

УДК 378:[519.2+519.67] (Высшее образование. Университеты. Академическое обучение. Теория вероятностей и математическая статистика. Машинные, графические и другие методы вычислительной математики)

СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ ПО РАЗДЕЛУ «ЛИНЕЙНАЯ АЛГЕБРА И АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ» В СРЕДЕ MOODLE

© 2019 Л.В. Кайдалова, Ю.В. Гуменникова, Р.Х. Черницына

Кайдалова Людмила Витальевна, кандидат физико-математических наук, доцент, доцент кафедры «Прикладная математика, информатика и информационные системы». E-mail: ludmila.kaid@gmail.com

Гуменникова Юлия Валерьевна, кандидат физико-математических наук, доцент, доцент кафедры «Прикладная математика, информатика и информационные системы». E-mail: gumennikov@yandex.ru

Черницына Рузилья Нябиуллоевна, старший преподаватель кафедры «Прикладная математика, информатика и информационные системы». E-mail: y-abc@mail.ru

Самарский государственный университет путей сообщения

Статья поступила в редакцию 10.03.2019

В статье проводится математико-статистический анализ качества теста по разделу «Линейная алгебра и аналитическая геометрия» с помощью следующих характеристик: надежность, валидность, дискриминативность. Тестирования проводилось в среде дистанционного обучения (LMS) Moodle. Постановка задачи формулируется следующим образом: отбраковать задания, не удовлетворяющие указанным требованиям. Анализ валидности проводился на основании проверки вида распределения баллов. Проверка гипотезы о нормальном распределении тестовых баллов проводилась по критерию Пирсона. Надежность теста проверялась с помощью регрессионного анализа путем расщепления теста по четным и нечетным заданиям и определения коэффициента корреляции между этими группами. Далее коэффициент корреляции корректировался по формуле Спирмана-Брауна и проверялась значимость полученного коэффициента корреляции. *Ключевые слова* тестирование; Moodle; линейная алгебра и аналитическая геометрия; анализ тестовых заданий; надежность, валидность, дискриминативность.

В связи с постоянным сокращением часов для преподавания курса «Математика» необходимо активнее внедрять электронные системы управления обучением. В СамГУПС такой системой является система on-line обучения Moodle – программный комплекс, обеспечивающий организацию учебного процесса, посредством сети Internet. Система Moodle дает возможность моделировать этапы обучения и контроль знаний. Тестирование с использованием Moodle позволяет проводить контрольные замеры вне аудиторных занятий, так как обучающиеся проходят тестирование on-line в точке доступа Internet. В системе управления обучением Moodle имеется набор средств для автоматизации подсчетов статистических показателей качества тестовых заданий, дающих реальную оценку пригодности тестовых заданий с точки зрения измерения уровня подготовки обучаемых, т. е. процесс тестирования становится более технологичным.

Тестирование проводилось по курсу «Линейная алгебра и аналитическая геометрия» на первом курсе специальностей 24.03.05 «Эксплуатация железной дороги» и 23.05.06 «Строительство железных дорог, мостов и транспортных тоннелей».

В ходе обучения студенты имели доступ к примерам, аналогичным тестовым заданиям. Для максимального снижения вероятности угадывания верного ответа количество дистракторов (от англ. distract – отвлекать – неверный, но правдоподобный ответ с выбором одного или нескольких вариантов ответов) в тестовых заданиях заложено не менее пяти. Для снижения воздействия факторов, связанных с применением подсказок, списывания, сторонних лиц, использовалась функция смешивания вариантов ответов, т. е. варианты ответов не связывались с порядковым номером дистрактора. Средства анализа Moodle позволяют анализировать дистракторы. Если дистрактор используется редко (менее чем в 10 % случаев), то его заменяли на более подходящий ответ. Для ограничения вероятности списывания верные ответы показывались только после закрытия теста.

В данной работе проведен математико-статистический анализ качества тестовых заданий, реализованных в электронной образовательной системе Moodle, с помощью следующих характеристик: надежность, валидность, дискриминативность. Электронная система управления

обучением Moodle позволяет сохранять результаты тестирования в виде электронных таблиц в MS Excel и использовать все возможности данного пакета для расчета статистических числовых характеристик и построения графиков.

В тестировании принимали участие 216 студентов, результаты 11 человек были исключены из рассмотрения, так как эти студенты либо по техническим причинам выбыли из процесса, либо не справились с заданиями и не решили ни одной задачи. Такие работы исключались из выборки, так как информацией о качестве теста не обладают.

