

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭМПИРИЧЕСКИХ ДАННЫХ РЕЗУЛЬТАТОВ ПРОВЕДЕНИЯ МАССОВЫХ ПРОЦЕДУР ОЦЕНКИ КАЧЕСТВА ОБРАЗОВАНИЯ В ТОМСКОЙ ОБЛАСТИ

К.Р. Балыкбаева, А.А. Зобнина
Томский политехнический университет
aaz47@tpu.ru

Введение

В науке эмпирические данные требуются для того, чтобы гипотеза получила признание научного сообщества. После сбора эмпирических данных начинается этап статистической обработки, которая проводится, как правило, с помощью специального программного обеспечения.

Система анализа правильности составления педагогических тестов представляет собой совокупность отдельных компонент. На данный момент существуют программные средства, которые решают только некоторые отдельные задачи из всей системы [3]. В результате возникла необходимость наличия средства для представления общей картины результатов обработки эмпирических данных.

Поэтому целью нашей работы стала разработка средства автоматизации имеющихся данных результатов проведения массовых процедур оценки качества образования и представления результатов для дальнейшего анализа разработчикам педагогических тестов.

Описание алгоритма обработки

Для обработки эмпирических данных был создан набор макросов в среде Microsoft Excel на языке Visual Basic for Applications (VBA). Среда и язык разработки были выбраны исходя из того, что с линейкой продуктов Microsoft Office работают даже неопытные пользователи. Таким образом, разработанное нами средство обработки данных эксперты могут сразу использовать без необходимости длительного освоения.

В качестве исходных данных для обработки нами использовались матрицы результатов тестирования пятых и десятых классов по математике и русскому языку, проведенных на базе общеобразовательных организаций Томской области в 2016 году. Обработываемые нами матрицы результатов содержат порядка 20-30 заданий и от 5 тыс. до 8,5 тыс. результатов испытуемых, в зависимости от теста.

Работа с данными осуществляется не на всей матрице тестовых результатов, из диапазона обработки устраняются все строки и столбцы, состоящие только из нулей и только из единиц. Причиной этому является непригодность как теста в целом (излишняя сложность или легкость, не позволяющая определить истинные знания испытуемого), так и отдельных заданий теста соответственно.

Эмпирические результаты тестирования представляются в виде полигонов частотных

распределений сгруппированных и не сгруппированных данных. Частоты необходимы для сокращения списка учеников с полученными тестовыми баллами. Для большой группы – скажем, в 100 или более учеников – используется сгруппированное частотное распределение. Количество групп вычисляется по правилу Стёрджеса.

Данные о результатах участников являются дихотомическими величинами. Между всеми парами заданий теста был вычислен коэффициент корреляции (коэффициент «фи»), результаты сведены в матрицу коэффициентов корреляции. Определенные (выходящие за границы допустимого) значения коэффициента корреляции отмечаются цветом, чтобы визуально выделить задания, содержания которых в меньшей степени связаны с содержаниями других заданий теста.

Для определения уровня дифференцирующей способности задания теста вычисляется точечно-бисериальный коэффициент. Задания теста, по которым значения вычисленного коэффициента близки к нулю, имеют низкую дифференцирующую способность. Такие задания требуют доработки. Все задания теста, где точечно-бисериальный коэффициент меньше нуля, необходимо удалить из теста [1].

Тестирование алгоритма обработки

Для оценки алгоритма обработки эмпирических данных было проведено тестирование его работы на различных входных данных. Для демонстрации работы средства обработки рассмотрена матрица результатов для четырех вариантов теста по математике учащихся пятого класса.

Графическая интерпретация эмпирических данных представлена в виде следующих полигонов частотного распределения:

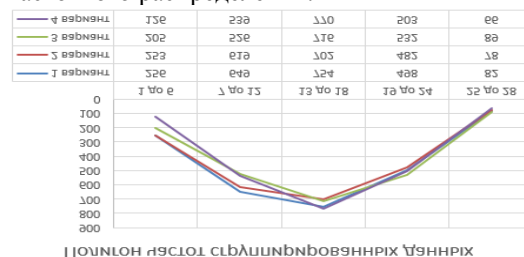


Рис. 1. Полигон частотного распределения сгруппированных данных

Полигон частот несгруппированных данных

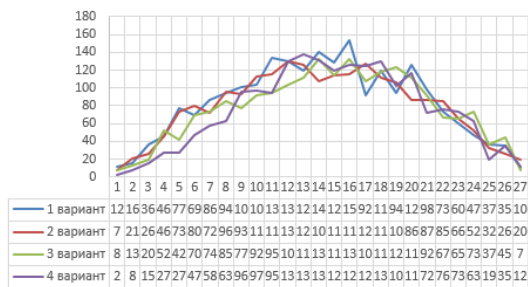


Рис. 2. Полигон частотного распределения не сгруппированных данных

Из графиков видно, что в среднем учащиеся набирают от 13 до 18 баллов из 28 (максимальный балл данного теста). Отчетливо видно сходство в результатах тестирования у всех четырех вариантов теста, имеющих почти совпадающие полигоны распределения частот.

Между каждыми заданиями теста была построена матрица коэффициентов корреляции.

