

УДК: 004.9:37, 519.237.8

Применение статистических методов для выявления аномалий в результатах экзаменов на уровне образовательной организации

Е. В. Шлипаков^{1,2,a}, И. А. Утешев^{1,2}, М. М. Аркушин¹,
В. А. Грянченко^{1,2}, Д. Е. Щербаков^{1,2}, И. В. Ященко^{1,2}

¹Московский физико-технический институт (национальный исследовательский университет),
Россия, 117303, г. Москва, ул. Керченская, д. 1а, корп. 1

²Центр педагогического мастерства,
Россия, 119270, г. Москва, ул. Хамовнический вал, д. 6

E-mail: ^a shlipakov.ev@phystech.edu

Получено 11.02.2026, после доработки — 08.04.2026.

Принято к публикации 08.04.2026.

В работе рассматривается задача составления методики выявления аномалий с помощью методов математической статистики в результатах экзаменов на примере Основного государственного экзамена (ОГЭ) по математике 2023–2024 учебного года. Актуальность исследования обусловлена отсутствием обязательного видеоконтроля на ОГЭ в рассматриваемый период, что создает предпосылки для потенциальных процедурных нарушений как со стороны отдельных учащихся, так и в масштабах целых образовательных организаций. На основе анализа распределений первичных баллов были выявлены характерные особенности учебных заведений, свидетельствующие о возможной нечестности при проведении экзамена, в частности резкие скачки функции распределения в области перехода между неудовлетворительной и удовлетворительной оценками. С целью определения наиболее подозрительных результатов были построены два критерия аномальности. Первый критерий основан на сравнении величины скачка эмпирической функции распределения результатов школы с общерегиональным уровнем и позволяет выделить 47 организаций с аномально высокими значениями. Для построения второго (общего) критерия проведено сравнение баллов, полученных учениками школы на ОГЭ и диагностической работе по математике, проведенной в 8 классе с использованием видеоконтроля. Проведение такой аналогии является корректным, так как контингент обучающихся, принимавших участие в каждой из работ, практически совпадает. Данный подход позволяет сузить число выявленных аномалий, отделив те, которые более похожи на нарушения протоколов, от вызванных особенностями конкретного набора обучающихся и их подготовки к экзаменам в определенном учебном заведении. В результате применения одноклассового метода опорных векторов выявлены 12 образовательных организаций с нехарактерными аномальными результатами. Предложенная методика позволяет обнаруживать потенциальные случаи нечестного поведения при проведении экзамена и может быть использована для предотвращения нарушений протоколов с помощью проведения адресной профилактической работы с учебными заведениями.

Ключевые слова: выявление аномалий, статистический анализ, эмпирическая функция распределения, одноклассовый метод опорных векторов, ОГЭ, выявление нарушений

UDC: 004.9:37, 519.237.8

Statistical methods for detecting anomalies in examination results at the institutional level

E. V. Shlipakov^{1,2,a}, I. A. Uteshev^{1,2}, M. M. Arkushin¹, V. A. Gryanchenko^{1,2},
D. E. Shcherbakov^{1,2}, I. V. Yashchenko^{1,2}

¹Moscow Institute of Physics and Technology,
1a/1 Kerchenskaya st., Moscow, 117303, Russia

²Centre for Teaching Excellence,
6 Khamovnicheskij val, Moscow, 119270, Russia

E-mail: ^a shlipakov.ev@phystech.edu

Received 11.02.2026, after completion — 08.04.2026.

Accepted for publication 08.04.2026.

This study proposes a methodology for anomaly detection in educational assessment data, demonstrated on the case of the 2023–2024 Basic State Exam (BSE) in mathematics in Russia. The relevance of the study is related to the absence of mandatory video surveillance during the examination period, which creates a risk of potential rule violations both by individual students and by entire educational institutions. By analyzing the distribution of primary scores, we identify a big spike in the area between grades 2 and 3 as a specific pattern in results that may indicate cases of cheating during the exam. To determine the most suspicious results, two anomaly criteria were constructed. The first criterion relies on comparing the magnitude of the spike in empirical distribution function in school's results with the corresponding regional average level. This criterion made it possible to identify 47 educational institutions with abnormally high values of the spike. The second (general) criterion was derived from comparing students' scores on the examination with their performance on a diagnostic mathematics test conducted in grade 8 under video surveillance. This comparison is appropriate because almost the same group of students took part in both assessments. This approach helps reduce the number of detected anomalies by distinguishing those more likely to reflect actual protocol violations from those arising due to the specific characteristics of a particular student population and their exam preparation within a given educational institution. The application of the one-class support vector machine method enabled the identification of 12 schools with atypical anomalous results. The proposed methodology could be useful for the detection of potential cases of cheating during exams and the development of methods for preventing such behavior. In particular, it can be used to support targeted preventive work with specific schools in order to reduce the risk of exam rule violations.

