

УДК [37.091.26]:004.9

Диховичний Олександр Олександрович
кандидат фізико-математичних наук, доцент
КПІ ім. Ігоря Сікорського, Київ, Україна
ORCID ID 0000-0001-8505-2878
a.dyx@ukr.net

Круглова Наталія Володимирівна
кандидат фізико-математичних наук, старший викладач
КПІ ім. Ігоря Сікорського, Київ, Україна
ORCID ID 0000-0002-9143-9398
natahak@ukr.net

ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ Й АНАЛІЗ МАТРИЦЬ ПЕРВИННИХ БАЛІВ ПЕДАГОГІЧНОГО ТЕСТУВАННЯ ЗА ДОПОМОГОЮ МОВИ R

Анотація. У статті досліджуються методи імітаційного моделювання матриць первинних балів педагогічного тестування засобами мови статистичного програмування R. Розроблено власні алгоритми і функції, що дозволяють генерувати матрицю первинних балів, яка відповідає тесту заданої структури. Важливість цього підходу зумовлена низкою причин, а саме, необхідністю: створення еталонних вибірок, аналізу первинних балів засобами КТТ (Класичної Теорії Тестів) та IRT; прогнозування основних статистичних характеристик тесту; уточнення параметрів «каліброваних» завдань; незалежного моделювання параметрів іспитників; підвищення компетентності викладача.

Зауважимо, що вхідні параметри для моделювання матриць можна як генерувати, так і задавати вручну.

Проведено порівняльний аналіз створених функцій з уже існуючими функціями пакетів *eRm*, *ltm*, *mcIRT*, а також статистичний аналіз згенерованих матриць. Аналіз передбачав такі процедури: перевірка гіпотези про адекватність згенерованих матриць заданим параметрам тестових завдань за критерієм χ^2 -квадрат; перевірка гіпотези про рівність середніх балів за всією матрицею за критерієм Стьюдента; перевірка гіпотези про рівність векторів відносних частот правильних відповідей по стовпцях за критерієм Хоттелінга; порівняння теоретичних характеристичних кривих з емпіричними ймовірностями; порівняння заданих параметрів складності завдань й оцінених за згенерованими матрицями з урахуванням похибок оцінювання.

Побудовано експериментальну систему імітаційного моделювання й аналізу результатів тестування, яка об'єднує методи сучасної IRT і КТТ і дає можливість генерувати матриці первинних балів тестування та проводити аналіз результатів тестування: обчислювати основні статистичні характеристики тесту, оцінювати латентні параметри, будувати характеристичні криві та інформаційні функції. Графічна оболонка системи побудована за допомогою пакету *Shiny*. У системі реалізовано моделювання та аналіз результатів тестування за базовими моделями IRT: Раша; Бірнбаума, Су-Болта, Раша-Мастерса.

Проведено перевірку роботи алгоритмів і функцій, імplementованих у систему, за допомогою низки вказаних статистичних методів і процедур, які підтвердили коректність роботи цих алгоритмів.

Ключові слова: IRT; модель Раша; модель Су-Болта; модель Раша-Мастерса; матриця первинних балів; латентні параметри; тестування.

1. ВСТУП

Постановка проблеми. Дослідження у галузі педагогічного тестування у вищій школі передбачають багато напрямків, одним з яких є імітаційне моделювання результатів тестування.

Під імітаційним моделюванням результатів педагогічного тестування ми будемо розуміти генерацію матриць первинних балів іспитників (матриць відповідей) за

заданою структурою тесту, а саме: загальною кількістю завдань і іспитників, типами завдань, кількістю завдань певного типу, параметрами відповідних завдань і іспитників. За такі параметри нами обрано латентні параметри, на підставі яких побудовано моделі Item Response Theory (IRT)[1]. Ці параметри вважаються такими, що найбільш адекватно визначають властивості як окремих тестових завдань, так і тесту у цілому. Попри це, на нашу думку, є доречним поєднання функцій моделювання результатів тестування з функціями оцінки параметрів та аналізу тестів як для штучно згенерованих, так і реальних матриць у єдину систему.

Актуальність такої постановки проблеми зумовлена поданими нижче факторами.

✓ **Створення еталонних вибірок.** Потреба у цьому зумовлена необхідністю перевірки (тестування) програм оцінки латентних параметрів. Оскільки реальне тестування є витратним процесом, то безумовно важливим є створення репрезентативних вибірок шляхом їх комп'ютерної імітації.

✓ **Аналіз первинних балів засобами класичної теорії тестів (КТТ).** КТТ, як і IRT, для дослідження якості тестів використовує матрицю первинних балів, тому, цікавим є встановлення зв'язку між результатами КТТ-аналізу, а саме: надійністю, валідністю тощо, з даними, створеними за IRT-моделями. Про аспекти такого зв'язку наголошено у роботі [2].

✓ **Прогнозування основних статистичних характеристик тесту.** Під час проведення масштабного тестування, наприклад, ЗНО, тести мають бути складеними з «каліброваних» завдань, про що свідчить, наприклад, сайт Київського регіонального центру оцінювання якості освіти: «Проведення якісних моніторингових досліджень передбачає використання каліброваних завдань з високими психометричними характеристиками» [3]. Ми теж під «калібруванням» завдань розуміємо визначення їх психометричних характеристик. Під час проведення таких масштабних і відповідальних тестувань доречно змоделювати їхні результати, за якими можна спрогнозувати базові статистичні показники тесту.

✓ **Уточнення параметрів «каліброваних» завдань.** Під час проведення «калібрування» завдань шляхом апробації часто виникає проблема вибірок малого обсягу. Для розв'язання цієї проблеми існують різні методи, наприклад, «метод складного ножа», або «бутстреп-методи» [4, 5]. Але тим же автором стверджується, що застосування вказаних методів не є ефективним і пропонується моделювати матрицю відповідей. Дійсно, як відомо з теоретичних основ IRT, оцінки параметрів складності завдань не повинні залежати від вибору діапазону зміни параметра підготовленості. Тому, генеруючи параметри підготовленості іспитників з різних діапазонів шкали зміни параметрів, шляхом імітаційного моделювання матриць первинних балів можна збільшити обсяг вибірки, сподіваючись на отримання додаткової інформації, і за допомогою оцінювання уточнити відповідні психометричні параметри тестових завдань. Але в рамках цієї роботи детальне обговорення цієї теми не планувалось.

