

# ОБ ОДНОЙ ОЦЕНКЕ СТЕПЕНИ КОРРУМПИРОВАННОСТИ РЕЗУЛЬТАТОВ ЕДИНОГО ГОСУДАРСТВЕННОГО ЭКЗАМЕНА 2009 ГОДА

В.В. Кромер

Аннотация. На основе исследования распределений тестовых баллов результатов Единого Государственного Экзамена 2009 года, отличающегося от нормального распределения, сделано предположение о неоднородности генеральной совокупности испытуемых. Генеральная совокупность представлена двумя однородными выборками, из которых одна, с меньшим значением математического ожидания тестового балла, считается отражающей истинные неискаженные результаты испытуемых, а другая, с большим значением математического ожидания тестового балла, предполагается искаженной на основе нарушения регламента процедуры тестирования. Вычислены влияние и массовость искажения тестовых результатов по каждому из 10 учебных предметов.

Целью настоящей работы является оценка степени коррумпированности результатов первой волны Единого государственного экзамена (ЕГЭ) 2009 года [5]. Одно из значений латинского слова *corrumpo* – искажать. Есть все основания полагать, что обсуждаемые результаты ЕГЭ существенно искажены, и степень искажения (коррумпированности) результатов по большинству предметов такова, что у не вовлеченного в процесс коррупции выпускника (= абитуриента) существенно снижены шансы быть зачисленным в более-менее престижный вуз.

Результаты ЕГЭ искажены отклонениями от регламентированной процедуры проведения ЕГЭ. Одно из нарушений уже названо – разрешение выпускникам пользоваться в ходе сдачи ЕГЭ мобильными телефонами [4]. А также подсказки выпускникам со стороны лиц, обязанных следить за соблюдением регламента ЕГЭ. Технически для этого достаточно собрать в отдельной аудитории выпускников, согласно определенным критериям выделенным для оказания вышеописанной помощи.

Итак, источник искажения результатов выделен – генеральная совокупность испытуемых неоднородна. Следуя принципу "Бритвы Оккама", требующему не множить сущности без необходимости, сделаем предположение, что генеральная совокупность состоит из всего двух однородных выборок. Первая выборка образована испытуемыми, результаты которых не искажены, вторая составлена из испытуемых, результаты которых искажены (доступом испытуемых при тестировании к источникам информации, необъективностью экспертов, оценивающих части С КИМ ЕГЭ).

Искажение результатов у части испытуемых должно проявиться в отклонении измеренного распределения результатов по всей генеральной совокупности от истинного. А каково оно, истинное распределение? Вопрос достаточно тщательно изучался рядом исследователей, пришедшими к внешне парадоксальному выводу – форма распределения результатов со-

циальных измерений распределения оценена быть не может, человек постулирует ее исходя из своих представлений об измеряемой латентной (скрытой) переменной, при этом у него выбор невелик – распределение должно быть устойчивым. Устойчивое распределение при свертке с таким же распределением приводит к распределению того же вида. Устойчивые распределения могут быть разделены на гауссовы и негауссовы. Первые сходятся к распределению Гаусса (нормальному), вторые нет [6, с. 101-102].

Наиболее для описания длиннохвостовых социальных распределений подходит ципфовое негауссовое распределение, но человеку в силу привычки удобнее описывать социальные переменные более сжатым нормальным распределением. Это распределение положено в основу шкалы интеллекта Стэнфорда-Бине, и автоматически применено к шкале уровня знаний. При этом к шкале выдвигается требование линейности.

При оценке знаний испытуемого КИМ'ами ЕГЭ ему предъявляется ряд заданий, в зависимости от успешности выполнения которых ему начисляются т.н. первичные баллы. Шкала первичных баллов в общем случае нелинейна, процедуры подготовки заданий КИМ ЕГЭ не предполагают линеаризации этой шкалы, и по распределению первичных баллов придти к заключению о форме распределения измеряемого качества невозможно. Но, наоборот, постулируя соответствие распределения измеряемого качества одному из стандартных (например, нормальному), мы линеаризуем шкалу первичных баллов, меняя в тесте соотношение между заданиями легкими, средними и трудными [7, с. 240]. При этом мы, разумеется, прежде всего должны быть уверены в неискаженности результатов измерения.