Статистическая обработка результатов таких испытаний позволяет получить числовые характеристики, дающие возможность объективно оценить как результаты испытуемых, так и качество отдельных тестовых заданий и теста в целом. При применении автоматизированных тестовых систем возникает возможность применения формализованных подходов, основанных на методологии теории педагогических измерений. Стати-

стический анализ качества отдельных вопросов и теста в целом осуществляли с использованием заложенного функционала электронной образовательной среды Moodle.

Данный тест относится к нормативно-ориентированным тестам. Для этого вида тестов при исследовании качества важной является следующая характеристика: вариация тестовых заданий p_j, q_j , где $p_j = m_j / n$ – доля правильных ответов на j -ое задание; $q_j = 1 - p_j$ – доля неправильных ответов на j -ое задание; p_j, q_j – дисперсия j -ого задания. Здесь m_j – количество правильных ответов на j -ое задание; n – общее количество испытуемых. Отсортируем номера заданий в порядке убывания количества правильных ответов и построим ее график (рис. 1). Эксперты считают, что величина вариативности должна примерно равняться 0,25 [1 – 5]. Менее всего удовлетворяют этим требованиям 1–5, 9, 13 и 23 задания. Поэтому рекомендуется их изменить.

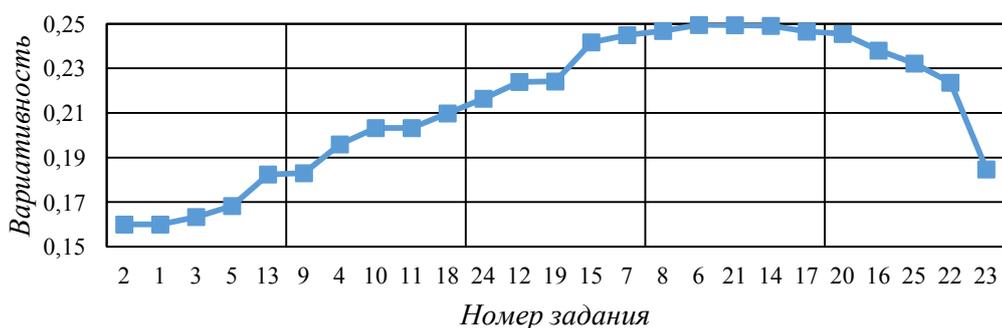


Рис. 1 Значения вариативности заданий теста (The values of variability of the test)

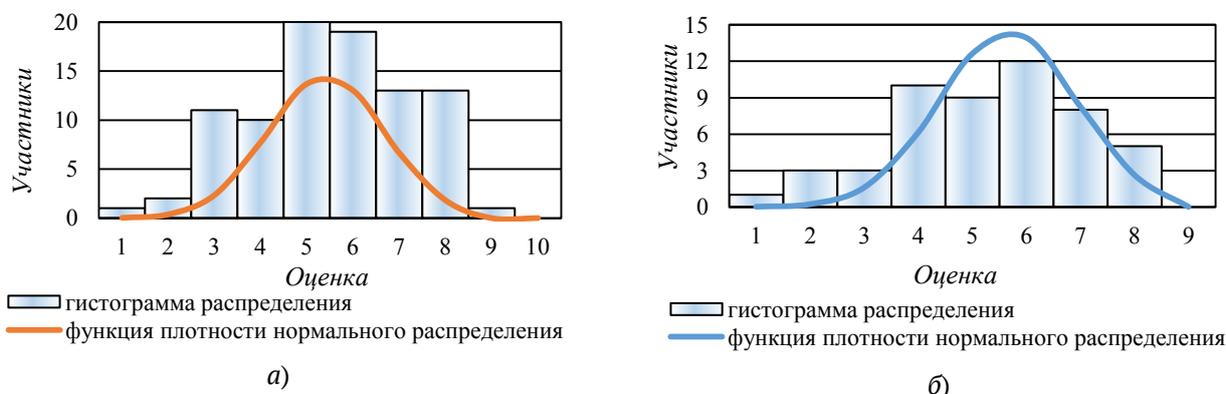


Рис. 2 Гистограмма и функция нормальной плотности распределения оценок при прохождении теста студентами специальностей СЖД (а) и ЭЖД (б) (The histogram and the density function of the normal distribution estimated for this test, the students of SDS (a) and Estonian railway (b))

Распределение оценок за тест по специальностям представлено на рис. 2. На представленных гистограммах по горизонтальной оси откладываются оценки в баллах, по вертикальной – количество студентов с указанной оценкой.