Keywords: anomaly detection, statistical analysis, empirical distribution function, one-class support vector machine, Basic State Examination, detection of cheating

Citation: *Computer Research and Modeling*, 2026, vol. 18, no. 2, pp. 537–552 (Russian).

© 2026 Egor V. Shlipakov, Ivan A. Uteshev, Maksim M. Arkushin, Vladimir A. Gryanchenko, Dmitrii E. Shcherbakov, Ivan V.

Yashchenko

This work is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/3.0/>
or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1. Введение

Выпускные экзамены являются неотъемлемой частью современного школьного образования в России. В настоящее время централизованно проводятся два этапа государственной итоговой аттестации: Основной государственный экзамен (ОГЭ) после окончания 9 класса и Единый государственный экзамен (ЕГЭ) после 11 класса. Результаты этих экзаменов во многом определяют дальнейшую образовательную траекторию выпускников: поступление в профильные классы и школы, организации среднего профессионального и высшего образования.

Централизованная сдача экзаменов с последующей двойной слепой (анонимизированной) проверкой обеспечивает сопоставимость результатов учащихся в равных условиях и снижает влияние субъективных факторов при выставлении оценок. Вместе с тем остается актуальной задача предотвращения процедурных нарушений, которые могут происходить как со стороны отдельных обучающихся, так и со стороны организаторов и учебных заведений. Одним из эффективных способов снижения соответствующих рисков является организация видеоконтроля в пунктах проведения экзамена. Данная практика на протяжении нескольких лет применяется при проведении ЕГЭ, но вплоть до 2025 года не была востребована на ОГЭ. Вследствие этого можно ожидать более высокой частоты нарушений при проведении ОГЭ — от индивидуального списывания до возможных массовых искажений результатов.

Выявление аномалий в данных — одно из актуальных направлений применения статистики и методов машинного обучения в различных сферах науки и жизни. При решении подобных задач выделяют объекты, поведение которых существенно отличается от ожидаемого. Такая информация может быть использована как для исключения подобных выбросов с целью повышения качества анализа системы в целом, так и для изучения аномалий как таковых и выявления их природы.

В педагогическом контексте случаи нечестного поведения во время экзамена (при условии, что их количество невелико по сравнению с общим объемом данных) можно трактовать как аномальные. Их возникновение зачастую связано с нарушениями процедуры проведения экзамена, например списыванием. Однако следует отметить, что статистические методы выявления аномалий не позволяют определить точные причины возникновения подобных отклонений в распределении. По этой причине их результаты стоит использовать как основание для проведения дополнительной проверки, а не как констатацию факта наличия нарушений. Существует ряд исследований, которые выявляют причины такого поведения [Borge, 2004; Borisova, Peresetsky, 2016; Шмелева, 2016]. Согласно им, важную роль играют условия выполнения работы и формат ее проведения [Borisova, Peresetsky, 2016] (очно или дистанционно, с наличием видеонаблюдения или без), уровень успеваемости и стресса обучающихся [Borge, 2004], а также их мотивации и заинтересованности в получаемой профессии [Шмелева, 2016]. Подобные случаи списывания во время контрольных работ и экзаменов отрицательно сказываются на уровне образования учащихся. Это позволяет им получить более высокие результаты, которые не соответствуют их реальным знаниям. Поэтому важно разработать методы, которые позволят эффективно выявлять случаи академической нечестности, что позволит принимать меры по предотвращению подобных ситуаций в дальнейшем.

