ITEM RESPONSE THEORY APPLICATIONS FOR SOCIAL PHENOMENA MODELING

Aleksandras Krylovas

Mykolas Romeris University, Faculty of Social Informatics,
Department of Mathematical Modelling
Ateities 20, LT-08303, Vilnius, Lithuania
Telephone (+370 5) 2714 731
E-mail krylovas@mruni.eu

Natalja Kosareva

Vilnius Gediminas Technical University, Faculty of Fundamental Sciences,
Department of Mathematical Modelling
Saulėtekio 11, LT-10223 Vilnius, Lithuania
Telephone (+370 5) 274 4827
E-mail natalja.kosareva@vgtu.lt

Abstract. *The article reviews Item Response Theory facilities in modeling social phenomena, its advantages and disadvantages comparing with Classical Test Theory and mathematical statistics methods. It discusses the application of IRT in social sciences specifics. The extension of Item Response Theory model where item characteristic functions have been approximated not only with logistic functions, but also with functions from other wide classes of parametric functions is presented. The procedure of item calibration—choosing of the best fitting item characteristic function and parameter's estimation—is described in the paper. When choosing the item characteristic function for the model from wider class of functions as the result we obtain better agreement between observed data and the formal model. The example of IRT application in measuring the Environmental Performance Index is examined. The technology of extending the class of logistic functions used for simulation was*

Socialinių mokslų studijos/Societal Studies
©Mykolo Romerio universitetas, 2011
©Mykolas Romeris University, 2011

ISSN 2029–2236 (print), ISSN 2029–2244 (online)
http://www.mruni.eu/lt/mokslo_darbai/SMS/
http://www.mruni.eu/en/mokslo_darbai/SMS/

applied in this case study. Calculated test information function's values for the proposed mixed model were higher comparing with corresponding values for 2 parameters logistic model. The clusterization procedure performed by EPI values calculated as the total score of 9 items test was accomplished. Kendall's τ and Spearman's ρ rank correlation coefficients for measuring the relation between estimations of EPI accomplished by the total test score value and the EPI value calculated as the weighted sum of 25 indicators are presented for different values of test items n. It was shown that the new methodology allows to perform rather precise norms-referenced evaluation of objects with respect to latent feature which couldn't be measured directly with a significantly less amount information. The proposed methodology could be used in various areas of human activities where the actual problem is norms-referenced evaluation.

Keywords: *Item Response Theory, mathematical modeling, test information function, Environmental Performance Index, latent feature, norms-referenced evaluation.*

Introduction

Methodology for social research is something different from methodology of research in physical or technological sciences. The researcher meets with difficulties concerned with impact of human factor when gathering empirical data.

The first problem arises when we define the object of investigation. Very often it is latent, i.e. not directly observed feature, for example, anxiety, nervousness, intelligence or other mental property. The question is—how one can measure this feature? One other question—is it possible to measure something that we cannot see, hear or feel? It is clear that we must use indirect methods of measuring, i.e. instruments that measure other related, but directly observable features.

The second problem arises when we are choosing a proper measuring instrument—a questionnaire. The construction of a good questionnaire is a rather responsible job. It can't be too short, because in this case the conclusions will be not reliable and also it can't be too long, because people cannot concentrate their attention to answer the excess number of questions (test items). Test items cannot be too easy or too difficult for the investigated group of testees, because responses to such items (questions) are predictable and not informative. It is known that the maximum information is received from such item when approximately one half of testees will respond it correctly. For the construction of a proper questionnaire we must be able to choose the most informative subset of items from the whole item bank. As different groups of testees have different levels of the investigated feature the question arises—if we can apply different questionnaires to different groups of persons to measure this feature? Fortunately with Item Response Theory coming we can answer positively to this question.

The third problem is test validity—the test should be constructed particularly for measuring mental feature under investigation.

1. Modeling in Social Sciences

Statistical modeling for social sciences solves the tasks of creating the equations which link factors to responses or independent variables to dependent variables[1]. The aim of statistical modeling is to predict values or levels of dependent variable and evaluate error of estimation. The well known statistical models which are used in social sciences are the log-linear model for discrete dependent variables[2] and the general linear model[3] and Item Response Theory models[4] for continuous dependent variables. Our field of interest is related to continuous latent dependent variables, so we'll take a look at the two latter—the general linear model and Item Response Theory models. General linear models could be applied when the normality assumption and some other assumptions about the data (i.e. homoscedasticity[5]) are satisfied. Classical statistical procedures for latent variable analysis are—factor analysis, linear discriminant analysis, covariance structure models, etc. But in real life normality assumptions are rarely satisfied, so the more universal tools have been created. When continuous latent variable isn't normally distributed, Item Response Theory could be applied for the modeling processes.

