

О. О. ДИХОВИЧНИЙ (канд. фіз.-мат. наук, доцент),

А. Ф. ДУДКО (аспірант)

Національний технічний університет України «КПІ», м. Київ

АВТОМАТИЗОВАНА СИСТЕМА АНАЛІЗУ РЕЗУЛЬТАТІВ КОМП'ЮТЕРНОГО ТЕСТУВАННЯ З ВИЩОЇ МАТЕМАТИКИ

Інформується про розробку в НТУУ «КПІ» автоматизованої системи статистичного аналізу результатів комп'ютерного тестування з вищої математики. Систему побудовано на базі як класичних статистичних методів, так і методів сучасної теорії параметризації тестових завдань, Item Response Theory (IRT). Серед відомих моделей IRT переважає модель Раша, Бірнбаума, Раша-Мастерса та Тіссена-Стейнберга. Система має розвинений графічний інтерфейс, який надає об'єктивну і наглядну візуалізацію результатів тестування, а також в системі передбачена можливість формування та обслуговування бази калібркованих завдань.

Ключові слова: тестування з вищої математики, статистичний аналіз тестових завдань, латентні параметри, IRT-моделі.

Постановка проблеми. Сучасний підхід до навчання математики, в тому числі і вищої, передбачає всебічне застосування тестового підходу в оцінюванні знань. Якщо на рівні середньої освіти проведення ЗНО та обробка його результатів є прерогативою Українського центру оцінювання якості освіти, то у ВНЗ проведення комп'ютерних тестів та їх обробка є справою самих ВНЗ. В НТУУ «КПІ» комп'ютерне тестування з вищої математики організовано кафедрою математичного аналізу та теорії ймовірностей з використанням створеного викладачами кафедри комплекту дистанційних курсів «Вища математика» [4].

Останнім часом коло факультетів, залучених до проведення комп'ютерного тестування, суттєво розширяється. Це зумовлює необхідність розширення та вдосконалення бази тестових завдань з вищої математики. Сучасний підхід до аналізу якості тестових завдань передбачає застосування достатньо складного математичного апарату, який разом зі збільшенням обсягів тестування стає можливим тільки з використанням інформаційних технологій. У зв'язку з цим виникає необхідність автоматизації процесу аналізу якості тестових завдань, що і зумовлює розробку відповідної автоматизованої системи.

Аналіз актуальних досліджень. Серед методів дослідження достовірності результатів педагогічних тестів виділяються класичні статистичні методи і методи Item Response Theory. Класичні статистичні методи представлені у роботах Крокер Л. і Алгіни Дж. [7], Нейман Ю.М. і Хлебнікова В.А. [10] та ін.

Методи сучасної теорії параметризації тестових завдань, яка була заснована у 50-х роках ХХ століття датським математиком Георгом Рашем [2], знайшли свій розвиток у роботах Лорда Ф.М. [1], Бейкера Ф.Б. [1], Бірнбаума А. [1], Вілсона М. [1], Мастерса Г.Н. [1], Бокка Р.Д. [1], Семеїми Ф. [1], Тіссена Л. і Стейнберга А. [1] та ін. Вітчизняні дослідження в цій галузі були практично відсутні. Лише частково моделі IRT розглядалися Лісовою Т.В. [8], Челишковою М.Б. [12], Аванессовим В.С. [3], Кардановою Є.Ю [5], Карпінським В.Б. [6] та Федоруком П.І. Застосування методів IRT в аналізі результатів «ЕГЭ» досліджувалось в роботах Монахова В.В. [9].

Зумовленість використання Item Response Theory пов'язана з тим, що IRT останнім часом привертає найбільшу увагу фахівців галузей педагогічних і психологічних вимірювань. Одна з можливих причин популярності цієї теорії — її застосування в широко відомих тестових службах. Серед них: Національна Рада розвитку освіти в США (National Assessment of Educational Progress, NAEP), американська служба по розробці тестів для оцінки здатності до оволодіння освітніми програмами різного рівня складності (Scholastic Aptitude Tests, SAT), а також американська атестаційна служба the Graduate Record Examination (GRE).

