

О подходах к формированию групп субъектов Российской Федерации для обмена экзаменационными работами в рамках перекрёстной проверки развёрнутых ответов участников ЕГЭ

**Левинская
Мария Александровна**

кандидат физико-математических наук, ведущий научный сотрудник ФГБНУ «ФИПИ», доцент ФГБОУ Московский авиационный институт (национальный исследовательский университет),
levinskaja@fipi.ru

Ключевые слова: перекрёстная проверка, согласованность экспертных отметок, кластеризация.

Введение

По решению Рособрнадзора при проверке развёрнутых ответов участников ЕГЭ (для части работ) организуется обмен экзаменационными работами между субъектами Российской Федерации — межрегиональная перекрёстная проверка. При этом заполненные бланки ответов на задания с развёрнутым ответом участников экзамена одного региона проверяют эксперты другого региона. Такая перекрёстная проверка работ в рамках ЕГЭ позволяет повысить объективность оценки.

В данной статье рассматривается вопрос, связанный с формированием групп субъектов Российской Федерации, допустимых для обмена экзаменационными работами в рамках перекрёстной проверки развёрнутых ответов участников ЕГЭ.

Цель исследования — выделить группы регионов, однородные по одному или нескольким признакам, которые смогут наиболее эффективно проводить перекрёстную проверку работ ЕГЭ внутри своей группы, обеспечивая объективность и согласованность. Для этого необходимо выявить ключевые показатели (параметры) качества проверки и разработать методику кластеризации регионов.

При правильном формировании групп субъектов Российской Федерации (кластеров) для обмена экзаменационными работами перекрёстная проверка даёт наиболее согласованную оценку развёрнутых ответов ЕГЭ, минимизируя работу конфликтных комиссий. Это и определяет актуальность работы.

Методика кластеризации регионов для перекрёстной проверки

После проведения ЕГЭ ежегодно проводится анализ качества работы предметных комиссий. В разработанной методике статистического анализа качества работы предметных комиссий были выделены ключевые показатели:

- среднее расхождение первого и второго экспертов по работе;

- среднее расхождение первого и второго экспертов на оцениваемый балл;
- доля третьей проверки;
- структура расхождений по заданиям¹.

Основным показателем, который позволяет оценить качество экспертной проверки и ранжировать регионы для выделения однородных групп участников перекрёстных проверок (с близкими по значению показателями *согласованности* проверки), является показатель «среднее расхождение первого и второго эксперта на оцениваемый балл», который вычисляется для каждого учебного предмета.

Данный показатель вычисляется как *среднее значение по региону* следующей величины:

$$\frac{\text{Расхождение по работе}}{\text{Оцениваемый балл по работе}}$$

Расхождение по работе — это модуль суммы по всем заданиям разностей баллов первого и второго эксперта:

$$\left| \sum_{i=1}^N \text{Оценка}_{\text{Эксперта1}_{\text{Задание } C_i}} - \text{Оценка}_{\text{Эксперта2}_{\text{Задание } C_i}} \right|$$

Оцениваемый балл по работе — сумма максимальных баллов за оцениваемые (присутствующие) задания в работе:

$$C_{1-\text{макс. балл}} + C_{2-\text{макс. балл}} + \dots + C_{N-\text{макс. балл}}$$

Разбив субъекты РФ на группы с учётом числа участников ЕГЭ, внутри каждой группы можно вычислить ранг региона по данному показателю *согласованности* первого и второго эксперта. При этом можно заметить, что качество проверки работ в регионах может значительно различаться. На рис. 1 показаны ранги² двух различных по показателю согласованности регионов. Минимальный ранг — 1 — соответствует минимальным расхождениям в кластере, максимальный ранг, равный числу регионов в кластере, соответствует максимальному расхождению. «Регион 1» (рис. 1)

¹ Методические материалы для председателей и членов предметных комиссий субъектов Российской Федерации по проверке выполнения заданий с развернутым ответом экзаменационных работ ЕГЭ 2021 года [Электронный ресурс] / ФГБНУ «ФИПИ». — URL: <https://fipi.ru/ege/dlya-predmetnyh-komissiy-subektov-rf>.

² Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. — новое изд. — М.: МЦНМО, 2016. — 368 с.

занимает «верхние» позиции (в своём кластере) по большинству учебных предметов. По показателю согласованности в учебном предмете «Английский язык» (АЯ) ранг «Региона 1» равен 2, по предметам «Математика» (МА), «Химия» (ХИ), «Биология» (БИ) ранг «Региона 1» в своём кластере равен 3. Число регионов в кластере, к которому принадлежит «Регион 1», равно 32. Соответственно, и максимальный ранг в этом кластере равен 32. По учебному предмету «Русский язык» (РУ) ранг рассматриваемого «Региона 1» равен 24, и по данному учебному предмету «Регион 1» находится в нижней части шкалы. В подписях (АЯ-11, МА-6, ХИ-7, БИ-6) цифры указывают ранг «Региона 1» среди всех субъектов РФ (из 85 возможных) и позволяют сравнить регионы различных кластеров.

«Регион 2» принадлежит кластеру из 15 субъектов РФ. Ранг регионов с максимальным расхождением в этом кластере равен 15. Данный максимальный ранг в «Регионе 2» получается по нескольким учебным предметам: «Русский язык», «Биология», «История» (ИС). На рисунке видно, что в Регионе 2 по показателю согласованности «нижние» позиции занимает большинство учебных предметов.

В каждом кластере регионов, выделенном для обмена работами ЕГЭ при перекрёстной проверке, должна быть обеспечена сопоставимость показателя качества проверки и согласованности экспертов, основанная на показателе «Среднее расхождение первого и второго эксперта на оцениваемый балл».

При кластеризации регионов для перекрёстной проверки в отдельный кластер объединяются регионы с экстремальными значениями данного показателя согласованности в определённом учебном предмете. Экстремальными значениями считаются значения, выходящие за три межквартильных размаха — от 75-го процентиля по данному показателю (рис. 2). Межквартильный размах — разница между третьей квартилью (X_{75} — 75-й процентиль) и первой квартилью (X_{25} — 25-й процентиль). На рис. 2 показана медиана показателя согласованности по всем субъектам РФ ~0.07 (приведён учебный предмет «Физика»), X_1 и X_2 — минимальное

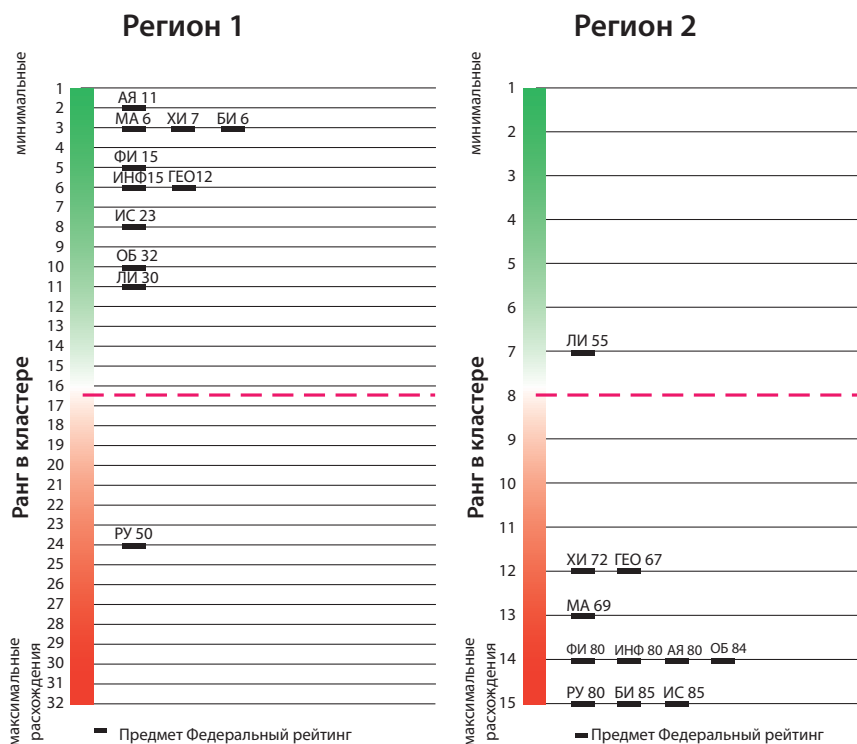


Рис. 1. Ранги регионов по показателю согласованности проверки

и максимальное значение показателя соответственно (за исключением «выбросов»). Экстремальные значения показателя отмечены «звёздочками» (Регион 1 и Регион 2), «кружочками» отмечены «выбросы» (точки 3, 5, 8) — значения показателя, превышающие 1,5 межквартильных размаха, но не являющиеся экстремальными.

