

ИНФОРМАЦИОННАЯ МОДЕЛЬ РАСЧЕТА БАЛЛОВ И ВЫСТАВЛЕНИЯ ОЦЕНОК ПО ТЕХНОЛОГИИ ЕГЭ

Объект измерения в ЕГЭ и его статистическая природа.

Л.Е.Смушкевич (профессор кафедры математического анализа Магнитогорского государственного университета (МаГУ), директор Лицея при МаГУ) в статье «ЕГЭ как элемент современной школьной реформы», говоря о ЕГЭ как измерительном инструменте, ставит три вопроса:

- Что мы хотим измерять в педагогических технологиях?
- Что значит «измерять»?
- Все ли можно измерить?

Ответ на первый вопрос вполне очевиден – мы хотим измерять уровень подготовленности учащихся в конкретной предметной области. Прежде начинать разговор о свойствах уровня подготовленности как объекта измерения, необходимо выяснить, что означает термин «измерять».

Ответ на этот вопрос хорошо иллюстрирует мультфильм «38 попугаев», герои которого искали способ измерить удава и в конце концов выяснили, что его длина равна «38 попугаям и еще одному попугайскому крылышку». Т.е. измерить – значит определить, сколько раз эталон (или его части) укладывается в измеряемый объект. Но это возможно только при наличии эталона. Если же эталон отсутствует? Например, в измерении уровня подготовленности нет эталона и баллы за ЕГЭ в 2000 г. немного не те, что баллы в 2003 г. Т.е. стобалльная шкала за один год отличается от стобалльной шкалы другого года. В таком случае мы можем производить только относительное сравнение объектов, т.е. сравнить объекты между собой. Мы ведь всегда можем сказать какой из двух людей выше, даже не зная их роста. Так, на уроках физкультуры класс выстраивается по росту, и мы получаем группу учащихся ранжированных по росту. Наличие такого списка – это уже некоторая оценочная шкала.

Вообще в тестировании используется 4 основных вида шкал:

- Номинальные шкалы – позволяют разделить объекты на разные, строго регламентируемые, группы. Например, деление учащихся по половому признаку.
- Порядковые шкалы – позволяют ранжировать объекты по какому-то признаку. Например, выстроить учащихся по росту. Первичные баллы (число верно выполненных заданий) откладываются на порядковой шкале.
- Интервальные шкалы – позволяют, кроме ранжирования, определять на сколько у одного объекта признака больше чем у другого. Самый яркий пример – летоисчисление. Или измерение температуры по шкале Цельсия. Первичная оценка уровня подготовленности учащихся измеряемая в логитах откладывается на интервальной шкале.

▪ Нормированные шкалы – это по своей сути интервальные шкалы, на которых определено начало отсчета. Данная шкала позволяет определить во сколько раз признака у одного объекта больше чем у другого. Примером служит измерение роста, но не сравнением двух людей, а с использованием метра. Первичные баллы переводятся на нормированную шкалу, и мы получаем тестовые баллы.

А теперь перейдем к третьему вопросу «все ли можно измерить?» (на самом деле нас интересует вопрос – можно ли измерить уровень подготовленности).

Если говорить о наблюдаемых признаках (длина, вес, цвет ...), то тут не встает вопроса о возможности измерения признака (в соответствующей шкале). Но как быть с измерением уровня подготовленности? Думаю, каждому известна фраза учителя «Не молчи, отвечай на вопрос, я же не могу заглянуть тебе в голову и определить – знаешь ты или нет». Эта фраза включает подсказку о том, как измерить уровень подготовленности. Сам уровень подготовленности мы наблюдать не можем, но мы можем наблюдать его проявление. Такие признаки (наличие которых можно определить только по их проявлению) называют латентными. Таким образом, в случае измерения латентных признаков нужно измерять проявления этих признаков, и по их проявлению уже судить о самих признаках.