A very important stage is the creation of the proper questionnaire that allows getting maximum information about a measured feature. For the construction of a "good" questionnaire the appropriate technique is choosing items with the biggest absolute value of correlation with the measured feature. Items could be dichotomous with yes/no or agree/disagree responses or polytomous which have limited number of ordered categories. When using a dichotomous model we lose much more information comparing with polytomous case and this results in bigger measurement error. But the advantages of a dichotomous model are that such items very often fit real data and that we avoid the problem of insufficient number of responses in some categories of polytomous items.[6] So, further we will deal with the dichotomous model.

2. Item Response Theory in Social Sciences

In the dichotomous Item Response Theory model, *k* subjects are responding to *n* test items. Then test results appear as *k* x *n* matrix, where rows represent subjects and co-

1   Clogg, C. The Impact of Sociological Methodology on Statistical Methodology. *Statistical Science*. 1992, 7(2): 183−207.

2   Christensen, R. *Log-Linear Models and Logistic Regression*. NY: Springer-Verlag Inc, 1997.

3   Friston, K.; Holmes, A.; Worsley K.; Poline, J.-P., Frith, C.; Frackowiak, R. Statistical Parametric Maps in functional imaging: A general linear approach. *Human Brain Mapping*. 1995, 2: 189–210.

4   Baker, F.; Kim, S.-H. *Item response theory: Parameter estimation techniques* (2nd ed.). NY: Marcel Dekker, 2004.

5   Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*. California: Sage Publications, 1997.

6   Molenaar, I. Parametric and nonparametric item response theory models in health related quality of life measurement. *Statistical Methods for Quality of Life Studies: Design, Measurements and Analysis*. 2007, p. 143−154.

lumns represent responses to the items—a matrix with 0's or 1's in the cells (see Table 1).

Table 1. Dichotomous Item Response Theory Model

	Item 1	Item 2	Item 3	Item 4	Item 5	Average person's ability estimation
Person 1	1	1	1	1	1	1
Person 2	0	1	1	1	1	0.8
Person 3	0	0	1	1	1	0.6
Person 4	0	0	0	1	1	0.4
Person 5	0	0	0	0	1	0.2
Average item's difficulty	0.8	0.6	0.4	0.2	0	

The well known model in latent trait testing theory is the One Parameter Logistic model (1PL) initially described by Rasch[7]. According to this model we can predict the probability of responding a particular item correctly $k_i(p) = P(x_i = 1|p)$ given the examinee ability level p by the following equation (1):

$$k_i(p) = \frac{1}{1 + e^{-(p-b_i)}}, \tag{1}$$

here b_i is the difficulty parameter of item i. The function above, which is called *item characteristic function*, belongs to the class of logistic functions. Rasch model is taking place if some constraints on the model are satisfied[8]:

1. Unidimensionality of latent trait p. This means that at a time one mental property is measured, the influence of other latent traits is treated as negligible.

2. Conditional independence of items given person and conditional independence of persons given item. Given the person's ability p the elements of the response vector are independent. Person's response to the item is independent of other respondent's responses to this item.

3. The item response function $k_i(p) = P(x_i = 1|p)$ is non-decreasing function of p.

4. The total test score $\sum_i x_i$ is a sufficient statistic for p.

In Fig. 1 graphs of 1PL item characteristic functions $k_i(p)$ with various difficulty parameter b_i values are presented.

7 　Rasch, G. *Probabilistic Models for some intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960; Expanded edition: Chicago, The University of Chicago Press, 1980.
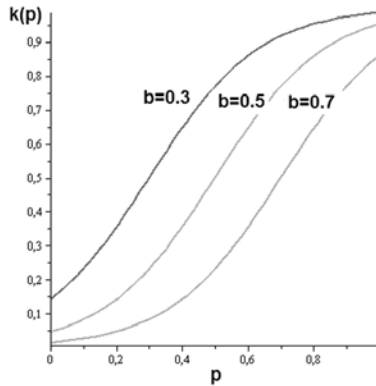
8 　Molenaar, I., *supra* note 6.

Figure 1. Graphs of functions $k(p)$ with various difficulty parameter values in 1PL.

Formula (2) presented below is generalization of Rasch model[9] where supplementary parameters are discrimination parameter a_i (2PL model) and additionally the probability of random guessing c_i of the item i (3PL model):

$$P\left(x_i = 1 \mid p\right) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta - b_i)}};$$

(2)

In 1PL, 2PL and 3PL models probabilities of correct responses to the items are logistic family functions. 2PL and 3PL logistic models satisfy only 1-3 conditions. In Figure 2 graphs of 2PL item characteristic functions $k_i(p)$ with various difficulty b_i and discrimination a_i parameters values are presented.
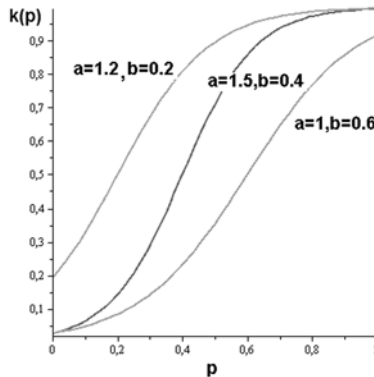

Figure 2. Graphs of functions $k(p)$ with various difficulty and discrimination parameters values in 2PL.