Крім цих служб IRT використовується в таких масштабних міжнародних порівняльно—оціочніх дослідженнях, як третє міжнародне дослідження рівня підготовленості з математики і природничих наук (Third International Math and Science Survey, TIMSS), а також у міжнародній

програмі оцінки якості підготовленості студентів (the Programme of International Student Assessment, PISA). Інші можливі причини підвищеної уваги до IRT — це широке коло нових статистичних розробок даної теорії, що спираються на можливості їх активного застосування в обчислювальних і освітніх технологіях.

Реалізація методів і моделей IRT здійснена у ряді комп'ютерних програм, таких як LOGIST (перша версія 1976 р.), BILOG (1984 р.), BICAL (1979 р.), RUMM (1990 р.), MULTILOG (1991 р.), WINSTEPS (1991 р.), WINMIRA (2001 р.) та ін. У вільному доступі немає придатних для оцінювання якості тестів з вищої математики, які проводяться в КПІ.

Вітчизняні безкоштовні комп'ютерні програми для дослідження якості тестових завдань практично відсутні.

Метою статті є висвітлення функціональних можливостей авторської автоматизованої системи статистичного аналізу результатів комп'ютерного тестування з вищої математики.

Виклад основного матеріалу. Система автоматизованого аналізу тестових завдань забезпечує:

- 1) зчитування інформації з бази результатів тестування середовища MOODLE;
- 2) гнучкий вибір алгоритму математичної обробки результатів тестування;
- 3) наочність графічного представлення результатів;
- 4) формування та обслуговування бази каліброваних завдань.

Схема роботи системи представлена на рис 1.

В основу статистичного аналізу системи покладено як класичні статистичні, так і IRT-методи.

Класичні методи передбачають:

- первинний аналіз вибірок результатів тестування;
- розрахунки основних вибіркових статистичних характеристик;
- аналіз розподілу вибірок результатів тестування;
- порівняння вибірок результатів тестування;
- елементарний кореляційний і дисперсійний аналіз.



Рис. 1. Схема роботи системи автоматизованого аналізу тестових завдань

Приклад роботи системи по проведенню первинного аналізу результатів тестування наведено на рис. 2.