Итак, мы выяснили, что уровень подготовленности является латентным признаком и для его измерения нужно измерять его проявление. Проявлением же уровня подготовленности, очевидно, будет вероятность правильного ответа учащегося на поставленный вопрос. Если группе учащихся предложить для решения тест из определенной предметной области, то результаты будут следующими:

▪ в случае репрезентативного теста (подготовленного с учетом математико-статистических требований, предъявляемых к тесту) больше правильных ответов будет у учащегося с лучшей подготовкой;

▪ в случае нерепрезентативного теста, результаты будут вызывать сомнения.

Вообще говоря, ответ на каждый отдельный вопрос довольно непредсказуем. На ответ кроме уровня подготовки учащегося влияет множество других факторов – физическое состояние, наличие внутренней мотивации, психологические особенности учащегося, наличие отвлекающих факторов, формулировка вопроса и др. Часть факторов должна сниматься организационным моментом (отвлекающие моменты, психологическая настройка на экзамен). Такие же факторы как корректная формулировка задания должны нейтрализоваться репрезентативностью теста. Поэтому при обработке результатов исходят из того, что на конечный результат влияет только уровень подготовленности учащихся и сложность заданий.

Но каким образом определить уровень подготовки по ответам учащихся? Очевидно, для этого необходим соответствующий математический аппарат, позволяющий оценить признак по его проявлению. И такой аппарат, безусловно, дает математическая статистика. Если теория

вероятностей занимается вопросами вычисления вероятности события исходя из известных параметров, то математическая статистика занимается обратной задачей — вычислением значений параметров по известному распределению вероятности.

В тестировании выделяют два подхода к обработке результатов тестирования (читай к оценке уровня подготовленности учащихся):

- классический подход;
- современный подход.

Эти подходы различаются используемым математическим аппаратом. Нельзя говорить, что один из подходов лучше другого, они просто разные. И выбор способа обработки результатов должен определяться рациональностью метода в каждом конкретном случае.

Обработка результатов тестирования в рамках обоих подходов начинается с построения матрицы ответов и вычисления первичных баллов.

№ испытуе мого i	№ задания j										Первичный балл испытуемо го X _i
	1	2	3	4	5	6	7	8	9	10	
4	0	0	0	0	0	0	0	1	0	0	1
11	1	0	1	0	0	0	0	0	0	0	2
12	1	1	0	1	0	0	0	0	0	0	3
8	1	1	1	0	0	0	0	0	0	0	3
10	1	1	1	0	1	0	0	0	1	0	5
6	0	0	1	1	1	1	1	0	0	0	5
13	1	1	1	1	1	0	0	0	0	0	5
5	1	1	1	0	1	1	1	0	0	0	6
1	1	1	1	1	1	1	1	0	0	0	7
2	1	1	1	1	1	1	0	1	0	0	7
7	1	1	0	1	1	1	1	1	1	1	9
Первичн ый балл задания R _j	9	8	8	6	7	5	4	3	2	1	53

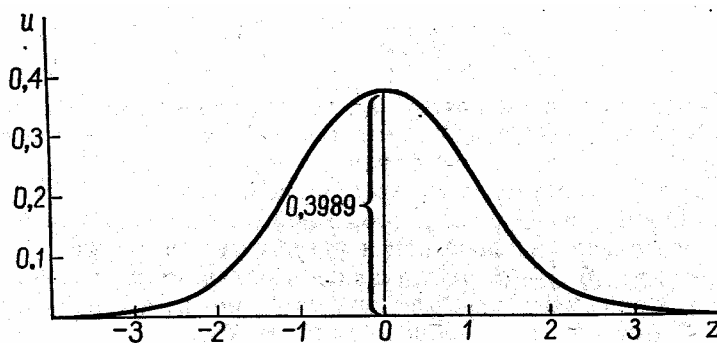
Первичный балл учащегося – это сумма баллов полученных за каждое задание. Первичный балл задания – это сумма баллов по столбцам таблицы.