✓ **Незалежне моделювання параметрів іспитників.** Під час прогнозування результатів тестування великого обсягу параметри іспитників особливо у новій групі іспитників, як правило, невідомі, тому логічним здається окреме генерування параметрів підготовленості іспитників за певним імовірнісним законом та використання їх як вхідні параметри для генерації матриці первинних балів з цими параметрами.

✓ **Підвищення компетентності викладача.** Швидке і ефективне генерування матриці первинних балів тесту довільної структури з довільними параметрами разом з методами аналізу цих результатів надає викладачеві можливість набуття практичного досвіду з аналізу й інтерпретації результатів тестування, що підвищує його компетентність у галузі тестування.

Аналіз останніх досліджень і публікацій. Питання імітаційного моделювання результатів педагогічного тестування досліджувались у роботах [6, 7, 8]. У роботі [6] моделювання проводилось за спеціальним алгоритмом для моделі Раша і за певних обмежень на розмірність матриць, в [7] розглядалось моделювання складності завдань та її впливу на оцінку знань, у [8] основна увага приділяється знаходженню диференціюючої спроможності завдань шляхом імітаційного моделювання відповідей студентів, але алгоритм моделювання лише згадується. Також за цією тематикою створено низку програм, написаних мовою статистичного програмування R і представлених на сайті cran.r-project.org у розділі «Psychometric Models and Methods» у таких пакетах як: **eRm**, **ltm**, **mcIRT** [9], і за допомогою яких генерують дані за моделями Раша, Бірнбаума, Су-Болта та Раша-Мастерса [1]. Загальна характеристика цих пакетів є занадто громіздкою й об'ємною, тому наводити її недоречно.

Невирішені аспекти проблеми. Існуючі програми мають низку поданих нижче недоліків.

✓ Програми генерують матриці первинних балів, що відповідають окремим типам тестових завдань. Але важливим є генерація матриці первинних балів, яка відповідає тесту, складеному із завдань різних типів з різними латентними параметрами. До того ж, існуючі програми надають користувачеві результати моделювання у різних форматах, для використання яких потрібні спеціальні програми, а іноді формат даних є просто незручним і не відповідає формату реальних результатів тестування, зокрема, у випадку моделі Су-Болта завдань з множинним вибором.

✓ В існуючих програмах параметри підготовленості іспитників або можна задавати вручну, або моделювати тільки за нормальним розподілом, що має обмежувальний характер.

✓ У програмах недостатньо прокоментовані: вхідні параметри, код програми, формат вихідних даних, що часто ускладнює їх коректне використання.

✓ Програми моделювання мають надто універсальний характер і забезпечують надлишковий набір функцій і супровідних опцій, що призводить до зайвого використання комп'ютерних ресурсів та уповільнення моделювання.

Виходячи з поставленої проблеми та вказаних вище особливостей існуючих програм, ми можемо визначити мету дослідження.

Метою статті є дослідження методів імітаційного моделювання матриць первинних балів відповідей іспитників у педагогічному тестуванні за заданими структурою тесту, параметрами тестових завдань й іспитників відповідно до моделей IRT, а також створення власного програмного комплексу, зорієнтованого на реалізацію саме такого моделювання, оцінювання параметрів тестів і тестових завдань, аналізу результатів моделювання та оцінювання.

2. МЕТОДИ ДОСЛІДЖЕННЯ

Дослідження проводилось у декількох напрямках.

А. Проаналізовано існуючі методи генерації матриць первинних балів, які реалізовано у названих вище пакетах сайту CRAN. Основним методом, покладеним в основу генерації матриці первинних балів функцію **sim.rasch**, є метод обернених функцій розподілу [10].

В. Створено власні алгоритми генерації та реалізовано програмно. На базі створених програм було побудовано систему моделювання результатів тестування. В алгоритмах застосовувався принцип безпосереднього розрахунку ймовірності для основних моделей IRT. За таку модель тестових завдань обрано моделі: Раша,

Бірнаума, Су-Болта та Раша-Мастерса [1]. В основу всіх цих моделей покладено латентні параметри тестів і іспитників. Далі всюди:

- N – кількість іспитників,
- K – кількість завдань в тесті.

Модель Раша (G. Rasch). Застосовується для дихотомічних завдань (завдань, у яких можливі тільки два варіанти відповіді: правильний або неправильний). Ця модель визначається такими латентними параметрами:

- підготовленістю іспитника - $\theta_i, i = \overline{1, N}$,
- складністю завдання тесту - $\beta_j, j = \overline{1, K}$.

Імовірність правильної відповіді i -го іспитника на j -те дихотомічне завдання тесту визначається за формулою Раша:

$$P_{ij} = \frac{1}{1 + \exp(-(\theta_i - \beta_j))}, i = \overline{1, N}, j = \overline{1, K}.$$

Модель Бірнаума (A. Birnbaum). Розвитком моделі Раша є модель Бірнаума, у якій введено додатковий параметр диференціюючої спроможності. Латентні параметри:

- $\theta_i, i = \overline{1, N}$ - підготовленості іспитника,
- $\beta_j, j = \overline{1, K}$ - складності завдання тесту ,
- $\alpha_j, j = \overline{1, K}$ - диференціююча спроможність завдань тесту .

Імовірність правильної відповіді i -го іспитника на j -те дихотомічне завдання тесту визначається за формулою:

$$P_{ij} = \frac{1}{1 + \exp(-\alpha_j(\theta_i - \beta_j))}, i = \overline{1, N}, j = \overline{1, K}.$$

Модель Су-Болта (Y. Suh-D.Bolt). Для завдань множинного вибору використовують модель Су-Болта.

