ITEM RESPONSE THEORY APPLICATIONS FOR SOCIAL PHENOMENA MODELING

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

Received 28 January, 2011; accepted 8 March, 2011

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
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 \(k \times n\) matrix, where rows represent subjects and co-

---

Columns 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**

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Average person’s ability estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Person 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Person 3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>Person 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Person 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>Average item’s difficulty</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

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):

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

where \(b_i\) 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 \(\sum x_i\) 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.

---


8 Molenaar, I., *supra* note 6.
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):

\[
P(x_i=1|p) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(b_i - p)}}; \tag{2}
\]

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.

---

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 $k(p) = k_i(a; b; p)$. 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):

---

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 \( \sum x_i \) —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 \( k_1, k_2, \ldots, k_n \). 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 \( p_i = P(S = i), i = 0, 1, \ldots, n \) is received by method of generating functions. The test information (entropy) function \( I \) is described as follows:

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

\[
k_2(p; a) = \frac{2}{\pi} \arccot \left( \frac{a \ln(p)}{\ln(1-p)} \right), a \geq 0
\]

\[
k_3(p; a; b) = \begin{cases} 0, & p < a \\ \frac{p-a}{b-a}, & a \leq p \leq b \\ 1, & p > b \end{cases}
\]

\[
k_4(p; a; b) = \begin{cases} 0, & p \leq a \\ \frac{2 \left( \frac{p-a}{b-a} \right)^2}{\frac{a+b}{2}}, & a < p \leq \frac{a+b}{2} \\ 1 - \frac{2 \left( \frac{b-p}{b-a} \right)^2}{\frac{a+b}{2}}, & \frac{a+b}{2} < p \leq b \\ 1, & p > b \end{cases}
\]
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. 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. CO2 emissions per electricity generation (g CO2 per kWh) (<394).

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. Standardized values $p_j$ of EPI are calculated for each country:

$$p_j = \frac{p_j^{fact} - p_{\min}}{p_{\max} - p_{\min}},$$

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.

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 \(p = p_j = \frac{r_j}{m_j}\). 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):

\[
d(k_i) = \sum_{j=1}^{24} w(p_j) \left( k_i(p_j; \alpha; b) - P(p_j) \right)^2, i = 1, 2, 3, 4, w(p_j) = \frac{m_j}{163}.
\]

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

<table>
<thead>
<tr>
<th>Item</th>
<th>(d(k_1))</th>
<th>(d(k_2))</th>
<th>(d(k_3))</th>
<th>(d(k_4))</th>
<th>(\frac{d_{\min}}{d(k_i)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item1</td>
<td>0.00939</td>
<td>0.00928</td>
<td>0.00915</td>
<td>0.00957</td>
<td>0.97444</td>
</tr>
<tr>
<td>Item2</td>
<td>0.01084</td>
<td>0.01031</td>
<td>0.01243</td>
<td>0.01099</td>
<td>0.95111</td>
</tr>
<tr>
<td>Item3</td>
<td>0.00743</td>
<td>0.01016</td>
<td>0.01052</td>
<td>0.00797</td>
<td>1.00</td>
</tr>
<tr>
<td>Item4</td>
<td>0.01185</td>
<td>0.01278</td>
<td>0.01302</td>
<td>0.01213</td>
<td>1.00</td>
</tr>
<tr>
<td>Item5</td>
<td>0.01971</td>
<td>0.02352</td>
<td>0.02157</td>
<td>0.01970</td>
<td>0.99949</td>
</tr>
<tr>
<td>Item6</td>
<td>0.01287</td>
<td>0.01420</td>
<td>0.01367</td>
<td>0.01292</td>
<td>1.00</td>
</tr>
<tr>
<td>Item7</td>
<td>0.02055</td>
<td>0.02221</td>
<td>0.02167</td>
<td>0.02075</td>
<td>1.00</td>
</tr>
<tr>
<td>Item8</td>
<td>0.00807</td>
<td>0.00800</td>
<td>0.00865</td>
<td>0.00800</td>
<td>0.99133</td>
</tr>
<tr>
<td>Item9</td>
<td>0.01813</td>
<td>0.02591</td>
<td>0.01795</td>
<td>0.01805</td>
<td>0.99007</td>
</tr>
</tbody>
</table>

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.
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 we are calculating the probability distribution of total test score $S$ in the whole population $\hat{p}_{ml}, j = 0,1,\ldots,n$ 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):

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

The test information function $I$ is described as follows:

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

In formula (10) we substitute $p_i$ with the estimations of probabilities under the mixed model ($\hat{p}_{ml}$) and under the logistic model ($\hat{p}_{jl}$) 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 $P(S = i) = \frac{1}{n+1}$, $i = 0,1,2,\ldots,n$ are calculated for the test consisting of 5, 7 and 9 items respectively.

---

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 ($\hat{p}_r$) and mixed ($\hat{p}_m$) models.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$I(\hat{p}_r)$</th>
<th>$I(\hat{p}_m)$</th>
<th>$I_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.776</td>
<td>1.778</td>
<td>1.792</td>
</tr>
<tr>
<td></td>
<td>(0.991)</td>
<td>(0.993)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>7</td>
<td>2.049</td>
<td>2.053</td>
<td>2.079</td>
</tr>
<tr>
<td></td>
<td>(0.985)</td>
<td>(0.987)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>9</td>
<td>2.256</td>
<td>2.2614</td>
<td>2.303</td>
</tr>
<tr>
<td></td>
<td>(0.980)</td>
<td>(0.985)</td>
<td>(1.0)</td>
</tr>
</tbody>
</table>

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 $\tau$ and Spearman’s $\rho$ rank correlation coefficients and are presented in the Table 4.

Table 4. Kendall’s $\tau$ and Spearman’s $\rho$ 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.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Kendall’s tau</th>
<th>Spearman’s rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.564</td>
<td>0.726</td>
</tr>
<tr>
<td>7</td>
<td>0.584</td>
<td>0.752</td>
</tr>
<tr>
<td>9</td>
<td>0.608</td>
<td>0.782</td>
</tr>
</tbody>
</table>

All the values of rank correlations are rather high and the correlation coefficients are significant at $\alpha = 0.05$ 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 result of clusterization performed by the total score $S$ of 9 items test.

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

Krylovas, A.; Kosareva, N. Mathematical Modelling of Diagnostic Tests. Knowledge-Based Technologies and OR Methodologies for Strategic Decisions of Sustainable Deve-
\textsuperscript{17} Harris, J.; Stocker, H. Handbook of Mathematics and Computational Science. New York: Springer-Verlag, 1998.
Aleksandras Krylovas, Natalja Kosareva. Item Response Theory Applications for Social Phenomena Modeling


Pagal EPI įverčio, apskaičiuoto kaip 9 klausimų testo bendrasis balas, reiškmes buvo atlikti klasterizacijos procedūra, suskirstant šalis į 6 klasterius ir pateiktas trumpas klasterių atpibūdinimas. Pasirinktas klasterizavimo metodas – Euklido atstumas ryšiui tarp klasterių
išmatuoti. Pasiūlyta metodika gali būti pritaikyta įvairiose žmogaus veiklos srityse, kuriose aktuali problema yra tiesiogiai neišmatuojanų objektų norminis vertinimas.

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

Aleksandras Krylov


Aleksandras Krylov, 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 modelling, asymptotic analysis, didactics of mathematics.

Natalja Kosareva


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.