LIETUVOS MATEMATIKOS RINKINYS. LMD DARBAI ISSN 0132-2818
51 tomas, 2010, 273–278 www.mii.lt/LMR/

Diagnostinio testo matematinio modelio taikymas vertinant studentų aerobinį pajėgumą

Natalja Kosareva[1], Aleksandras Krylovas[2], Povilas Tamošauskas[1], Stanislavas Dadelo[1]

[1] *Vilniaus Gedimino technikos universitetas*
Saulėtekio al. 11, LT-10223 Vilnius
[2] *Mykolo Romerio universitetas*
Ateities 20, LT-08303, Vilnius
E. paštas: natalja.kosareva@vgtu.lt; krylovas@mruni.eu; hinst@hi.vgtu.lt; sta.da@takas.lt

Santrauka. Šiame darbe autorių anksčiau pasiūlytas diagnostinio testo modelis pritaikytas konstruojant testą, skirtą įvertinti pirmo kurso studentų vaikinų aerobinį pajėgumą. Atliktas pilotinis tyrimas, kuris rodo, kaip iš dvireikšmių diagnostinių operatorių sudaryti informatyviausią testą, pritaikytą duotai tiriamųjų grupei. Atliekama kalibravimo procedūra – pagal esamus empirinius duomenis parenkami modelio parametrai, po to iš skirtingų diagnostinių operatorių išrenkamas tinkamiausias. Pagal siūlomą metodiką skaičiuojamas testo rezultato tikimybinis skirstinys ir testo teikiamos informacijos kiekis. Testo rezultatas – tiriamųjų norminis įvertinimas latentinio kintamojo (aerobinio pajėgumo VO_2max) atžvilgiu. Didžiausio tikėtinumo metodu gaunamas maksimalaus deguonies sunaudojimo (VO_2max) įvertis. Siūlomo mišraus modelio rezultatai yra tikslesni atliekant aerobinio pajėgumo norminį vertinimą.

Raktiniai žodžiai: užduoties sprendimo teorija, diagnostinis operatorius, entropijos funkcija, maksimalus deguonies sunaudojimas, matematinis modeliavimas.

Įvadas

Įvairiose srityse atsiranda poreikis sutvarkyti tiriamus objektus pagal tam tikrą tiesiogiai neišmatuojamą (latentinį) parametrą arba įvertinti šio parametro reikšmes. Vienas iš plačiai taikomų šių uždavinių sprendimo būdų yra užduoties sprendimo teorija, angl. Item Response Theory (IRT). Tikimybė teisingai atsakyti į testo klausimą nagrinėjama kaip latentinio parametro p funkcija. Ji modeliuojama logistinių funkcijų pagalba, parenkant i-jam klausimui tinkamiausius parametrus: b_i – klausimo sunkumas, a_i – klausimo skiriamoji geba. Sąlyginė tikimybė teisingai atsakyti į klausimą, priklausomai nuo p, vadinama *diagnostiniu operatoriumi*, turi tokią analizinę išraišką [10]:

$$k_i(p) = \mathbf{P}(x_i = 1|p) = \frac{1}{1 + e^{-a_i(p-b_i)}}.$$

Šiuo atveju klausimai yra dvireikšmiai (dichotominiai), t.y. į juos galima atsakyti tik teisingai ($x_i = 1$) arba klaidingai ($x_i = 0$) (arba tik „taip“ arba „ne“). Papildomi IRT reikalavimai – testuojamųjų atsakymai turi būti nepriklausomi ir klausimai taip pat turi būti nepriklausomi atsitiktiniai dydžiai. Tam, kad užtikrinti klausimo bei viso testo validumą [1], diagnostinių operatorių funkcijos $k_i(p)$ turi būti nemažėjančiosios p atžvilgiu.

Autorių darbuose [6, 8] buvo pasiūlyta praplėsti duotąjį modelį, leidžiant pasirinkti diagnostinius operatorius iš plačių parametrinių funkcijų klasių. Pavyzdžiui, galime nagrinėti tokias funkcijas:

$$k_1(p;a) = \frac{2}{\pi}\mathrm{arccot}\left(\frac{a\ln(p)}{\ln(1-p)}\right), \quad a > 0 \text{ – arkkotangento funkcija},$$

$$k_2(p;a;b) = \frac{1}{1+e^{-a(p-b)}}, \quad a > 0,\ b \in (0;1) \text{ – logistinė funkcija},$$

