

Асаналиев Б.О.

ШКАЛИРОВАНИЯ РЕЗУЛЬТАТОВ С ПРИМЕНЕНИЕМ УРОВНЯ ТРУДНОСТИ ЗАДАНИЙ

Во всех известных теориях тестирование рассматривается как процесс противоборства испытуемого с предлагаемым ему заданием. Если испытуемого обозначить символом i , а задание – символом j , то исход противоборства оценивается баллом x_{ij} . Обычно рассматриваются два исхода: победа или поражение. Если испытуемый справляется с заданием, ему за победу дается один балл. В таких случаях пишут $x_{ij}=1$. Если не справляется, то дается ноль баллов, соответственно $x_{ij}=0$. Для такого сопоставления тестовый балл т.е. число правильно выполненных заданий при дихотомической оценке (оценка 0 или 1) переводится в производный показатель с помощью процедуры, называемой шкалированием. Чтобы отличить необработанные тестовые баллы от производных показателей, необработанные баллы нередко называются сырыми или наблюдаемыми.

Таблица №1

Правило дихотомического оценивания теста

Ответ x_{ij}	Оценка
Правильный ответ i -го ученика на j -е задание	1
Неправильный ответ i -го ученика на j -е задание	0

Тогда, если обозначить как i -испытуемый, j -задание, то можно представить следующий макет матрицы наблюдаемых результатов выполнения теста.

Таблица №2

Макет матрицы результатов

Испытуемые	Задания $1, \dots, j, \dots, n$	Индивидуальный балл X_i
$1, \dots$ i, \dots N	$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}$	$X_i = \sum_{j=1}^n x_{ij}$
Количество правильных ответов на задания (R)	$R_j = \sum_{i=1}^N x_{ij}$	$\sum_{i=1}^N X_i = \sum_{j=1}^n R_j$

Общий вид матрицы наблюдаемых результатов выполнения N учащимися n заданий приведен в табл. №2.

Справа в матрице, в вертикальном столбце, содержатся индивидуальные баллы учеников X_i ($i=1, 2, \dots, N$), которые получаются суммированием единиц по горизонтали в каждом профиле ответов учащегося. Сложение единиц в столбцах по профилям ответов на n заданий теста позволяет получить числа R_j ($j=1, 2, \dots, n$), соответствующие количеству правильных ответов на задания теста.

После этого пояснения из дидактических соображений для иллюстрации методов обработки выбрана упорядоченная небольшая матрица, когда 13 учеников отвечали всего на 10 заданий теста (табл.3). Однако, все формулы и подсчеты могут быть распространены на любые выборки испытуемых и применимы к тестам любой длины.[2, стр. 128,129]

В начале определим меры трудности заданий. В данном случае определим эмпирически трудности заданий с подсчетом доли неправильных ответов по каждому из них.

Таблица №3

Матрица тестовых результатов

Испытуемые	Задания										X_i	Z_i
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}		
1	1	1	1	0	1	1	1	1	1	1	9	1,760
2	1	1	0	1	1	1	1	1	1	0	8	1,320
3	1	1	1	1	0	1	1	0	1	0	7	0,880
4	1	1	1	1	1	1	0	0	0	0	6	0,440
5	1	1	1	1	0	1	0	1	0	0	6	0,440
6	1	1	1	1	1	0	0	0	0	0	5	0
7	1	1	1	1	0	0	1	0	0	0	5	0
8	1	1	0	1	1	0	1	0	0	0	5	0
9	1	1	1	1	0	0	0	1	0	1	4	-0,440
10	1	0	1	0	1	1	0	0	0	0	4	-0,440
11	1	1	1	0	0	0	0	0	0	0	3	-0,880
12	1	1	0	0	0	0	0	0	0	0	2	-1,320
13	1	0	0	0	0	0	0	0	0	0	1	-1,760
R_j	12	11	9	7	6	6	5	4	3	2		
W_j	1	2	4	6	7	7	8	9	10	11		
p_j	0,92	0,846	0,692	0,538	0,46	0,462	0,385	0,30	0,231	0,154		
g_i	0,07	0,154	0,308	0,462	0,53	0,538	0,615	0,69	0,769	0,846		

В таблице №3 приводятся основные статические данные, принимаемые во внимание на первом этапе эмпирической проверки качества заданий.

