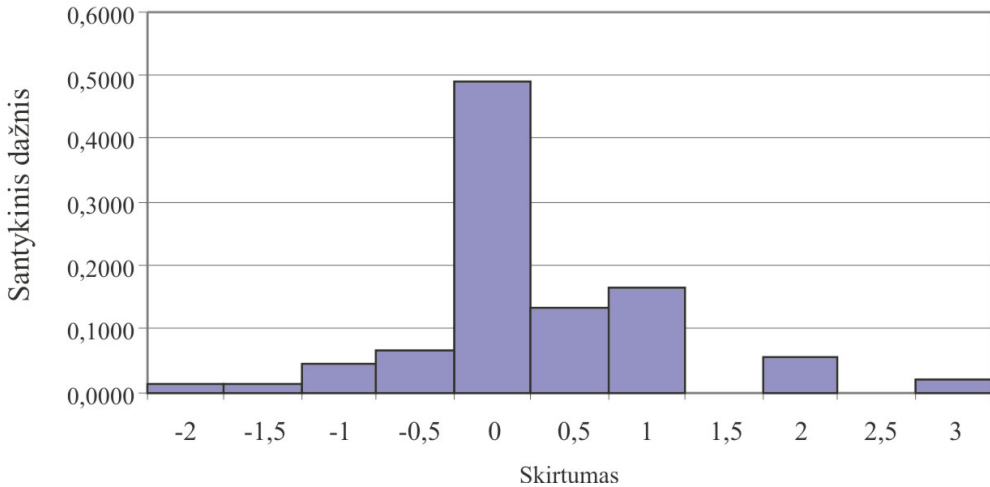
Iš lentelės matyti, jog tikimybė gauti kitą įvertinimą yra 0,43.

4 užduotis. Užduoties maksimalus galimų surinkti taškų skaičius yra 5. Skirtumų dažnių lentelė atrodo taip:

5 lentelė. Ketvirtos užduoties įverčių skirtumai

| Skirtumas | -2 | -1,5 | -1 | -0,5 | 0 | 0,5 | 1 | 1,5 | 2 | 2,5 | 3 |
|-----------|----|------|----|------|----|-----|----|-----|---|-----|---|
| Dažnis | 1 | 1 | 4 | 6 | 44 | 12 | 15 | 0 | 5 | 0 | 2 |

Duomenis pateikiame histograma:



3 pav. Ketvirtos užduoties įverčių skirtumų histograma

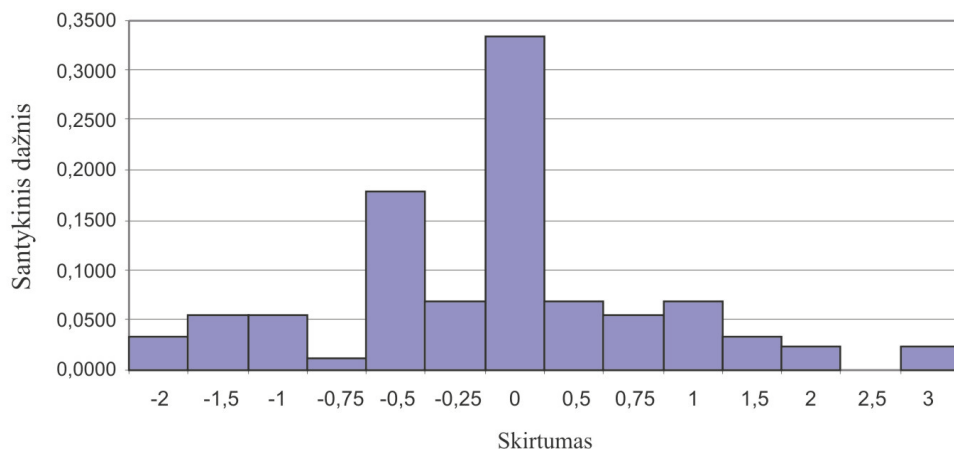
Iš pateiktos dažnių lentelės ir duomenų histogramos matyti, jog tikimybė gauti kitą įvertinimą už šią užduotį yra 0,51. Uždavinio sąlyga buvo tokia: „Išspręskite sistemą Gauso metodu“ ir maksimalus galimų surinkti taškų skaičius buvo 5, todėl ir išsibars-tymas didelis. Kiekvienas dėstytojas tas pačias aritmetines klaidas įvertina skirtingai, todėl būtų galima sugriežtinti vertinimo kriterijus. Pavyzdžiui, jei tik suvedė lygčių sis-temą į trapecijos pavidalą – 2 taškai, jei suvedamas padarė klaidą, bet iki galo sprendė gerai – 4 taškai ir pan. Duomenų išsibars-tymas yra didelis, tačiau dėstytojų nuomonė apie testuojamuosius yra gana vienoda, tai parodė atliktas statistinis tyrimas.