$$k_3(p;a;b) = \begin{cases} 0, & p < a, \\ \frac{p-a}{b-a}, & p \in [a;b], \\ 1, & p > b, \end{cases} \qquad a < b \text{ – atkarpomis tiesinė funkcija},$$

$$k_4(p;a;b) = \begin{cases} 0, & p < a, \\ 2\left(\frac{p-a}{b-a}\right)^2, & p \in [a;\frac{a+b}{2}], \\ 1-2\left(\frac{b-p}{b-a}\right)^2, & p \in [\frac{a+b}{2};b], \\ 1, & p > b, \end{cases}$$

$$a < b \text{ – dviejų parabolių segmentų funkcija.}$$

Konstruojami dvireikšmiai testo klausimai, kurie būtų informatyviausi duotai tiriamųjų grupei. Testo teikiamos informacijos kiekį matuojame testo informacijos (entropijos) funkcija $I(k_1,k_2,\dots,k_N;f) = -\sum_{i=0}^{N} p_i \ln p_i$, čia $k_1 k_{,2},\dots,k_N$ – klausimų diagnostiniai operatoriai, $f(p)$ – kintamojo p tikimybinio tankio funkcija populiacijoje, N – testo klausimų skaičius, p_i, $i = 0,1,\dots,N$ – tikimybė, kad populiacijos atstovas teisingai atsakys į i testo klausimų. Straipsnyje [7] pateiktos formulės tikimybėms p_i skaičiuoti, kurios buvo gautos generuojančiųjų funkcijų metodu. Šio darbo tikslas – parodyti, kaip matematinis modelis gali būti pritaikytas sprendžiant praktinį uždavinį.

1 Uždavinio formulavimas

Buvo tiriami 115 pirmo kurso VGTU studentų vaikinų, atrinktų atsitiktine tvarka. Vertinamas parametras – aerobinės ištvermės fizinio gebėjimo charakteristika – maksimalus deguonies sunaudojimas ($\mathrm{VO_2max}$). Šį parametrą tiesiogiai išmatuoti galima tik laboratorijose, naudojant specialią įrangą (dujų analizatorių), matuojant deguonies kiekio skirtumą įkvėptame ir iškvėptame ore [5]. Kadangi tiesioginis $\mathrm{VO_2max}$ matavimo būdas yra brangus ir sunkiai praktiškai įgyvendinamas, naudojami netiesioginiai matavimo būdai. Metodiką apskaičiuoti $\mathrm{VO_2max}$ pasiūlė K. Cooper [4] pagal 12 min. bėgimo rezultatą (atstumą), panašią metodiką pasiūlė ir J. Carlstedt [3] pagal 3000 m. bėgimo rezultatą (laiką). Yra ir kitų apytikslių $\mathrm{VO_2max}$ nustatymo būdų, pavyzdžiui, PWC_{170} testas, kuris remiasi laiptinės ergonometrijos rezultatais [5]. Šiame tyrime žmogaus aerobinę ištvermę p vertiname pagal $\mathrm{VO_2max}$ reikšmę.

Tyrimo tikslas – sukonstruoti informatyviausią testą, pritaikytą duotai tiriamųjų grupei, įvertinti tiriamuosius latentinio kintamojo ($\mathrm{VO_2max}$) atžvilgiu, atlikti maksimalaus deguonies sunaudojimo ($\mathrm{VO_2max}$) kriterinį įvertinimą.

2 Diagnostinių operatorių parinkimas

Iš visų turimų matavimų buvo pasirinkti dešimt, kurių koreliacijos koeficientai su VO_2max buvo reikšmingi:

1. 100 metrų bėgimas (s) (<13,8);
2. Prisitraukimai prie skersinio (kartai) (>8);
3. Forsuoto iškvėpimo tūris (forced vital capacity – FVC) (litrai) (>4,22);
4. Abiejų plaštakų jėgos ir kūno masės indeksas PJI (procentai) (>100);
5. Rufje testo 3-asis mėginys (pulso dažnis per 15 s) (<26);
6. Šuolis iš vietos tolyn (cm) (>234);
7. Sėstis – gultis (30 s/kartai) (>32);
8. PWC_{170} indekso reikšmė (vnt) (>11,86);
9. Pulso dažnis per 10 s, atlikus PWC_{170} testo stepergometrijos pirmąjį pratimą (pulso dažnis per 10 s) (<26);
10. Rufje testo 2-asis mėginys (pulso dažnis per 15 s) (<32).