В зависимости от условий исследовательской задачи для ее решения могут применяться различные методы выявления аномального поведения. Важной особенностью данного раздела машинного обучения является, что в общем случае при анализе реальных данных заранее предельно не известно, какие именно объекты проявляли аномальное поведение. В таком случае необходимо либо экспертным образом разметить данные для последующего применения классических методов классификации, либо использовать алгоритмы машинного обучения без учителя.

В работе [Erdem, Karabatak, 2025] исследовались результаты сдачи нескольких онлайн-экзаменов в Университете Битлис Эрен (BEU) в Турции. Все задания были решены группой

экспертов, после чего на основании среднего затраченного ими времени был зафиксирован факт списывания: если студент потратил на задание меньше времени, чем вычисленное критическое значение, то результаты данного задания считались подозрительными. Авторы статьи обучили и сравнили различные модели машинного обучения (глубокая нейронная сеть, машина опорных векторов, решающие деревья, случайный лес, бустинг, логистическая регрессия и метод ближайших соседей). Лучшие результаты решения задачи классификации работ показали нейронная сеть и бустинг, их точность составила выше 96%. В статье [Alsabhan, 2023] авторы работали в основном с искусственно сгенерированными и уже размеченными данными о результатах нескольких контрольных работ студентов. Для предсказания факта списывания использовалась рекуррентная нейронная сеть с двумя LSTM-слоями, которая достигла точности предсказания 92% на валидационной выборке. В работах [Cavalcanti et al., 2012; Alexandron et al., 2019] также содержатся результаты применения и сравнение методов машинного обучения с учителем для выявления аномалий в образовательных данных, которые определенным образом заранее были размечены экспертами. Помимо возможности явного определения случаев списывания, данные методы позволяют оценить, какие признаки и статистики являются наиболее важными и информативными при решении данной задачи [Ranger et al., 2020].

В работе [Jacob, Levitt, 2003] исследовались результаты теста ITBS (Iowa Test of Basic Skills) для учеников 3–8 классов за семь лет. Аномальные результаты учащихся заранее никак не размечались. С целью выявить нечестное поведение со стороны преподавателей (например, диктовка ответов в классе) авторы разработали два статистических критерия определения аномальности результатов. Первый из них учитывает только полученные учащимися баллы и сравнивает результаты одного класса за два последовательных года. Высокие значения статистики, превосходящие посчитанные выборочные квантили, свидетельствуют о заметном снижении уровня результатов класса по сравнению с предшествующим годом, что в рамках данной работы считалось аномальной флуктуацией. Также авторы исследования всесторонне рассмотрели ответы учащихся на предмет наличия нетипичного поведения и сильных корреляций внутри класса и составили формулу статистики аномальности ответов из четырех слагаемых. В результате применения данных статистических методов были получены выводы, что потенциальные нарушения процедуры со стороны преподавателей наблюдаются в 4–5% классов ежегодно. Задача получения надежной статистики для выявления случаев возможного списывания и копирования ответов во время проведения экзамена ставится также и в других исследованиях последних лет [Belov, 2011; Belov, Wollack, 2021].

Статистические методы выявления нечестного поведения во время экзамена также рассматривались в статье [Dee et al., 2019]. Авторы анализируют распределение результатов выпускного экзамена из старшей школы в Нью-Йорке, отметив необычный скачок в районе порогового значения для успешной сдачи. Для оценки формы ожидаемого распределения данная аномальная зона была исключена, а по оставшимся данным построена ее полиномиальная аппроксимация. В результате проведенного сравнения полученной оценки с реальными результатами авторы сделали вывод о том, что около 40% результатов в данной зоне порогового значения были искусственно завышены при проверке. Также было отмечено, что после внесения существенных изменений в процедуру проверки (например, организация внешней проверки экзаменационных работ) доля работ с потенциально завышенными результатами сократилась на 80%, что также подтвердило эффективность рассмотренных методов выявления аномалий.

В исследовании [Кузнецов, Семенов, 2017] также рассматривались образовательные данные, в которых авторы заранее не производили разметку аномальных результатов. В рамках исследования было проведено сравнение работы трех различных методов выявления аномалий на нескольких независимых выборках учащихся: метод эллиптической огибающей, одноклассовый метод опорных векторов и метод изолирующих деревьев. В качестве метрики качества

работы моделей было выбрано среднее квадратичное отклонение полученного набора выбросов от полной выборки. Методы опорных векторов и изолирующих деревьев показали схожие высокие результаты по выявлению аномалий в выборках среднего балла, общего числа оценок, числа неудовлетворительных оценок и пропусков занятий.