Иная картина вырисовывается при оценке результатов ЕГЭ, которые оцениваются не в рамках классической тестовой теории (т.е. первичными баллами), а в рамках модели датского математика Раша. Модель Раша – самая простая из семейства т.н. логистических моделей, где единицей измерения служит логарифмическая единица – логит. При постулируемой в рамках модели Раша логистической зависимости между уровнем подготовки испытуемого и вероятностью выполнения испытуемым задания определенной трудности [7, с. 264] шкала логитов линейная. А линейность шкалы позволяет нам постулировать стандартное распределение измеряемой переменной, нормальность в нашем случае.

Итак, для генеральной совокупности испытуемых распределения логитов знания должны быть нормальными. Согласно принятой методике шкалирования результатов ЕГЭ [3] тестовые баллы участников ЕГЭ вычисляются с помощью линейного преобразования, переводящего оценку на шкале логитов в оценку на шкале тестовых баллов, что не меняет формы распределения.

Анализ реальных распределений [5] показывает, что распределения значительно отклоняются от нормального, что фиксируется значениями коэф-

фициента асимметрии (скоса) распределений (таблица 1). Исследовались данные по 10 предметам, за пределами исследования оказались данные по 4 иностранным языкам вследствие малочисленности тестируемых генеральных совокупностей.

Таблица 1. Средние значения, стандартные отклонения, коэффициенты асимметрии и эксцесса распределений тестовых баллов по 10 предметам

Предмет	Среднее значение	Стандартное отклонение	Коэффициент асимметрии	Коэффициент эксцесса
Русский язык	56,3	12,6	0,04	-0,18
Математика	42,9	15,4	0,00	-0,24
Литература	52,5	16,0	-0,22	0,86
Обществознание	56,6	11,1	-0,07	-0,37
История	48,0	14,9	0,45	-0,63
Химия	54,8	16,4	0,13	-0,89
Физика	48,8	12,7	0,38	0,11
Информатика	56,3	15,9	-0,07	-0,40
Биология	52,3	14,0	0,40	-0,57
География	49,6	12,9	0,33	-0,49

Стандартное отклонение коэффициента асимметрии равно  $\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$ , а коэффициента эксцесса  $\sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}$ , где  $n$  – размер выборки [1, с. 282]. При достаточно большом  $n$  стандартное отклонение коэффициента асимметрии равно  $\sqrt{\frac{6}{n}}$ , а коэффициента эксцесса  $\sqrt{\frac{24}{n}}$ .

Из перечисленных 10 предметов минимальная выборка (для географии) составляет 33256 испытуемых, что дает стандартное отклонение коэффициента асимметрии 0,013, а коэффициента эксцесса 0,027. Для остальных 9 предметов стандартное отклонение коэффициентов асимметрии и эксцесса еще меньше. Для нас представляет интерес то, что для предметов с близким к нулю коэффициентом асимметрии (русский язык, математика, обществознание и информатика) коэффициенты эксцесса значимо отрицательны, что указывает на более низкий и плоский характер вершины распределения по сравнению с нормальной кривой.

Таким образом, из анализов коэффициентов асимметрии и эксцесса вытекает, что ни одно из распределений тестовых баллов не является нормальным. Сделаем предположение, что распределение отклоняется от нормального вследствие неоднородности генеральной совокупности – ее смешанности. Известно, что если генеральная совокупность испытуемых существенно неоднородна, распределение данных бимодально [7, с. 244]. Генеральная совокупность с менее выраженной неоднородностью приво-

дит к уплощению вершины распределения, что отмечается для большинства распределений по таблице 1. Но физиологических или психологических предпосылок к неоднородности генеральной совокупности нет. Тогда объяснить неоднородность можно искажением распределенных нормально результатов у части испытуемых, и генеральную совокупность в первом приближении можно представить состоящей из двух однородных выборок – неискаженной и искаженной (соответственно с неискаженными и искаженными результатами). Зададимся целью параметризовать эти выборки по имеющимся данным. Для этого для каждого из распределений (по каждому предмету) необходимо решить методом итераций систему из 6 уравнений (по 3 параметра для каждой выборки – математическое ожидание  $m$ , стандартное отклонение  $\sigma$  и доля выборки в генеральной совокупности по численности испытуемых  $A$ ). Приближение осуществляем по методу наименьших квадратов. Результаты решения приведены в таблице 2.