Эмпирически трудность заданий определяется сложением элементов матрицы по столбцам, что указывает на число правильных ответов, полученных по каждому заданию (R_j). Чем больше правильных ответов на задание, тем оно легче для данной группы испытуемых. Больше правильных ответов оказалось в первом задании ($R_1=12$), это означает, что оно самое легкое в матрице. Первая колонка таблицы 3 представляет номера (или фамилии) испытуемых. Последняя колонка показывает сумму сырых баллов испытуемого.

Обычно, чем больше испытуемых, тем больше значения R_j . Поэтому для получения сопоставительных

характеристик R_j делят на число испытуемых [1, стр.159-163].
$$p_j = \frac{R_j}{N} \quad (1)$$

В результате получается нормированный (числом испытуемых) статистический показатель, или короче, статистика – доля правильных ответов p_j . p_j долго использовалась в качестве показателя трудности в так называемой классической теории тестов. Позже было осознана содержащаяся в ней смысловая неточность: ведь увеличение значения p_j направлена не на возрастание трудности, а наоборот, на возрастание легкости, если можно применить такое слово. Поэтому в последние годы с показателем трудности заданий стали ассоциировать противоположную статистику –долю неправильных ответов q_j . Эта доля вычисляется из отношения числа неправильных ответов к числу испытуемых N .

$$\text{Коэффициент трудности задания } q_j = \frac{W_j}{N} \quad (2)$$

На основе методики шкалирования результатов [2], было разработано программное обеспечение Z-шкалирования результатов тестирования с применением уровня трудности заданий.

Z-шкала, применяя количество верных ответов при переводе дихотомических оценок испытуемого в стандартные показатели, не учитывает трудности заданий. Возможно, что испытуемый, выполнивший меньшее количество более трудных заданий получит балл ниже, чем испытуемый, выполнивший большее количество легких заданий или возможно испытуемые получают одинаковое количество сырых баллов при ответе на тестовые задания разной сложности. Таким образом, оценка знаний этих испытуемых получена не объективно.

Для определения более точного уровня подготовленности испытуемого надо свести коэффициент уровня знаний каждого испытуемого с коэффициентом уровня трудности каждого задания в общую шкалу оценок.

В нашем исследовании мы используем уровень сложности самого теста для каждого испытуемого суммированием коэффициентов трудности из 10 заданий.

Наша задача данной статьи –это получение новой Z шкалы для определения более точного уровня подготовленности испытуемого.

В рассматриваемом методе шкалирования в таблице №3 правильные ответы т.е. единицы заменяем коэффициентами трудностей заданий, результаты будут приставлены на таблице №6.

Таблица №4.

Правила использования уровня трудности заданий при оценивании ответа

Ответ x_{ij}	Оценка
Ответ i-го ученика на j-е задание правильный	q_{ij}
Ответ i-го ученика на j-е задание неправильный	0

Таблица №5

Макет матрицы результатов с применением коэффициента уровня трудности заданий

Испытуемые	Задания 1,...,j,...,n	Индивидуальный балл X_i
1... i... N	$x_{ij} = \begin{cases} q_{ij} \\ 0 \end{cases}$	$X_i = \sum_{j=1}^n x_{ij}$

Матрица результатов тестирования с применением коэффициента уровня трудности заданий q_{ij}

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_i
1	0,077	0,154	0,308	0	0,538	0,538	0,615	0,692	0,769	0,846	4,534
2	0,077	0,154	0	0,462	0,538	0,538	0,615	0,692	0,769	0	3,842
3	0,077	0,154	0,308	0,462	0	0,538	0,615	0	0,769	0	2,919
4	0,077	0,154	0,308	0,462	0	0,538	0	0,692	0	0	2,227
5	0,077	0,154	0,308	0,462	0,538	0,538	0	0	0	0	2,073
6	0	0,154	0,308	0	0	0	0	0,692	0	0,846	1,998
7	0,077	0,154	0	0,462	0,538	0	0,615	0	0	0	1,843
8	0,077	0,154	0,308	0,462	0	0	0,615	0	0	0	1,612
9	0,077	0,154	0,308	0,462	0,538	0	0	0	0	0	1,535
10	0,077	0	0,308	0	0,538	0,538	0	0	0	0	1,459
11	0,077	0,154	0,308	0	0	0	0	0	0	0	0,536
12	0,077	0,154	0	0	0	0	0	0	0	0	0,229
13	0,077	0	0	0	0	0	0	0	0	0	0,076

В таблице №6 определим среднее значение $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$ (3)

Среднее арифметическое индивидуальных баллов испытуемых для рассмотренного выше примера матрицы (таблица 5) равно

$$\bar{X} = \frac{4,534+3,842+2,919+2,227+2,073+1,998+1,843+1,612+1,535+1,459+0,536+0,229+0,076}{13} = \frac{24,883}{13} = 1,914 \quad (4)$$

Вычислим дисперсию по формуле $S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N-1}$ (5).

