

СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ ГРУПП ИСПЫТУЕМЫХ

О. В. Паркалова

Анализ результатов тестирования – это важная составляющая процесса контроля знаний. Статистическая обработка материалов по результатам тестирования решает две основные задачи: объективно оценивает результаты испытуемых и позволяет сделать вывод о качестве тестов. Поэтому важно иметь инструментарий, позволяющий оценить качество тестовых заданий и сделать вывод о целесообразности использования каждого отдельного задания.

Характеристики, вычисляемые по результатам выполнения тестов испытуемыми можно разделить на две категории: характеристики всего теста и характеристики отдельных вопросов.

Существует различный инструментарий, позволяющий вычислять эти характеристики. Это можно сделать в каком-либо математическом пакете, в табличном процессоре или с использованием специальных программных средств. Система управления обучением Moodle, например, имеет широкий набор средств по оценке результатов тестирования.

Удобным решением представляется включение модуля анализа результатов тестирования в систему компьютерного тестирования. Структурная схема модуля представлена на рис. 1.

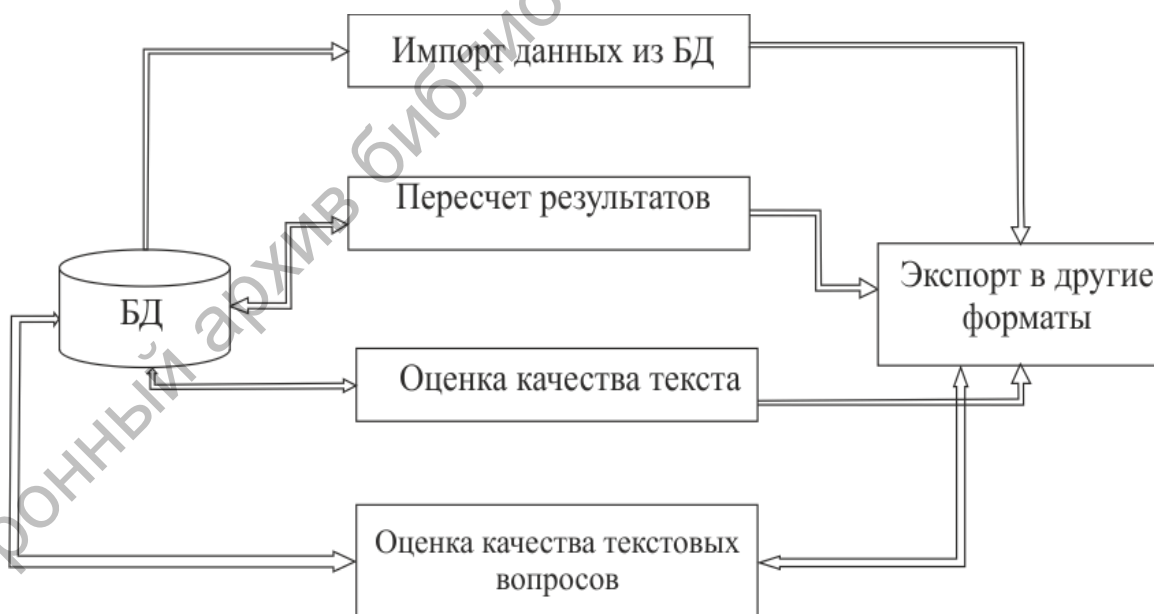


Рис 1. Структурная схема

Рассмотрим элементы схемы «Оценка качества теста» и «Оценка качества тестовых заданий». Они соответствуют двум категориям характеристик оценки результатов тестирования, упомянутым выше.

Для анализа взяты результаты 8 учебных тестов. По 4 теста проводились среди двух групп испытуемых. Условно обозначим тесты T1_1, T2_1, T3_1, T4_1, T1_2, T2_2, T3_2, T4_2, где идентичными являются пары тестов T1_1 и T1_2, T2_1 и T2_2, T3_1 и T3_2, T4_1 и T4_2.

В таблице представлены значения характеристик, с помощью которых можно оценить качество теста в целом.

Тестовые баллы обычно группируются близко к некоторым значениям – мерам центральной тенденции. Это характеристики мода, медиана и среднее выборочное (математическое ожидание) [1].

В таблице видно, что мода, то есть значение тестового балла, которое встречается в выборке испытуемых наиболее часто, смещена в сторону больших баллов. Эта признак того, что тесты не соответствуют уровню знаний студентов, обладают пониженной трудностью.