Tyrimai atlikti taikant standartinę metodiką [5]. Parametras p, kurio reikšmės $p \in [0;1]$, parodo santykinę VO_2max reikšmę. Pirmiausia reikėjo dichotomizuoti diagnostinius operatorius, t.y. kiekvienam klausimui rasti tokią reikšmę arba slenkstį, kad būtų išsaugotas didžiausias informacijos kiekis. Šiuos slenksčius radome kiekvienam iš nepriklausomų kintamųjų ir atitinkamam priklausomam kintamajam VO_2max pritaikę klasifikavimo medžio radimo procedūrą (SPSS paketo procedūra Classification and Regression Trees). Šios reikšmės nurodytos skliausteliuose prie atitinkamų kintamųjų. Pavyzdžiui, pirmam kintamąjam, kurio koreliacijos koeficientas su VO_2max yra neigiamas ir slenkstis lygus 13,8 s, klausimas formuluojamas taip: Ar 100 metrų bėgimo rezultatas mažesnis už 13,8 sekundes? Klausimas gali būti įvertintas teigiamai arba neigiamai. Tikimybė, kad į klausimą bus atsakyta teigiamai ($k_1(p)$) yra nemažėjančioji funkcija kintamojo p (VO_2max) atžvilgiu. Kai koreliacijos koeficientas teigamas, pavyzdžiui, antrojo klausimo atveju, klausimą formuluojame taip: Ar prisitraukimų prie skersinio skaičius didesnis už 8?

Po to, kai testo klausimai yra suformuluoti, parenkame tinkamiausią šiai tiriamųjų grupei parametro p tikimybinio tankio funkciją. Geriausia empirinių duomenų aproksimacija buvo gauta normaliuoju skirstiniu $N(0,57;0,21)$. Šis skirstinys buvo tinkamas turimiems duomenims patikrinus suderinamumo hipotezes χ^2 kriterijumi su reikšmingumo lygmeniu $\alpha = 0,05$. Taip sukonstruotam testui apskaičiuota santykinė testo informacijos funkcijos reikšmė (procentais nuo didžiausios reikšmės) $I(k_1, k_2, \dots, k_N; f) = 0,96$. Parinkus kitas klausimų slenksčių reikšmes, buvo gautos mažesnės testo informacijos funkcijos reikšmės.

3 Diagnostinių operatorių parametrų kalibravimas

Kiekvienam iš 10 diagnostinių operatorių buvo apskaičiuotos *empirinės diagnostinės funkcijos.* Tiriamieji grupuojami į intervalus pagal VO_2max reikšmes. Intervalas [0;1] buvo padalintas į 31 vienodo ilgio intervalą su m_j tiriamųjų j-jame intervale. Bendras tiriamųjų skaičius $\sum_{j=1}^{31} m_j = 115$. Laikome, kad tiriamieji, kurie pateko į tą patį intervalą turi tą pačią kintamojo p reikšmę. Tarkime, kad r_j tiriamųjų iš j-jo

1 lentelė. Dešimties diagnostinių operatorių aproksimacijų $k_1(p)$–$k_4(p)$ parametrai ir jų atstumai nuo empirinių diagnostinių funkcijų $d(k_1)$–$d(k_4)$.

Kl.	$a(k_1)$	$d(k_1)$	$a(k_2)$	$b(k_2)$	$d(k_2)$	$a(k_3)$	$b(k_3)$	$d(k_3)$	$a(k_4)$	$b(k_4)$	$d(k_4)$
Kl1	1,30	0,203	0,92	0,54	**0,195**	0,12	0,98	0,196	−0,17	1,26	0,195
Kl2	1,44	0,110	0,86	0,55	0,094	0,14	0,97	**0,09**	−0,05	1,19	0,35
Kl3	0,35	0,129	0,50	0,12	0,100	−0,6	0,92	0,103	−1,09	1,37	**0,099**
Kl4	0,89	0,240	0,37	0,40	0,166	−0,67	1,44	**0,165**	−1,24	2,04	0,167
Kl5	1,01	0,188	0,54	0,46	**0,144**	−0,21	1,14	0,145	−0,65	1,59	0,145
Kl6	1,28	0,161	1,17	0,54	0,162	0,21	0,88	**0,156**	0,02	1,06	0,159
Kl7	6,20	0,178	0,58	0,87	**0,125**	0,23	1,53	0,127	−0,06	1,76	0,477
Kl8	0,48	0,151	0,49	0,23	**0,129**	−0,93	1,15	0,133	−1,39	1,72	0,132
Kl9	1,06	0,193	0,49	0,48	**0,126**	−0,36	1,30	0,131	−1,07	1,99	0,129
Kl10	1,42	0,223	0,40	0,54	**0,085**	−0,24	1,33	**0,085**	0,88	1,97	0,086