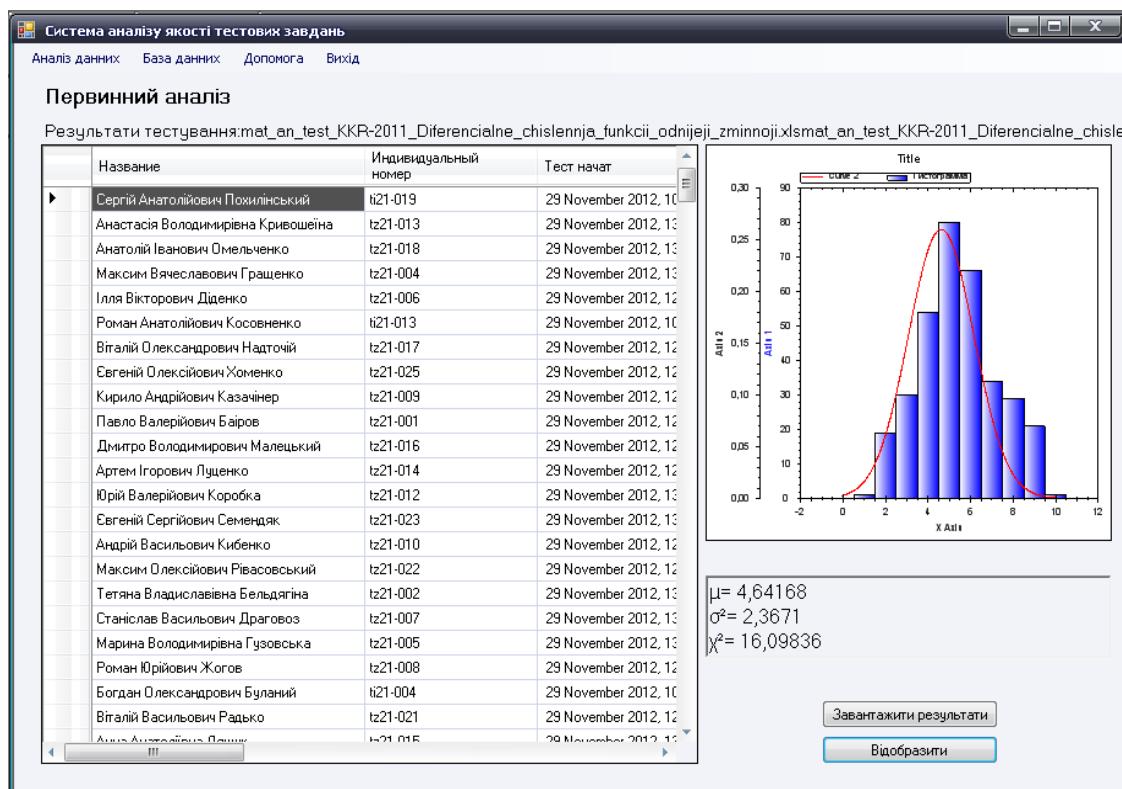


Рис. 2. Первинний аналіз вибірки

В основу IRT-моделей, реалізованих у системі, покладено ідею Г. Раша [2] впровадження наступних латентних параметрів:

- підготовленості іспитника θ_i , $i = \overline{1, N}$, де N – кількість іспитників;
- складності завдання тесту β_j , $j = \overline{1, K}$, де K - кількість завдань в тесті.

Імовірність правильної відповіді i -го іспитника на j -те завдання тесту визначається наступним чином:

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

Залежність імовірності від неперервного параметру θ при фіксованому значенні β_j називають характеристичною кривою j -го завдання тесту (рис. 3):

$$P_j(\theta) = \frac{1}{1 + \exp(-(\theta - \beta_j))}, j = \overline{1, K}.$$

Аналогічно визначають характеристичну криву i -го іспитника (рис. 4).

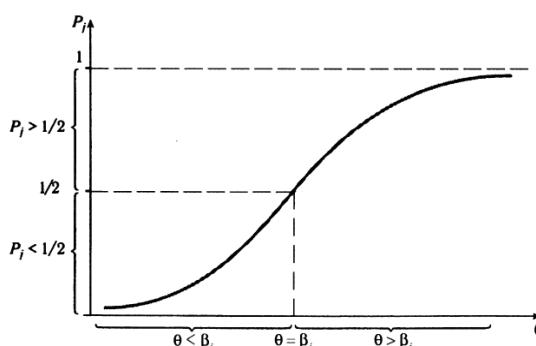


Рис.3. Характеристична крива завдання

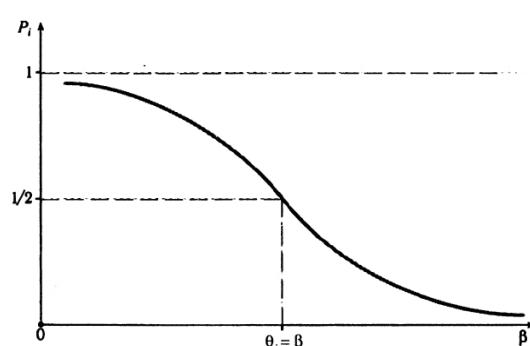


Рис. 4. Характеристична крива іспитника

Розвитком дихотомічної моделі Раша є двопараметрична модель Бірнбаума [1,2], політомічна модель Раша-Мастерса [1], модель завдань з множинним вибором Тіссена-Стейнберга [1], які обрані як основні математичні моделі тестових завдань. Згідно з цими моделями латентними параметрами тестового завдання є його складність, диференціюча спроможність, складність підрівній для політомічної моделі та параметри категорій відповідей для моделі з множинним вибором.