Согласно [5] тест считается *валидным*, если средний результат тестирования свойственен большей части тестируемых, а сами оценки распределяются по нормальному закону. Из рис. 2 можно сделать вывод о том, что распределение тестовых баллов близко к нормальному. Проверка гипотезы о нормальном распределении указанных баллов проводилась по критерию Пирсона. Статистический анализ результатов 205 тестируемых показал, что для рассматриваемой вы-

борки можно принять гипотезу о том, что оценки являются случайными величинами, распределенными по нормальному закону.

Вычислим оценки связи тестовых заданий как между собой, так и с суммой тестовых баллов испытуемых. Проведем оценку надежности теста с помощью коэффициента корреляции Спирмана-Брауна. Метод относится к классу методов раздельного коррелирования и является наиболее часто используемым методом этого класса [8]. Задания теста делятся на две части по четным и нечетным номерам, затем определяется коэффициент корреляции r_{xy} между этими двумя группами (X и Y) с последующим вычислением поправки по формуле Спирмана-Брауна [5, 8]:

$$r = \frac{2r_{xy}}{1 + r_{xy}}$$

Взяв в качестве представителей разрядов их середины, получим коэффициент надежности $r_{xy} = 0,77$, а с учетом поправки $r = 0,87$. Тогда согласно [5] надежность теста хорошая. Далее проверим значимость полученного коэффициента корреля-

ции при уровне значимости 0,05 и установим, что задания теста по четным и нечетным номерам взаимосвязаны. Для этого подсчитаем наблюдаемое значение статистического критерия по формуле

$$t_{набл} = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}},$$

которая имеет распределение Стьюдента с $k = n - 2$ степенями свободы и сравнивается с

критическим значением. Результаты расчетов показывают, что коэффициент r_{xy} значим.

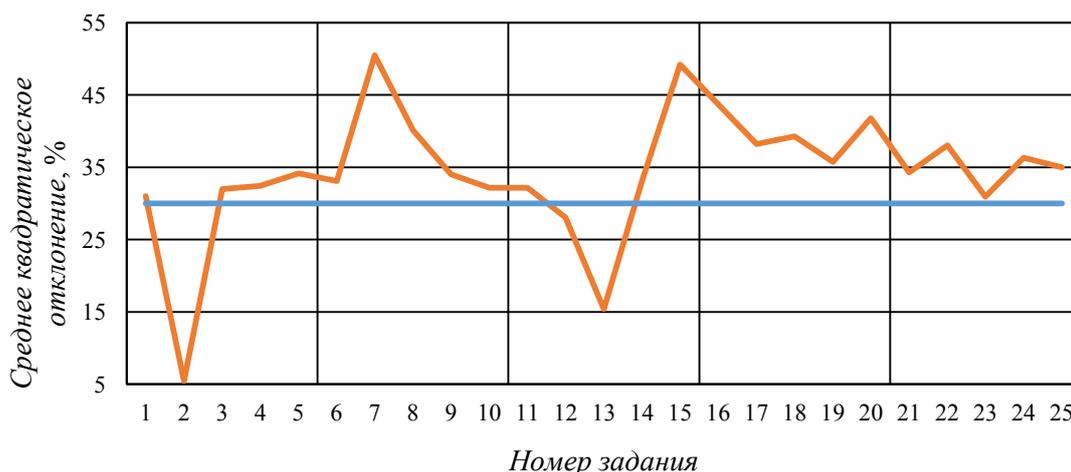


Рис. 3 Среднее квадратическое отклонение результатов выполнения заданий (The average square deviation of the results of the tasks)

На рис. 3 представлен график изменения средних квадратических отклонений итогов выполнения заданий. Среднее квадратическое отклонение σ характеризует разброс значений оценок, полученных за данное задание теста. Если для какого-то задания $\sigma = 0$, это означает, что все проверяе-

мые получили за этот вопрос одинаковую оценку. Такое задание следует признать не удовлетворяющий задачам тестирования. В литературе отмечается [9], что в соответствии с требованиями педагогической теории измерений, задания со значением среднего квадратического отклонения

менее 0,3 лучше исключать из теста, так как они не обладают в достаточной мере дифференцирующей способностью, т. е. не способны разделить сильных и слабых студентов. Такими заданиями являются 2 и 13. Для большинства тестовых вопросов оно имеет значение выше 0,3, что соответствует требованиям педагогической теории измерений [5, 6, 9] и является довольно хорошим индикатором их дифференцирующей способности.