5 užduotis. Už šią užduotį buvo galima surinkti 3 taškus. Pateikiame lentelę su skirtumais:

6 lentelė. Penktos užduoties įverčių skirtumai

| Skirtumas | -2 | -1,5 | -1 | -0,75 | -0,5 | -0,25 | 0 | 0,5 | 0,75 | 1 | 1,5 | 2 | 2,5 | 3 |
|-----------|----|------|----|-------|------|-------|----|-----|------|---|-----|---|-----|---|
| Dažnis | 3 | 5 | 5 | 1 | 16 | 6 | 30 | 6 | 5 | 6 | 3 | 2 | 0 | 2 |

Šiuos duomenis pavaizduojame histograma:



4 pav. Penktos užduoties įverčių skirtumų histograma

Matome, jog tikimybė gauti kitą įvertinimą yra 0,67. Šią užduotį ekspertai vertino gana skirtingai, o tai galėjo atsitikti dėl per mažo balo skaičiaus, skirto teisingai atlikus šią užduotį, nes kiti klausimai, kuriems daugiausia buvo skirta 5 balai, parodė didesnį ekspertų nuomonių suderinamumą nei šiuo atveju, kai maksimumas – 3 balai. Juk skirtumas tarp 2 ir 3 balų trijų balų sistemoje yra santykinai didesnis nei skirtumas tarp 3 ir 4 balų keturių balų sistemoje. Tačiau tai yra tik hipotezė, kurią reiktų tikrinti.

Išvados ir rekomendacijos

1. Dėstytojais, tyrime vadinami ekspertais, testuojamuosius vertino vienodai. Tai rodo gauti statistiniai duomenys (1 lentelė), t. y. konkordacijos koeficiento W ir Frydmano statistikos Q reikšmės yra didelės. Tačiau penktosios užduoties statistikų reikšmės yra mažiausios, tai reiškia, jog vertinant šią užduotį ekspertų nuomonės išsiskyrė labiausiai.

2. Antrojo uždavinio sprendimui, t. y. studentų norminiam vertinimui (1), buvo pasirinktas Page'o L testas trendui nustatyti. Šis testas leidžia išdėstyti testuojamuosius pažymių vidutinių rangų didėjimo tvarka.

3. Trečiasis uždavinys – įvertinti testuojamųjų žinias balais. Šio kriterinio vertinimo uždavinio sprendimui galime naudoti (2) nelygybes, kurias gavome apskaičiavę bendrųjų balų suminių rangų mažiausius reikšminius skirtumus. Pavyzdžiui, testuojamieji 14, 10, 2 ir 4 gauna 10 balų, testuojamasis 12 – 9 balus, 9, 15, 7 ir 11 – 8 balus ir t. t. Arba būtų galima ir kitaip vertinti, t. y. testuojamiesiems 14, 10, 2, 4 rašyti 9, 12 – 8 balus ir t. t.

Pažymėsime, jog pirmoji užduotis nėra tinkama testavimui, nors dėstytojų vertinimai už šią užduotį buvo beveik identiški ($W=0,957$). Tai paaiškinama tuo faktu, kad šią užduotį didžioji testuojamųjų dauguma atliko teisingai, todėl ir vertintojų nuomonės neišsiskyrė. Užduotys, kurios yra per lengvos arba, atvirkščiai, per sunkios, netinka nurodytai testuojamųjų grupei, nes tokių užduočių rezultatai yra gerai prognozuojami.