Двопараметрична модель Бірнбаума передбачає впровадження параметра диференціючої спроможності α_j , який показує, наскільки добре завдання може розрізняти іспитників з різним рівнем підготовленості. За цією моделлю ймовірність правильної відповіді i -ого іспитника на j -те завдання тесту визначається наступним чином:

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

Параметр диференціючої спроможності визначається за формулою:

$$\alpha_j = \frac{r_{bis}(j)}{\sqrt{1 - r_{bis}^2(j)}}, j = \overline{1, K},$$

де $r_{bis}(j)$ — коефіцієнт кореляції між балами, отриманими за кожне тестове завдання і балами за тест в цілому.

У політомічній моделі Раша-Мастерса припускають, що j -те завдання тесту має m_j підрівнів. Тоді ймовірність досягнення i -м іспитником g -го рівня j -ого завдання означається наступним чином:

$$P_{ijg} = \frac{e^{\sum_{l=0}^g (\theta_i - \beta_{jl})}}{\sum_{k=0}^{m_j} e^{\sum_{l=0}^k (\theta_i - \beta_{jl})}}, j = \overline{1, K}, i = \overline{1, N}, g = \overline{1, m_j},$$

де β_{jg} - складність переходу з $(g-1)$ -го рівня j -го завдання на g -ий,

θ_i – рівень підготовленості i -ого іспитника.

Відповідні залежності ймовірностей від рівня підготовленості іспитника зображені за допомогою характеристичних кривих підрівнів завдання. На рис. 5 представлена криві для завдань з трьома підрівнями.

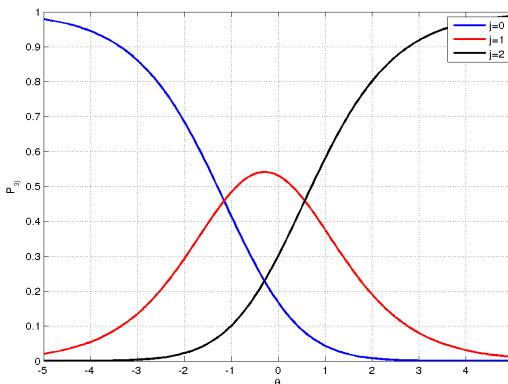


Рис.5. Характеристичні криві підрівнів завдання

Для аналізу якості завдань з множинним вибором обрано модель Тіссена-Стейнберга, в якій j -те завдання тесту має h_j варіантів відповіді. Ймовірність вибору i -м іспитником k -ого варіанту j -ого завдання задається наступним чином:

$$P(k) = \frac{e^{(a_{jk}\theta_i + c_{jk})} + d_{jk}e^{(a_{j0}\theta_i + c_{j0})}}{\sum_{s=0}^{h_j} e^{(a_{js}\theta_i + c_{js})}}, i = \overline{1, N}, j = \overline{1, K},$$

де a_{jk} – диференціюча спроможність k -ого варіанту відповіді j -ого завдання,

c_{jk} – складність k -ого варіанту відповіді j -ого завдання,

d_{jk} – відносна вага k -ого варіанту відповіді j -ого завдання,

θ_i – рівень підготовленості i -ого іспитника.

Відповідні залежності ймовірностей від рівня підготовленості іспитника зображені за допомогою характеристичних кривих варіантів відповіді (рис. 6).

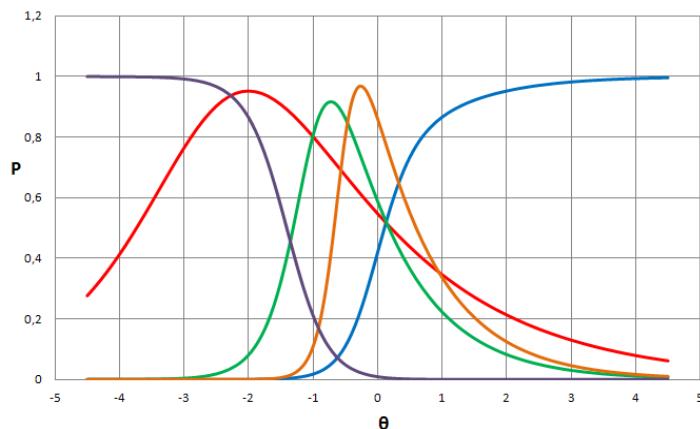


Рис.6. Характеристичні криві варіантів відповіді завдання

Процес оцінки значень латентних параметрів завдань полягає у використанні методу максимальної правдоподібності, на підставі якого будується система нелінійних рівнянь, та у розв'язанні цієї системи ітераційними методами.

