

## ОБРАБОТКА И АНАЛИЗ РЕЗУЛЬТАТОВ КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ

Немцева С.К.

БНТУ, г. Минск, Республика Беларусь, anlana19@mail.ru

Благодаря развитию тестологии и информационных технологий сегодня не только автоматизирован процесс разработки и проведения тестирования, но и появилось множество методик обработки результатов тестирования с применением современных статистических методов и программных продуктов. Item Response Theory (IRT) – английское название теории, которая достаточно широко используется как в психологических, так и педагогических измерениях. Существует несколько переводов этого термина: «теория латентных черт», «теория тестовых заданий», «теория пунктуальных ответов», сегодня используют более общее понятие – «современная теория тестов». Условно процесс шкалирования в IRT можно подразделить на три этапа. Первый предполагает построение шкалы логитов для латентного параметра подготовленности испытуемых, второй — шкалы логитов для оценок латентного параметра трудности заданий. Третий этап позволяет свести две шкалы в общую шкалу стандартных оценок для обоих латентных параметров.

Для обработки результатов тестирования целесообразно воспользоваться табличным процессором MS Excel. Для проверки тестовых свойств заданий на листе книги составляется бинарная матрица результатов тестирования.

Матрица тестовых результатов представляет результаты испытуемых по всем проверяемым заданиям. В нашем случае тестировалось 8 студентов по десяти заданиям – матрица  $X_{8 \times 10}$ . В случае правильного ответа студенту выставляется в матрице 1, в случае неправильного ответа – 0. Необходимо упорядочить бинарную матрицу, сначала подсчитывается количество набранных баллов и количество правильных ответов.

Таблица 1 – Матрица тестовых результатов

Номер компьютера	Номера заданий										Набрано баллов
	1	2	3	4	5	6	7	8	9	10	
1	0	1	1	0	1	1	0	0	1	0	5
2	1	0	0	1	1	0	1	0	0	0	4
4	0	0	0	0	1	1	0	1	0	0	3
5	1	1	1	0	1	1	1	1	1	0	8
3	1	1	0	1	0	1	0	0	1	1	6
6	1	0	1	0	1	1	1	0	1	1	7
7	0	0	1	1	1	0	1	1	0	1	6
8	1	1	1	1	0	1	1	1	1	1	9
К-во правильных ответов	5	4	5	4	6	6	5	4	5	4	48

Один из способов определения меры трудности заданий – эмпирический, путем опробования задания, с подсчетом доли неправильных ответов по каждому из них. Эмпирически трудность заданий определяется сложением элементов матрицы по столбцам, что указывает на число правильных ответов, полученных по каждому заданию ( $R_j$ ). Для проверки качества тестовых заданий в матрице проводятся два упорядочения: сортировка тестируемых по количеству набранных баллов, начиная с самого успешного студента. Другое упорядочение

проведено для тестовых заданий: на первом месте стоит самое легкое задание, по которому имеется наибольшее число правильных ответов, далее – сортировка по убыванию.

Для получения сопоставимых характеристик и определения показателя трудности заданий используют долю правильных ( $p_j$ ) и неправильных ответов ( $q_j$ ), которые вычисляются по формуле:

$$q_j = \frac{W_j}{N},$$

где  $W_j$  – число неправильных ответов,

$N$  – число тестируемых.

Целесообразно определить средний арифметический тестовый балл в данной группе испытуемых

$$M = 47 / 8 = 5,875$$

В современных технологиях адаптивного обучения и контроля используется показатель меры трудности задания – логит трудности задания, который равен натуральному логарифму отношения доли неправильных ответов на это задание к доле правильных ответов –  $\ln q_j/p_j$ . Симметрично введена и логарифмическая оценка уровня знаний, так называемый логит уровня знаний, равный  $\ln p_i/q_i$ , где  $p_i$  – доля правильных ответов студента, рассчитываемая по формуле:

$$p_i = Y_i / k,$$

где  $Y_i$  – число правильных ответов испытуемого  $i$ ,

$k$  – общее число заданий.

