

ПРИМЕНЕНИЕ КЛАСТЕРНОГО И ДИСПЕРСИОННОГО АНАЛИЗА ДЛЯ ОЦЕНКИ КАЧЕСТВА ОБУЧЕНИЯ ПО ДИСКРЕТНОЙ МАТЕМАТИКЕ

Чернета Д.С.

Научный руководитель к.т.н.: КацманЮ.Я.

Томский политехнический университет, Институт кибернетики
e-mail:dimano1993@rambler.ru

Введение

Оценка равнозначности контрольно-измерительных материалов, используемых для оценивания знаний студентов, является актуальной задачей как для новых дисциплин, так и для тех, обучение по которым производится в течение ряда лет. Контрольно-измерительные материалы как правило представлены в нескольких вариантах, из-за чего возникает проблема их равнозначной сложности, что может усложнить оценку знаний студентов и ее объективность [1].

Обычно при анализе качества контролирующих материалов большое внимание уделяется обеспечению параллельности вариантов задания [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 позволяет получить оценки параметров, ограничиваясь значительно меньшим количеством опытов.

Целью работы является изучение, анализ и исследование результатов экзамена по дискретной математике, используя кластерный и дисперсионный анализ.

Постановка задачи

По результатам проведенного экзамена можно было получить 40 баллов за выполнение 6 заданий. За первые 2 задание максимальный балл составляет 6 баллов, за оставшиеся 4 задания максимальный балл составляет 7 баллов. Всего экзамен был пройден 71 раз. Следовательно первоначальный объём выборки составляет $6 \cdot 71 = 426$ значений.

Проведение анализа

Для того, чтобы можно было сравнивать результаты, баллы были переведены в процентный вид, в результате чего за каждое задание стало начисляться от 0 до 100%.

Далее, из всех вариантов заданий были найдены одинаковые задания, которые находились в разных вариантах. Затем было найдено количество баллов, которые получили студенты за каждое задание. Для дальнейшего анализа были оставлены только те задания, на которые имелось не менее 4х ответов. После этого шага объём выборки составил 281 значение.

Предполагалось, что сложности вариантов равнозначны, в этом случае оценки студентов

будут адекватны их знаниям, а не уровню сложности билетов. Для этого для каждого варианта задания 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.

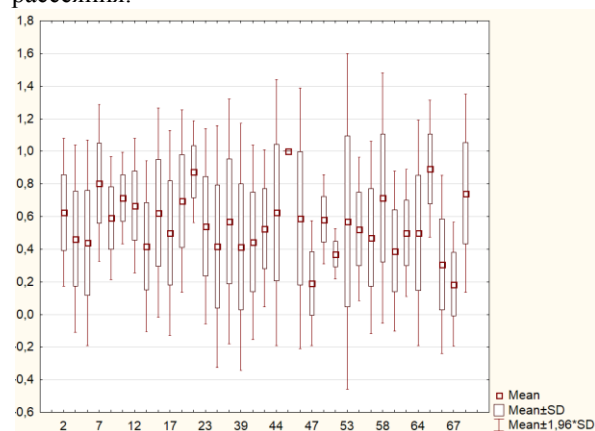


Рис. 1. Диаграммы рассеяния для различных вариантов заданий

Результаты наглядно свидетельствуют о неодинаковой сложности разных вариантов.

Далее было решено разделить задания на 3 кластера, при помощи метода к-средних, чтобы определить каждое задание по уровню сложности, как лёгкой, средней, либо высокой сложности. Для этого было найдено среднее арифметическое каждого вопроса и далее была произведена кластеризация.

После чего было выяснено, что кластер 1 состоит из вопросов средней сложности, в него попали вопросы: 2, 3, 8, 12, 14, 17, 23, 38, 43, 44, 46, 50, 53, 55, 56, 60, 64 вопросы.

Кластер 2 состоит из вопросов лёгкой сложности, в него попали вопросы: 7, 9, 18, 22, 45, 58, 65, 87 вопросы.

Кластер 3 состоит из вопросов высокой сложности, в него попали вопросы: 4, 13, 24, 39, 40, 47, 52, 59, 66, 67.

Далее было необходимо проверить 3 кластера на неоднородность, то есть принадлежат ли они к одному и тому же распределению или к разным. Для этого сформулируем нулевую гипотезу H_0 - исходные выборки однородны, соответственно гипотеза H_1 – выборки не однородны.