Таблица 2. Результаты параметризации данных по предметам

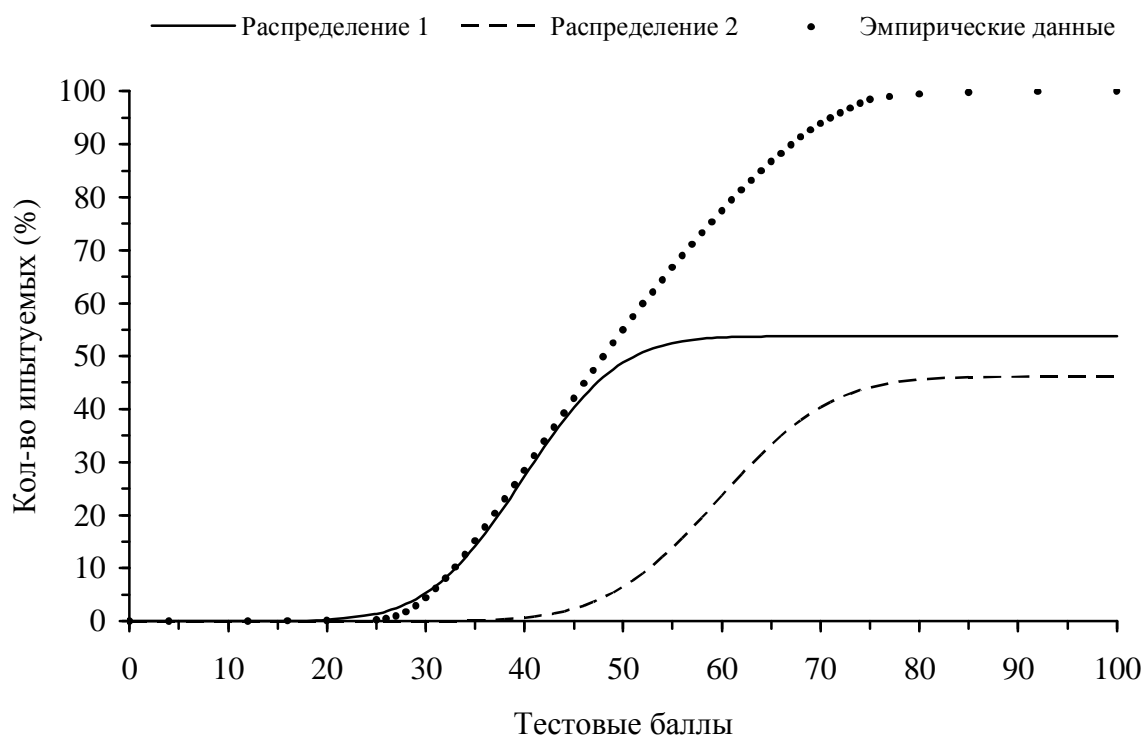
Предмет	Распределение 1			Распределение 2			Параметры коррупции	
	$m_1$	$\sigma_1$	$A_1$	$m_2$	$\sigma_2$	$A_2$	Влияние	Массовость
Русский язык	53,3	11,9	85,6	69,0	5,6	14,1	15,7	0,14
Математика	40,4	15,2	93,8	60,4	12,8	6,5	20,0	0,06
Литература	50,9	17,4	74,4	54,5	7,8	25,8	3,6	0,26
Обществознание	55,0	11,9	54,4	58,3	11,5	47,1	3,3	0,46
История	36,7	7,9	54,6	60,3	11,2	45,5	23,6	0,45
Химия	40,5	9,6	49,0	67,4	10,5	51,3	26,9	0,51
Физика	40,9	8,8	62,0	59,2	8,5	37,3	18,3	0,38
Информатика	41,3	11,1	40,0	64,4	11,1	59,7	23,1	0,60
Биология	41,2	7,3	46,0	61,1	12,2	54,3	19,9	0,54
География	39,8	7,6	53,8	59,7	9,0	46,2	19,9	0,46

Распределение с меньшим значением математического ожидания назовем Распределением 1, а с большим – Распределением 2. Поскольку в силу характера процедуры тестирования и ее значения для испытуемых искажение заключается в увеличении математического ожидания результатов тестирования, Распределение 1 считаем неискаженным, а Распределение 2 – искаженным. Степень искажения – разницу между математическими ожиданиями искаженного и неискаженного результата – назовем влиянием коррупции. Массовостью коррупции назовем долю испытуемых в генеральной совокупности, чьи результаты подверглись искажению в результате коррупции. Эти значения также приведены в таблице 2. Встает вопрос о мере значимости сделанного вывода о неоднородности генеральной совокупности. По каждому из предметов выдвигаем нулевую гипотезу о равенстве Распределения 1 и Распределения 2 и вычисляем значение

критерия Стьюдента для проверки средних значений. Все эти значения существенно превышают границу критерия при разумных уровнях значимости, что позволяет отвергнуть нулевую гипотезу и считать генеральную совокупность испытуемых неоднородной по всем предметам, т.е. часть результатов – искаженной.