В статьях [Kamalov et al., 2021; Feng, Zhang, 2024; Bulut et al., 2024] авторы также работают с методом изолирующих деревьев для выявления аномалий в данных, содержащих результаты образовательного процесса учащихся. Во всех трех работах данные были предварительно преобразованы с помощью рекуррентных нейронных сетей для снижения размерности до нескольких значимых характеристик (например, предсказания итоговой оценки за учебный курс), которые в дальнейшем и подавались на вход алгоритму изолирующих деревьев. В работе [Dillon, De Marsico, 2025] авторы создают искусственный датасет для обучения моделей, способных определить аномальное поведение по характеру набора текста во время онлайн-экзаменов. Было проведено сравнение работы изоляционного леса ($\text{recall} = 83\%$) и машины опорных векторов ($\text{recall} = 67\%$). При этом при совместном использовании обоих методов авторы исследования добились полного отсутствия ложноотрицательных ответов. Данные методы машинного обучения без учителя также используются в работах [Pan et al., 2022; Smith et al., 2025] для выявления случаев академической нечестности, связанных с заранее известными для участников ответами на тест.

Как уже было отмечено выше, задача поиска аномалий актуальна для очень многих научных сфер. Рассмотренные методы машинного обучения с учителем и без него активно применяются для выявления редких событий и избежания аномального поведения систем в биологии [Grekov et al., 2023; Peleg et al., 2025], физике элементарных частиц [Crispim Romao et al., 2021], астрономии [Shumilo, 2020; Ishida et al., 2021], сенсорной и спутниковой навигации [Шолохова, 2017; Ладанова, Юдин, 2023] и многих других разделах науки.

2. Анализируемые данные

Для проведения работы были собраны и обработаны результаты ОГЭ выпускников 9 класса 2023/2024 учебного года. В данных имеется информация о предмете сдачи, первичном и вторичном баллах, а также учебном заведении Москвы, в котором обучается экзаменуемый. Последнее необходимо для проведения сравнительного анализа образовательных организаций друг с другом и выявления аномалий в результатах обучающихся на уровне учебных заведений.

Общее распределение первичных баллов по математике в Москве представлено на рис. 1. Максимальный балл, который можно было набрать на экзамене, равен 31. Пунктирными линиями обозначены минимальные баллы, которые необходимо было набрать, чтобы вторичный балл составил 3, 4 и 5 соответственно. Стоит отметить, что пик распределения находится на 19 первичных баллах. Этот результат соответствует максимуму за тестовую часть данного экзамена. Данный факт можно объяснить тем, что не все экзаменуемые пожелали приступить ко второй части экзамена, требующей предоставления подробного решения. Также отличительной чертой данного распределения является скачок в виде ступеньки на 8 баллах (минимальный балл на оценку 3 за экзамен). Среди вероятных причин такого поведения графика может быть специальная подготовка образовательными организациями учеников к решению определенного числа самых простых заданий для успешной сдачи экзамена. Однако это также может свидетельствовать и о какой-либо нечестности при проведении работы. По этой причине в дальнейшем нас будет интересовать именно такой паттерн распределения результатов.

Рассмотрения исключительно результатов одного экзамена недостаточно для получения информации о всей образовательной организации. Для того чтобы сделать полноценные выводы об аномальности результатов учащихся заведения, необходимо провести сравнения с результатами других работ по данному предмету. С этой целью были исследованы результаты сдачи

