

СТАТИСТИЧЕСКИЙ АНАЛИЗ ВАРИАНТОВ ИНДИВИДУАЛЬНЫХ КОНТРОЛЬНЫХ ЗАДАНИЙ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Костенко К.А.

науч. рук. к.т.н. Ю.Я. Кацман

Томский политехнический университет, Институт кибернетики

kak31@tpu.ru

Введение

На сегодняшний день оценка качества контрольно-измерительных материалов, используемых для тестирования знаний и умений студентов, является достаточно трудоемкой и сложной задачей, которая актуальна как для новых дисциплин, так и для тех, обучение по которым производится в течение ряда лет. Контрольно-измерительные материалы обычно представлены в нескольких вариантах, вследствие чего возникает проблема их параллельности, что может усложнить оценку знаний студентов и ее объективность.

Обычно при анализе качества контролируемых материалов большое внимание уделяется обеспечению параллельности вариантов задания [1, 2]. При этом, если применение современной теории тестов – Item Response Theory (IRT) [3] для оценки латентных факторов требует обеспечить для одного теста минимальную выборку от 200 до 1000 наблюдений, то классическая статистическая теория позволит получить оценки параметров, ограничиваясь значительно меньшим количеством опытов.

Цель работы

Главной целью данной работы являлся статистический анализ параллельности вариантов индивидуальных контрольных заданий по теории вероятностей для оценки качества знаний, усвоенных студентами.

Постановка задачи

По результатам проведенной контрольной работы (тестирования) минимальная оценка (3 балла) давалась за попытку решить хотя бы одну задачу, максимальная оценка (15 баллов) – за правильное решение трех задач. Все результаты контрольной работы по теории вероятностей были обработаны в лицензионной программе Statistica. Кроме того, требовалось определить равносильность вариантов с помощью использования тестов Крускала–Уоллиса, Шеффе и медианного теста.

Проведение анализа

Перед анализом параллельности вариантов был использован модуль описательных статистик и исключены нерепрезентативные варианты (4 и менее наблюдений).

Далее предполагалось, что эти варианты параллельны (равносильны), и тогда оценки студентов должны быть адекватны их знаниям, а не сложности билетов. С этой целью для каждого варианта были рассчитаны точечные и

интервальные оценки, что с учетом случайных факторов предполагало приблизительное равенство средних баллов и дисперсий для каждого варианта. Реальные оценки для каждого варианта представлены на рис. 1 в виде диаграмм рассеяния.

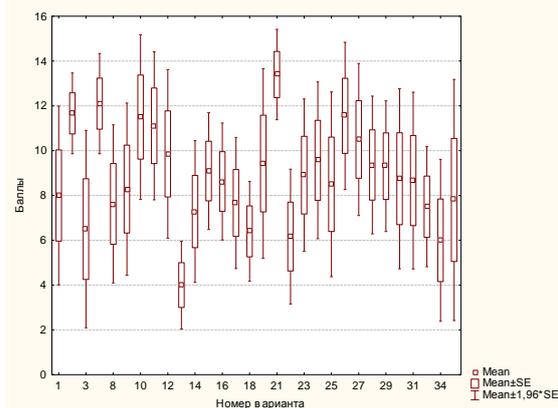


Рис. 1. Диаграммы рассеяния для различных вариантов контрольной работы

Представленные результаты наглядно свидетельствовали о неодинаковой сложности (не параллельности) различных вариантов контрольной работы.

Разбиение вариантов на группы сложности проводилось с помощью метода кластерного анализа k-средних (k-means) и только по одной переменной – баллам, поэтому для обеспечения примерного равенства количества наблюдений в каждой группе и однородности наблюдений внутри группы все варианты сортировались по среднему баллу:

- со сложными заданиями, средний балл которых составил менее 7,8;
- с заданиями средней сложности, средний балл которых находился в пределах от 7,8 до 9,3;
- с простыми заданиями, средний балл которых превысил 9,3.

Модуль описательных статистик способствовал получению точечных оценок для всех наблюдений и каждого кластера отдельно. Благодаря этому модулю были получены следующие результаты:

- максимальное различие в оценках для первого и третьего кластеров составило менее трех баллов;
- практически все точечные характеристики для второго кластера и всей совокупности наблюдений совпадают;

– для всех наблюдений второго кластера 50 % полученных результатов превысили 8,6 балла, в то же время для первого кластера 50 % результатов не превысили 6,6 баллов, а для третьего кластера 50 % оценок превысили 10,5 баллов;

– дисперсии для всех наблюдений и кластеров считались практически равными (отношение дисперсий менее 2);

– анализ коэффициентов скоса и эксцесса свидетельствовал, что распределения баллов в каждой группе несимметричны и существенно отличаются от гауссова распределения.