Для рассматриваемого примера дисперсия равна $S_x^2 = \frac{20,285}{12} = 1,69$ отсюда стандартное отклонение вычисляется путем извлечения квадратного корня из дисперсии $S_x = \sqrt{S_x^2}$ (6). Для рассматриваемого примера стандартное отклонение равно $S_x \approx \sqrt{1,69} \approx 1,3$.

При выборе метода шкалирования обращаемся к стандартным показателям, указывающим отличие индивидуального результата испытуемого от среднего балла по выборке в единицах стандартного отклонения. Эти показатели используются для установления места первичного балла каждого испытуемого в сравнении с результатами других на основе подсчета нормированных отклонений и называются z оценками. Результат отображения z оценок на числовую ось образует Z-шкалу [2].

Для перевода в Z-шкалу первичный балл испытуемого преобразуется по формуле $Z_i = \frac{X_i - \bar{X}}{S_x}$ (7), где X_i - первичный балл i-испытуемого; \bar{X} - среднее значение индивидуальных баллов N испытуемых группы; S_x - стандартное отклонение [1, стр. 111].

Таблица №7

Таблица результатов с z оценками

N	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X _i	Z _i
1	0,077	0,154	0,308	0	0,538	0,538	0,615	0,692	0,769	0,846	4,534	2,015
2	0,077	0,154	0	0,462	0,538	0,538	0,615	0,692	0,769	0	3,842	1,483
3	0,077	0,154	0,308	0,462	0	0,538	0,615	0	0,769	0	2,919	0,773
4	0,077	0,154	0,308	0,462	0	0,538	0	0,692	0	0	2,227	0,241
5	0,077	0,154	0,308	0,462	0,538	0,538	0	0	0	0	2,073	0,122
6	0	0,154	0,308	0	0	0	0	0,692	0	0,846	1,998	0,065
7	0,077	0,154	0	0,462	0,538	0	0,615	0	0	0	1,843	-0,055
8	0,077	0,154	0,308	0,462	0	0	0,615	0	0	0	1,612	-0,232
9	0,077	0,154	0,308	0,462	0,538	0	0	0	0	0	1,535	-0,292
10	0,077	0	0,308	0	0,538	0,538	0	0	0	0	1,459	-0,350
11	0,077	0,154	0,308	0	0	0	0	0	0	0	0,536	-1,060
12	0,077	0,154	0	0	0	0	0	0	0	0	0,229	-1,296
13	0,077	0	0	0	0	0	0	0	0	0	0,077	-1,414

Например: испытуемый №6 на первое задание получил $X_1=0$ балл, на второе задание получил $X_2=0,154$, и т. д. сырой балл равен $X_i=1,998$, стандартное отклонение $S_k=1,3$, то где среднее значение баллов $\bar{X}=1,914$ и z-оценок равно $z_6=0,065$.

Методы перевода z-оценок в область положительных целых чисел подробно описано в литературе [2, стр.182].

Если сравнить таблицу №3 и таблицу №7, то в таблице №3 испытуемый, стоявший на девятом месте, в таблице №7 получил шестое место $Z_6=0,065$, так как ответил на 4 трудных вопроса. По таблице №3 четвертый и пятый (N=4, N=5) испытуемые, правильно ответив на 6 заданий, получили $Z_4=0,440$, $Z_5=0,440$ оценок соответственно по таблице №7 испытуемые N=4 и N=5 правильно ответив на 6 заданий и получили $Z_4=0,241$, $X_5=0,122$ оценок.

Это показывает, что предлагаемый метод более точно оценивает уровня подготовленности испытуемых.

Литература

1. Аванесов В.С. Композиция тестовых задач. Москва: АДЕПТ, 1998.
2. Звонников В.И., Чельщикова М. Б. Современные средства оценивания результатов обучения. Москва: Издательский центр "Академия", 2008.