Латентні параметри:

- $\alpha_j, j = \overline{1, K}$ - диференціююча спроможність k -го варіанту відповіді j -го завдання ,
- $\beta_j, j = \overline{1, K}$ - складність k -го варіанту відповіді j -го завдання,
- $\gamma_j, j = \overline{1, K}$ - імовірність угадування правильного варіанту відповіді j -го завдання .

$$P(U_{ji} = 1 | \theta_i) = P_j(\theta_i) = \gamma_j + (1 - \gamma_j) \left[\frac{1}{1 + \exp\{-(\beta_j + \alpha_j \theta_i)\}} \right], i = \overline{1, N}, j = \overline{1, K}.$$

- імовірність вибору k -го дистрактора (неправильного варіанта) відповіді j -го завдання .

$$P_{jk}(\theta_i) = \left\{ 1 - \left[\gamma_j + (1 - \gamma_j) \frac{1}{1 + \exp\{-(\beta_j + \alpha_j \theta_i)\}} \right] \right\} \left[\frac{\exp\{Z_{jk}(\theta_i)\}}{\sum_{s=1}^{h_j} \exp\{Z_{js}(\theta_i)\}} \right], i = \overline{1, N}, j = \overline{1, L}, k = \overline{1, h_j}.$$

Модель Раша-Мастерса (G. Rasch- G.Masters). Для політомічних завдань(завдань на встановлення відповідності) застосовують модель Раша-Мастерса.

Латентні параметри:

- $\theta_i, i = \overline{1, N}$ - підготовленості іспитника,
- $\beta_{jk}, j = \overline{1, K}, k = \overline{0, m_j}$ - параметри складності k-го рівня j-го завдання,
- m_j - кількість рівнів завдання .

У політомічній моделі Андерсена ймовірність досягнення і-м іспитником k-го рівня j-ого завдання означається так:

$$P_{jk}(\theta_i) = \frac{\exp \sum_{s=0}^k (\theta_i - \beta_{js})}{\sum_{k=0}^{m_j} \exp \sum_{s=0}^k (\theta_i - \beta_{js})}, i = \overline{1, N}, j = \overline{1, K}, k = \overline{1, m_j}.$$

С. Проведено статистичний порівняльний аналіз матриць первинних балів, що згенеровані існуючими функціями сайту CRAN, з матрицями, які згенеровані авторськими функціями. З метою спрощення викладення матеріалу аналіз проілюстровано на прикладі моделі Раша. Аналіз передбачав такі процедури:

- 1) перевірялась гіпотеза про адекватність обох видів згенерованих матриць заданим параметрам тестових завдань за критерієм хі-квадрат [11, с. 77];
- 2) перевірялась гіпотеза про рівність середніх балів за всією матрицею за критерієм Стьюдента [11, с. 87];
- 3) перевірялась гіпотеза про рівність векторів відносних частот правильних відповідей по стовпцях за критерієм Хоттелінга [12];
- 4) порівнювались теоретичні характеристичні криві (побудованих за заданими параметрами) з емпіричними ймовірностями (оціненими за згенерованими матрицями) [11];
- 5) порівнювались задані параметри складності завдань й оцінені за згенерованими матрицями з урахуванням похибок оцінювання [11].

3. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

1) Створення алгоритмів генерації матриць первинних балів

У середовищі програмування R існує достатньо розгалужений набір пакетів з моделювання й аналізу результатів тестування. Але вони мають низку вже перелічених недоліків. Методом, покладеним в основу моделювання результатів тестування для функції sim.rasch, пакет eRm, є метод обернених функцій розподілу [10], а саме, наступний алгоритм генерації матриці первинних балів:

- 1) за допомогою датчика псевдовипадкових чисел генерується випадкова величина U рівномірно розподілена на відріжку $[0, 1]$. Зауважимо, що базовим датчиком псевдовипадкових чисел для середовища R є *Mersenne twister* [13];

2) обчислюється ймовірність неправильної відповіді за формулою:
 $p_{ij} = 1 / (1 + \exp(\theta_i - \beta_j))$;

3) якщо $U > p_{ij}$, то елементом матриці буде 1, в іншому випадку – 0.

Але цей алгоритм має обмеження в застосуванні, тому що з його допомогою можна створювати лише дихотомічні матриці. У випадку політомічної моделі Раша-Мастерса або моделі множинного вибору Су-Болта цей алгоритм не спрацює. Тому ми пропонуємо інший підхід, **підкреслюємо**, до моделювання саме матриць, а не моделювання розподілів дискретних випадкових величин, який полягає в іншій схемі використання функцій середовища R. Ми виконуємо розрахунок імовірностей варіантів відповідей відповідних моделей IRT і робимо вибір потрібного варіанту згідно розрахованої ймовірності за допомогою функції **sample** (пакет **base**). Функція **sample** викликає функцію **do_sample**, яка входить до ядра системи R і написана мовою C++. Ця функція забезпечує генерацію дискретної випадкової величини за допомогою алгоритма Уолкера, який у даному випадку є більш ефективним, принаймні, за швидкістю. Ми не вважаємо за доцільне у рамках цієї роботи наводити цей відомий алгоритм і його переваги над стандартним методом моделювання дискретної випадкової величини, оскільки з алгоритмом Уолкера можна ознайомитись в оригіналі [14] або у перекладі [15, с. 28].

2) Розробка системи моделювання й аналізу результатів тестування

Створені програми генерації матриць первинних балів було об'єднано разом з уже існуючими програмами розрахунку латентних параметрів у програмну систему, написану мовою статистичного програмування R. Це дозволяє забезпечити проведення статистичного аналізу як змодельованих даних, так і результатів реального тестування.