– Для проверки гипотезы о незначительном влиянии фактора был проведен однофакторный ранговый анализ в пакете Statistica с установленным уровнем значимости $\alpha=0,05$. Анализ суммы рангов по группам (кластерам) полученным в результате проведения теста Крускала–Уоллиса подтвердил, что максимальная оценка наблюдалась в кластере с вариантами, имевшими легкие задания, а минимальная – в кластере с вариантами, имевшими сложные задания.

Анализ результатов еще одного вида рангового тестирования – медианного теста, представленного в виде таблицы, показал, что:

– верхняя половина содержала максимальное значение кластера, которое соответствовало задачам с высоким уровнем сложности и полученными по ним минимальными оценками;

– нижняя половина включала максимальное значение кластера, которое соответствовало задачам с малым уровнем сложности и полученными по ним максимальными оценками.

Гипотеза о влиянии фактора, проверенная при проведении медианного теста, как и при анализе критерия Крускала–Уоллиса, показала, что влияние фактора существенное.

Далее, с помощью анализа критерия Манна – Уитни была проверена гипотеза на однородность двух разных выборок (кластеров).

Оценки эффектов обработки – рис. 2.

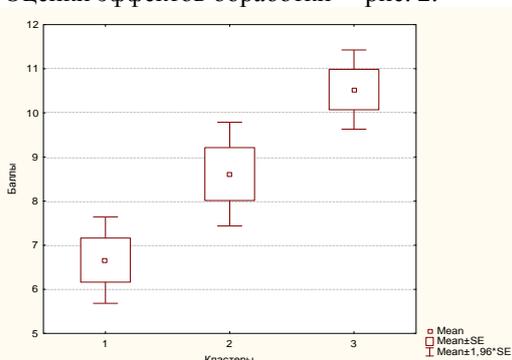


Рис. 2. Диаграммы рассеяния для всех кластеров

Приведенные на рис. 2 результаты свидетельствуют о существенном различии точечных и интервальных характеристик для различных групп. Анализируя их, можно было

попытаться ответить на вопрос: какие пары групп вариантов заданий можно считать значимо различными? Для ответа на него было проведено сравнение средних по методу Шеффе для различных пар уровней факторов, анализ которого показал значительное различие средних баллов для различных пар кластеров, что доказывает справедливость альтернативной гипотезы о значительном влиянии фактора.

Выводы

Статистический анализ проведенного мониторинга индивидуальных заданий по теории вероятностей доказал, что варианты тестовых заданий не параллельны. На основании проведенных исследований был сделан окончательный вывод о качестве предлагаемых тестовых заданий: из 39 имевшихся вариантов индивидуальных заданий 4 - были исключены в связи с нерепрезентативной выборкой, 9 вариантов содержали задачи сложного уровня с минимальными полученными по ним оценками, 9 вариантов – с задачами среднего уровня и 12 вариантов с задачами легкого уровня.

Проведенные исследования показали, что первый и второй кластеры можно расширить за счет добавления новых вариантов, исключив из третьего кластера варианты с двумя или тремя легкими заданиями, учитывая средний набранный балл. И при этом заменить их более сложными заданиями из первого и второго кластеров, так же учитывая набранный средний балл.

Заключение

Мониторинг качества преподавания дисциплин в большой степени определяется качеством методических индивидуальных контролирующих материалов (тестов). Одной из важнейших характеристик вариантов тестов является их параллельность. В данной работе показано, что даже для заданий, используемых на протяжении ряда лет, задача обеспечения параллельности является актуальной. Предложенные в работе статистические методы позволяют решить эту задачу, что продемонстрировано на примере контрольных заданий по теории вероятностей.

Используемые источники

1. Suen H.K., Lei P.W. Classical versus Generalizability theory of measurement. Available at: <http://suen.educ.psu.edu/~hsuen/pubs/Gtheory.pdf> (дата обращения 12.10.2015).

2. Пюкхин В.В., Пермяков О.Е. Проблемы обеспечения качества приема и направления совершенствования системы конкурсного отбора поступающих в ВУЗы Российской Федерации [Quality assurance issues and directions in perfecting the system of competitive selection of applicants in the Russian Federation] – Томский Политехнический университет, 2007, 310 с.

3. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960. 216 p.