9   Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In Lord, F.; Novick, M. (eds.). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley, 1968, p. 397–479.

Before the testing process begins it is necessary to estimate the values of parameters a_i and b_i for each test item. The process of parameters estimation is called item *calibration*. Parameters are calculated by minimizing the distances between empirical diagnostic functions and item characteristic functions $k_i(p) = k_i(a_i;b_i;p)$. When item parameters are estimated the examinees testing stage begins. Each examinee responds all test items. The probability of each possible response vector under the model given the ability level p is a function of latent variable p and item parameters called the *likelihood function*. This function is used to estimate examinees ability levels. Likelihood function is a function of actually observed responses, latent variables and item parameters. The values of latent variable p which maximize likelihood function are the ability estimates. It is important that it is possible to predict responses of a given respondent to a given set of items. This feature could be used for imputation of missing observations or distinguishing those respondents whose responses were unexpected for various reasons—cheating, misunderstanding of questions, etc.

IRT has an attractive property—it enables the proposition of different items to different respondents and obtaining estimates of ability which are independent from the set of items given to respondent. So, we can apply different questionnaires to different groups of persons and obtain the same result. This possibility is particularly important in social processes modeling because very often questionnaire which is perfectly fitted for one group of persons, countries or areas of human activity is meaningless for other group. This feature is the advantage of IRT comparing with Classical Test Theory[10] (CTT). The other advantage is that in IRT the latent variable is not necessarily a normally distributed random variable.

3. The Extension of the Item Response Theory Model

The first stage is to postulate a formal measurement model, estimate item characteristic function's parameters, and establish whether the observed data are in agreement with the model. The measurement quality heavily depends on the applied model and how precisely we have accomplished the parameters calibration procedure. For increasing the accuracy of calibration we have proposed to supplement the class of logistic functions, described by formulas (1) and (2) with other parametric functions classes.[11] In the proposed model the best fitting item characteristic function $k(p)$ is selected from one of the 4 classes of parametric functions depending on one or two parameters. These functions are: 2 parameters logistic function when restricted in the interval [0; 1] described by (3); arc-cotangent function (4); segments of linear functions (5); segments of two parabolas (6):

10   Lord, F.; Novick, M., *supra* note 9.

11   Krylovas, A.; Kosareva, N. Mathematical Modelling of Diagnostic Tests. *Knowledge-Based Technologies and OR Methodologies for Strategic Decisions of Sustainable Development (KORSD-2009): 5th international conference*, 2009, p. 120–125.

$$k_1(p;a;b) = \frac{1}{1 + e^{-a(p-b)}}, a \geq 0, b \in [0;1] \tag{3}$$

$$k_2(p;a) = \frac{2}{\pi} arc\cot\left(\frac{a\ln(p)}{\ln(1-p)}\right), a \geq 0 \tag{4}$$

$$k_3(p;a;b) = \begin{cases} 0, p < a \\ \dfrac{p-a}{b-a}, a \leq p \leq b \\ 1, p > b \end{cases} \tag{5}$$

$$k_4(p;a;b) = \begin{cases} 0, p \leq a \\ 2\left(\dfrac{p-a}{b-a}\right)^2, a < p \leq \dfrac{a+b}{2} \\ 1 - 2\left(\dfrac{b-p}{b-a}\right)^2, \dfrac{a+b}{2} < p \leq b \\ 1, p > b \end{cases} \tag{6}$$

The interpretation of parameters for models described by equations (5) and (6) seems to be the difficulty for $\frac{a+b}{2}$ and the discrimination for $\frac{1}{b-a}$ similarly to the parameters a and b for the Rasch model. The interpretation of parameter a in the model (4) isn't obvious.

Clearly, when choosing the item characteristic function for our model from wider class of functions, as the result we will obtain better agreement between observed data and our formal model. This will enable to improve the estimate of latent ability p. It is notable, that only 3 constraints on the model 1-3 are satisfied, while the fourth—the sufficiency of total test score $\sum_i x_i$ —has not. As a result, the location of items and respondents is determined only at the ordinal level.

