Liet. matem. rink., **47**, *spec. nr.*, 2007, 254–258

Studentų žinių įverčių statistinė analizė

Aleksandras KRYLOVAS, Olga SUBOČ, Natalja KOSAREVA (VGTU)
el. paštas: {akr,os}@fm.vtu.lt

1. Šio tyrimo tikslas – palyginti dvi diskrečiosios matematikos žinių tikrinimo metodikas: uždarojo tipo testą su pateiktais atsakymų variantais (iš viso 20 klausimų) ir rašomąjį darbą, kuriame buvo pateikta 17 klausimų. Testo ir rašomojo darbo klausimai parinkti iš tų pačių diskrečiosios matematikos temų. Eksperimente dalyvavo 45 VGTU Fundamentinių mokslų ir Elektronikos fakultetų studentai. Išspręstas užduotis nepriklausomai patikrino du dėstytojai. Kiekvieną rašomojo darbo klausimą dėstytojas galėjo įvertinti nuo 0 iki 1 taško, tačiau intervalo $[0,1]$ skaidymo žingsnis nebuvo reglamentuojamas.

2. Naudojami tekste žymenys: TP – skirtumas tarp teisingai $(+)$ ir klaidingai $(-)$ atsakytų testo klausimų skaičių. $TESTAS$ – testo pažymys 10 – balėje skalėje, skaičiuojamas pagal formulę $TESTAS = 1 + TP/2$. Jei $TESTAS$ reikšmė viršija 10, ji pakeičiama į 10, neigama $TESTAS$ reikšmė pakeičiama į 0. $DEST1$, $DEST2$ – pirmojo ir antrojo dėstytojo rašomojo darbo įverčiai 10-balėje skalėje. 1 lentelėje pateiktas pradinių duomenų fragmentas.

3. Ekperimento metu gauti rezultatai buvo tiriami, naudojant statistinių programų paketą SPSS ([3]). Iš aprašomųjų statistikų 2 lentelėje galime daryti pirmas išvadas. Testo pažymio $TESTAS$ vidurkis yra žemesnis už $DEST1$, $DEST2$ vidurkius. Visų trijų atsitiktinių dydžių medianos sutampa. Abiejų dėstytojų rašomojo darbo pažymių vidurkų skirtumas yra $0,44$ balo. $TESTAS$ standartinis nuokrypis viršija $DEST1$, $DEST2$ standartinius nuokrypius.

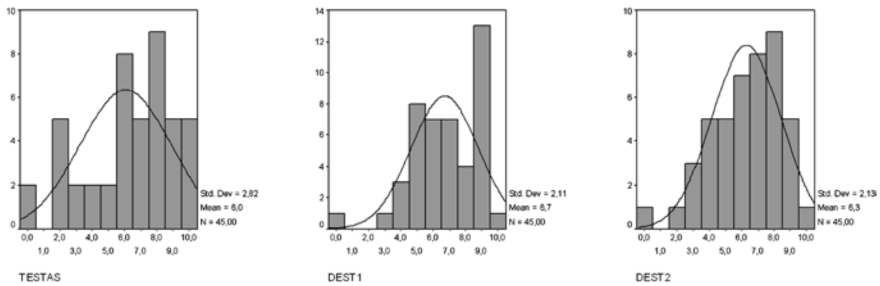
Iš pateiktų 1 pav. histogramų matome, kad atsitiktiniai dydžiai $TESTAS$ ir $DEST1$ turi kelias modas (reikšmes, kuriuose dažniai pasiekia lokalųjį maksimumą), tuo tarpu $DEST2$ yra unimodalusis atsitiktinis dydis.