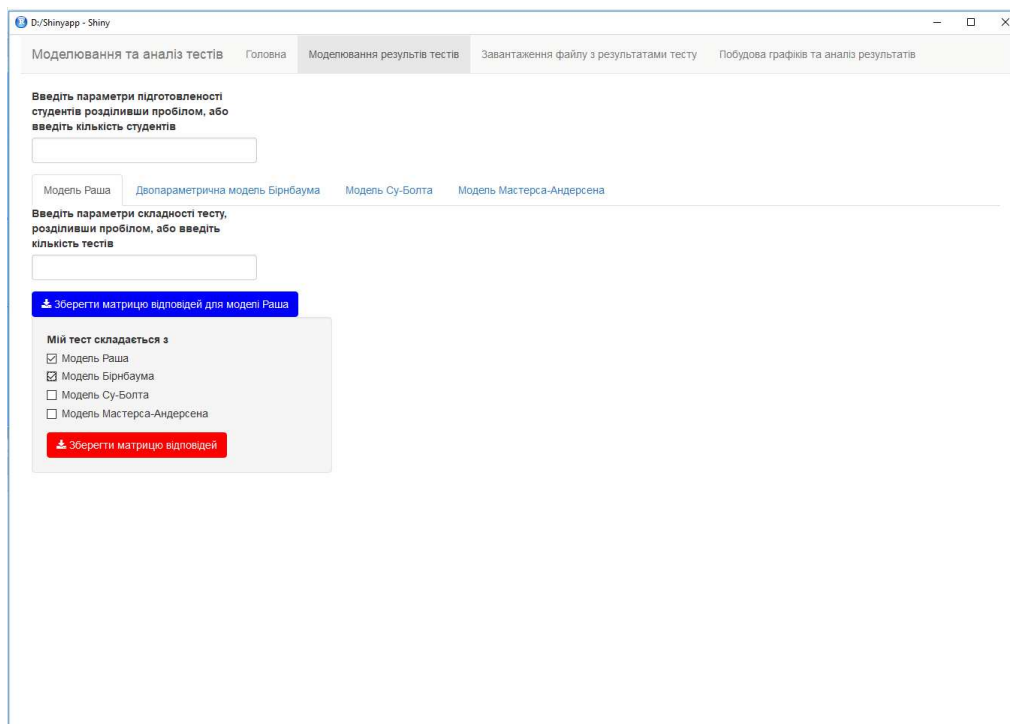


Рис. 1. Приклад формування структури тесту

Важливим є те, що у системі користувач може згенерувати матрицю, яка відповідає тесту, складеному з тестових завдань різних типів для довільної кількості тестових завдань та іспитників. Причому у системі реалізовано можливість як задавати

вручну латентні параметри, так і генерувати їх за певним імовірнісним законом: нормальним, рівномірним, експоненційним, логнормальним.

Система дозволяє провести найпростіший статистичний аналіз на основі КТТ, а саме: обчислити вибіркові характеристики, оцінити валідність та надійність тесту.

Графічну оболонку системи побудовано з використанням пакету Shiny.

На рис. 1 представлено приклад формування структури і параметрів завдань тесту, результати якого буде змодельовано.

У системі передбачено й аналіз тестів на основі КТТ (рис. 2).

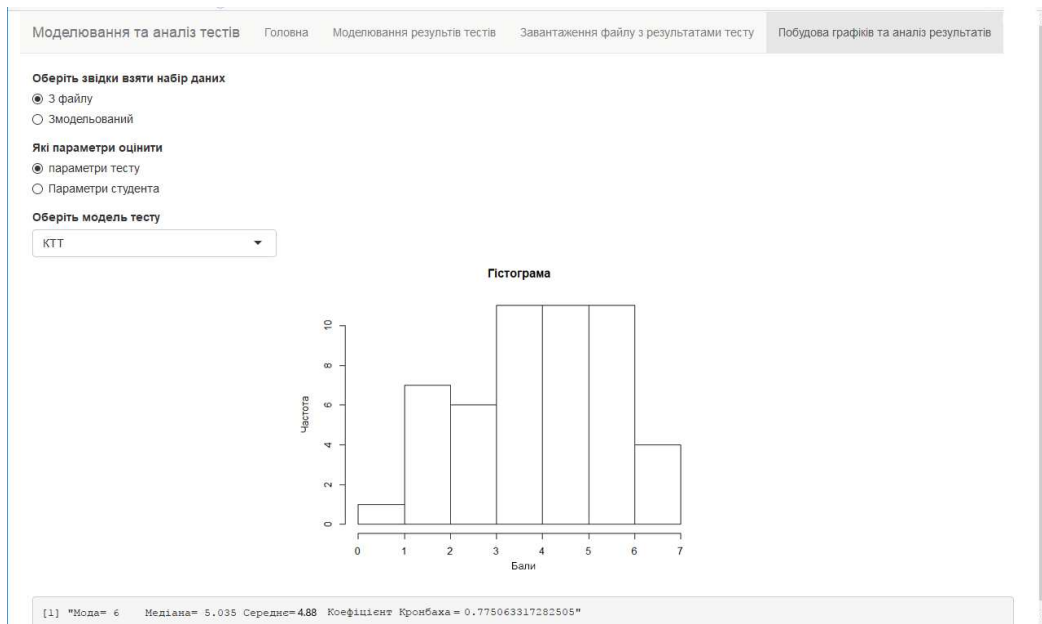


Рис. 2. Аналіз тесту на основі КТТ

Разом з цим у системі можливий аналіз ансамблів характеристичних кривих і інформаційних функцій (рис. 3).