Системою автоматизованого аналізу передбачено обробку таблиць відповідей і формування відповідної системи рівнянь та її розв'язання. Також передбачена перевірка побудованої моделі на адекватність.

На підставі розрахованих латентних параметрів будується різний ансамблі характеристичних кривих, що дозволяє достатньо наглядно інтерпретувати результати. На рис.7 зліва наведені характеристичні криві підрівнів політомічного завдання, а справа – характеристичні криві усіх завдань тесту.

Основним результатом роботи системи є формування бази каліброваних завдань, в якій зберігаються окремі тестові завдання, систематизовані за темами, призначенням (контрольна, іспит) та формою тестового завдання (True/False, Yes/No, Multiple Choice — Single Answer, Multiple Choice — Multiple Answer, Matching). Для кожного завдання існує відповідний паспорт, який містить основні характеристики тестового завдання, оцінені системою (рис. 8). Робота системи по обслуговуванню бази каліброваних тестових завдань передбачає:

- підбір та перегляд завдань за заданими параметрами;
- редактування та оновлення бази.

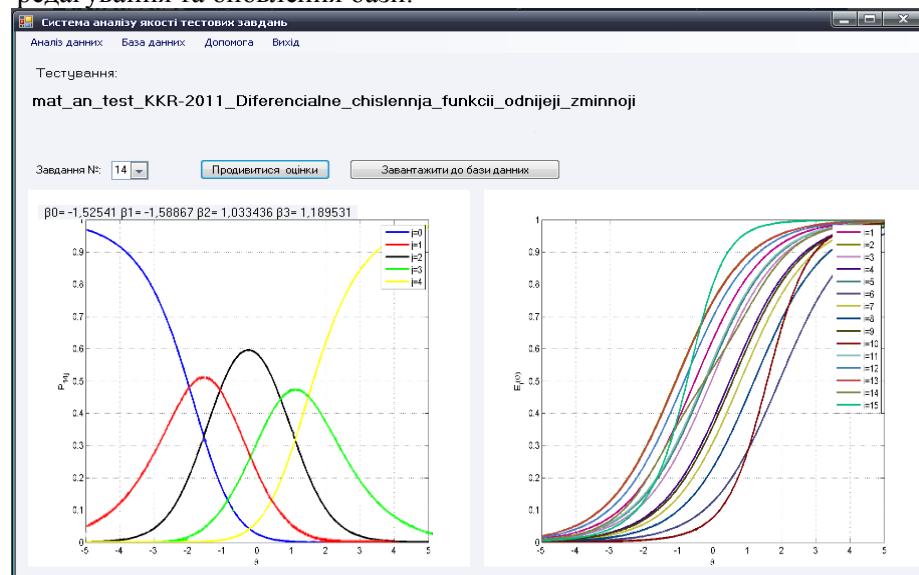


Рис.7. Ансамблі характеристичних кривих

Формування бази каліброваних тестових завдань дозволяє суттєво підвищити ефективність та гнучкість складання комп’ютерних контрольних та іспитів різних рівнів складності для студентів НТУУ «КПІ».

The screenshot shows a Windows application window titled 'Система аналізу якості тестових завдань'. The menu bar includes 'Аналіз даних', 'База даних', 'Допомога', and 'Вихід'. Below the menu is a toolbar with icons for back, forward, search, and other functions. The main area is titled 'Перегляд' (Preview) and displays a table of 8 rows. The columns are labeled: id, Номер завдання (Task Number), Тема тестування (Test Topic), Середнє квадратичне (Mean square), Індекс легкості (Difficulty Index), Коефіцієнт диференціації (Differential coefficient), and Латентні параметри (Latent parameters). The data in the table is as follows:

