

Palyginamoji pažangumo analizė: panelinės logistinės regresijos modelis

Marijus Radavičius^{1,2}, Pavel Samusenko²

¹ *Matematikos ir informatikos institutas*
Akademijos g. 4, LT-08663 Vilnius

² *Vilniaus Gedimino technikos universitetas*
Saulėtekio al. 11, LT-10223 Vilnius

E. paštas: mrad@ktl.mii.lt; pavels.vgtu@gmail.com

Santrauka. Studentų pažangumo duomenims tirti buvo parinktas sąlyginės logistinės regresijos modelis. Šis modelis leidžia palyginti skirtingų egzaminų sudėtingumą ir identifikuoti susijusius veiksnius.

Raktiniai žodžiai: IRT, binarinis užduoties atsakas, maišantieji parametrai, Rasch modelis, sąlyginė logistinė regresija, studentų pažangumas.

1 Įvadas

Socialiniuose ir humanitariniuose moksluose dažnai tenka tirti požymius ir savybes, kurie negali būti tiesiogiai išmatuoti, pagal prigimtį yra nekiekybiniai. Pavyzdžiui, studento matematiniai sugebėjimai, kokio nors dalyko žinios, darbuotojo profesionalumas, partnerio patikimumas ir pan. Tokiems požymiams apibrėžti kiekybiškai sudaromi įvairūs testai, kriterijai, kurie matuoja skaitines charakteristikas, potencialiai susijusias su tiriamu požymiu, atspindinčias vieną ar kitą tiriamo požymio aspektą. Šiuos klausimus nagrinėja *matavimų teorija* (angl. *measurement theory*).

Šiuolaikinės matavimų teorijos, kuri paprastai vadinama *apklausos arba testo atsakymų teorija* (laisvas angliško termino *item response theory* vertimas; trumpai IRT), matematinį pagrindą sudaro Rasch modelis [3]. Jo esmę sudaro prielaida apie tam tikrą matuojamo objekto (testuojamo individo ar asmens) ir pačio matuoklio (testo) simetriją: tiriamojo požymio išreikštumas, jo lygis ir testo klausimų sudėtingumas matuojamas toje pačioje skalėje. Tokiu būdu, tiriamieji patys yra tarsi naudojamų testų „matuokliai“ (testai), nes tik jų atsakymai leidžia ištirti ir palyginti naudojamų testų tinkamumą, sudėtingumą, patikimumą ir kitas savybes.

Matematinė prasme klasikinis Rasch modelis yra atskiras logistinės regresijos atvejis. Šiame darbe VGTU studentų pažangumo duomenims taikomas jo apibendrinimas įvedant papildomus aiškinančiuosius kintamuosius. Remiantis *panelinės* logistinės regresijos su fiksuotais veiksniais modeliu siekiama išsiaiškinti egzaminuojančių dėstytojų vertinimo sistemų ypatumus.

Kitame skyrelyje pateikiamas trumpas pagrindinių IRT modelių aprašymas bei šiame darbe taikytas logistinės regresijos modelis [4]. Trečiame skyrelyje aprašomi naudoti duomenys ir aptariami gauti rezultatai. Pabaigoje pateiktos išvados.

2 IRT modeliai

Tegu $y_{ij} \in \{0, 1\}$, j yra *testo užduoties* (angliškai *item*, verčiant pažodžiui „apklausos punktas“) numeris, $j = 1, \dots, n$, i yra tiriamojo (testuojamojo individo) numeris, $i = 1, \dots, m$. Įvykis $\{y_{ij} = 1\}$ reiškia, kad i -asis tiriamasis sėkmingai atliko j -ąją testo užduotį.

IRT remiasi prielaida, kad i -ojo individo sėkmę atliekant testo užduotis nulemia jo gebėjimų lygis, kuris traktuojamas kaip latentinis (paslėptasis) kintamasis θ_i [5].

