Liet. matem. rink., T. 42, spec. nr., 397–401

Matematikos žinių įvairių vertinimų suderinamumas

Aleksandras KRYLOVAS, Juozas RAULYNAITIS, Janina JAURIENĖ (VGTU)
el. paštas: akr@fm.vtu.lt

1. Įvadas

Minimalusis įskaitomas pažymys (5) rašomas VGTU studentui, įsisavinusiam 50–55% mokymo dalyko žinių. Tačiau žinios yra sunkiai išmatuojamos (žr., pavyzdžiui, [1]) ir tą patį studentą skirtingi dėstytojai dažnai vertina skirtingai. Šiame straipsnyje mes pateikiame kelis tokios įvairovės tyrimo rezultatus.

Mes nagrinėjame ryšius tarp įstojusių į VGTU vieno[1] fakulteto pirmojo kurso studentų mokyklinės matematikos pažymių ir jų pirmosios sesijos matematikos egzamino rezultatų. Taip pat įvairiomis priemonėmis tirtas studentų mokymosi rezultatų skirtingų įverčių suderinamumas.

Priimant studentus į Lietuvos aukštąsias mokyklas, jiems pridedami balai, atsižvelgiant į mokymosi lygį, laikomą egzaminą (valstybinį ar mokyklinį) ir pan. Todėl mūsų tyrimų rezultatai gali būti naudingi tokių priedų dydžiams optimizuoti.

Kitas tyrimas – bandymas suvienodinti įvairių dėstytojų vertinimus. Ribotas nemokančių už mokslą studentų skaičius neišvengiamai paaštrins nevienodų įvertinimų problemą ir neišvengiamai pareikalaus spręsti šį klausimą.

Mokyklinis metinis pažymys ir pirmoji sesija

Nagrinėjami vieno VGTU fakulteto pirmojo kurso dviejų srautų (imčių didumai $n_1 = 84$ ir $n_2 = 104$) studentų mokymosi rezultatai, kai pirmą srautą egzaminavo vienas dėstytojas, o antrą – kitas. Mes apsiribojame tik pirmo laikymo rezultatais, manydami, kad jie yra patikimesni, negu galutiniai. Koreliacinės lentelės (1 lentelė) langeliuose, kaip paprastai, nurodomi porų absoliutieji dažniai n_{ij} ($i = 1,2,3,4,5,6,7,8$; $j = 1,2,3,4,5,6,7$); dvitaškiu atskirti pirmojo ir antrojo srauto duomenys (vietoje 0:0 paliekame tuščią langelį). Lentelėje pažymėta: X – pirmosios sesijos matematikos egzamino pažymys, Y – mokyklinės matematikos metinis pažymys, neišskiriant mokymosi lygių. Lentelės paskutinėje eilutėje nurodytos dažnių sumos pagal stulpelius:

$$k_1^{(1)} = 2, \quad k_1^{(2)} = 5, \quad k_2^{(1)} = 21, \quad k_2^{(2)} = 22, \quad k_3^{(1)} = k_3^{(2)} = 25, \quad \ldots$$

[1] Mes neminėsime fakultetų pavadinimų bei dėstytojų pavardžių, vadovaudamiesi etikos samprotavimais.

1 lentelė. Universiteto ir mokyklos įvertinimai

X \ Y	10	9	8	7	6	5	4
10	1:0		1:0				
9	0:3	0:8	0:5	0:3			
8		3:5	0:6	1:5	0:1	0:1	0:1
7	1:1	2:8	0:2	2:4	1:1	0:3	
6		4:1	4:6	3:6	1:3	0:2	0:1
5	0:1	3:0	7:4	1:2	1:3	2:2	0:1
4		1:0		3:2		1:0	
3,2,1		8:0	13:2	8:1	3:4	5:4	5:2
k_j	2:5	21:22	25:25	18:23	5:12	8:12	5:5

Analizuojant pateiktus duomenis, gauti tokie rezultatai:

- pirmosios sesijos matematikos pažymių vidurkiai $\overline{x}^{(1)} = 3,94$; $\overline{x}^{(2)} = 6,46$;
- mokyklinių metinių pažymių vidurkiai $\overline{y}^{(1)} = 7,44$, $\overline{y}^{(2)} = 7,34$;
- koreliacijos koeficientai $r^{(1)} = 0,41$, $r^{(2)} = 0,51$.