Let the test or the questionnaire consist of n items with corresponding item characteristic functions $k_1, k_2,...,k_n$. Total test result S would be the number of correctly responded items. Random variable S is gaining values from 0 to n. The probability distribution of total test result S in the whole population $p_i = P(S = i), i = 0,1,...,n$ is received by method of generating functions.[12] The test information (entropy) function I is described as follows:

---

12 Krylovas, A.; Kosareva, N. Mathematical modelling of forecasting the results of knowledge testing. *Technological and Economic Development of Economy*. 2008, 14(3): 388–401.

$$I(k_1, k_2, ..., k_n) = -\sum_{i=0}^{n} p_i \ln p_i \qquad (7)$$

The normalized value of the function I is the percentage of the test information function (7) from the maximum value, which is reached when all probabilities are equal to $\frac{1}{n+1}$. Our purpose is to choose test items from the whole set of items (items bank) that maximize the value of the test information function (7).

4. Case Study: Evaluation of Environmental Performance Index

According to the data gathered in 2010 from 163 world countries Yale University Center for Environmental Law and Policy performed a research for evaluation of the Environmental Performance Index.[13] The 2010 Environmental Performance Index (EPI) ranks 163 countries on 25 performance indicators tracked across ten well-established policy categories covering both environmental public health and ecosystem vitality. These indicators provide a gauge at a national government scale of how close countries are to established environmental policy goals.

EPI is a latent, i.e. not directly observed feature depending on a number of various factors, which measures the effectiveness of national environmental protection efforts of the countries. The 2010 EPI is calculated as weighted sum of 25 indicators presented in Figure 3. As a result countries were assigned scores from 32.1 (Sierra Leone) to 93.5 (Iceland).

The goal of our research was to choose the most informative subset from those 25 performance indicators and to fit each item characteristic function with the most appropriate model 3-6 with accomplishing parameters calibration procedure. We have chosen 9 indicators which have the biggest absolute values of correlation coefficient with EPI:

1. Environmental burden of disease, i.e. disability life adjusted years per 1000 population (<40);
2. Percentage of population having access to sanitation (>80);
3. Indoor air pollution, i.e. percentage of population using solid fuels (<30.1);
4. Outdoor air pollution (urban particulates) (<36.25);
5. Water quality index (WQI) (>72.73);
6. Forest cover change in percents (>-0.05)
7. Agricultural subsidies (>0);
8. Pesticide regulation (>17);
9. CO_2 emissions per electricity generation (g CO_2 per kWh) (<394).

---

13   Environmental Performance Index 2010 [interactive]. [accessed 08-09-2010]. <http://epi.yale.edu>.
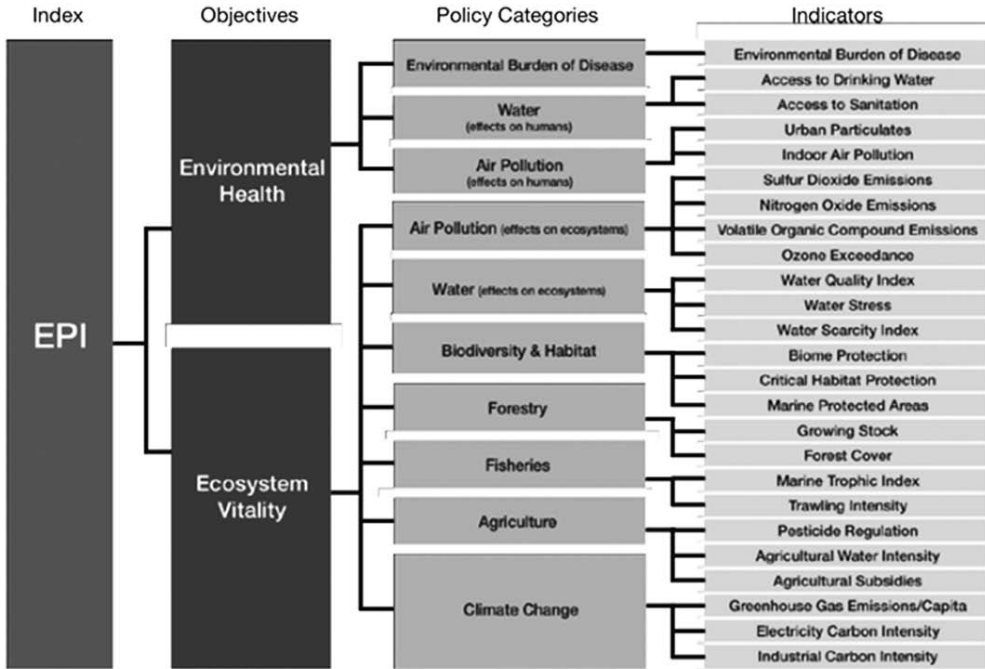
Figure 3. Construction of the Environmental Performance Index

Source: <http://epi.yale.edu>

Values in brackets are the thresholds for item dichotomization. These values were selected with the purpose, that approximately one half of countries will positively respond the given item. For example, the first item could be formulated as follows: "Is the environmental burden of disease less than 40?" The second item: "Is the percentage of population having access to sanitation greater than 80?" There could be only two responses to these items—"yes" or "no".