Užduočių sudėtingumas matuojamas toje pačioje skalėje ir aprašomas parametru b : j -osios užduoties parametras b_j galima interpretuoti kaip gebėjimų lygį, kurio reikia tam, kad tą užduotį sėkmingai atlikti su tikimybe $1/2$. Tokiu būdu, i -ojo individo tikimybę sėkmingai atlikti j -ąją užduotį nusako skirtumas $\theta_i - b_j$, $i = 1, \dots, m$, $j = 1, \dots, n$. Vienas iš būdų apibrėžti tą tikimybę yra *trijų parametru logistinis modelis* [1, 2]

$$\mathbf{P}(y_{ij} = 1) = \frac{1 - c_j}{1 + e^{-a_j(\theta_i - b_j)}}. \quad (1)$$

Čia įvesti du papildomi testo užduoties parametrai, a_j ir c_j , kurių nebuvo klasikiniam *Rasch modelyje* (jame $a_j = 1$, $c_j = 0$) [3]. Parametras a_j aprašo j -osios užduoties jautrumą gebėjimų pokyčiams, jos skiriamąją gebą, o *spėjimo parametras* c_j nusako tikimybę sėkmingai atlikti j -ąją užduotį *visiškai atsitiktinai*, net ir jeigu tiriamasis nieko apie ją nežino.

Skirtingai negu klasikinis Rasch modelis trijų parametru logistinis modelis nepriklauso apibendrintųjų tiesinių modelių klasei, kas ženkliai apsunkina parametru įverčių apskaičiavimą ir modelio parinkimo uždavinį. Standartiniai paketai šiuo atveju nepritaikomi, ir reikia specialių programų. Kita IRT modelių problema yra sąlygota didelio nežinomų parametru skaičiaus, kuris yra proporcingas $m + n$. Paprastai užduočių skaičius n būna nedidelis, bent jau gerokai mažesnis už tiriamųjų skaičių m , ir tuomet nežinomų parametru skaičius yra proporcingas imties dydžiui $N = m \cdot n$. Vadinasi, neegzistuoja pagrįsto individo gebėjimų lygio θ_i įvertinio, ir todėl θ_i , $i = 1, \dots, m$, traktuojami kaip *maišantieji* parametrai, kuriuos koku nors būdu reiktų eliminuoti, kad būtų galima gauti pagrįstuosius užduočių parametru įvertinius. Vienas iš tam skirtų modelių aprašytas žemiau.

Panelinės logistinės regresijos su fiksuotais efektais modelis yra ekonometrijoje plačiai taikomo panelinių duomenų su fiksuotais veiksniais analogas binariniam aiškinamajam kintamajam. Jis dar vadinamas *sąlyginiu logistinės regresijos modeliu* [4].

Tegu $x_{ij} \in \mathbf{R}^k$ žymi su apklausa susijusių *aiškinančiųjų kintamųjų* vektorių ($j = 1, \dots, n$, $i = 1, \dots, m$). Šiame darbe tariama, kad sąlyginį y_{ij} skirstinį, kai žinomos aiškinančiųjų kintamųjų reikšmės x_{ij} , nusako logistinės regresijos modelis

$$\mathbf{P}\{y_{ij} = 1 \mid x_{ij} = x\} = \frac{\exp\{\theta_i - b_j + \beta^\top x_{ij}\}}{1 + \exp\{\theta_i - b_j + \beta^\top x_{ij}\}}. \quad (2)$$

Kad modelis būtų identifikuojamas, dalis parametru b turi būti lygūs 0. Nemažinant bendrumo imama $b_j = 0$, $j = n_1 + 1, \dots, n$, su $n_1 + k < n$. Užduotis su numeriais $j = n_1 + 1, \dots, n$ vadinsime *kontrolinėmis*. Tokiu būdu, narys $\beta^\top x_{ij}$ aprašo papildomų aiškinančiųjų kintamųjų x_{ij} sąryšį su kontrolinių užduočių rezultatais. Klasikiniam Rasch modelyje šio nario nėra. Kadangi x_{ij} , kaip rodo jo indeksai, gali būti susijęs

ties su individo, tiek ir su apklausos kontrolinės užduoties savybėmis, tai (2) modeliu galima aprašyti ir tam tikras tų savybių sąveikas. Parametras b_j parodo j -osios užduoties ir kontrolinių užduočių su tomis pačiomis savybėmis x_{ij} sudėtingumo skirtumą ($j = 1, \dots, n_1$).