Таблица 2 – Результаты расчета логитов

Номер компьютера	Номера заданий										Набрано баллов (X)	Логит уровня знаний
	5	6	1	3	7	9	2	4	8	10		
4	1	1	0	0	0	0	0	0	1	0	3	-0,847
2	1	0	1	0	1	0	0	1	0	0	4	-0,405
1	1	1	0	1	0	1	1	0	0	0	5	0,000
3	0	1	1	0	0	1	1	1	0	0	5	0,000
7	1	0	0	1	1	0	0	1	1	1	6	0,405
6	1	1	1	1	1	1	0	0	0	1	7	0,847
5	1	1	1	1	1	1	1	0	1	0	8	1,386
8	0	1	1	1	1	1	1	1	1	1	9	2,197
К-во правильных ответов (Y)	6	6	5	5	5	5	4	4	4	3	47	
Логит трудности задания	-1,099	-1,099	-0,511	-0,511	-0,511	-0,511	0,000	0,000	0,000	0,511		

Уровень трудности задания и уровень подготовленности тестируемых – латентные (невяные) показатели, которые невозможно измерить в процессе тестирования. Для оценки этих параметров необходимо использовать тесно связанные с ними индикаторы. При тестировании знаний студентов в качестве индикатора выступают сами тестовые задания. Проблема состоит в преобразовании значения индикаторов в значения латентных параметров. Существуют различные подходы к решению этой задачи, один из которых – однопараметрическая модель Г. Раша. Георг Раш разместил на одной шкале и уровень знаний тестируемого и уровень трудности задания, введя для них общую единицу измерения – логит. [] На гистограмме (рисунок 1) по оси абсцисс откладываются значения латентных переменных (в логитах)

уровня подготовленности студентов и уровня трудности тестовых заданий, т. е. на одной и той же шкале откладываются значения двух латентных переменных. В этом состоит особенность использования модели Раша. По оси ординат с левой стороны указывается число испытуемых (вверху) и число заданий (внизу). В идеальном случае гистограмма распределения трудностей заданий должна быть близка к равномерному закону распределения. Это означает, что представленный набор тестовых заданий с одинаковой точностью позволяет оценить уровень подготовленности студентов на всем диапазоне изменения трудностей тестовых заданий.

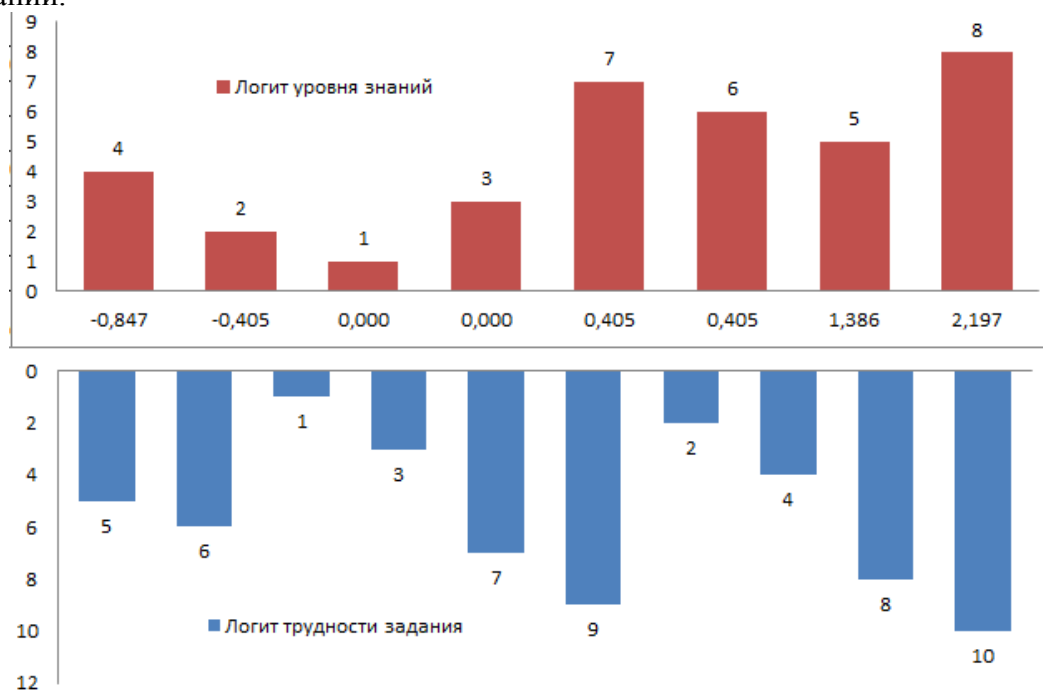


Рисунок 1 – Соответствие между уровнем подготовленности студентов и трудностью тестовых заданий

Анализируя полученную гистограмму, делаем вывод, что трудность тестовых заданий не полностью соответствует нормальному закону распределения. Значения трудностей заданий заполняют весь диапазон их изменений. Средняя трудность заданий не должна отличаться от среднего уровня подготовленности студентов более чем на 0,5 логит, в нашем случае средний уровень подготовленности студентов на 0,821 логита выше среднего уровня трудности заданий: трудность теста соответствует уровню подготовленности студентов, можно даже усложнить тестовые задания. Диапазон изменения уровня подготовленности студентов составил: от -0,847 до 2,197 логит, длина всего интервала составила 3,045 логит, что подтверждает дифференцирующую способность теста. Уровень трудности заданий изменяется от -1,099 до 0,511, что свидетельствует о необходимости добавить как более легкие, так и задания потруднее.