Now we'll calculate empirical diagnostic functions by application the technique described by Baker.[14] Standardized values p_j of EPI are calculated for each country:

$$p_j = \frac{p_j^{fact} - p^{min}}{p^{max} - p^{min}}, \tag{8}$$

here p_j^{fact} is the observed EPI value, p^{min} and p^{max} are minimum and maximum values of EPI—respectively 32.1 for Sierra Leone and 93.5 for Iceland.

---

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

After transformation (8) values of EPI are in the interval [0; 1]. Then countries are grouped to the intervals according to the EPI values. The interval [0; 1] was divided to 24 equal length intervals with m_j countries in j-th interval. The total number of countries is $\sum_{j=1}^{24} m_j = 163$. Let r_j countries from j-th interval responded the item positively. Then the observed proportion of positive responses in this interval where $p=p$ is $P(p_j) = \frac{r_j}{m_j}$. Our purpose is to select the function from 4 function classes, which will provide the best accuracy of the approximation to the observed proportions of correct responses to the item.

The distances between empirical diagnostic functions $P(p_j)$ and each function from 4 functions classes (3)-(6) are minimized by fitting functions parameters. The distances are calculated by the formula (9):

$$d(k_i) = \sum_{j=1}^{24} w(p_j) \cdot \left(k_i(p_j; a; b) - P(p_j) \right)^2, i = 1,2,3,4, w(p_j) = \frac{m_j}{163}. \tag{9}$$

The results of fitting 9 empirical item characteristic functions are presented in Table 2.

Table 2. The minimum values of distances $d(k_1) - d(k_4)$ between diagnostic operators approximations $k_1(p) - k_4(p)$ and 9 item's characteristic functions and the ratio of minimum distance value d_{min} and the distance to the logistic approximation $d(k_1)$.

Item	$d(k_1)$	$d(k_2)$	$d(k_3)$	$d(k_4)$	$\frac{d_{min}}{d(k_1)}$
Item1	0,00939	0,00928	**0,00915**	0,00957	0,97444
Item2	0,01084	**0,01031**	0,01243	0,01099	0,95111
Item3	**0,00743**	0,01016	0,01052	0,00797	1,00
Item4	**0,01185**	0,01278	0,01302	0,012213	1,00
Item5	0,01971	0,02352	0,02157	**0,01970**	0,99949
Item6	**0,01287**	0,01429	0,01367	0,01292	1,00
Item7	**0,02055**	0,02221	0,02167	0,02075	1,00
Item8	0,00807	0,00809	0,00865	**0,00800**	0,99133
Item9	0,01813	0,02591	**0,01795**	0,01805	0,99007

The minimum values of distances are highlighted in bold.

Having data about 163 countries EPI values we selected the best fitting probability distribution density function. The best fitting was normal distribution when normalized in the interval [0;1] with average 0.43 and standard deviation 0.22. The best fitting approximations to the empirical data from two probability distribution functions classes—normal distribution and Beta distribution—are presented in Fig. 4.
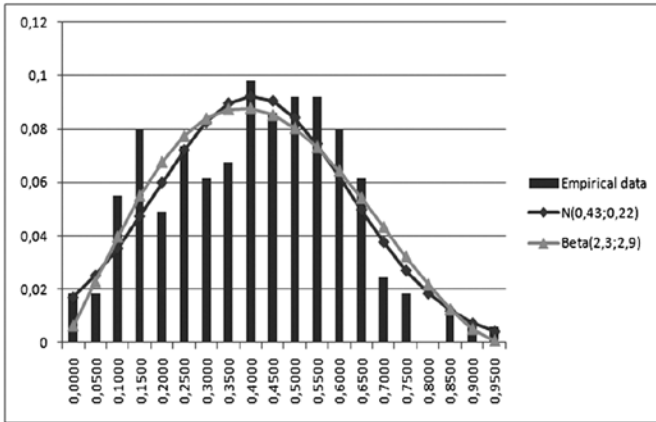
Figure 4. Approximation of the EPI distribution

We have got the results of approximation of 9 empirical diagnostic functions with functions (3)- (6) (Table 2) and their parameter's estimations. The next step is to design mixed model from these diagnostic functions. Let n to be the number of test items. By method of generating functions[15] we are calculating the probability distribution $\hat{p}_{jm}, j = 0,1,...,n$ of total test score S in the whole population $\hat{p}_{jm} = P(S = j), j = 0,1,...,n$ for the mixed model. Similarly we are calculating the same probabilities under the logistic model when all approximations of 9 empirical diagnostic functions are selected from one class of logistic functions (3):

$$\hat{p}_{jl} = P(S = j), j = 0,1,...,n \cdot$$

The test information function I is described as follows:

$$I(k_1, k_2,...,k_n) = -\sum_{i=0}^{n} p_i \ln p_i \ . \tag{10}$$

In formula (10) we substitute p_i with the estimations of probabilities under the mixed model $(\hat{p}_m)$ and under the logistic model $(\hat{p}_l)$ respectively. In Table 3 the values of the test information function I, normalized values of the function I as the percentage of the test information function (10) from the maximum value, which is reached when all probabilities are equal to $P(S = i) = \dfrac{1}{n+1}, i = 0,1,2,...,n$ are calculated for the test consisting of 5, 7 and 9 items respectively.