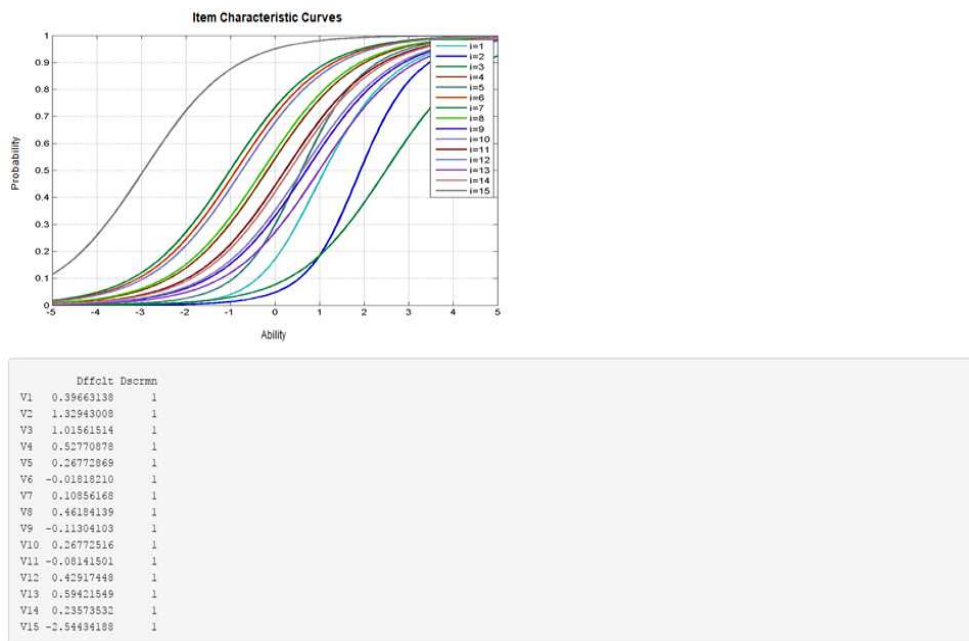


Рис. 3. Характеристичні криві

3) Тестування системи

Наведемо результати тестування функцій генерації матриць на прикладі моделі Раша. Нова розроблена функція називається **Rasch.sim**, а вже існуюча функція – **sim.rasch** з пакету **eRm**. Було згенеровано два види матриці, один – функцією **Rasch.sim**, другий – **sim.rasch**, для тесту з 10 дихотомічних тестових завдань для різної кількості іспитників.

Нижче у пунктах 1-4; параметри складності тестових завдань:

$$\beta_1 = 2.2, \beta_2 = -1.1, \beta_3 = 0.1, \beta_4 = 0.3, \beta_5 = 0.6, \beta_6 = 0.8, \beta_7 = 2.2, \beta_8 = 2.0, \beta_9 = -1.5, \beta_{10} = 2.7,$$

а параметри іспитників генерувались як 1000 випадкових величин, розподілених за гауссівським законом зі середнім – 0.0, і дисперсією – 1.0.

1. *Перевірка гіпотез про адекватність обох видів згенерованих матриць заданим параметрам тестових завдань за критерієм χ^2 -квадрат.* За формулами (4.3.5) з [11] було розраховано статистики χ^2 . Для функції **sim.rasch** $\chi_1^2 = 78,098$, а для функції **Rasch.sim** - статистика $\chi_2^2 = 77,141$. Критичне значення для 110 степенів свободи і рівня значущості $\alpha = 0.05$ $\chi_k^2 = 135,48$. Отже, приймається гіпотеза про адекватність обох матриць заданим параметрам тестових завдань.

2. *Перевірка гіпотези про рівність середніх балів за всією матрицею за критерієм Стьюдента.* За формулами (4.6.1) [11] обчислено середні значення $\bar{B}_1 = 0.4554$ для функції **sim.rasch** і $\bar{B}_2 = 0.4581$ – для функції **Rasch.sim**, статистика Стьюдента $t = 0,5084$, при рівні значущості $\alpha = 0.05$ і 1998 степенях свободи критичне значення: $t_{kp} = 1.88354$. Отже, гіпотеза про рівність середніх приймається.

3. *Перевірка гіпотези про рівність векторів відносних частот по стовпцях за критерієм Хомтелінга.* Для цього порівнювались відносні частоти правильних відповідей по стовпцях матриць: якщо n_j — кількість правильних відповідей у стовпці, тоді $V_j = \frac{n_j}{N}$ – відносна частота правильної відповіді у стовпці. Розраховані частоти наведено у таблиці 1.

Таблиця 1

Відносні частоти правильних відповідей у стовпцях, обчислені функцією пакету eRm (sim.rasch) і розробленою функцією (Rasch.sim)

sim.rasch	0,132	0,697	0,468	0,445	0,383	0,321	0,113	0,149	0,781	0,336
Rasch.sim	0,121	0,72	0,472	0,426	0,368	0,342	0,128	0,167	0,756	0,358

Обчислено статистику $F = 0,0131$, а при рівні значущості $\alpha = 0.05$ критичне значення $F_{kp(0.05;10;1989)} = 1,645$. Отже, гіпотеза про рівність векторів відносних частот по стовпцях приймається.

4. *Порівняння теоретичних характеристичних кривих(побудованих за заданими параметрами) з емпіричними ймовірностями [11] (розрахованими як відносні частоти правильних відповідей по рядках за згенерованими матрицями: якщо n_i – кількість правильних відповідей у рядку, тоді $V_i = \frac{n_i}{K}$ – відносна частота правильної відповіді у рядку правильних відповідей також відображено на рисунках 4 і 5. Неперервна крива – характеристична крива моделі Раша, побудована для параметру $\beta = 0.1$, а крапками*

позначено розраховані ймовірності правильних відповідей для різних значень підготовленості іспитників. Як видно з малюнку, емпіричні ймовірності достатньо точно відповідають теоретичній кривій.