А теперь опишем различия классического и современного подходов.

В классическом подходе конечный балл получается путем линейных или нелинейных преобразований первичных баллов.

Например, можно получить процентную шкалу следующими преобразованиями: $B_i = \frac{X_i}{k} \cdot 100$, где k – число заданий в тесте.

Или можно нормализовать первичные баллы, т.е. привести к распределению достаточно близкому к нормальному:



На практике график распределения первичных баллов не симметричен относительно вершины.

Но в любом случае, тестовый балл при использовании классического подхода, будет откладываться на порядковой шкале. А расстояние между точками на порядковой шкале изменяется нелинейно. Т.е. если V_1, V_2, V_3 – это баллы учащихся, и $V_1 > V_2 > V_3$, то $V_1 - V_2 \neq V_2 - V_3$. И если мы используем несколько вариантов тестов по одному предмету, то вопрос сопоставимости результатов по разным вариантам довольно сложен.

В современном же подходе все операции по подготовке тестов и обработке результатов осуществляются в рамках выбранной модели, описывающей зависимость вероятности правильного ответа на задание от уровня подготовки учащегося и уровня сложности задания. Зная результаты тестирования, можно методами математической статистики оценить вероятности правильных ответов учащихся на задания теста. И опираясь на эти оценки можно получить оценку уровня подготовки учащихся и уровня сложности заданий.

Причем модель строится из предположения, что и уровень подготовленности учащегося и уровень сложности задания располагаются на одной интервальной шкале. И если мы получим оценки уровня подготовки учащихся на разных вариантах теста в рамках такой модели, то в силу того, что шкала у нас интервальная, задача сопоставимости результатов выглядит уже гораздо проще. По крайней мере, степень достоверности единой для всех вариантов шкалы баллов полученной в современном подходе гораздо выше, чем у шкалы полученной в классическом подходе.

Выделяют три модели, используемые в современном подходе.

Однопараметрическая модель Г. Раша

$$p = \frac{1}{1 + e^{-(\theta - \beta)}};$$

Модель названа однопараметрической логистической моделью в силу того, что функция зависит от одного параметра – разности θ (уровень подготовленности учащегося) и β (уровень сложности задания). Если уровни подготовки двух учащихся различаются на такое же число что и сложность двух заданий, то шанс слабого ученика решить правильно легкое задание равен шансу сильного ученика при решении сложного задания. В

этой модели у учеников с одинаковым первичным баллом будут одинаковые тестовые баллы.

Двухпараметрическая модель А. Бирнбаума

$$p = \frac{1}{1 + e^{-d(\theta - \beta)}} ,$$

где d – дополнительный параметр тестового задания, характеризующий дифференцирующую способность задания. Эта модель учитывает, какие задания решены учащимся правильно. Т.е. при одинаковых первичных балла можно получить различные тестовые баллы. Тестовые баллы в данном случае отражают процент выполнения содержательной составляющей теста.

Все модели показывают что

θ гораздо больше β	$p \rightarrow 1$
θ равно β	$p = \frac{1}{2}$
θ гораздо меньше β	$p \rightarrow 0$

Интерпретация тестового балла ЕГЭ.

Так в технологии ЕГЭ при обработке результатов используется двухпараметрическая модель Бирнбаума, то:

- Между первичным и тестовым баллами нельзя устанавливать прямую пропорциональность. Единственно, что мы можем сказать исходя из первичных баллов, это: если первичный балл учащегося А больше первичного балла учащегося В, то тестовый балл учащегося А будет по крайней мере не меньше, чем у В.

- Если два ученика имеют одинаковые первичные баллы, то их тестовые баллы могут различаться.

- Тестовый балл выражает рейтинг учащегося по данному предмету.

- Тестовый балл выражает процент успешно выполненного объема теста (в семантическом смысле).

И, т.к. тестовые баллы принадлежат нормированной шкале, то мы можем складывать, вычитать, делить баллы.