1 lentelė. Pradinių duomenų fragmentas

	$+$	$-$	tp	testas	dest1	dest2
studentas1	7	4	3	2,5	6,8	6,4
studentas2	17	1	16	9	9,1	8,7
studentas3	11	1	10	6	5,9	6,3
…	…	…	…	…	…	…

2 lentelė. Aprašomųjų statistikų lentelė

		TESTAS	DEST1	DEST2
N	Valid	45	45	45
	Missing	0	0	0
Mean		6,0444	6,6956	6,2600
Median		6,5000	6,5000	6,5000
Mode		8,00	9,10	7,20
Std. Deviation		2,82405	2,10745	2,13226
Skewness		–,622	–,745	–,775
Std. Error of Skewness		,354	,354	,354
Kurtosis		–,593	,760	,492
Std. Error of Kurtosis		,695	,695	,695
Sum		272,00	301,00	281,70



1 pav. Testo ir dviejų dėstytojų vertinimų histogramos.

Visų trijų atsitiktinių dydžių moda yra didesnė už medianą ir vidurkį, todėl visi atsitiktiniai dydžiai yra asimetriniai su neigiamais asimetrijos koeficientais. $TESTAS$ eksceso koeficientas yra neigiamas (plokščiaviršūniškumas), $DEST1$ ir $DEST2$ eksceso koeficientai yra teigiami (smailaviršūniškumas). Iš asimetrijos ir eksceso koeficientų dydžių, taip pat jų standartinių paklaidų galime spręsti apie stebimo atsitiktinio dydžio skirstinio normališkumą. Normaliojo atsitiktinio dydžio asimetrijos ir eksceso koeficientai yra lygūs nuliui. Kai asimetrijos ir eksceso koeficientų dydžių absoliučios reikšmės neviršija jų standartinių paklaidų, galime daryti išvadą apie tai, kad asimetrijos ir eksceso koeficientai reikšmingai nesiskiria nuo nulio, t.y. atsitiktinio dydžio skirstinys reikšmingai nesiskiria nuo normaliojo skirstinio. Atsitiktiniams dydžiams $TESTAS$, $DEST1$ ir $DEST2$ šios normališkumo sąlygos nėra išpildytos.

Pastebėsime, kad patikrinus normališkumą Kolmogorovo–Smirnovo kriterijumi, visiems nagrinėjamiems atsitiktiniams dydžiams gauname pakankamai dideles p reikšmes ($p > 0,05$). Tai reiškia, kad pagal turimus imties duomenis, esant reikšmingumo lygmeniui $\alpha = 0,05$, hipotezės apie atsitiktinių dydžių $TESTAS$, $DEST1$ ir $DEST2$ skirstinių normališkumą atmesti negalime. Ši prieštara gali būti paaiškinta tuo, kad Kolmogorovo–Smirnovo kriterijus nėra pakankamai jautrus esant santykinai nedideliam stebėjimų skaičiui ($n < 50$).

Apibendrinant galima teigti, kad nagrinėjamų atsitiktinių dydžių skirstiniai statistiškai reikšmingai skiriasi nuo normaliojo skirstinio, stebėjimų imtis nėra didelė, matavimai atlikti rangų skalėje. Todėl vietoje parametrinių kriterijų tikslinga ir korektiška taikyti neparametrinius kriterijus. Tiriant kintamųjų tarpusavio ryšius, vietoje Pirsono koreliacijos koeficiento naudosime Spirmeno ranginės koreliacijos koeficientą.

Viena vertus, atsisakydami parametrinių kriterijų, prarandame dalį pradinių duomenų, be to parametriniai kriterijai yra galingesni už neparametrinius. Antra vertus, neparametriniai kriterijai turi tam tikrų privalumų. Pirma, ranginiai koreliacijos koeficientai nėra tokie jautrūs išskirtims, t.y. ekstremaliai mažoms arba ekstremaliai didelėms stebėjimų reikšmėms, kaip Pirsono koreliacijos koeficientas. Antra, ranginiai koreliacijos koeficientai yra bet kokio monotoninio, ne tik tiesinio ryšio tarp atsitiktinių dydžių matas.

Spirmeno ranginės koreliacijos koeficientas ρ skaičiuojamas pagal formulę:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},$$

čia $d_i = R_{xi} - R_{yi}$ – i-jo stebėjimo rangų skirtumas, n – rangų porų skaičius (stebėjimų skaičius). Koreliacinė analizė rodo, kad ryšys tarp tiriamų atsitiktinių dydžių yra stiprus, visi koreliacijos koeficientai reikšmingai skiriasi nuo nulio $(p < 0,01)$.