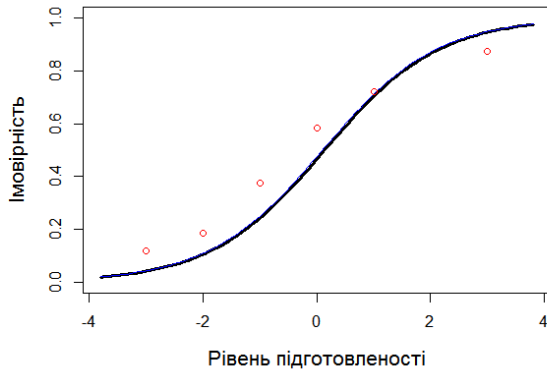


Рис.4. Порівняльний аналіз для функції Rasch.sim

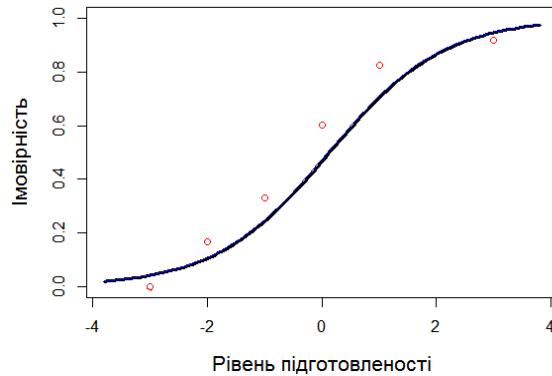


Рис.5 Порівняльний аналіз для функції sim.rasch

5. Порівняння заданих параметрів складності завдань з оціненими і за згенерованими матрицями з урахуванням похибок оцінювання. Результати порівняння заданих латентних параметрів складності завдань та їхніх оцінок, отриманих за згенерованою матрицею результатів тестування, наведено у таблиці 2 з урахуванням середньоквадратичних похибок оцінювання. Кількість іспитників змінювалась від 100 до 100000. Як видно з таблиці, оцінені параметри є близькими до заданих, особливо це помітно для великих об'ємів вибірок.

Таблиця 2

Порівняння параметрів складності завдань тесту

К-сть іспитників	Вхідні параметри складності завдань, β	Функція sim.rash (eRm), β^*	sim.rash (eRm) σ	Функція Rasch.sim (створена), β^*	Rasch.sim σ
100	2.2	1.8552231	0.2929	1.7766026	0.2928
	-1.1	-0.8446752	0.2700	-1.0025117	0.2692
	0.1	0.1373498	0.2418	0.0997455	0.2416
	0.3	0.3300621	0.2442	0.3889647	0.2445
	0.6	0.5771397	0.2508	0.8983361	0.2460
	0.8	0.7830952	0.2465	0.844814	0.2474
	2.2	2.5400339	0.3497	2.2058975	0.3522
	2	1.5566183	0.2883	2.1110822	0.2882
	-1.5	-1.4906020	0.2876	-1.3591294	0.2819
	0.7	0.2333351	0.2521	0.4379226	0.2512
1000	2.2	2.18326066	0.0981	2.2242199	0.1050
	-1.1	-1.10810611	0.0828	-1.1210688	0.0822
	0.1	0.07098132	0.0769	0.0446658	0.0767
	0.3	0.22689251	0.0772	0.2779495	0.0768
	0.6	0.58591472	0.0793	0.6173313	0.0786
	0.8	0.76811939	0.0802	0.8534236	0.0799
	2.2	2.23196498	0.1024	2.1662261	0.1073

	2	1.86770972	0.0982	1.9689541	0.0980
	-1.5	-1.61812261	0.0878	-1.4493953	0.0868
	0.7	0.55530740	0.0784	0.6999082	0.0781
10000	2.2	2.19055413	0.0327	2.2179892	0.0328
	-1.1	-1.09126786	0.0263	-1.0919610	0.0261
	0.1	0.09551229	0.0242	0.1136219	0.0242
	0.3	0.30510750	0.0243	0.3360016	0.0243
	0.6	0.60497623	0.0248	0.5879345	0.0248
	0.8	0.80444974	0.0253	0.8037771	0.0253
	2.2	2.17997242	0.0329	2.1508123	0.0328
	2	1.96993333	0.0316	2.0494279	0.0314
	-1.5	-1.47519084	0.0282	-1.4710350	0.0280
	0.7	0.68418397	0.0251	0.6790485	0.0250
	100000	2.2	2.2130397	0.0103	2.19948368
-1.1		-1.0769315	0.0083	-1.10675397	0.0083
0.1		0.1063861	0.0077	0.09699357	0.0077
0.3		0.3028872	0.0077	0.30575558	0.0077
0.6		0.5857575	0.0078	0.59349047	0.0078
0.8		0.8002250	0.0080	0.80564333	0.0080
2.2		2.2023219	0.0104	2.19107029	0.0104
2		2.0066365	0.0098	2.00263885	0.0099
-1.5		-1.5145542	0.0088	-1.49313174	0.0089
0.7		0.7099893	0.0079	0.70282177	0.0079

4. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

1. Мова R, як інструмент статистичного програмування, надає гнучкі і ефективні засоби для створення як спеціалізованих функцій, так і програмних комплексів для імітаційного моделювання та аналізу результатів педагогічних тестів.

2. У роботі запропоновано нові алгоритми генерації матриць первинних балів відповідей іспитників за заданою структурою тесту, здійснено їх програмну реалізацію у вигляді самостійної системи, у яку додано функції оцінки параметрів тестів та їх аналізу. Таке об'єднання, на думку авторів, є цілком логічним.

3. Проведено перевірку роботи алгоритмів за допомогою низки статистичних методів, які підтвердили коректність роботи цих алгоритмів.

Перспективи подальших досліджень:

1. Створення єдиної системи моделювання й аналізу результатів тестування дозволить у перспективі вирішувати проблеми, про актуальність яких наголошено у вступі.

2. Розроблена система стане у нагоді викладачам у проектуванні, конструюванні тестів та у проведенні аналізу їх результатів, тим самим підвищуючи їх компетентність.

3. Створена система може застосовуватись безпосередньо в навчальному процесі для прогнозування результатів тестування, наприклад, у навчальних підрозділах, там, де тестування ще не проводилось, а також в аналізі результатів тестування.

4. На думку авторів, перспективним є розширення функціональних можливостей системи шляхом розширення переліку доступних моделей IRT, а також доповнення новими статистичними функціями, наприклад, перевірки статистичних гіпотез тестування.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] W. Linden and R. Hambleton, *Handbook of modern item response theory*. New York: Springer, 1997. 503 с.