Оптимальным считается такое среднее квадратическое отклонение результатов выполнения заданий σ , при котором значение среднего \bar{x} равно утроенному значению стандартного отклонения $\bar{x} = 3\sigma$. В данном тесте по модулю «Линейная алгебра и аналитическая геометрия» опти-

мальное значение среднего квадратического отклонения пока не получено, отношение \bar{x} к σ составляет для специальности СЖД 2,61, а для специальности ЭЖД – 2,7.

Дифференцирующая способность задания (эффективность дискриминации) рассчитывается как разность двух показателей: сложности задания для сильной группы с высокой продуктивностью тестируемых и сложности задания для группы с низкой продуктивностью тестируемых. Чем выше дифференцирующая способность задания, тем лучше оно разделяет испытуемых по уровню знаний.

Значения индекса эффективности дискриминации для исследуемой выборки заданий теста приведены на рис. 4.

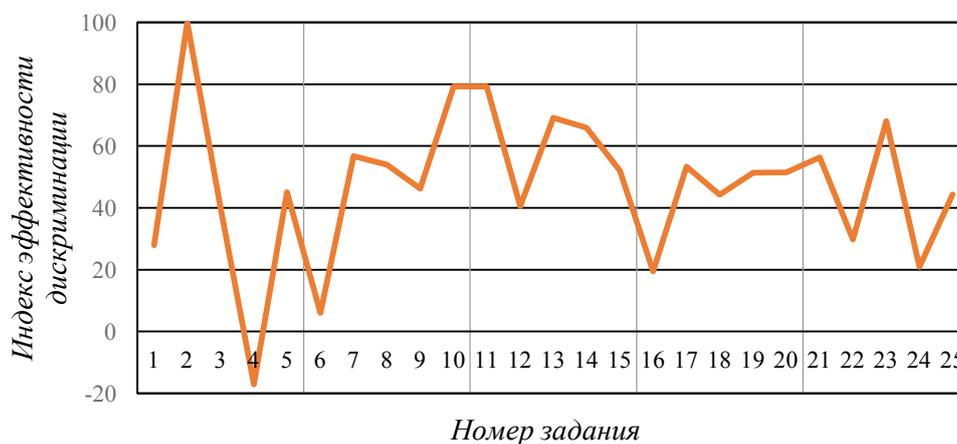


Рис. 4 Значения индекса эффективности дискриминации заданий теста (Values of the index of efficiency of discrimination of tasks of the test)

Выводы. 1. Разработка тестов по разделам математики требует длительного подготовительного этапа, охватывающего все стадии создания теста. Применение электронные системы управления обучением Moodle позволяет стандартизировать процедуру апробации теста, выявления и корректировки заданий, способствует повышению точности и созданию качественных тестов.

2. Анализ результатов тестирования двух специальностей показал, что выборка подчинена нормальному закону распределения статистиче-

ских данных, поэтому разработанные тесты можно использовать в качестве инструмента измерения знаний.

3. Надежность тестов (76 %) попадает в допустимую полосу значений, но может быть увеличена при дальнейшей целенаправленной работе над тестовыми заданиями.

4. Необходим дальнейший анализ характеристик используемых тестовых заданий, а также составление новых заданий, направленных на повышение валидности теста.

1. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. М., «Логос», 2002. 431 с.
2. Ronald K.H. Jones Russel W. Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development // Educational Measurement: Issues and Practice. 1993. P. 38 – 47.
3. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., Прогресс, 1976. 496 с.
4. Майоров А.Н. Теория и практика создания тестов для системы образования. М., Народное образование, 2000. 352 с.
5. Ким В.С. Тестирование учебных достижений. Уссурийск, Изд-во УГПИ, 2007. 214 с.

6. Олейник Н.М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: уч. пособ. Донецк, Донецкий госуд. ун-т. [http:// opentest.com.ua/test-kak-instrument-izmereniya-urovnnya-znanij/](http://opentest.com.ua/test-kak-instrument-izmereniya-urovnnya-znanij/)
7. Аванесов В.С. Содержание тестов и тестовых заданий // Педагогические Измерения, 2007. № 3. <http://testolog.narod.ru/Theory63.html>
8. Карпенко А.П., Домников А.С., Белоус В.В. Тестовый метод контроля качества обучения и критерии качества образовательных тестов. Обзор // Наука и образование (МГТУ им. Н.Э. Баумана) (электронный журнал). 2011. № 4. С. 1 (28 с.).
9. Толстобров А.П., Коржик И.А. Возможности анализа и повышения качества тестовых заданий при использовании сетевой системы управления обучением MOODLE // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2008. № 2. С. 100 – 106.