Согласно таблице 2, наименьшая массовость коррупции характерна для математики и русского языка. Наибольшая массовость коррупции характерна для информатики, биологии и химии (более 50% результатов искажены). Наиболее типичное (медианное) влияние коррупции составляет около 20 тестовых баллов. По литературе и общественному мнению влияние коррупции невелико, менее 4 тестовых баллов. И литературе же и русскому языку свойственно более чем двукратное уменьшение стандартного отклонения результатов искаженной выборки по сравнению с неискаженной.

На рисунке ниже в качестве примера приведена эмпирическая интегральная кривая распределения тестовых баллов по географии и интегральные кривые распределений для неискаженной (Распределение 1) и искаженной (Распределение 2) выборок. Сумма двух распределений и дает эмпирические данные в пределах погрешности метода.



В данном исследовании нами распространен на социальные явления процесс, аналог которому существует в физическом мире. Так, известно, что средний рост американцев (мужчин) составляет 179 см, а женщин 167 см. При этом распределение ростов и мужчин, и женщин нормальное,

но стандартное отклонение роста мужчин больше, чем у женщин. При обследовании смешанной выборки без учета пола мы получаем несимметричное уплощенное распределение (разница в средних ростах мужчин и женщин с учетом стандартных отклонений недостаточна для проявления бимодальности), вид которого зависит от соотношения численности мужчин и женщин в выборке.

В отношении результатов 1-й волны ЕГЭ 2009 года нами обследовались данные по России, что позволило сделать обобщающие выводы.

1. Результаты ЕГЭ 2009 года искажены вследствие несоблюдения регламента процедуры, при этом по отдельным предметам искажено до 60% всех результатов.

2. Вследствие искажения результатов ЕГЭ тестовые баллы по отдельным предметам у части испытуемых смещены в сторону увеличения от истинных, соответствующих уровню их подготовки, в среднем на 27 баллов, что вследствие общеизвестности явления вызывает недоверие руководителей образования к высоким тестовым баллам [2].

3. Влияние и массовость коррупции при проведении процедуры ЕГЭ не позволяют считать результаты ЕГЭ 2009 года пригодными для ранжирования абитуриентов вузов с целью профессионального отбора, поскольку в соответствии с принятым порядком отбора абитуриентов преимущество получают абитуриенты с несоответствующими истинному уровню знаний тестовыми баллами.

4. Анализ распределений предметных тестовых баллов по отдельным регионам России предположительно позволит выявить региональные особенности распределения степени коррупции при проведении ЕГЭ.

5. В основу выводов положены определенные предположения, в частности о линейности шкалы логистической модели Раша и о нормальности распределения уровня подготовки выпускников в логитах. Отказ от этих предположений и выдвижение более слабых, в частности предположение о симметричности распределения уровня подготовки с отказом от нормальности, в силу робастности модели не повлияет на сделанные качественные выводы относительно массового искажения результатов ЕГЭ.

### **Библиографический список**

1. Варден ван дер, Б.Л. Математическая статистика. М.: ИИЛ, 1960.
2. Кириллова С. Сдвиг по шкале. Ректоры не доверяют механизму выставления оценок на ЕГЭ. Первое сентября. 2005. № 44. 21 июня. <http://ps.1september.ru/articlef.php?ID=200504406>
3. Методика шкалирования результатов ЕГЭ в 2008 году. Официальный информационный портал Единого Государственного Экзамена: <http://www1.ege.edu.ru/content/view/431/166/>

4. Письмо Рособнадзора о выявленных нарушениях Порядка проведения ЕГЭ, связанных с наличием сотовых телефонов у участников экзамена. Официальный информационный портал Единого Государственного Экзамена: <http://www1.ege.edu.ru/content/view/583/161/>

5. Соответствие первичных и тестовых баллов. Официальный информационный портал Единого Государственного Экзамена: [http://www1.ege.edu.ru/images/stories/ege2009/shkalir/tab1\\_perev\\_23.06.09.xls](http://www1.ege.edu.ru/images/stories/ege2009/shkalir/tab1_perev_23.06.09.xls)

6. Хайтун С.Д. Мои идеи. М.: Агар, 1998.

7. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002.