---

15 Bagdonavičius, V.; Kruopis, J. *Matematinė statistika.* 1 dalis [Mathematical statistics. First Part]. Vilnius, 2007.

The values in the brackets describe the proportion of the test information function from its maximum value, given in the last column. We can see from the Table 3 that mixed model gives higher information function's values for all test length *n* cases. Naturally test information function value increases when test length *n* is increasing.

Table 3. Test information function's values calculated for different values of test items *n* for the logistic $(\hat{p}_l)$ and mixed $(\hat{p}_m)$ models

n	$I(\hat{p}_l)$	$I(\hat{p}_m)$	I_{max}
5	1,776 (0,991)	1,778 (0,993)	1,792 (1,0)
7	2,049 (0,985)	2,053 (0,987)	2,079 (1,0)
9	2,256 (0,980)	2,2614 (0,985)	2,303 (1,0)

The next task is to compare the norms-referenced evaluation of Environmental Performance Index accomplished a) by the total test score value *S* for the test consisting of 5, 7 and 9 items and b) by EPI values calculated as the weighted sum of 25 indicators[16] (represented in Fig. 3). The latter value will be taken as the basic value in our comparisons because we expect that it is the most precise value of EPI. The accuracy of the norms-referenced evaluation of EPI by the total test score value *S* could be measured with Kendall's τ and Spearman's ρ rank correlation coefficients and are presented in the Table 4.

Table 4. Kendall's τ and Spearman's ρ rank correlation coefficients for measuring the relation between estimations of EPI accomplished by the total test score value *S* for the test consisting of 5, 7 and 9 items and the basic EPI value calculated as the weighted sum of 25 indicators

n	Kendall's tau	Spearman's rho
5	0,564	0,726
7	0,584	0,752
9	0,608	0,782

All the values of rank correlations are rather high and the correlation coefficients are significant at α = 0,05 level. It's clear that the proposed methodology enables to get sufficiently accurate norms-referenced evaluation of Environmental Performance Index having significantly less information—5, 7, or 9 items from the whole bank of 25 indicators.

In the next step the clusterization procedure was accomplished by the total test score value *S* for the test consisting of 9 items. All 163 countries were divided into 6 clusters. The clusterization method is Euclidean distance for between-groups linkage measuring. The clusterization results are presented in Table 5.

---

16   Environmental Performance Index 2010, *supra* note 13.

Table 5. The resulst of clusterization performed by the total score *S* of 9 items test.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Sierra Leone	Angola	Turkmenistan	United Arab Emirates	Kuwait	Belgium
Central African Republic	Togo	Mali	Bahrain	Cyprus	Greece
Mauritania	Chad	Benin	Oman	South Korea	USA
Niger	Cambodia	Iraq	Qatar	Ukraine	Slovenia
Haiti	Equatorial Guinea	Uzbekistan	Libyan Arab Jamahiriya	Turkey	Netherlands
Nigeria	Senegal	Burundi	South Africa	Argentina	Canada
Botswana	Mongolia	Rwanda	Tajikistan	Russia	Ireland
North Korea	Ethiopia	Indonesia	Ghana	Bulgaria	Luxembourg
Guinea	Bangladesh	Cameroon	Trinidad and Tobago	Poland	Lithuania
Guinea-Bissau	Papua New Guinea	Tanzania	Saudi Arabia	Brazil	Croatia
Pakistan	Bolivia	Yemen	Bosnia and Herzegovina	Estonia	Hungary
Honduras	Zambia	India	Jordan	Malaysia	Denmark
Solomon Islands	Sudan	China	Kazakhstan	Australia	Spain
Eritrea	Burkina Faso	Uganda	Lebanon	Romania	Czech Republic
	Zimbabwe	Gambia	Moldova	Ecuador	Japan
	Madagascar	Mozambique	Viet Nam	Serbia and Montenegro	Portugal
	Dem. Rep. Congo	Myanmar	Uruguay	Singapore	Italy
	Nicaragua	Kenya	Guyana	Latvia	Germany
	Sao Tome and Principe	Malawi	Kyrgyzstan	Chile	New Zealand
	Namibia	Congo	Iran	Malta	UK
		Guatemala	Armenia	Cuba	Slovakia
		Côte d'Ivoire	Tunisia	Costa Rica	Finland
		Swaziland	Macedonia		Austria
		Gabon	Brunei Darussalam		France
		Jamaica	Thailand		Norway
		Azerbaijan	Israel		Sweden
		Laos	Venezuela		Switzerland
		Djibouti	Syria		Iceland
		Egypt	Belarus		
		Paraguay	Morocco		
		Georgia	Philippines		
		Sri Lanka	Fiji		