Для двух групп испытуемых наблюдается одинаковая картина изменения среднего выборочного от теста к тесту: увеличение среднего балла во втором тесте, а затем уменьшение в третьем, и еще большее снижение в четвертом тесте. Это может говорить с одной стороны, об уменьшении интереса к дисциплине и соответственно снижению успеваемости студентов, а с другой стороны, о недочетах в методике преподавания дисциплины или качестве тестовых заданий.

Дисперсия и стандартное отклонение позволяют провести оценку нормальности распределения и сделать вывод о том, что распределения тестовых баллов в первом приближении можно считать близкими к нормальным.

Таблица

Статистические характеристики

| | T1_1 | T2_1 | T3_1 | T4_1 | T1_2 | T2_2 | T3_2 | T4_2 |
|------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| число участников | 76 | 78 | 74 | 50 | 82 | 87 | 85 | 61 |
| мода | 70 | 70 | 55 | 50 | 70 | 60 | 60 | 70 |
| медиана | 70 | 70 | 60 | 50 | 60 | 60 | 60 | 50 |
| среднее выборочное | 61,45 | 68,08 | 60,81 | 49,00 | 54,51 | 61,72 | 59,65 | 52,95 |
| дисперсия | 431,21 | 215,73 | 300,70 | 250,00 | 392,21 | 220,00 | 393,44 | 387,68 |
| стандартное отклонение | 20,77 | 14,69 | 17,34 | 15,81 | 19,80 | 14,83 | 19,84 | 19,69 |
| асимметрия | -0,53 | -1,22 | -0,06 | -0,05 | -0,34 | -0,46 | -0,26 | -0,17 |
| эксцесс | -0,36 | 2,68 | -0,12 | -0,43 | -0,48 | 0,74 | -0,40 | -0,50 |

Значение асимметрии во всех тестах отрицательное, но в третьем и четвертом тесте для каждой группы испытуемых происходит наибольшее движение распределения в сторону нормального.

По эксцессу наблюдаемых распределений видно, что во втором тесте каждого курса эксцесс положительный. А в остальных случаях он имеет отрицательные значения. Это означает, что на тестировании недостаточное количество испытуемых получило баллы, близкие к среднему значению баллов.

Для оценки качества отдельных тестовых заданий нужно работать с матрицей результатов каждого вопроса теста для каждого испытуемого.

Рассмотрим некоторые характеристики тестовых заданий, которые используются для корректировки заданий или их исключения из теста.

Индекс легкости – это процент испытуемых, которые ответили на вопрос теста правильно. Если каждый из испытуемых ответил на задание правильно или неправильно, то такое задание не является тестовым, так как не обладает дифференцирующей способностью.

Стандартное отклонение характеризует разброс значений оценок испытуемых при ответе на конкретный вопрос теста. Если значение равно нулю, то задание не является пригодным, так как все получили за него одинаковую оценку.

Предполагаемый вес – вес, который преподаватель назначил тестовому заданию при формировании сценария теста.

Коэффициент дифференциации – это характеристика, позволяющая связать результаты за конкретное задание с результатами за весь тест. Положительное значение коэффициента показывает, что хорошо подготовленные испытуемые отвечают на конкретное задание лучше, чем слабо подготовленные. Вопросы с отрицательными коэффициентами должны исключаться из списка вопросов, так как они понижают точность результата тестирования [2].

Список характеристик для оценки качества тестовых заданий не ограничивается только перечисленными выше параметрами.

В классической и современной теории тестов существует большое количество характеристик, позволяющих проводить анализ тестовых заданий и тестов в целом. Это нужно для совершенствования процесса тестирования, улучшения качества отдельных тестовых заданий. Кроме того, на основании результатов тестирования можно вносить поправки в организацию процесса обучения.

Выводы по каждой из характеристик позволяют контролировать процесс создания тестов и не пропустить необходимость совершенствования и доработки заданий и всего теста в целом.

Принципы теории тестов можно использовать в готовом виде с помощью уже существующих средств для анализа тестовых заданий. Но для получения более широкой функциональности целесообразной может быть разработка собственной системы по анализу результатов тестирования. Это наиболее актуально в случае, когда в учебном процессе используется оригинальная система компьютерного тестирования.

Литература

1. <http://clipperkim.narod.ru/test/monotest/index.html>
2. http://www.institutemvd.by/images/materials/umo/seminars/2013_06_03/metod_recom.pdf