Kai prie metinių A lygio pažymių pridėtas 1 balas, o prie metinių S lygio pažymių pridėti du balai (taip buvo, pavyzdžiui, 2001 m. priimant į septynias Lietuvos aukštąsias mokyklas), gauti tokie rezultatai:

- metinių pažymių vidurkiai $\overline{y}^{(1)} = 8,52$, $\overline{y}^{(2)} = 8,29$;
- koreliacijos koeficientai $r^{(1)} = 0,50$, $r^{(2)} = 0,58$.

Buvo modeliuojami įvairūs A lygio ir S lygio metinių pažymių priedai: atestato A lygio pažymys $+0,1i_A$ ir atestato S lygio pažymys $+0,1i_S$, kai i_A ir i_S įgyja reikšmes nuo 1 iki 45, norint gauti didžiausią koreliacijos koeficientą. Taigi didžiausias $r^{(1)} = 0,5276$ gautas, jei prie S lygio metinio pažymio būtų pridėta 4,0 balo, o prie A lygio pažymio būtų pridėta 1,4. Šis $r^{(1)}$ 0,0001 tikslumu dar būtų gaunamas, jei prie S lygio metinio pažymio būtų pridėta 3,9 balo, o prie A lygio pažymio būtų pridėta 1,4 arba jei prie S lygio metinio pažymio būtų pridėta 3,8 balo, o prie A lygio pažymio būtų pridėta 1,3. Antrojo srauto atveju didžiausias $r^{(2)} = 0,5976$ gautas, jei prie S lygio metinio pažymio būtų pridėta 3,5 balo, o prie A lygio pažymio būtų pridėta 2,5.

Matome, kad nors abiejų srautų studentų mokyklinių žinių pažymių vidurkiai yra beveik lygūs, dviejų dėstytojų sesijos įvertinimų vidurkiai skiriasi 2,52 balo. Akivaizdu, kad minėtus 50% žinių šie du dėstytojai supranta skirtingai. Tačiau koreliacijos koeficientų skirtumai yra ne tokie ryškūs. Kadangi Spirmeno ir Kendalo ranginės koreliacijos koeficientai šiuo atveju buvo beveik lygūs r, mes taikėme dar vieną pažymių suderinamumo priemonę: *inversijos intensyvumo matą* I_{XY}. Šis rodiklis buvo taikomas vietoje koreliacijos koeficiento, kai savo laiku Vilniaus inžinerinio statybos instituto matematikos katedros bendradarbiai irgi analizavo brandos egzaminų pažymius ir juos lygino su stojamųjų egzaminų ir pirmosios sesijos egzaminų rezultatais [2, 3].

Pažymių skirtumas $|i-j|$ vadinamas inversija arba žingsniu. Apskaičiuojamas bendrasis žingsnių skaičius (iš pradžių imami pirmojo srauto duomenys)

$$z=\sum_{i=1}^{8}\sum_{j=1}^{7}|i-j|\cdot n_{ij}=256$$

ir maksimalus žingsnių skaičius

$$\max z=\sum_{j=1}^{7}\max_{1\leqslant i\leqslant 8}|i-j|\cdot k_j=7k_1+6k_2+5k_3+4k_4+4k_5+5k_6+6k_7=427.$$

Tuomet inversijos intensyvumo matas

$$I_{XY}^{(1)}=\frac{z}{\max z}\approx 0,60.$$

Kai $0\leqslant I_{XY}<0,2$, sakoma, kad inversijos intensyvumo laipsnis *nežymus*, o požymių X ir Y ryšys *stiprus*; kai $0,2\leqslant I_{XY}<0,4$ – inversijos intensyvumo laipsnis *pastebimas*, o ryšys *esminis*; kai $0,4\leqslant I_{XY}<0,8$ – laipsnis *žymus* (stiprus), o ryšys *silpnas*; kai $0,8\leqslant I_{XY}\leqslant 1,0$, sakoma, kad inversijos intensyvumo laipsnis *labai stiprus*, o *ryšio* tarp požymių *nėra*.