- [2] О. О. Диховичний, А. Ф. Дудко, "Комплексна методика аналізу якості тестів з вищої математики", *Науковий часопис НПУ імені М. П. Драгоманова. Серія 2 : Комп'ютерно-орієнтовані системи навчання.*, № 15, с. 139-144, 2015.
- [3] "Що передбачає апробація інструментів дослідження? Київський регіональний центр оцінювання якості освіти: веб-сайт". [Електронний ресурс]. Режим доступу: <https://kievtest.org.ua/2017/04/моніторинг-початкової-освіти-запита/>.
- [4] И. Н. Елисеев, "Калибровка заданий теста с использованием бутстреп-метода" *Программные продукты и системы.* № 2. С. 96-99. 2010.
- [5] И. Н. Елисеев, И. И. Елисеев, "Уточненный расчет параметров латентных переменных по выборкам малого объёма", *Развитие тестовых технологий в России: тезисы доклада на IX Всероссийской научно-методической конференции.* М., С. 117–118, 2007.
- [6] И. Н. Елисеев, "Модель дихотомической матрицы результатов тестирования", *Программные продукты и системы,* № 3. С. 80-86, 2011.
- [7] Р. Х. Сафаров, О. Ю. Панищев, " Численное моделирование инвариантности оценки знанияотносительно трудности тестовых заданий в рамках модели Г. Раша" , *Образовательные технологии и общество,* Т.15, №1, С. 424-435, 2012
- [8] А. В. Саяпин, К. В. Сафонов, "Оценка дифференцирующей способности компьютерного теста методами имитационного моделирования" *КиберЛенинка.*[Електронний ресурс]. Режим доступу: <https://cyberleninka.ru/article/n/otsenka-differentsiruyushey-sposobnosti-kompyuternogo-testa-metodami-imitatsionnogo-modelirovaniya>, 2012.
- [9] Available Packages. The R Project for Statistical Computing: Website. .[Електронний ресурс]. Режим доступу: <https://cran.us.r-project.org>.
- [10] С. М. Ермаков, Г. А. Михайлов, *Статистическое моделирование*, М.: Наука, 1982, 296с.
- [11] Ю. М. Нейман, В. А. Хлебников , *Введение в теорию моделирования и параметризации педагогических тестов*, М., 2000, 168с.
- [12] С. Рао , *Линейные статистические методы и их применения.* пер. с англ., М. : Наука., 1968, 547 с.
- [13] М. Matsumoto, Т. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator»" *ACM Trans. on Modeling and Computer Simulations*, Vol. 8, № 1, pp. 3-30, 2017.
- [14] A. J Walker, "New fast method for generating discrete random numbers with arbitrary frequencydistributions",. *Electronics Letters.* 10 (8), p . 127, 1974.
- [15] Г. А. Михайлов, А. В. Войтишек. *Численное статистическое моделирование. Методы Монте-Карло.* М.: «Академия», 2006, 368с.

Матеріал надійшов до редакції 15.02.2018 р.

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ И АНАЛИЗ МАТРИЦ ПЕРВИЧНЫХ БАЛЛОВ ПЕДАГОГИЧЕСКОГО ТЕСТИРОВАНИЯ С ПОМОЩЬЮ ЯЗЫКА R

Дыховичный Александр Александрович
кандидат физико-математических наук, доцент
КПИ им. Игоря Сикорского, Киев, Украина
ORCID ID 0000-0001-8505-2878
a.dux@ukr.net

Круглова Наталья Владимировна
кандидат физико-математических наук, старший преподаватель
КПИ им. Игоря Сикорского, Киев, Украина
ORCID ID 0000-0002-9143-9398
natahak@ukr.net

Аннотация. В статье исследуются методы имитационного моделирования матриц первичных баллов педагогического тестирования средствами языка статистического программирования R. Разработаны собственные алгоритмы и функции, которые позволяют генерировать матрицу первичных баллов, соответствующей тесту заданной структуры. Важность этого подхода предопределена рядом причин, а именно, необходимостью: создания эталонных выборок, анализа первичных баллов средствами КТТ (Классической Теории Тестов) и IRT; прогнозирования основных статистических характеристик теста;

уточнення параметров "откалиброванных" заданий; независимого моделирования параметров испытуемых; повышения компетентности преподавателя.

Заметим, что входные параметры для моделирования матриц можно как генерировать, так и задавать вручную.

Проведен сравнительный анализ созданных функций с уже существующими функциями пакетов eRm, ltm, mcIRT, а также статистический анализ сгенерированных матриц. Анализ предусматривал такие процедуры: проверка гипотезы об адекватности сгенерированных матриц заданным параметрам тестовых заданий по критерию хи-квадрат; проверка гипотезы о равенстве средних баллов за всей матрице по критерию Стьюдента; проверка гипотезы о равенстве векторов относительных частот правильных ответов по столбцам по критерию Хоттелинга; сравнение теоретических характеристических кривых с эмпирическими вероятностями; сравнение заданных параметров сложности заданий и оцененных по сгенерированным матрицам с учетом погрешностей оценивания.

Построена экспериментальная система имитационного моделирования и анализа результатов тестирования, которая объединяет методы современной IRT и КТТ (Классической Теории Тестов), позволяет генерировать матрицы первичных баллов тестирования и проводить анализ результатов тестирования: вычислять основные статистические характеристики теста, оценивать латентные параметры, строить характерные кривые и информационные функции. Графическая оболочка системы построена с помощью пакета Shiny. В системе реализовано моделирование и анализ результатов тестирования в соответствии с базовыми моделями IRT: Раша; Бирнбаума, Су-Болта, Раша-Мастерса.

Проведена проверка работы алгоритмов и функций, имплементированных в систему, с помощью ряда указанных статистических методов и процедур, которые подтвердили корректность работы этих алгоритмов.