STATISTICAL ANALYSIS OF TESTING RESULTS BY SECTION «LINEAR ALGEBRA AND ANALYTICAL GEOMETRY» IN THE MOODLE ENVIRONMENT

© 2019 L.V. Kaidalova, Yu.V. Gumennikova, R.N. Chernytsyna

Ljudmila V. Kaidalova, Candidate of Physical and Mathematical Sciences, Associate Professor, Associate Professor of the Department «Applied mathematics, computer science and information systems». E-mail: ludmila.kaid@gmail.com

Julia V. Gumennikova, Candidate of Physical and Mathematical Sciences, Associate Professor, Associate Professor of the Department «Applied mathematics, computer science and information systems». E-mail: gumennikov@yandex.ru

Ruzilya N. Chernytsyna senior lecturer of the department «Applied mathematics, computer science and information systems». E-mail: y-abc@mail.ru

Samara state University of railway engineering. Samara, Russia

The article provides a mathematical and statistical analysis of the quality of the test in the section «Linear Algebra, Analytical Geometry» with the help of the following characteristics: reliability, validity, diskriminativnost. Testing was carried out in the environment of distance learning (LMS) Moodle. The problem statement: as a result, tasks that did not meet the specified requirements were rejected. The analysis of validity was carried out on the basis of checking the type of distribution of points. Testing the hypothesis of the normal distribution of test scores was carried out by Pearson's criterion. The reliability of the test was checked by regression analysis by splitting the test for even and odd jobs and determining the correlation coefficient between these groups. Further, the correlation coefficient was corrected by the Spearman-Brown formula and the significance of the obtained correlation coefficient was checked.

Key words: testing; Moodle; Linear Algebra, Analytical Geometry, analysis of test tasks; reliability, validity, diskriminativnost.

1. Chely`shkova M.B. Teoriya i praktika konstruirovaniya pedagogicheskix testov (Theory and practice of construction of pedagogical tests). М., «Logos», 2002. 431 s.
2. Ronald K.H. Jones Russel W. Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. Educational Measurement: Issues and Practice. 1993. P. 38 – 47.
3. Glass Dzh., Ste`nli Dzh. Statisticheskie metody` v pedagogike i psixologii (Statistical methods in pedagogy and psychology). М., Progress, 1976. 496 s.
4. Majorov A.N. Teoriya i praktika sozdaniya testov dlya sistemy` obrazovaniya (Theory and practice of creating tests for the education system). М., Narodnoe obrazovanie, 2000. 352 s.
5. Kim V.S. Testirovanie uchebny`x dostizhenij (Testing of educational achievements). Ussurijsk, Izd-vo UGPI, 2007. 214 s.
6. Olejnik N.M. Test kak instrument izmereniya urovnya znaniy i trudnosti zadaniy v sovremennoj texnologii obucheniya (Test as a tool to measure the level of knowledge and difficulty of tasks in modern learning technology): uch. пособ. Doneczk, Doneczkij gosud. un-t. <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnnya-znanij/>
7. Avanesov V.S. Soderzhanie testov i testovy`x zadaniy (Content of tests and test tasks). *Pedagogicheskie izmereniya*, 2007. № 3. <http://testolog.narod.ru/Theory63.html>
8. Karpenko A.P., Domnikov A.S., Belous V.V. Testovy`j metod kontrolya kachestva obucheniya i kriterii kachestva obrazovatel`ny`x testov. Obzor (Test method of training quality control and quality criteria of educational tests. Review). *Nauka i obrazovanie* (MGTU im. N.E`. Baumana) (e`lektronny`j zhurnal). 2011. № 4. S. 1 (28 s.).
9. Tolstobrov A.P., Korzhik I.A. Vozmozhnosti analiza i povы`sheniya kachestva testovy`x zadaniy pri ispol`zovanii setевой sistemy` upravleniya obucheniem MOODLE (Possibilities of analysis and improvement of quality of test tasks at use of network control system of training of MOODLE). *Vestnik Voronezhskogo gosudarstvennogo universiteta*. Seriya: Sistemy`j analiz i informacionny`e texnologii. 2008. № 2. S. 100 – 106.