Mūsų apskaičiuotas $I_{XY}^{(1)}=0,60$, todėl teigiame, kad ryšys tarp pirmosios sesijos pirmojo laikymo matematikos egzamino pažymių ir matematikos metinių pažymių yra silpnas. Antrojo srauto atveju inversijos intensyvumo matas $I_{XY}^{(2)}=0,29$. Taigi ryšys tarp pažymių yra esminis. Pastebėkime dar, kad antruoju atveju studentai buvo egzaminuojami testais, turinčiais daug klausimų.

Pratybų pažymiai ir sesijos rezultatas

Koreliacinėje lentelėje (2 lentelė) pažymėta: X – pirmosios egzaminų sesijos pirmojo laikymo matematikos pažymiai ir Y – pratybų pažymiai. Turimų imčių didumai $n_1=98$, $n_2=51$. Kaip ir anksčiau dvitaškiu atskirti dviejų srautų duomenys, kai kiekvieną srautą egzaminavo vienas dėstytojas, o pratybas vedė du dėstytojai.

Taikydami jau išdėstytą metodiką, gavome tokius rezultatus:
$\overline{x}^{(1)}=3,40;\ \overline{y}^{(1)}=6,95;\ \overline{r}^{(1)}=0,66;$
$\overline{x}^{(2)}=4,49;\ \overline{y}^{(2)}=4,80;\ \overline{r}^{(2)}=0,90;$
$I_{XY}^{(1)}=0,53;\ I_{XY}^{(2)}=0,15.$

Matome, kad antrojo srauto pratybas vedančio ir paskaitas skaitančio dėstytojų vertinimai yra pakankamai gerai suderinti, ko negalima pasakyti apie pirmąjį srautą.

Įvairūs žinių vertinimai turi teigiamas koreliacijas ir, be abejo, teisingai atspindi tendenciją aukštesniais balais vertinti stipresnius studentus ir žemesniais – silpnesnius. Tačiau esminiai žinių vertinimų vidurkių skirtumai rodo, kad vertinimų skalių „pradiniai taškai" nesutampa ir 50% žinių „matavimo" tikslumas nėra patenkinamas.

2 lentelė
Dėstytojų įvertinimai

X \ Y	10	9	8	7	6	5	4	3	2	1
10		0:2								
9			0:4							
8		0:1	1:1		0:1		0:1			
7	1:0	0:1	0:1	0:2						
6	6:0	4:0		2:1	0:3	0:3	0:1			
5	1:0	4:0	6:0	0:2	2:0	0:1	0:1			
4	2:0	6:0	2:0	4:0	7:0	2:2				
3	1:0	3:0	2:0	3:0	2:1	3:1	0:1	0:1		
2		1:0	3:0	3:0		2:0	3:2	0:3	1:1	
1		1:0		2:0	5:0	2:0	5:0	4:2	1:6	1:5
k_j	11:0	19:4	14:6	14:5	16:5	9:7	8:6	4:6	2:7	1:5

Tarpinis egzaminas

Mes siūlome vieną minimalių žinių lygio suderinimo, t.y., vertinimų skalės pradinio taško nustatymo, metodą. Buvo nagrinėjami dviejų pirmojo kurso studentų tarpinio egzamino (25 klausimų testo) rezultatai ir pratybų dėstytojų pažymiai. Užpildome tokią santykinių dažnių lentelę:

		testo rezultatas	
		neišlaikyta	išlaikyta
pratybų pažymys	neišlaikyta	LL	LG
	išlaikyta	GL	GG

Taigi LL yra dalis studentų, kurių žinias ir pratybų dėstytojas įvertino mažiau kaip 5 balais, ir testo rezultatas buvo blogas (mažiau 13 teisingų atsakymų), GG reiškia, kad pratybų pažymys yra 5 arba daugiau ir testo rezultatas 13 ir daugiau. Gauti tokie rezulatai:

pirmoji grupė: $n = 36, LL = 0,25, LG = 0,25, GL = 0,08, GG = 0,42$;

antroji grupė: $n = 27, LL = 0,30, LG = 0,04, GL = 0,11, GG = 0,55$.

Čia n yra studentų skaičius.

Skaičius $LG + GL$ rodo dalį studentų, kurių žinios testu ir pratybų dėstytoju įvertintos skirtingai, t.y., rodo tam tikrą įvertinimų sistemų defektą. Matome, kad tokių studentų buvo 33% pirmoje grupėje ir 15% antroje. Pastebėkime dar, kad atlikus panašius tyrimus su dviem antrojo kurso grupėmis, buvome gavę 42% ir net 48%.