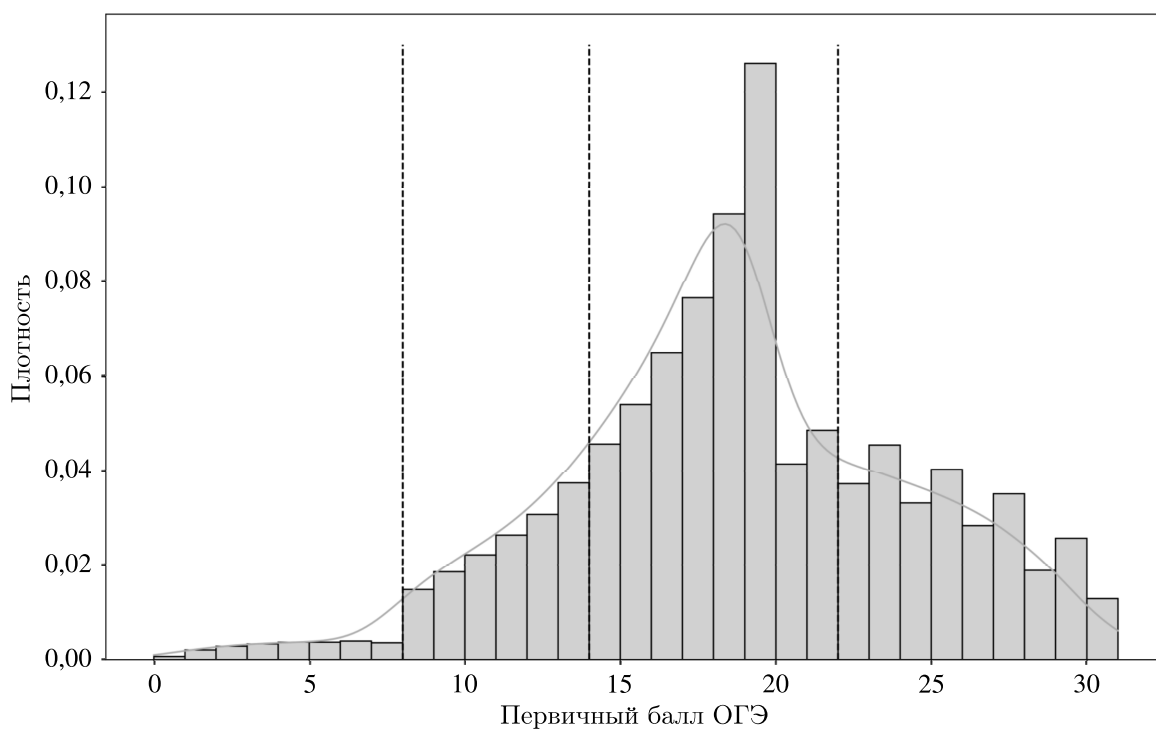


Рис. 1. Распределение первичных баллов ОГЭ по математике в Москве

данным набором учеников общей диагностической работы по математике в 8 классе. Во время проведения данной работы было организовано видеонаблюдение, исключающее возникновение аномалий по причине массовых нарушений протоколов проведения. Аналогично результатам ОГЭ данные содержат предмет сдачи, первичные и вторичные баллы и информацию об образовательной организации обучающегося.

На рис. 2 представлена гистограмма распределения первичных баллов данной диагностической работы в регионе. Суммарно участники могли набрать не более 10 первичных баллов. На данном графике пунктирные линии также обозначают минимальные баллы, необходимые для получения итоговых оценок 3, 4 и 5. Пик распределения приходится на 3 балла, что является граничным баллом на удовлетворительную оценку. Отчетливого резкого скачка на границе 2 и 3 баллов по пятибалльной шкале в данной работе не наблюдается. Этот факт является еще одной предпосылкой к проведению подробного анализа, так как мы видим достаточно сильное различие в характере распределения результатов экзамена без видеоконтроля и диагностической работы с ним.

3. Обработка и кластеризация результатов образовательных организаций

Для составления корректных критериев аномальности результатов экзамена необходимо провести кластеризацию образовательных организаций, чтобы в первом приближении рассмотреть отличия в поведении и выделить самые яркие отклонения от общей картины.