Taip pat netinkama reikėtų pripažinti ir penktąją užduotį, kurios vertinimai bendrajam testo balui įtakos neturėjo ir ekspertai priėmė sprendimą dėl bendrojo testo balo remdamiesi tik pirmosiomis keturiomis užduotimis. Šią išvadą priėjome apskaičiavę mažiausius reikšminius skirtumus pagal kiekvieną iš penkių užduočių. Gavome, jog penktoje užduotyje nėra nė vieno reikšminio skirtumo tarp gretimų testuojamųjų suminių rangų. Taip atistiko dėl to, kad dėstytojų nuomonės ties šia užduotimi labai išsiskyrė. Jie tas pačias testuojamųjų klaidas vertino nevienareikšmiškai. Taip galėjo atsitikti ir dėl to, kad ši užduotis nėra tinkama šiam dalykui testuoti, t. y. dėl validumo stokos. Galimi keli šios problemos sprendimo būdai: pirmasis – nustatyti aiškius klaidų vertinimo kriterijus, jei tai nepadėtų, t. y. rezultatai būtų tokie pat, tuomet antrasis – pakeisti penktąją užduotį kita užduotimi, kuri būtų tinkamesnė šiam dalykui vertinti.

Metodika, kurią mes aprašėme, gali būti pritaikyta ne tik matematikos, bet ir kitų gamtos mokslų ar socialinių dalykų žinioms vertinti.

Pateiktoji metodika leidžia pagal žinių lygį išskirstyti testuojamuosius į grupes, taip pat pasiūlytas būdas įvertinti testuojamųjų žinias balais.

Atlikę statistinius skaičiavimus, pastebėjome, kad dėstytojų nuomonės, interpretuojant tas pačias klaidas, išsiskyrė, ypač tai parodė penktoji testo užduotis. Šiai problemai išspręsti galima siūlyti žinioms vertinti taikyti testus, kurie būtų sudaryti iš uždarojo tipo klausimų. Tokie testai padėtų sumažinti subjektyvumo įtaką žinių vertinimui, kadangi formuluojant testo klausimus vieną sudėtingą užduotį galima išskaidyti į klausimų blokus, kuriuose kiekvienas klausimas vertinamas atskirai. Vertinant išvengiama ir dviprasmiškumo. Testai gali būti tikrinami kompiuteriu. Tai padėtų išvengti žmogiškojo faktoriaus įtakos, o kartu ir subjektyvumo problemos vertinant studentų žinias. Kitas būdas išvengti subjektyvumo problemos – vertinimo kriterijų sugriežtinimas, tačiau tai reikalauja nemažai laiko, nes reikia atkreipti dėmesį į studentų klaidų priežasčių analizę: ar tos klaidos atsiranda dėl studentų tam tikrų žinių bei įgūdžių trūkumo (Krylovas et

al, 2002; Krylovas et al, 2007), ar dėl to, kad jie nesuprato konkrečios užduoties. Čia labiausiai pasireiškia vertinimo subjektyvumas. Vienas iš galimų šios problemos sprendimo būdų – vertinant atsižvelgti į tipines klaidas. Tokių klaidų žinojimas galėtų būti vienas iš vertinimo kriterijų.

Literatūra

- Andziulienė, B. *Žinių ir gebėjimų testavimas*. Klaipėda, 2004.
- Bagdonavičius, V.; Kruopis, J. *Matematinė statistika. 1 dalis*. Vilnius, 2007.
- Bulajeva, T. *Žinių ir kompetencijų vertinimas: kaip sukurti studentų pasiekimų vertinimo metodiką*. Vilnius, 2007.
- Conover, W. J. *Practical Nonparametric Statistics*. 3d Editon. New York, USA, 1999.
- Gage, N. L.; Berliner, D. C. *Pedagoginė psichologija*. Vilnius, 1994.
- Girdzijauskas, S. *Studentų žinių kontrolės ir vertinimo sistema*. Vilnius, 1997.
- Kriauzienė, R.; Krylovas, A.; Kosareva, N. Neparimetrinių statistinių metodų taikymas vertinant studentų žinias. *Tarptautinės konferencijos „Socialinės technologijos’10: iššūkiai, galimybės, sprendimai“*, vykusios 2010 m. lapkričio 25–26 d. Vilniuje, konferencijos medžiaga. Vilnius-Net. Vilnius: Mykolo Romerio universitetas, 2010, p. 63–68.
- Krylovas, A.; Raulynaitis, J. Studentų matematikos žinių kompleksinis vertinimas semestro metu. Lietuvos matematikos rinkinys. 2005, 44 (spec. nr.): 477–481.
- Krylovas, A.; Raulynaitis, J. VGTU pirmakursių mokyklinės matematikos žinių bei įgūdžių tikrinimas. Lietuvos matematikos rinkinys. 2005, 45 (spec. nr.): 261–265.
- Krylovas, A.; Raulynaitis, J.; Suboč O. Dėstytojų nuomonės apie studentų klaidas. Lietuvos matematikos rinkinys. 2002, 46 (spec. nr.): 158–162.
- Krylovas, A.; Suboč, O. Diskrečiosios matematikos dėstymas VGTU Elektronikos fakulteto pirmakursiams. Lietuvos matematikos rinkinys. 2005, 45 (spec. nr.): 266–269.
- Krylovas, A.; Suboč, O. Diskrečiosios matematikos testo klausimų analizė. Lietuvos matematikos rinkinys. 2006, 46 (spec. nr.): 163–166.
- Krylovas, A.; Suboč, O.; Kosareva, N. Studentų žinių įverčių statistinė analizė. Lietuvos matematikos rinkinys. 2007, 47 (spec. nr.): 254–258.
- Krylovas, A.; Suboč, O.; Kosareva, N. Diskrečiosios matematikos žinių tikrinimo testų lygiagrečiųjų variantų lygiavertiškumo tyrimas. Lietuvos matematikos rinkinys. 2007, 47 (spec. nr.): 249–253.
- Page, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*. 1963, 58: 216–230.
- Sprent, P.; Smeeton, Nc. *Applied Nonparametric Statistical Methods*. 4th edition. London, UK., 2007.