Minimalus teisingų testo atsakymų skaičius – *kriterijus* (13 iš 25) buvo praneštas studentams iš anksto ir yra nustatytas remiantis straipsnio pradžioje minėtų 50% įgytų žinių lygiu. Akivaizdu, kad sunkesniam testui kriterijus galėtų būti mažesnis negu pusė teisingai išspręstų uždavinių. Antra vertus, nevisada sėkmingai pavyksta prognozuoti testo sunkumą. Ir svarbiausia: skirtingi vertintojai skirtingai įvertina kas yra 50% žinių.

Nustačius kitą tarpinio egzamino kriterijų, (pavyzdžiui, ne 13, o 15 balų), gaunamas ir kitas įvertinimų defektas $LG + GL$. 3 lentelėje patekiami dydžių $LG + GL$ skaičiavimo rezultatai, esant skirtingoms kriterijaus reikšmėms.

3 lentelė
Kriterijaus nustatymas

Kriterijus	11	12	13	14	15	16
Procentinis rangas	37	37	40	46	52	63
$n = 36$	0.333	0.361	0.333	0.278	0.361	0.333
$n = 27$	0.185	0.111	0.148	0.148	0.185	0.296

Taigi, norint minimizuoti įvertinimų nesutapimų dalį $LG+GL$, pirmuoju atveju reikia nustatyti kriterijų 14 teisingai atsakytų klausimų iš 25, o antruoju – 12 klausimų. Surašyti antroje lentelės eilutėje procentiniai rangai rodo studentų, gavusių atitinkamą testo balą, dalį. Pavyzdžiui, kriterijų 12 atitinkantis procentinis rangas 37 reiškia kad 12 arba mažiau balų yra gavę 37% visų abiejų grupių ($n = 36 + 27$) studentų, t.y., tiek studentų neišlaikė testo, kai kriterijus buvo 13.

Išvados ir pasiūlymai

Straipsnyje dėstomi keli žinių įverčių pertvarkymų metodai, kurių teorinis pagrindimas išplaukia iš žinomos prielaidos (žr., pavyzdžiui, [4]), kad tokie įverčiai sudaro tik intervalinę skalę (dažnai net dar mažiau – tik tvarkos skalę) bei neturi objektyviai nustatyto pradžios taško. Taigi mes ir siūlome skirtingų vertinimų sistemų pradžios taškų suderinimo metodus. Tai gali būti naudinga informacija dėstytojams apie jų reikalavimus studentams. Atlikus panašius tyrimus, esant didesniam studentų skaičiui, galima pagrįsti įvairių priedų įvedimą bei jų dydį, priimant studentus į skirtingas specialybes.

Egzaminuojant studentus galima rekomenduoti plačiau taikyti normomis pagrįstus vertinimus (žr., [5]), t.y., atsižvelgti į visų studentų rezultatus, pavyzdžiui, skaičiuojant procentinius rangus. Aišku, kad studentų turi būti pakankamai daug, tarkime, visas srautas, o ne mažos atskiros grupės.

Literatūra

[1] N.L. Gage, D.C. Berliner, *Pedagoginė psichologija*, Alma litera, Vilnius (1994).

[2] V. Liutikas, Inversijos intensyvumo matas ir jo taikymas mokinių mokėjimui įvertinti, *Mokymo ir auklėjimo klausimai*, PMTI, Vilnius (1983), pp. 133–136.

[3] J. Jaurienė, E. Dagienė ir kt., Inversijos intensyvumo mato taikymas lyginant studentų mokymosi pažangumo rezultatus, *Mokymo ir auklėjimo klausimai*, PMTI, Vilnius (1983), pp. 136–138.

[4] G.V. Glass, J.C. Stanley, *Statistical Methods in Education and Psychology*, Prentice–Hall, Inc., Englewood Cliffs, New Jersey (1970).

[5] S. Girdzijauskas, E. Stumbrys, Studentų žinių norminis vertinimas, *Aukštojo mokslo sistemos ir didaktika*, Technologija, Kaunas (2000), pp. 46–57.

Adjustment of different results for mathematics examinations

A. Krylovas, J. Raulynaitis, J. Jaurienė

The article deals with the adjustment of the evaluation of knowledge by different professors of mathematics.