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
		Maldives	Mexico		
		Algeria	Bhutan		
		Nepal	Suriname		
		Peru	Dominican Republic		
		Belize	El Salvador		
			Antigua and Barbuda		
			Panama		
			Albania		
			Colombia		
			Mauritius		

According to the clusterization performed by the total score *S* of 9 items test in the bottom of EPI rating in Cluster 1 we see some African countries which are among the poorest in the world and other countries with very low economical level.

Countries in Cluster 2 are mostly poor African and Asian countries with low economical level.

In Cluster 3 there are former southern Soviet Union republics and some developing countries from Europe and Asia. This cluster involves some countries with large population and rapidly growing industrial base.

Cluster 4 is the biggest one and represents more diverse set of countries. There are many countries of Central and South America, Asia and North Africa.

Cluster 5 contains countries with high developed economics and some countries of Central Europe —the new members of Europe Union which are middle-income countries with low level of industrial pollution.

Cluster 6 contains Scandinavian countries and other developed countries with sufficient financial resources.

Conclusions

The main implications of the research on the Environmental Performance Index evaluation are:
- The values of test information function are higher when probabilities in (10) formula are calculated by the proposed mixed model. The essentiality of this model is the extension of item characteristic functions approximation with 3 additional parametric functions classes (4)- (6). This methodology let us to obtain higher test information function's values for the mixed model comparing with the logistic model for 5, 7 and 9 items tests.

- We have got sufficiently high Kendall's τ and Spearman's ρ rank correlation coeficients between the EPI values obtained from the total test score *S* for the test consisting of 5, 7 and 9 items and EPI values calculated as the weighted sum of 25 indicators. By increasing the number of test items rank correlation coeficients increase.
- With the proposed methodology it is possible to get accurate enough norms-referenced evaluation of latent variable (EPI) having significantly less information—5, 7, or 9 items from the whole bank of 25 indicators.
- Criterion-referenced evaluation of EPI could be made by applying maximum likelihood method.[17]
- The proposed mathematical model could be applied for solving diagnostic tasks in various fields of human activities—medicine, sports, geology, technical diagnostics and others.

References

Bagdonavičius, V.; Kruopis, J. *Matematinė statistika.* 1 dalis [Mathematical Statistics. First Part]. Vilnius, 2007.

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

Baker, F.; Kim, S.-H. *Item response theory: Parameter estimation techniques* (2nd ed.). NY: Marcel Dekker, 2004.

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In Lord, F.; Novick, M. (eds.). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley, 1968.

Christensen, R. *Log-Linear Models and Logistic Regression*. NY: Springer-Verlag Inc, 1997.

Clogg, C. The Impact of Sociological Methodology on Statistical Methodology. *Statistical Science*. 1992, 7(2): 183−207.

Environmental Performance Index. 2010 [interactive]. [accessed 08-09-2010]. <http://epi.yale.edu>.

Friston, K.; Holmes, A.; Worsley K.; Poline, J.-P., Frith, C.; Frackowiak, R. Statistical Parametric Maps in functional imaging: A general linear approach. *Human Brain Mapping*. 1995, 2: 189–210.

Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*. California: Sage Publications, 1997.

Harris, J.; Stocker, H. *Handbook of Mathematics and Computational Science*. New York: Springer-Verlag, 1998.

Lord, F.; Novick, M. (eds.). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley Publishing company, 1968.

Molenaar, I. Parametric and nonparametric item response theory models in health related quality of life measurement. *Statistical Methods for Quality of Life Studies: Design, Measurements and Analysis*. 2007.

Krylovas, A.; Kosareva, N. Mathematical modelling of forecasting the results of knowledge testing. *Technological and Economic Development of Economy*. 2008, 14(3): 388–401.

Krylovas, A.; Kosareva, N. Mathematical Modelling of Diagnostic Tests. *Knowledge-Based Technologies and OR Methodologies for Strategic Decisions of Sustainable Deve-*

---

17 Harris, J.; Stocker, H. *Handbook of Mathematics and Computational Science*. New York: Springer-Verlag, 1998.

lopment (KORSD-2009): 5th international conference, 2009.

Rasch, G. *Probabilistic Models for some intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960; Expanded edition: Chicago, The University of Chicago Press, 1980.