Для регионов с неэкстремальными показателями согласованности при выделении кластеров рассматривается доверительный интервал этого показателя (95%)³. К регионам с низким расхождением относятся регионы, в которых доверительный интервал показателя согласованности находится ниже (левее) среднего расхождения по РФ (рис. 3, Регион 1). Регионы, доверительный интервал которых включает в себя значение среднего расхождения по РФ, относятся ко второму кластеру (со средним показателем расхождений). К регионам с высокой рассогласованностью относятся регионы, в которых доверительный интервал показателя согласованности находится

выше (правее) среднего показателя по РФ (рис. 3, Регион 2).

На рисунке 3 среднее значение показателя согласованности региона обозначено $\bar{x}$, значения показателя согласованности μ принадлежат интервалу $(\bar{x} - E; \bar{x} + E)$.

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ — ширина доверительного}$$

интервала, где $z_{\frac{\alpha}{2}}$ — Z-оценка для соответствующего уровня статистической значимости $\frac{\alpha}{2}$. В расчётах согласованности берётся 95%-ный доверительный интервал и $z_{\frac{\alpha}{2}} = 1,96$.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \text{ — стандартное}$$

отклонение по выборке (по проверяемым работам в регионе), x_i — расхождение экспертов на оцениваемый балл по i -й работе, n — объём выборки, или количество проверяемых работ в регионе.

При исследовании зависимостей показателей результатов ЕГЭ и показателя согласованности первого и второго экспертов была выявлена его сильная связь (корреляция) со средним региональным

³ Пугачёв В.С. Теория вероятностей и математическая статистика: учеб. пособие. — 2-е изд., испр. и доп. — М.: Физматлит, 2002. — 496 с.

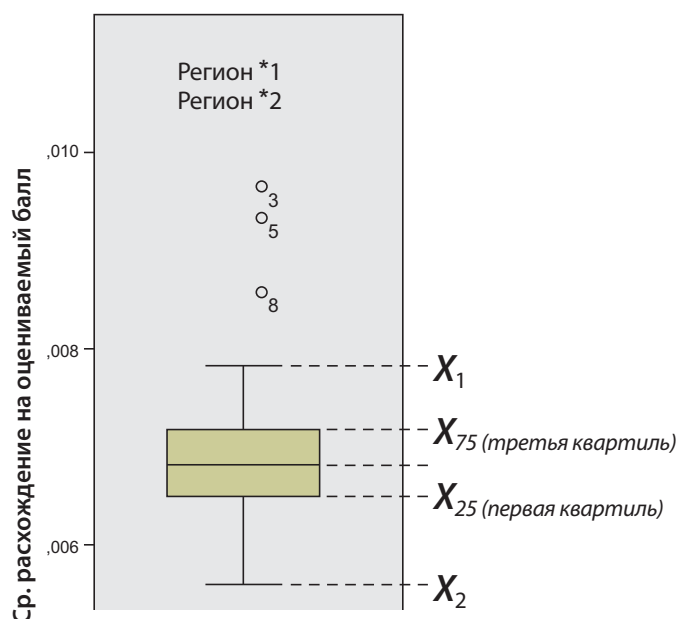


Рис. 2. Диаграмма размаха (ящичковая) по показателю среднего расхождения экспертов на оцениваемый балл

баллом за задания с развёрнутым ответом. Таким образом, показатель «средний балл за выполнение заданий части 2 (с развёрнутым ответом)» должен быть включён в переменные кластеризации. Разбиение субъектов РФ на группы в зависимости от среднего балла за выполнение заданий части 2 — первый этап кластеризации.