Dviejų dėstytojų rašomojo darbo įverčių koreliacijos koeficientas yra $0,887$, testo įverčio $TESTAS$ koreliacijos koeficientas su pirmojo bei antrojo dėstytojų rašomojo darbo įverčiais $DEST1$ ir $DEST2$ yra lygūs atitinkamai $0,745$ ir $0,784$.

Pažymių skirstinių palyginimui buvo panaudotas Vilkoksono ženklų kriterijus (Wilcoxon signed rank sum test) priklausomoms imtims ([2]). Jei turime dvi dydžio n imtis $(X_1, X_2, \ldots, X_n)$ ir $(Y_1, Y_2, \ldots, Y_n)$, apskaičiuojame $d_i = X_i - Y_i$, $i = 1, 2, \ldots, n$. Randame visus $|d_i|$ ir juos ranguojame. Rangų, atitinkančių teigiamas d_i reikšmes, sumą pažymėsime T^+. T^+ yra Vilkoksono kriterijaus statistika. Mūsų atveju imtis yra pakankamai didelė $(n \geqslant 25)$. Todėl naudojama Vilkoksono statistikos standartinio normaliojo atsitiktinio dydžio aproksimacija, t.y. kriterijaus statistika:

$$Z = \frac{T^+ - \mu}{\sigma}, \quad \text{čia} \quad \mu = \frac{n(n+1)}{4}, \quad \sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}.$$

Pasirenkame reikšmingumo lygmenį α (pvz., $\alpha = 0,05$). Tikrinsime hipotezes apie testo ir rašomojo darbo įverčių skirstinių sutapimą:

H_0: Testo ir rašomojo darbo įverčių skirstiniai sutampa (kiekvienam dėstytojui atskirai).

Alternatyvi hipotezė:

H_1: Testo ir rašomojo darbo įverčių skirstiniai skiriasi.

Taip pat tikrinsime hipotezę apie abiejų dėstytojų rašomojo darbo įverčių skirstinių sutapimą. Rezultatai pateikti 3 lentelėje.

Esant pasirinktam reikšmingumo lygmeniui $\alpha = 0,05$, statistiškai nesiskiria kiekvieno dėstytojo rašomojo darbo įverčio ir testo įverčio $DEST1$ ir $TESTAS$ $(p = 0,069)$, $DEST2$ ir $TESTAS$ $(p = 0,536)$ skirstiniai. Tačiau, abiejų dėstytojų rašomojo darbo įverčių skirstiniai reikšmingai skiriasi $(p = 0,005)$. Todėl buvo atliktas

3 lentelė. Vilkoksono ženklų kriterijaus rezultatai

	DEST1 – TESTAS	DEST2 – TESTAS	DEST2 – DEST1
Z	–1,818[a]	–,619[a]	–2,834[b]
Asymp. Sig. (2-tailed)	,069	,536	,005

[a] Based on negative ranks.
[b] Based on positive ranks.

4 lentelė. Kryžminių dažnių lentelė

		ISL2		Total
		,00	1,00	
ISL1	,00	7	1	8
	1,00	4	33	37
Total		11	34	45

tolimesnis tyrimas. Buvo sukurti du nauji dichotominiai atsitiktiniai dydžiai $ISL1$ ir $ISL2$, įgyjantys reikšmę 0, kai atitinkamai pirmasis arba antrasis dėstytojas studento rašomąjį darbą įvertino neigiamai (žemiau, nei 5 balai), ir įgyja reikšmę 1, kai gautas balas už rašomąjį darbą yra nemažesnis už 5. T.y. nulinė reikšmė reiškia, kad studentas neišlaikė egzamino pas tam tikrą dėstytoją, o reikšmė 1 reiškia, kad egzaminas pas šį dėstytoją išlaikytas. Šiems naujiems atsitiktiniams dydžiams buvo sudaryta dvireikšmių požymių dažnių lentelė (4 lentelė) ir pritaikytas Maknemaro kriterijus [1]. Tai yra χ^2 kriterijaus hipotezes atsitiktinių dydžių $ISL1$ ir $ISL2$ imčių homogeniškumui tikrinti analogas priklausomoms dvireikšmėms imtims.