Kaip jau buvo minėta, sąlyginiame logistinės regresijos modelyje (analogiškai kaip panelinių duomenų su fiksuotais veiksniais modeliuose) laikoma, kad tiriamųjų skaičius m yra didelis (proporcingas imties dydžiui). Taigi, parametrai θ_i , $i = 1, \dots, m$, yra *maišantieji*, ir jiems eliminuoti vietoje pilnosios tikėtinumo funkcijos imama *sąlyginė* tikėtinumo funkcija su sąlyga, kad yra žinoma maišančiųjų parametrų pakankamos statistikos reikšmė. Nagrinėjamu atveju (2) ta pakankama statistika yra sudaryta iš kiekvieno individo i sėkmingai atliktų užduočių kiekio. Kadangi individui i sėkmingo j -osios užduoties ir n -osios užduoties atlikimo *galimybių* (arba *šansų*, angl. *odds*) O_j ir O_n ,

$$O_j := \frac{\mathbf{P}\{y_{ij} = 1 \mid x_{ij} = x\}}{\mathbf{P}\{y_{ij} = 0 \mid x_{ij} = x\}}, \quad j = 1, \dots, n, \quad (3)$$

santykis

$$\frac{O_j}{O_n} = \exp \{ -b_j + \beta^\top (x_{ij} - x_{in}) \}$$

nepriklauso nuo maišančiojo parametro θ_i , tai nuo jų nepriklauso ir sąlyginė tikėtinumo funkcija [4]. Kai n_1 ir k yra fiksuoti, maksimizuojant sąlyginio tikėtinumo funkciją gaunami pagrįsti ir asimptotiškai normalieji nežinomų parametrų b_j , $j = 1, \dots, n_1$, ir β įvertiniai. Tačiau jų apskaičiavimo procedūros sudėtingumas didėjant $n_1 + k$ gali augti labai greitai (eksponentiškai). Todėl sąlyginė logistinė regresija gali būti praktiškai nepritaikoma didelio užduočių kiekio n_1 atveju.

3 Rezultatai

3.1 Duomenys

Tyrime naudojami duomenys apie VGTU inžinerinės informatikos (II) specialybės studentų 2000–2005 metų egzaminų pažymius bei bendra nekonfidenciali informacija apie egzaminavusius dėstytojus. Bendras studentų (*tiriamų individų*) skaičius $m = 1238$. Kiekvieno dėstytojo egzaminas traktuojamas kaip atskira *testo užduotis*. Taigi, užduočių yra tiek kiek yra egzaminuojančių dėstytojų ($n = 238$).

Darbe nagrinėjami du binariniai požymiai, nusakyti egzamino pažymiu ir parodontys, ar sėkmingai, ar nesėkmingai buvo atlikta atitinkama užduotis. Vienas jų yra pažangumas. Jis aprašomas aiškinamuoju kintamuoju *pažangus*, kuris lygus 1, jeigu studentas iš egzamino gavo 8 ar daugiau, ir lygus 0 priešingu atveju. Kitas požymis yra aprašomas aiškinamuoju kintamuoju *neišlaikė*, kuris lygus 1, jeigu studentas egzamino neišlaikė, ir lygus 0 likusiais atvejais. Naudojant sąlyginės logistinės regresijos modelį (2) buvo tirtas minėtų požymių statistinis ryšys su studijų metais (kintamasis *metai*), egzaminavusių dėstytojų (gal būt, ir egzaminuojamo dalyko) ypatybėmis bei bendromis jų savybėmis: dėstytojo mokslo vardas (kintamasis *m.vardas*), dėstytojo mokslo laipsnis (*m.laipsnis*), dėstytojo lytis (*d.lytis*), bei dėstytojų savybių sąveika su studentų savybėmis (pavyzdžiui, studento lytimi, kintamasis *s.lytis*).

Kadangi sąlyginės logistinės regresijos modelis reikalauja daug kompiuterinių resursų, buvo pasirinktas $n_1 = 20$. Kontrolinė dėstytojų grupė buvo sudaryta dviem

būdais. 1A grupė – tai 20 dėstytojų, kurie egzaminavo daugiausiai II specialybės studentų (5 moterys, 15 vyrų), o kontrolinę grupę (1B grupę) sudaro visi likę dėstytojai (69 moterys, 149 vyrai). 2A grupė – tai 20 dėstytojų (6 moterys, 14 vyrų), parašiusių informatikams daugiausia neigiamų įvertinimų, t.y. pažymių mažesnių už 5, o kontrolinė grupė (2B grupė) – visi likę dėstytojai (68 moterys, 150 vyrų). Dėstytojų iš 1A grupės (2A grupės) sąryšį su tiriamu požymiu y aprašo atitinkami kintamieji d_1 – d_{20} (kintamieji \bar{d}_1 – \bar{d}_{20}).