Так же как и при анализе качества проверки, при формировании групп субъектов РФ для перекрёстной проверки следует учитывать *объём выборки* (количество работ в регионе). Перекрёстная проверка происходит между регионами приблизительно одного объёма проверяемых работ. Для этого регионы разделены на три подгруппы. Верхняя и нижняя квартили количества проверяемых работ выделены в отдельные подгруппы: «крупные» и «малые» по объёму работ соответственно. Остальные регионы объединены в одну группу («средние» по объёму работ). Разбиение субъектов РФ на группы в зависимости от объёма работ — второй этап кластеризации.

Такая кластеризация проводится по каждому предмету ЕГЭ или группе предметов, объединённых на основании схожих средних значений показателя расхождения экспертов. К группе предметов с минимальными расхождениями относятся

такие предметы, как математика, химия. Учебные предметы обществознание, иностранный язык и русский язык относятся к группе с высокими расхождениями.

Опишем алгоритм выделения итоговых групп субъектов РФ для перекрёстной проверки.

1. Регионы разбиваются на две группы в зависимости от среднего балла за задания с развёрнутым ответом.

2. В этих группах (полученных в п. 1) отдельно рассматриваются регионы малые, средние и крупные по объёму (количеству) работ.

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

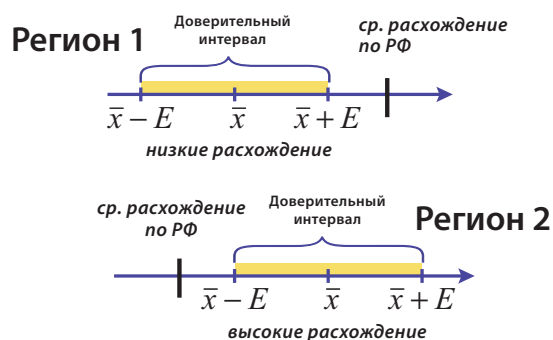


Рис. 3. Доверительные интервалы показателя согласованности

Таблица 1

Формирование групп субъектов РФ для перекрёстной проверки

Средний балл части с развёрнутыми ответами	Кластер количества участников	Кластер качества проверки (доверительный интервал расхождения экспертов)	Регион
Ниже среднего	Малые	Низкие расхождения	Регион 1 Регион 2
		Средние расхождения	Регион 3
	Средние	Низкие расхождения	...
		Средние расхождения	
		Высокие расхождения	
	Высокие	Средние расхождения	
		Высокие расхождения	
Выше среднего	Малые	Низкие расхождения	
		Средние расхождения	
		Высокие расхождения	
	Средние	Низкие расхождения	
		Средние расхождения	
		Высокие расхождения	
	Высокие	Низкие расхождения	
		Средние расхождения	
		Высокие расхождения	

3. В каждой полученной подгруппе регионы разбиваются по качественному показателю *среднего расхождения экспертов на оцениваемый балл*: низкие, средние и высокие расхождения в зависимости от доверительного интервала (см. табл. 1). При этом некоторые подгруппы в частном случае могут быть с нулевым количеством регионов. Например, в табл. 1 отсутствует подгруппа «высокие расхождения в малых регионах с низким средним баллом». Это означает, что показатель согласованности по РФ принадлежит доверительным интервалам расхождений экспертов во всех субъектах РФ этой подгруппы (малые регионы с низким средним баллом) либо превышает правую границу этих доверительных интервалов.

Итоговые кластеры формируются на основе полученных подгрупп разбиений (с возможным объединением малых сопоставимых подгрупп). Например, могут быть объединены малые и средние регионы с сопоставимым уровнем со-

гласованности экспертов внутри одной подгруппы по среднему баллу за задания с развёрнутым ответом.

Выводы

При кластеризации регионов для перекрёстной проверки используются следующие показатели:

- среднее расхождение экспертов на оцениваемый балл (доверительный интервал по региону);
- средний балл за часть с развёрнутым ответом;
- объём выборки (количество проверяемых работ в регионе).

Перекрёстная проверка обеспечивает объективность и прозрачность процедуры ЕГЭ. При использовании представленной кластеризации должна обеспечиваться сопоставимость и согласованность проверки работ, вследствие чего сокращается количество дополнительных экспертных проверок.