Nulinę hipotezę galima būtų suformuluoti taip:

H_0: Išlaikiusių studentų proporcija tarp visų rašiusių rašomąjį darbą nepriklauso nuo to, kuris dėstytojas tikrino rašomąjį darbą.

Alternatyvi hipotezė:

H_1: Išlaikiusių studentų proporcija tarp visų rašiusių rašomąjį darbą priklauso nuo to, koks dėstytojas vertino darbą.

Iš 4 lentelės matome, kad 7 studentai gavo neigiamus įvertinimus pas abu dėstytojus, 4 studentai buvo neigiamai įvertinti antrojo dėstytojo ir teigiamai įvertinti pirmojo dėstytojo, 1 studentas buvo neigiamai įvertintas pirmojo dėstytojo ir teigiamai įvertintas antrojo dėstytojo, 33 studentai abiejų dėstytojų buvo įvertinti teigiamai. Maknemaro χ^2 kriterijaus statistikai skaičiuoti yra naudojami tik duomenys apie destytojų skirtingai įvertintus rašomuosius darbus. Pažymėsime n_{ij} dažnį, esantį i-je eilutėje ir j-me stulpelyje. Maknemaro χ^2 kriterijaus statistika: $\chi^2 = \frac{(|n_{21}-n_{12}|-1))^2}{(n_{21}+n_{12})}$ asimptotiškai turi χ^2 skirstinį su 1 laisvės laipsniu. Kai $n_{12}+n_{21}<10$, vietoje χ^2 naudojama binominio atsitiktinio dydžio aproksimacija.

Maknemaro dvipusio kriterijaus p reikšmė yra lygi $0,375$ (5 lentelė). Esant pasirinktam reikšmingumo lygmeniui $\alpha = 0,05$, hipotezė apie gavusių teigiamą rašo-

5 lentelė. Maknemaro kriterijaus rezultatai

	Value	Exact Sig. (2-sided)
McNemar	,375[a]	
N of Valid Cases	45	

[a] Binomial distribution used.

mojo darbo įvertinimą proporcijų lygybę, vertinant dviem dėstytojams, negali būti atmesta.

4. Išvados

Tyrimas patvirtino žinomą faktą (žr. [5]), kad skirtingi tų pačių studentų žinių vertinimai nors ir rodo stiprias koreliacijas, tačiau neleidžia stabiliai gauti tų pačių įverčių. Du dėstytojai, net pakankamai detaliai suderinę vertinimo kriterijus, tuos pačius darbus įvertino gana skirtingai.

Literatūra

1. V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai*. I, TEV, Vilnius (2000).
2. V. Čekanavičius, G. Murauskas, *Statistika ir jos taikymai*. II, TEV, Vilnius (2004).
3. A.D. Nasledov, *Matematičeskije metody psichologičeskogo issledovanija. Analiz i interpretacija dannych*, Reč, SPb. (2004).
4. B. Bitinas, *Edukologinis tyrimas*: *sistema ir procesas*, Kronta, Vilnius (2006).
5. A. Krylovas, J. Raulynaitis, O. Subač, Dėstytojų nuomonės apie studentų klaidas, *Liet. matem. rink.*, **46**(spec. nr.), 158–162 (2006).

SUMMARY

A. Krylovas, O. Subač, N. Kosareva. Statistical analysis of student's marks obtained by different methods

Students' knowledge of discrete mathematics was tested using two methods – 'close-type' test and 'in-writing'. 'In-writing' was independently evaluated by two different teachers. These results were statistically analyzed in this article.

Keywords: discrete mathematics, tests, statistical analysis.