На первом этапе необходимо выделить признаки в данных, на основе которых будет производиться сравнительный анализ результатов образовательных организаций. В связи с тем что в Москве присутствует более пятисот школ, в некоторых из которых обучается большое число детей, их результаты могут сильно отличаться друг от друга по характеру распределения, образуя

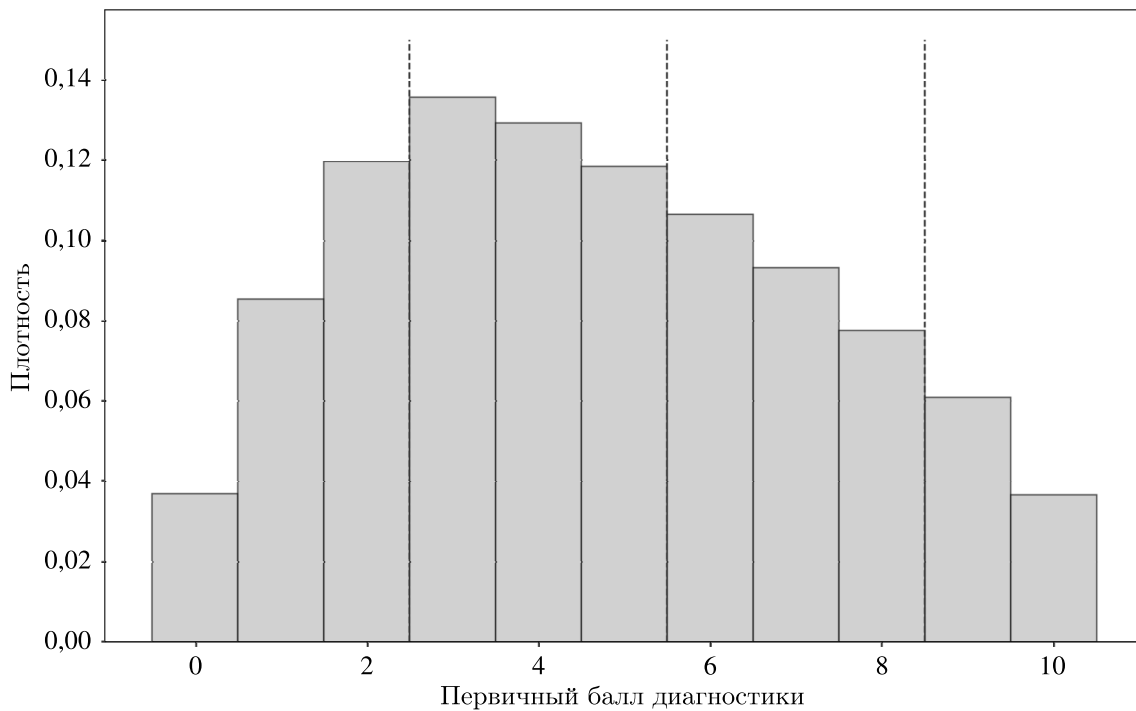


Рис. 2. Распределение первичных баллов диагностической работы за 8 класс по математике в Москве

несколько совершенно не похожих друг на друга небольших групп. Поэтому было решено разделить учебные заведения на достаточно больше число кластеров. Для корректного разделения на такое количество групп необходимо выделить достаточно много признаков, описывающих результаты отдельной образовательной организации. В качестве таких анализируемых признаков были выбраны значения эмпирической функции распределения (далее — ЭФР). То есть для каждой организации по результатам экзамена был посчитан набор из $M - 1$ параметров $\widehat{F}_k = \frac{1}{N} \sum_{n=0}^N I\{X_n \leq k\}$, где M — максимальный первичный балл на экзамене, $k = 0, \dots, M - 1$ — каждый возможный балл, кроме максимального, N — число экзаменуемых учеников, X_n — первичный балл n -го ученика, а функция I является индикатором выполнения неравенства, являющегося ее аргументом. Подбор таких признаков позволяет наиболее полно сравнить образовательные организации по виду распределения и разбросу результатов экзаменуемых.

Кластеризация школ производилась с помощью метода k-средних (KMeans), который итеративно выделяет сферические кластеры в данных на основе L2-расстояния между точками и центроидами. Отметим важный факт, что в нашей задаче кластеризация носит функцию выявления небольших групп образовательных организаций с нестандартным аномальным поведением для определения их основных отличий от общей массы. Таким образом, следует производить разделение школ на достаточно большое число кластеров, рассмотрев общую структуру данных, например, с помощью метода снижения размерности t-SNE (t-distributed Stochastic Neighbor Embedding), позволяющего наглядно разделить менее схожие объекты, при этом сохранив локальную структуру данных внутри выделенных скоплений. На рис. 3 представлена визуализация результатов снижения размерностей массива значений ЭФР всех образовательных организаций методом t-SNE со значением перплексии, равным 10. Перплексия отвечает за количество соседних точек, которые учитываются при анализе схожести, рекомендуемый диапазон значений — от 5 до 50. Таким образом, в рамках решаемой нами задачи необходимо установить достаточно низкое значение перплексии, которое позволит акцентировать внимание на локальной структу-