Задание в тестовой форме нельзя называть тестовым, если оно не коррелирует с суммой баллов по всему тесту. Коррелируемость задания с критерием ( $r_{xy}$ ) представляет собой более точную и технологичную меру дифференцирующей способности задания. Коррелируемость проверяется посредством расчета коэффициента корреляции  $r_{jy}$ , где символом  $r$  обозначается так называемый классический коэффициент корреляции Пирсона, или один из его вариантов.

Для расчета  $r_{xy}$  формируется два вектор-столбца, один из которых – задание ( $X_j$ ), другой – критерий ( $Y$ ). Между значениями этих двух векторов и устанавливается мера связи, если таковая существует.

При проверке тестовых заданий в качестве критерия, для начала, используется сумма баллов испытуемых, полученная по всем заданиям пробного варианта теста. Символ  $j$  пред-

ставляет номер коррелируемого задания, а символ  $Y$  – числовой вектор-столбец тестовых баллов испытуемых.

Чем выше значения  $r$ , тем больше вероятность превращения задания в тестовой форме в тестовое задание, то есть быть включенным в тест. Особенно заметно эта вероятность повышается при  $r > 0,4$ .

Если взять значение  $r^2 * 100\%$ , то получим значение так называемого коэффициента детерминации, выраженного в удобной для интерпретации процентной мере связи задания с суммой баллов. Результаты расчета в MS Excel с помощью встроенной функции PEARSON() представлены в таблице 3. Полученные результаты по третьему заданию можно интерпретировать так: 58% вариации суммы тестовых баллов испытуемых по всем заданиям связано с вариацией баллов по одному только третьему заданию, что указывает на очень высокий потенциальный вклад этого задания в общую дисперсию теста.

Нулевая корреляция свидетельствует об отсутствии у задания системных свойств, присущих тесту. Такие задания, равно как и задания с отрицательными значениями  $r_{xy}$  устраниваются из тестовых материалов, как не выдержавшие эмпирической проверки.

Таблица 3 – Результат расчета коэффициента корреляции

Номер компьютера	Номера заданий										Набрано баллов (X)	Ранг
	5	6	1	3	7	9	2	4	8	10		
4	1	1	0	0	0	0	0	0	1	0	3	1
2	1	0	1	0	1	0	0	1	0	0	4	2
1	1	1	0	1	0	1	1	0	0	0	5	3
3	0	1	1	0	0	1	1	1	0	0	5	3
7	1	0	0	1	1	0	0	1	1	1	6	5
6	1	1	1	1	1	1	0	0	0	1	7	6
5	1	1	1	1	1	1	1	0	1	0	8	7
8	0	1	1	1	1	1	1	1	1	1	9	8
К-во правильных ответов (Y)	6	6	5	5	5	5	4	4	4	3	47	
Коэффициент корреляции Пирсона	-0,34	0,27	0,49	0,76	0,63	0,63	0,46	0,07	0,33	0,59		
Коэффициент детерминации	12%	7%	24%	58%	40%	40%	21%	0%	11%	35%		

Из теста не исключаются тестовые задания, поскольку нет заданий, на которые ответили все студенты и не ответил ни один тестируемый, но необходимо переформулировать пятое и четвертое задания, так как они не выдержали эмпирической проверки. Кроме того, обработка и анализ результатов тестирования показали, что требуется дополнительная проработка учебного материала дисциплины.

Помимо отмеченных, в практике тестирования чаще других используются такие методы определения надежности тестов, как:

- повторное тестирование испытуемых, в одинаковых условиях, одним и тем же тестом, с последующим коррелированием результатов. Получаемое значение коэффициента корреляции указывает на надежность измерения в смысле устойчивости результатов испытуемых.

- использование параллельных тестов, с последующим коррелированием результатов. Получаемое при этом значение интерпретируется в смысле воспроизводимости, близости или параллельности тестовых результатов.

#### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Елисеев, И.Н. Методология оценки уровня компетенций студента // Информатика и образование. – 2012. – № 4. – С.80-85.
2. Ким, В.С. Тестирование учебных достижений. Монография / В.С. Ким - Уссурийск: Издательство УГПИ, 2007. – 214 с.: ил.
3. Майоров, А.Н. Теория и практика создания тестов для системы образования / А.Н. Майоров – М. – 2014.
4. Попов, Д.И., Попова, Е.Д. Экспертиза качества тестовых заданий. Режим доступа: <http://www.hi-edu.ru/e-books/xbook688/01/part-008.htm> – Дата доступа: 05.11.2016