SUBJECTIVITY PROBLEM IN STUDENT ASSESSMENT: THEORETICAL AND PRACTICAL ASPECTS

Rima Kriauziene

Mykolas Romeris University, Lithuania, kriauziene@mruni.lt

Aleksandras Krylovas

Mykolas Romeris University, Lithuania, krylovas@mruni.lt

Natalja Kosareva

Vilnius Gediminas Technical University, Lithuania, natalja.kosareva@vgtu.lt

Summary. *Objective evaluation is not an easy task when assessing student attainment even if we are evaluating such a well-measured subject as knowledge of mathematics. This subject is discussed in a number of papers. Their authors are of the same opinion that even when the evaluation criteria is matching in detail, different teachers evaluate the same work differently. Subjectivity in knowledge assessment is researched in this paper. Not only teachers' individual characteristics in assessing student attainment are analyzed, but also problems of designing objective evaluation criteria of evaluated object, for example, solved problem. Authors analyze their own works and researches of other scientists in this field. The paper describes the authors' original experiments and the results of their statistical analysis. As the solution of the evaluation subjectivity problem authors propose a methodology which allows dividing students to groups according to their attainment level. This methodology could be applied not only for evaluation of mathematical knowledge, but also for attainment evaluation in other disciplines.*

The experiment, when 15 accidentally selected student's works in higher mathematics were independently verified by 6 teachers, is described in the paper. The analysis of the total test result and the results of individual tasks was performed with nonparametric statistical methods. To check the compatibility of teachers assessment Kendall's coefficient of concordance W was calculated, nonparametric Friedman's test was applied. The conclusion of the investigation was that all 6 teachers' estimates were rather similar when evaluating students' attainment except the estimation of the fifth task. Page's L test was applied to the total test result to determine trend and consequently norm-referenced estimate of tested students. Least significant differences between sums of neighbouring ranks were calculated to establish 5 groups in which students received the same marks, so the criterion-referenced evaluation of student attainment was performed. The first task is not suitable for testing of this group of testees, because it is too easy for them. Tasks that are too easy or, conversely, too hard, are not suitable, because the results of such assignments are well predicted and do not distinguish between students.

The fifth task should be recognized as inappropriate, because it does not affect the total assessment score and the criterion-referenced evaluation of the total test score is performed based on only the first four tasks.

The applied nonparametric statistical methods were described in the paper in detail. The results were summarised in inferences and recommendations. The recommendations to escape or reduce the influence of subjectivity in attainment evaluation were formulated:

- *to assign more points to the task, where teachers disagree in their assessment of students knowledge,*
- *to use standard tests for attainment evaluation,*
- *to formulate more strict evaluation criteria and agreement on the most often occurring mistakes of evaluation.*

The proposed methods could be applied not only for attainment in mathematical assessment, but also in other natural sciences.

Keywords: *attainment estimation, teaching of mathematics, nonparametric statistical methods, least significant difference, subjectivity of assessment, knowledge assessment tests.*