intervalo į klausimą atsakė teigiamai. Tuomet stebima teigiamų atsakymų proporcija šiame intervale, kai $p = p_j$, yra $P(p_j) = \frac{r_j}{m_j}$. Taip apibrėžtai empirinei diagnostinei funkcijai $P(p)$ parenkami tinkamiausi parametrai iš kiekvienos iš keturių funkcijų klasių $k_1(p)$–$k_4(p)$ pagal metodiką, aprašytą [2]. Apskaičiuoti atstumai nuo empirinių diagnostinių funkcijų iki geriausiai jas aproksimuojančių funkcijų iš kiekvienos iš 4 aukščiau aprašytų funkcijų klasių:

$$d(k_i) = \sqrt{\sum_{j=1}^{31} w(p_j)\bigl(k_i(p_j; a; b) - P(p_j)\bigr)^2}, \quad i = \overline{1,4}, \ w(p_j) = \frac{m_j}{115}.$$

Rezultatai pateikti 1 lentelėje.

Paryškintos reikšmės lentelėje – mažiausi atstumai tarp empirinės ir atitinkamos ją aproksimuojančios diagnostinio operatoriaus funkcijos. Matome, kad geriausia empirinės diagnostinės funkcijos aproksimacija šešiais atvejais buvo iš logistinių funkcijų klasės, keturiais atvejais – iš atkarpomis tiesinių funkcijų ir vienu atveju iš dviejų parabolių segmentų funkcijų. Paskutinio klausimo atveju geriausiai aproksimuojančias funkcijas parinkome iš dviejų klasių – $k_2(p)$ ir $k_3(p)$. Funkcija $k_1(p)$ nė karto nebuvo geriausia aproksimacija. Tai paaiškinama tuo, kad funkcija priklauso tik nuo vieno parametro ir parinkimo laisvė yra apribota. Turėdami geriausiai empirines diagnostines funkcijas aproksimuojančias funkcijas ir jų parametrus, iš šių funkcijų konstruojame mišrųjį modelį.

4 Testų rezultatų palyginimas ir numatomi tolimesni tyrimai

Testo rezultatas $S = \sum_{i=1}^{N} x_i$ – bendras surinktų testo balų skaičius, čia

$$x_i = \begin{cases} 1, & \text{jei į } i\text{-ąjį klausimą atsakyta teigiamai,} \\ 0, & \text{jei į } i\text{-ąjį klausimą atsakyta neigiamai.} \end{cases}$$

Šis rodiklis yra vienas iš parametro p norminio vertinimo rodiklių. Turėdami kiekvieno studento testo rezultatų vektorius $(x_{1j}, x_{2j}, \ldots, x_{Nj})$, $j = \overline{1, 115}$, apskaičiavome parametrų p_j įverčius $\hat{p}_{1j}$ ir $\hat{p}_{2j}$, maksimizuojančius atitinkamus didžiausio tikėtinumo

2 lentelė. Koreliacijos koeficientų reikšmės, apskaičiuotos VO_2max, bendrajam testo balų skaičiui S bei didžiausio tikėtinumo metodu gautais logistinio ir mišraus modelio įverčiais $\hat{p}_1$ ir $\hat{p}_2$.

Koreliacijos koeficientas		S	$\hat{p}_1$	$\hat{p}_2$
Kendall's $\mathrm{tau_b}$	VO_2max	0,429	0,439	0,446
Spearman's rho	VO_2max	0,558	0,596	0,603
Pearson correlation	VO_2max	0,573	0,625	0,614

funkcijų logaritmus

$$L(k_1, k_2, \ldots, k_N, p) = \sum_{i=1}^{N} \bigl(x_i \ln k_i(p) + (1 - x_i) \ln \bigl(1 - k_i(p)\bigr)\bigr)$$

logistinio ($\hat{p}_{1j}$) ir mišraus ($\hat{p}_{2j}$) modelio atvejais.