ре небольших кластеров, а не общей картине. Проанализировав график, мы можем заметить, что в данных присутствует один небольшой кластер, сильно отличающийся от остальных. При этом в основном скоплении точек также присутствует некоторая локальная структура, в которой можно визуально выделить до 9 более плотных скоплений точек.

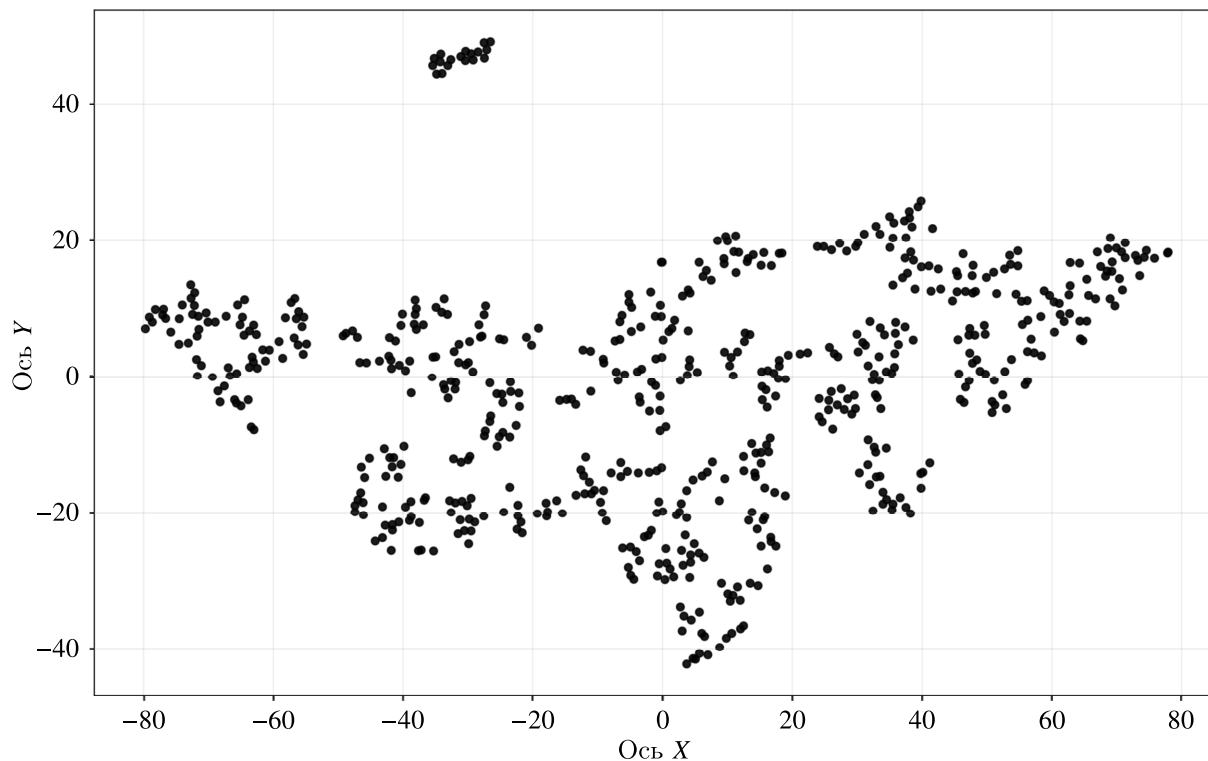


Рис. 3. Визуализация общей структуры массива значений ЭФР всех образовательных организаций с помощью метода снижения размерностей t-SNE