3.2 Rezultatų aptarimas

Studentų pažangumo tyrimas

Maždaug apie 41% egzaminų išlaikomi *gerai* (8–10). 1 lentelėje pateikti kintamajam *pažangus* sudaryto sąlyginės logistinės regresijos modelio rezultatai, atspindintys tiriamų veiksnių sąryšio su juo statistinį reikšmingumą.

Daugumos 1A grupės dėstytojų egzaminų sudėtingumas statistiškai reikšmingai skiriasi nuo kontrolinės grupės 1B dėstytojų (daugumos kintamųjų d_1 – d_{20} p -reikšmės yra mažesnės už 0,0001). Tiesiogiai 1A grupės dėstytojų tarpusavyje palyginti neišeina, tačiau 5 dėstytojų egzaminų sudėtingumas buvo statistiškai reikšmingai mažesnis už kontrolinės grupės, o 9 – statistiškai reikšmingai didesnis. Tai liudija apie ženklus egzaminų sudėtingumo *geriems studentams* skirtumus, o tuo pačiu ir apie didelę pasirinktos specialybės ir klausomų kursų įtaką studento egzaminų įvertinimams ir jo pažymių vidurkiui. Vadinasi, pažymių vidurkis apskritai nėra tinkamas rodiklis studentų pažangumui, jų gebėjimams ir žinių lygiui palyginti.

1B (taip pat ir 2B) grupės dėstytojai atskirai nebuvo analizuojami dėl sąlyginai mažo studentų skaičiaus, kuriuos jie egzaminavo, ir dėl to, kad reikėjo riboti parametrų skaičių, kad uždavinys būtų skaitiškai išsprendžiamas. Tačiau buvo tiriamas šios grupės dėstytojų savybių ryšys su egzamino sudėtingumu. Iš 1 lentelėje pateiktų rezultatų matosi, kad su egzaminų sudėtingumu yra statistiškai reikšmingai susijusi dėstytojo lytis (p -reikšmė 0,0164), dėstytojo mokslo vardas (p -reikšmė <0,0001), o taip pat kintamųjų sąveikos: $d.lytis*s.lytis$ (p -reikšmė <0,0001), $d.lytis*m.laipnsnis$ (p -reikšmė <0,0001), $m.vardas*m.laipnsnis$ (p -reikšmė <0,0001). Galima teigti, kad dėstytojų mokslo laipsnis ir kintamųjų sąveika $d.lytis*m.vardas$ su egzaminų sudėtingumu *geriems studentams* nėra susiję (atitinkamos p -reikšmės yra 0,7070 ir 0,5587).

1 lentelė. Požymio *pažangus* rezultatai.

Kintamieji ir jų sąveikos	Pasikliautiniai intervalai	Laisvės laipsniai	p -reikšmė	
$d.lytis$		1	0.0164	
$d.lytis*s.lytis$		1	<.0001	
$d_2, d_4, d_5, d_6, d_{10}, d_{13}, d_{14}, d_{19}, d_{20}$	0.138	0.869	9	<.0001
$d_9, d_{11}, d_{15}, d_{16}, d_{17}, d_{18}$	0.660	1.555	6	>.0512
$d_1, d_3, d_7, d_8, d_{12}$	1.348	2.357	5	<.0001
<i>metai</i>		1	<.0001	
<i>m.vardas</i>		2	<.0001	
$d.lytis*m.vardas$		2	0.5587	
<i>m.laipnsnis</i>		2	0.7070	
$d.lytis*m.laipnsnis$		2	<.0001	
$m.vardas*m.laipnsnis$		4	<.0001	

Verta atkreipti dėmesį į statistiškai reikšmingą santykinio egzaminų sudėtingumo *geriems studentams* augimo arba gerų studentų (jų gebėjimų) mažėjimo tendenciją laike (p -reikšmė $< 0,0001$). Šios tendencijos nėra kintamajam *neišlaikė*, kuris atspindi egzaminų sudėtingumą mažiau gabiems, *silpniesiems studentams*.

Studentų egzaminų išlaikymo tyrimas

Neišlaikoma maždaug apie 10% egzaminų. Skirtingai nuo kintamojo *pažangus* šiuo atveju visi 2A grupės dėstytojai, išskyrus dėstytoją $j = 12$, statistiškai reikšmingai skiriasi nuo grupės 2B dėstytojų pagal egzaminų sudėtingumą. Atitinkamų kintamųjų p -reikšmės yra mažesnės už pasirinktą reikšmingumo lygmenį 0.05. Verta paminėti, kad 19 dėstytojų egzaminų sudėtingumas yra statistiškai reikšmingai didesnis už kontrolinės grupės. Tai natūralu, nes į 2A grupę buvo atrinkti dėstytojai, kurie parašė daugiausiai neigiamų pažymių. Įdomu tai, kad ir atrinktos 2A grupės kai kurių dėstytojų egzaminų sudėtingumas ženkliai skiriasi tarpusavyje, kadangi atitinkamų šansų santykių pasikliautinieji intervalai tarpusavyje (beveik) nesikerta.