2 lentelėje pateiktos Spearman ir Kendall ranginės koreliacijos koeficientų reikšmės, Pearson koreliacijos koeficientai, apskaičiuoti priklausomam kintamajam VO_2max ir bendrajam testo balų skaičiui S bei didžiausio tikėtinumo metodu gautais logistinio ir mišraus modelio įverčiais $\hat{p}_1$ ir $\hat{p}_2$. Visi apskaičiuoti koreliacijos koeficientai yra reikšmingi esant reikšmingumo lygmeniui $0,01$.

Iš 2 lentelės matome, kad testuojamųjų norminiam vertinimui geriausiai tinka didžiausio tikėtinumo metodu gautas mišraus modelio įvertis $\hat{p}_2$, kuriam abiejų ranginės koreliacijos koeficientų reikšmės didžiausios. Tačiau Pearson koreliacijos koeficiento reikšmė didžiausia logistinio modelio atveju ($\hat{p}_1$). Norėdami paaiškinti šį rezultatą, turėtume atlikti išsamesnius tyrimus – apskaičiuoti testo informacijos funkcijos reikšmes parametro p kitimo intervalo $[0;1]$ vidiniuose taškuose. Žinoma [2], kad latentinio parametro įverčio $\hat{p}$ dispersija atvirkščiai proporcinga testo informacijos funkcijos reikšmei taške $\hat{p}$, o standartinis nuokrypis:

$$\sigma(\hat{p}) = \frac{1}{\sqrt{I(\hat{p})}} = 1 \Big/ \sqrt{\sum_{i=1}^{N} a_i^2 k_i(\hat{p})\bigl(1 - k_i(\hat{p})\bigr)},$$

todėl turėdami šių funkcijų reikšmes logistiniam ir mišriajam modeliams, galėtume palyginti įverčių $\hat{p}_1$ ir $\hat{p}_2$ standartines paklaidas įvairiom parametro p reikšmėm. Gali būti, kad vienu modeliu gauti įverčiai yra tikslesni intervalo $[0;1]$ vidiniuose taškuose, o kitu modeliu – marginaliais atvejais, t.y. intervalo galuose. Tai yra tolimesnių tyrimų objektas.

Literatūra

[1] A. Anastasi and S. Urbina. *Psychological Testing.* Prentice Hall, 7th edition, 1997.

[2] F. Baker. *The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation.* University of Maryland, College Park, MD, 2001.

[3] J. Carlstedt. *Tester för idrottare.* SISU, Idrottsböcker, Malmö, 1995.

[4] K. Cooper. A means of assessing maximal oxygen intake. *JAMA*, **203**:201–204, 1968.

[5] R. Dadelienė. *Kineziologija.* Lietuvos sporto informacijos centras, 2008.

[6] A. Krylovas, N. Kosareva. Žinių tikrinimo matematinis modelis. *Liet. mat. rink. LMD darbai*, **48/49**:217–221, 2008.

[7] N. Kosareva, A. Krylovas. Diagnostinio testo matematinio modelio tyrimas. *Liet. mat. rink. LMD darbai*, **50**:202–207, 2009.

[8] A. Krylovas and N. Kosareva. Mathematical modelling of forecasting the results of knowlegdge testing. technological and economic development of economy. *Baltic Journal on Sustainability*, **14**(3):388–401, 2008.

[9] J. Pfanzagl in cooperation with V. Baumann and H. Huber. *Theory of Measurement*. Physica-Verlag, Wurzburg–Wien, 1971.

[10] G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen, 1960. Expanded edition: The University of Chicago Press, Chicago, 1980.

SUMMARY

The applicacion of mathematical model for diagnostic test while estimating student's aerobic capacity

N. Kosareva, A. Krylovas, P. Tamošauskas, S. Dadelo

In the paper diagnostic test creation model proposed earlier by authors is applied for creation of the test designed to esimate aerobic capacity of VGTU students. The pilot research demonstrated how the most informative test for the observed group of testees could be constructed from dichotomous diagnostics operators. The calibration procedure of fitting model parameters according to the empirical data was accomplished. The best fitting diagnostic operator was chosen from four various function classes. Probability distribution of test result and test information value were calculated according to the proposed technique. The result of the test – normed-referenced estimation of testees depending on the latent parameter (aerobic capacity VO_2max). Maximum oxigen consumption was evaluated using maximum likelihood method. The proposed mixed model results were more accurate for normed-referenced estimation of aerobic capacity.

Keywords: item response theory, diagnostics operator, measurement, probability distribution, entropy function, maximum oxigen consumption, mathematical modelling.