Ключевые слова: IRT; модель Раша; модель Су-Болта; модель Раша-Мастерса; матрица первичных баллов; латентные параметры; тестирование.

IMITATIONAL MODELING AND ANALYSIS OF MATRIXES CONTAINING PRIMARY GRADING OBTAINED IN EDUCATIONAL TESTING BY THE MEANS OF LANGUAGE R

Oleksandr O. Dykhovychnyi

PhD (physical and mathematical sciences),

Associated professor of Mathematical analysis and Probability Theory Department

National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Kyiv, Ukraine

ORCID ID 0000-0001-8505-2878

a.dyx@ukr.net

Nataliia V. Kruglova

PhD (physical and mathematical sciences), Assistant professor

National Technical University of Ukraine "Igor Sikorsky Kiev Polytechnic Institute", Kyiv, Ukraine

ORCID ID 0000-0002-9143-9398

natahak@ukr.net

Abstract. The article researches methods of imitational modeling of matrixes containing primary grading obtained in educational testing by the means of statistical programming language R. Unique algorithms and functions were developed to allow generating of matrix of the primary grades according to corresponding test of the defined structure.

The importance of this approach is defined by several reasons, specifically the needs to: create reference samples; analyze primary grades by means of CTT (Clasic Test Theory) and IRT; predict basic statistical test characteristics; clarify parameters for the calibrated tasks; model independent parameters for the test takers; increase and development educator's competency.

It should be noted that input parameters could be generated or set up manually.

Comparable analysis was conducted for created functions against already existing function packages, such as eRm, ltm, mcIRT, as well as statistical analysis of the generated matrixes. This analysis took into consideration the following procedures: verification of hypothesis about compatibility between generated matrixes and set parameters for the testing tasks per χ^2 criteria;

verification of hypothesis about equivalence of average grades in entire matrix according Student criteria; verification of hypothesis about equality for vectors of correct answers relative frequencies by columns according to Hotteling criteria; comparison of theoretical characteristic curves with empirical probabilities; comparison of set parameters for task complexity and parameters graded by generated matrixes with consideration of errors in grading.

An experimental system of imitational modeling and analysis for testing results was created. Such system combines contemporary methods of IRT and methods of Classical Testing Theory (CTT). It allows generating matrixes of primary testing grades and performing test results analysis; permits computation of basic statistical characteristics of the test, estimation of the latent parameters, construction of characteristic curves and informational functions. The system graphic shell was generated with the help of the package Shiny. The system utilizes modeling and analysis for testing results according to basic IRT models: Rasch, Birnbaum, Suh-Bolt, Rasch-Masters.

A performance verification for algorithms and functions implemented into the system has been done by utilizing several noted statistical methods and procedures; and correct execution of these algorithms has been confirmed.

Keywords: IRT; Rasch's model; Suh-Bolt's model; Rasch-Masters model; matrix' modeling of the primary grades; testing.

REFERENCES (TRANSLATED AND TRANSLITERATED)

- [1] W. Linden and R. Hambleton, *Handbook of modern item response theory*. New York: Springer, 1997, 503p. (in English).
- [2] O. Dykhovychnyi and A. Dudko, "The comprehensive procedure of analysis of tests in higher mathematics ", *Kompiuterno-oriientovani systemy navchannia: Zb. nauk. prats NPUimeni M.P.Drahamanova, Volume 2, №15, P.139-144, 2015* (in Ukrainian).
- [3] What does testing research tools mean? Kyiv Regional Center for Educational Quality Assessment: Website. . [Online]. Available: <https://kievtest.org.ua/2017/04/моніторинг-початкової-освіти-запита/> (in Ukrainian).
- [4] I.N. Eliseev, «Calibration of items by the means Bootstrap methods», *Program products and systems.*, № 2. P. 96-99. 2010 (in Russian).
- [5] I.N. Eliseev , I.I. Eliseev «Refined calculation of latent variables parameters by small samplies», *Razvitie testovykh tehnologij v Rossii: tezisy doklada na IX Vserossijskoj nauchno-metodicheskoy konferenci*. M., P.117–118, 2007 (in Russian).
- [6] I.N. Eliseev, " Model of the dichotomius matrix of test results", *Program products and systems.*, № 3. P. 80-86, 2011 (in Russian).
- [7] R.H. Safarov and O.Ju. Panishhev, " Numerical simulation of the invariance of the estimation of the Knowledges relative to difficulty of the test items within the model of G. Rasch", *Obrazovatel'nye tehnologii i obshchestvo*, T. 15, № 1, P. 424-435, 2012 (in Russian).
- [8] A.V. Sajapin and K.V. Safonov, "Measuring of differentiation ability of computer testing by methods of imitation modelling", *Cyberleninka*. [Online]. Available: <https://cyberleninka.ru/article/n/otsenka-differentsiruyushey-sposobnosti-kompyuternogo-testa- metodami-imitatsionnogo-modelirovaniya>, 2012. (in Russian)
- [9] Available Packages. The R Project for Statistical Computing: Website. [Online]. Available: <https://cran.us.r-project.org>.
- [10] S.M. Ermakov, G.A. Mihajlov, *Statistical modeling, M., Nauka, 1982*, (in Russian)
- [11] Iu.M. Neiman and V.A. Khlebnykov, *Introduction in the theory of modeling and parametrization of pedagogical tests*, M., 2000, 168p. (in Russian).
- [12] C. Rao , *Linear statistical inference and its application*, M. Nauka, 1968, 547p. (in Russian).
- [13] Matsumoto, T. Nishimura, «Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator», *ACM Trans. on Modeling and Computer Simulations*, Vol.8, № 1, pp. 3—30, 2017, (in English).
- [14] A. J.Walker, «New fast method for generating discrete random numbers with arbitrary frequency distributions». *Electronics Letters*. 10 (8), p.127, 1974 (in English).
- [15] H.A. Mykhailov and A.V. Voityshek, *Numerical statistical modeling. Monte-Carlo methods*, M.: «Academia», 2006, 368p. (in Russian).