1 вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1,000	0,083	0,102	0,066	0,143	0,109	0,129	0,111	0,114	0,014	0,149	0,125	0,156	0,078	0,056	0,091
2	0,083	1,000	0,130	0,117	0,201	0,144	0,183	0,213	0,202	0,045	0,154	0,134	0,180	0,147	0,104	0,194
3	0,102	0,130	1,000	0,125	0,200	0,146	0,170	0,196	0,220	-0,002	0,179	0,142	0,153	0,155	0,096	0,154
4	0,066	0,117	0,125	1,000	0,098	0,135	0,090	0,134	0,144	0,006	0,132	0,122	0,103	0,120	0,088	0,105
5	0,143	0,201	0,200	0,098	1,000	0,186	0,221	0,197	0,215	0,019	0,179	0,212	0,217	0,188	0,092	0,205
6	0,109	0,144	0,146	0,135	0,186	1,000	0,168	0,182	0,161	-0,003	0,161	0,148	0,191	0,149	0,078	0,162
7	0,129	0,183	0,170	0,090	0,221	0,168	1,000	0,224	0,209	0,010	0,176	0,163	0,185	0,132	0,088	0,210
8	0,111	0,213	0,196	0,134	0,197	0,182	0,224	1,000	0,285	0,027	0,208	0,149	0,235	0,160	0,112	0,296
9	0,114	0,202	0,220	0,144	0,215	0,161	0,209	0,285	1,000	0,025	0,237	0,174	0,274	0,185	0,076	0,304
10	0,014	0,045	-0,002	0,006	0,019	-0,004	0,010	0,027	0,019	1,000	-0,006	0,012	0,051	0,037	0,015	0,017
11	0,149	0,154	0,179	0,132	0,179	0,161	0,176	0,208	0,217	-0,005	1,000	0,188	0,265	0,158	0,117	0,211
12	0,125	0,134	0,142	0,122	0,212	0,148	0,163	0,149	0,174	0,012	0,188	1,000	0,252	0,168	0,071	0,150
13	0,156	0,180	0,153	0,103	0,217	0,191	0,185	0,235	0,274	0,051	0,265	0,252	1,000	0,278	0,102	0,261
14	0,078	0,147	0,155	0,120	0,188	0,149	0,132	0,160	0,185	0,037	0,158	0,168	0,278	1,000	0,126	0,207
15	0,056	0,104	0,096	0,088	0,092	0,078	0,088	0,112	0,076	0,015	0,117	0,071	0,102	0,126	1,000	0,149
16	0,091	0,194	0,154	0,105	0,205	0,162	0,210	0,296	0,304	0,017	0,211	0,150	0,261	0,207	0,149	1,000

Рис. 3. Матрица коэффициентов корреляции

Из данной матрицы видно, что отрицательные значения коэффициента (выделено курсивом и рамкой) в десятом задании теста указывают на определенный просчет разработчиков в содержании задания. Наиболее распространенной причиной является отсутствие предметной чистоты содержания. Жирным шрифтом и рамкой в матрице выделяется высокий положительный коэффициент (больше 0,3), это говорит об одинаковом содержательном элементе заданий теста.

По значениям точечно-бисериальных коэффициентов определяются наиболее неудачные задания теста.

1 вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Номер задания	18	16	9	19	5	13	20	8	2	17	11	14	7	3	12	6
Коэффициент	0,599	0,595	0,574	0,539	0,515	0,509	0,495	0,494	0,492	0,460	0,444	0,421	0,416	0,402	0,377	0,354
2 вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Номер задания	9	18	13	16	19	10	2	5	17	20	14	11	8	3	12	6
Коэффициент	0,635	0,594	0,564	0,534	0,525	0,525	0,519	0,474	0,471	0,425	0,425	0,424	0,423	0,412	0,393	0,368
3 вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Номер задания	16	9	18	13	2	19	3	20	8	5	17	12	7	14	11	15
Коэффициент	0,611	0,601	0,598	0,556	0,535	0,521	0,490	0,483	0,476	0,475	0,470	0,453	0,426	0,414	0,399	0,376
4 вариант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Номер задания	18	9	16	13	19	8	5	20	17	11	14	3	6	2	15	7
Коэффициент	0,613	0,600	0,577	0,564	0,530	0,490	0,488	0,469	0,455	0,416	0,398	0,390	0,373	0,371	0,354	0,330

Рис. 4. Значения коэффициентов

Среди полученных коэффициентов нет отрицательных значений, значит, нет заданий, подлежащих выводу из теста.

В целом задание можно считать валидным, когда значение точечно-бисериального коэффициента около 0,5 [1]. Оценка валидности задания позволяет судить о том, насколько задание пригодно для работы в соответствии с общей целью создания теста. Так как целью исследуемого теста является дифференциация учеников по уровню подготовки, то валидные задания должны

четко отделять хорошо подготовленных от слабо подготовленных учеников тестируемой группы.

Заключение

Разработано средство, автоматически обрабатывающее эмпирические данные результатов тестирования учащихся. Представленный алгоритм обработки распространяется на любые выборки испытуемых и применим к тестам любой длины.

Полученные результаты могут быть использованы как специалистами по разработке педагогических тестов, так и учителями школ, преподавателями высших и других профессиональных учебных заведений, заинтересованных в овладении методами анализа тестов, правильном их составлении и использовании в учебном процессе для контроля и оценки уровня подготовки обучаемых.

Список используемых источников

1. Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432с.
2. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно-исследовательских работах: Учебное пособие. – М.: Экспресс, 2008. – 60с.
3. The automation of the procedures of educational progress assessment as a method of education quality improving [Electronic resources] / A. V. Lepustin [et al.] // The 9th International Forum on Strategic Technology (IFOST-2014), September 21-23, 2014, Cox's Bazar, Bangladesh : [proceedings]. — [S. l.]: IEEE, 2014. — [P. 109-112]. — Title screen. — Доступ по договору с организацией-держателем ресурса.
4. Обработка эмпирических данных исследований. [Электронный ресурс]. – URL: <http://mt-9.ru/wp-content/uploads/2015/09/Обработка-эмпирических-данных-исследования.doc> (дата обращения 19.03.2017).