2B grupės dėstytojų savybių sąryšį su pagrindiniu kintamuoju *neišlaikė* aprašo kintamieji *d.lytis*, *m.laipsnis* ir *m.vardas*. Egzamino sudėtingumas *silpniesiems studentams* nėra susijęs nei su dėstytojo lytimi (p -reikšmė lygi 0.2694), nei su egzaminavimo metais (p -reikšmė lygi 0.0761), nei su požymio *d.lytis* sąveika su *s.lytis* ar su *m.laipsnis* (atitinkamos p -reikšmės yra 0.9953 ir 0.1470), nei su požymiu *m.laipsnis* ir *m.vardas* sąveika (atitinkama p -reikšmė yra 0.9982). Kiti požymiai ir sąveikos: *m.laipsnis*, *m.vardas*, *d.lytis* ir *m.vardas* sąveika, yra statistiškai reikšmingi (atitinkamos p -reikšmės yra < 0.0001 , 0.0088, ir 0.0324).

4 Išvados

1. Atliktas tyrimas parodė ženklus egzaminų sudėtingumo *geriems studentams* skirtumus, o tuo pačiu didelę pasirinktos specialybės ir klausomų kursų įtaką studento egzaminų įvertinimams, vadinasi, ir jo pažymių vidurkiui. Todėl pažymių vidurkis apskritai nėra tinkamas kriterijus studentų pažangumui ir jų gebėjimams palyginti. Į tai svarbu atsižvelgti organizuojant studijų procesą ir sudarant studentų skatinimo sistemą, nes dabartinė tvarka palanki tiems studentams, kurie renkasi paprastesnius kursus ir dėstytojus, linkusius rašyti geresnius pažymius.
2. Nustatyta, kad, su egzaminų sudėtingumu *geriems studentams* yra statistiškai reikšmingai susijusi dėstytojo lytis, dėstytojo mokslo vardas, o taip pat kintamųjų sąveikos: dėstytojo lytis ir studento lytis, dėstytojo lytis ir dėstytojo mokslo laipsnis, dėstytojo mokslo vardas ir dėstytojo mokslo laipsnis. Egzamino sudėtingumas *silpniesiems studentams* yra statistiškai reikšmingai susijęs su požymiais: dėstytojo mokslo laipsnis ir dėstytojo mokslo vardas, bei požymių dėstytojo lytis ir dėstytojo mokslo vardas sąveika. Čia išvardinti statistiniai sąryšiai, matyt, nėra universalūs, jų negalima apibendrinti kitoms specialybėms, juo labiau kitoms aukštosios mokykloms. Tačiau jie parodo, koks subtilus ir subjektyvus (vertinami skirtingi gebėjimų aspektai) yra abstrakčių gebėjimų vertinimo procesas. Vienu parametru jų neaprašysi.

Literatūra

- [1] F.B. Baker. *The Basics of Item Response Theory*. ERIC, Wisconsin, second edition, 2001.
- [2] I. Partchev. *A visual guide to item response theory*. Available from Internet: <http://www.pabst-publishers.de/mpr/>. 2009.
- [3] G. Rach. *Probabilistic Models for Some Intelligence and Attainment Tests*. MESA, Chicago, 1960.
- [4] M.E. Stokes, C.S. Davis and G.G. Koch. *Categorical Data Analysis Using the SAS System*. NC: SAS Institute Inc, Cary, 2000.
- [5] H. Verstralen, T. Bechger and G. Maris. *The combined use of classical test theory and item response theory*. Available from Internet: <http://www.cito.nl/pok/poc/eind-fr.htm>. 2009.

SUMMARY

Analysis of comparable academic performance: panel logistic regression model

M. Radavičius, P. Samusenko

Conditional logistic regression model is fitted to data of student academic performance. This enables one to compare difficulties of different examinations and to identify related factors.

Keywords: Academic performance, binary item response, conditional logistic, IRT, nuisance parameters, Rasch model.