id	Номер завдання	Тема тестування	Середнє квадратичне	Індекс легкості	Коефіцієнт диференціації	Латентні параметри
2	1DC0Z02d	KKP-2011: Диференціальне числення функцій однієї змінної	0,48	0,64	0,49	-1,09466685
3	1DC0Z03d	KKP-2011: Диференціальне числення функцій однієї змінної	0,257	0,63	0,52	-0,05111996
4	1DC0Z04d	KKP-2011: Диференціальне числення функцій однієї змінної	0,492	0,61	0,74	0,489393008
5	1DC0Z05d	KKP-2011: Диференціальне числення функцій однієї змінної	0,412	0,79	0,79	-0,21452877
6	1DC0Z06d	KKP-2011: Диференціальне числення функцій однієї змінної	0,49	0,44	0,65	1,935749514
7	1DC0Z07d	KKP-2011: Диференціальне числення функцій однієї змінної	0,49	0,57	0,58	0,800779185
8	1DC0Z08d	KKP-2011: Диференціальне числення функцій однієї змінної	0,36	0,84	0,88	1,19973952

Рис.8. Калібрована база завдань

Програмно систему реалізовано у середовищі «Visual Studio C++» у вигляді окремої програми Windows. Система є легкою у користуванні та незалежною від мережевих ресурсів. Читування інформації з бази результатів тестування середовища MOODLE відбувається за допомогою імпорту Excel файлів та їх обробки за допомогою набору інтерфейсів OLE DB. Розв’язання системи та оцінювання латентних параметрів реалізовано у середовищі MATLAB і за допомогою набору динамічних бібліотек MATLAB Compiler Runtime інтегровано в середовище «Visual Studio C++». База каліброваних завдань зберігається у вигляді бази даних MySQL, завдяки чому забезпечується легкість підбору та сортування завдань.

Висновки. Пробне впровадження окремих компонентів автоматизованої системи статистичного аналізу результатів комп’ютерного тестування в дослідження аналізу якості тестів комплексу дистанційних курсів «Вища математика» підтвердило їхню дієздатність у проведенні такого аналізу. Об’єднання компонентів у єдину систему разом з додатковим підключенням до системи бази каліброваних завдань та реалізацією графічногоредставлення результатів тестування дозволило підвищити рівень аналізу тестів та зробити його істотно ефективнішим та якіснішим.

Список використаної літератури

1. Linden W. Handbook of Modern Item Response Theory / W. Linden, R.Hambleton. – NY: Springer-Verlag, 1997. – 510 p.
2. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests / G. Rasch. – Chicago: Univ. of Chicago Press, 1980. – 199 p.
3. Аванесов В.С. Item Response Theory: Основные понятия и положения. Статья первая / В.С. Аванесов // Педагогические Измерения. – 2007. - № 2. – С. 3-28.
4. Про розвиток та досвід експлуатації комплексу дистанційної освіти «Вища математика» / [І.В. Алексєєва, В.О. Гайдей, О.О. Диховичний та ін.] // Дидактика математики: проблеми і дослідження: міжнар. зб. наук. робіт. – 2009. – Вип. 31. – С. 49-56.
5. Карданова Е.Ю. О некоторых свойствах характеристической и информационной функций политомического тестового задания / Е.Ю. Карданова, Р. С. Карданов // Вестник Новгородского государственного университета. - 2010. – № 55. – С. 19–24.
6. Карпинский В.Б. Математическое и программное обеспечение оценки достоверности результатов массового тестирования: автореф. дис. на соиск. ученой степ. канд. техн. наук: спец. 05.13.18 / В.Б. Карпинский; Новгородский государственный университет имени Ярослава Мудрого. – Великий Новгород, 2009. – 21 с.
7. Крокер Л. Введение в классическую и современную теорию тестов / Л. Крокер, Дж. Алгина; под общей ред. В.И. Звонникова и М.Б. Чельышковой. – М.: Логос, 2012. – 668 с.
8. Лісова Т.В. Моделі та методи сучасної теорії тестів / Т.В. Лісова. – Ніжин: Видавець ПП Лисенко М.М., 2012. – 112 с.
9. Монахов В.В. Анализ результатов ЕГЭ по математике и физике и интернет-олимпиады по физике / В.В. Монахов // Компьютерные инструменты в образовании. - 2011. – №1. – С. 50 – 57.
10. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. – М.: Прометей, 2000. – 169 с.