Для определения, принадлежат ли баллы из трёх кластеров одному и тому же распределению

(Табл. 1) был использован ранговый критерий Краскела-Уоллиса.

Таблица 1. Результаты критерия Краскела-Уоллиса.

Kruskal-Wallis ANOVA by Ranks; Ball (Voprosi bolshe 4 NEW) Independent (grouping) variable: Cluster Kruskal-Wallis test: H (2, N=278)=52.37230 p =,0000				
Depend.: Ball	Code	Valid N	Sum of Ranks	Mean Rank
1	1	147	21091.00	143.4762
2	2	50	9945.00	198.9000
3	3	81	7745.00	95.6173

Нулевую гипотезу можно принять с вероятностью $p = 0.0000$. Поскольку заданный уровень значимости много больше $\alpha = 0.05$, то нулевую гипотезу следует отвергнуть в пользу альтернативной гипотезы H_1 – влияние фактора существенно, следовательно, баллы из трёх кластеров не принадлежат одному и тому же распределению.

Далее будем использовать критерий Манна – Уитни (Табл. 2, Табл. 3, Табл. 4).

Таблица 2. Результаты критерия Манна – Уитни для 1 и 2 кластеров.

Mann-Whitney U Test (Voprosi bolshe 4 NEW) By variable Cluster. Marked tests are significant at p<,05000					
Variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value
Ball	12962	6541	2084	-4.567	0.000005
Variable	Valid N Group 1	Valid N Group 2	p-value		Z adjusted
Ball	147	50	0.000004		4.60304

Таблица 3. Результаты критерия Манна – Уитни для 1 и 3 кластеров.

Mann-Whitney U Test (Voprosi bolshe 4 NEW) By variable Cluster. Marked tests are significant at p<,05000					
Variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value
Ball	19007	7099	3778	4.563	0.000005
Variable	Valid N Group 1	Valid N Group 2	p-value		Z adjusted
Ball	147	81	0.000004		4.5872

Таблица 4. Результаты критерия Манна – Уитни для 2 и 3 кластеров.

Mann-Whitney U Test (Voprosi bolshe 4 NEW) By variable Cluster. Marked tests are significant at p<,05000					
Variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-value
Ball	4679	3967	646	6.531	0.000000
Variable	Valid N Group 1	Valid N Group 2	p-value		Z adjusted
Ball	50	81	0.000000		4.5817

Во всех трёх тестах значение $p\text{-value} < 0.05$, значит баллы из трёх кластеров не принадлежат одному распределению и верна гипотеза H_1 .

Так как предварительный ранговый однофакторный анализ подтвердил гипотезу о

значимом влиянии фактора, попробуем оценить это влияние количественно в рамках дисперсионного анализа (Табл. 5).

Таблица 5. Результаты дисперсионного анализа

Variable	Analysis of Variance (Voprosi bolshe 4 NEW) Marked effects are significant at p<,05000			
	SS Effect	df Effect	MS Effect	SS Error
Ball	5.1478	2	2.5739	22.9265
Variable	df Error	MS Error	F	P
Ball	275	0.0833	30.874	0.000000

Статистика Фишера $F=30.87406$ незначимо отличается от единицы с вероятностью $p=0.000000$, что значительно меньше уровня значимости. Следовательно, нулевую гипотезу следует отвергнуть в пользу альтернативной гипотезы – влияние фактора существенно.

Выводы

В ходе выполнения работы из исходных данных, а именно из экзаменационных баллов студентов по дискретной математике, была получена информация для исследований, а именно таблица баллов студентов за каждый вопрос.

Далее, при использовании кластерного анализа, вопросы были разделены на 3 кластера, по уровню сложности, и к ним были применены критерии Краскела-Уоллиса, Манна – Уитни и однофакторный дисперсионный анализ которые показали, что 3 кластера неоднородны и принадлежат различным распределениям.

Заключение

Одним из главных параметров вариантов экзамена является их равнозначность. В проделанной работе показано, что даже для заданных, используемых в течение нескольких лет, задача обеспечения равнозначности вариантов является актуальной. Используемые в работе статистические методы позволяют решить эту задачу, что показано на примере контрольных заданий по дискретной математике.

Список использованных источников

1. Кацман Ю.Я. Статистический анализ индивидуальных заданий по теории вероятностей. // Известия Томского политехнического университета. – 2014. – Т. 325. – № 5. – С. 84-90.
2. 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 10.10.2016).
3. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960. 216 p.