UŽDUOTIES SPRENDIMO TEORIJOS TAIKYMAI, MODELIUOJANT SOCIALINIUS REIŠKINIUS

Aleksandras Krylovas

Mykolo Romerio universitetas, Lietuva

Natalja Kosareva

Vilniaus Gedimino technikos universitetas, Lietuva

Santrauka. Straipsnyje apžvelgiamos užduoties sprendimo teorijos (IRT – Item Response Theory) galimybės, modeliuojant socialinius reiškinius, jos privalumai bei trūkumai, palyginti su klasikinės testų teorijos ir matematinės statistikos metodais. Aptariama IRT taikymo socialiniuose moksluose specifika. Pristatomas IRT apibendrinimas, kai klausimo charakteristinė funkcija aproksimuojama ne vien logistinėmis funkcijomis, bet ir funkcijomis iš kitų plačių parametrinių funkcijų klasių: arkkotangento, atkarpomis tiesinių, dviejų parabolių segmentų funkcijų. Aprašoma klausimo charakteristinės funkcijos kalibravimo – geriausios aproksimuojančios funkcijos parinkimo ir parametrų apskaičiavimo – procedūra. Iš platesnių funkcijų klasių parinkus klausimo charakteristinės funkcijos modelį, galima geriau pritaikyti formalų modelį stebimiems duomenims.

Nagrinėjamas IRT taikymo, matuojant aplinkos sveikatingumo indeksą (Environmental Performance Index – EPI), pavyzdys. Iš šį indeksą formuojančių 25 indikatorių, išmatuotų 163 pasaulio šalims, buvo pasirinkti 9 informatyviausi. Šiems indikatoriams pritaikyta autorių pasiūlyta modeliavimo technologija, praplečiant empirinių duomenų aproksimavimui naudojamų funkcijų klasę. Mišrusis modelis pasirodė informatyvesnis, palyginti su IRT modeliu. Apskaičiuotos testo informacijos funkcijos reikšmės, esant pasiūlytam mišriam modeliui, buvo aukštesnės nei dviejų parametrų logistinio modelio atveju. Ryšiui tarp dviejų EPI įverčių – bazinio, apskaičiuoto kaip svertinė 25 indikatorių suma, ir bendrojo testo balo – nustatyti buvo apskaičiuoti Kendalo τ ir Spearmano ρ ranginės koreliacijos koeficientai skirtingoms testo klausimų skaičiaus reikšmėms n. Visais atvejais buvo gautos gana aukštos koreliacijos koeficientų reikšmės. Koreliacijos koeficientai buvo reikšmingi esant reikšmingumo lygmeniui 0,05. Parodyta, kad, remiantis daug mažesniu informacijos kiekiu, nauja metodika leidžia atlikti pakankamai tikslų objektų norminį vertinimą pagal tiesiogiai neišmatuojamo latentinio kintamojo reikšmes.

Pagal EPI įverčio, apskaičiuoto kaip 9 klausimų testo bendrasis balas, reikšmes buvo atlikta klasterizacijos procedūra, suskirstant šalis į 6 klasterius ir pateiktas trumpas klasterių apibūdinimas. Pasirinktas klasterizavimo metodas – Euklido atstumas ryšiui tarp klasterių

išmatuoti. Pasiūlyta metodika gali būti pritaikyta įvairiose žmogaus veiklos srityse, kuriose aktuali problema yra tiesiogiai neišmatuojamų objektų norminis vertinimas.

Reikšminiai žodžiai: *užduoties sprendimo teorija, matematinis modeliavimas, testo informacijos funkcija, aplinkos sveikatingumo indeksas, latentinis požymis, norminis vertinimas.*

Aleksandras Krylovas, Mykolo Romerio universiteto Socialinės informatikos fakulteto Matematinio modeliavimo katedros profesorius; Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakulteto Matematinio modeliavimo katedros profesorius. Mokslinių tyrimų kryptys: matematinis modeliavimas, asimptotinė analizė, matematikos didaktika.

Aleksandras Krylovas, Mykolas Romeris University, Faculty of Social Informatics, Department of Mathematical Modelling, professor; Vilnius Gediminas Technical University, Faculty of Fundamental Sciences, Department of Mathematical Modelling, professor. Research interests: mathematical modeling, asymptotic analysis, didactics of mathematics.

Natalja Kosareva, Vilniaus Gedimino technikos universiteto Fundamentinių mokslų fakulteto Matematinio modeliavimo katedros docentė; Mykolo Romerio universiteto Socialinės informatikos fakulteto Matematinio modeliavimo katedros docentė. Mokslinių tyrimų kryptys: matematinis modeliavimas, matematinė statistika ir jos taikymai, informacinės technologijos.

Natalja Kosareva, Vilnius Gediminas Technical University, Faculty of Fundamental Sciences, Department of Mathematical Modeling, associate professor; Mykolas Romeris University, Faculty of Social Informatics, Department of Mathematical Modeling, associate professor. Research interests: mathematical modeling, applications of mathematical statistics, information technologies.