11. Федорук П.І. Адаптивна система дистанційного навчання та контролю знань на базі інтелектуальних Інтернет-технологій / П.І. Федорук. – Івано-Франківськ: Видавничо-дизайнерський відділ ЦДТ Прикарпатського національного університету ім. Василя Стефаника, 2008. – 326 с
12. Чельшкова М.Б. Теория и практика конструирования педагогических тестов / М.Б. Чельшкова. – М.: Логос, 2002. – 431 с.

Стаття надійшла до редакції 30.09.2013

А. А. Дыховичный, А. Ф. Дудко

Национальный технический университет Украины «КПИ»

Автоматизированная система анализа результатов компьютерного тестирования по высшей математике

В статье сообщается о разработке в НТУУ «КПИ» автоматизированной системы статистического анализа результатов компьютерного тестирования по высшей математике. Система построена на базе как классических статистических методов, так и методов современной теории параметризации тестовых заданий, Item Response Theory (IRT). Среди известных моделей IRT предпочтение отдается моделям Раша, Бирнбаума, Раша-Мастерса и Тиссена-Стейнберга. Система имеет развитый графический интерфейс, который предоставляет объективную и наглядную визуализацию результатов тестирования, также в системе предусмотрена возможность формирования и обслуживания базы калиброванных задач.

Ключевые слова: тестирование по высшей математике, статистический анализ тестовых заданий, латентные параметры, IRT-модели.

O. Dykhovichnyi, A. Dudko

National Technical University of Ukraine "KPI"

Computer-based Analysis System of Results of Online Testing in Higher Mathematics

The paper informs about creation and development of computer-based statistical analysis system of results of online testing in NTUU "KPI". The system is based on both classic statistical methods and techniques of the Item Response Theory (IRT). Preference is given to the IRT, because it is a powerful tool for quality analysis of tests. It is applied in many widely known testing programs, such as the National Assessment of Educational Progress (NAEP, USA), the Scholastic Aptitude Tests (SAT, USA), the Graduate Record Examination (GRE, USA) and the Unified State Exam in Russia; and in large international assessment programs such as the Third International Math and Science Survey (TIMSS) and the Programme of International Student Assessment (PISA).

The main difference between IRT and classical methods of test analysis is an approach to the estimation of student ability and item difficulty. This approach is based on the introduction of two sets of latent parameters, namely set of person parameter and set of item parameters, that are related by certain probability functions and are determined based on test results.

Preferred models of IRT are models of Rush, Birnbaum, Rush-Masters and Thyssen-Steinberg. The models of Rush, Birnbaum are used for quality analysis of dichotomously-scored test items. The models of Andersen and Thyssen-Steinberg are used for quality analysis of polytomously-scored test items and multiple-choice items, respectively. Under these models the analogous probability depends on the following item parameters: the difficulty parameter, the discrimination parameter and the guessing parameter.

Empirical estimates of latent parameters are the maximum likelihood estimates and they are obtained as solutions of the system of nonlinear equations by iterative technique.

The procedure of evaluation of latent parameters requires complicated mathematical calculations. Because of this, it is necessary to create an appropriate computer program. Existing programs are not suitable for quality analysis of test conducted in NTUU "KPI".

In NTUU "KPI" online testing is conducted in the form of tests that cover the whole course of higher mathematics and involve all types of test items.

The above mentioned models are used for quality analysis of tests offered in the KPI. The validity of them is confirmed.

The computer-based statistical analysis system of results of online testing system has developed graphic interface that provides an objective graphic visualization of test results and the system provides the possibility of formation and maintenance of database of calibrated items.

Key words: *testing in higher mathematics, statistical analysis of tests, latent parameters, IRT-models.*