В результате проведенного анализа структуры данных было принято решение разделить образовательные организации с помощью метода k-средних на 10 кластеров. На рис. 4 представлены графики эмпирической плотности распределения результатов выделенных кластеров. Здесь также пунктирной линией обозначены минимальные баллы, соответствующие оценкам 3, 4 и 5 по пятибалльной шкале. По данному графику можно явно определить, что в образовательных организациях из кластеров 4 и 9 результаты экзаменов заметно выше и имеют пик в зоне оценки 5. Также отметим, что у данных кластеров в области высоких баллов наблюдается необычное поведение в виде «гребенки». Это объясняется тем, что во второй части экзамена все задачи оцениваются в два первичных балла. Однако для нас больший интерес представляют кластеры, которые имеют самые заметные скачки плотности на переходе между оценками 2 и 3 — кластеры 0 и 5. На рис. 5 отдельно представлены их эмпирические плотности распределений баллов на ОГЭ. У образовательных организаций, попавших в данные кластеры, наблюдается некоторое плато в зоне неудовлетворительных оценок, которое совершает резкий скачок в районе 8 первичных баллов, после чего происходит медленный (по сравнению с остальными кластерами) рост до пика в районе 18–19 баллов. Также отметим, что два подобных кластера с нетипичными скачками выделяются и в случае кластеризации на меньшее число групп (от 6 до 9). Поэтому рассмотрение максимально большого значения числа кластеров, на которые можно визуально разделить данные на рис. 3, никак не исказило выводы о характерном виде аномальных графиков эмпирической плотности распределения.

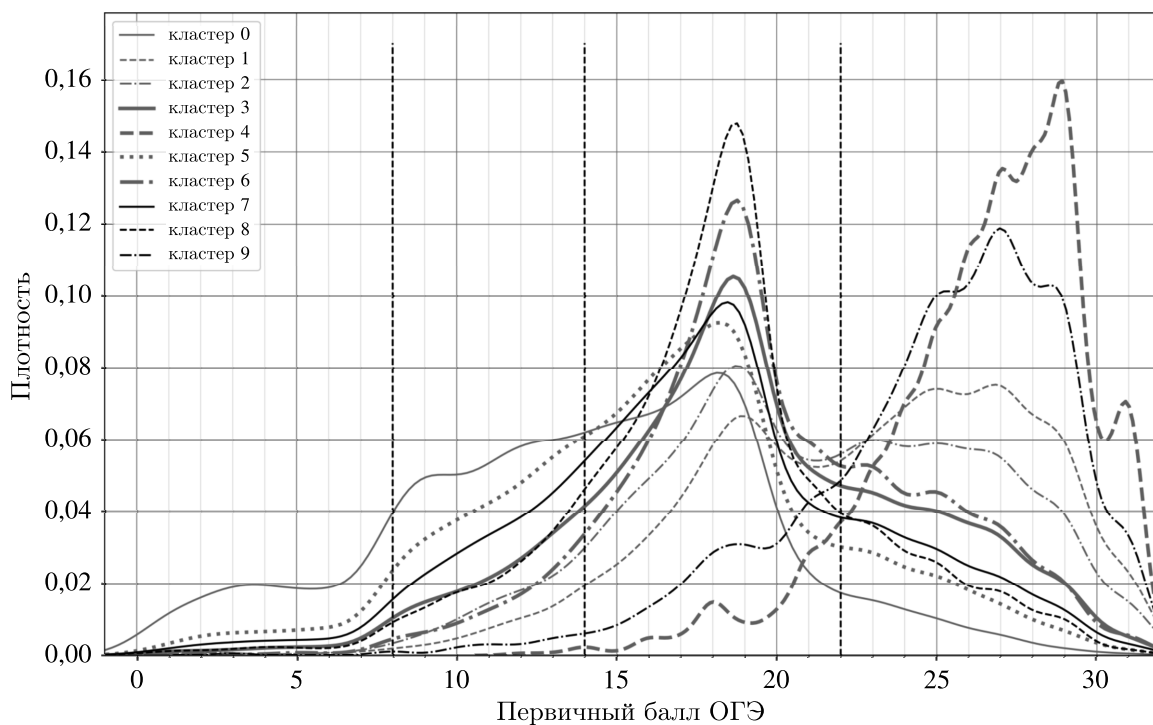


Рис. 4. Эмпирические плотности распределения первичных баллов ОГЭ по математике всех кластеров

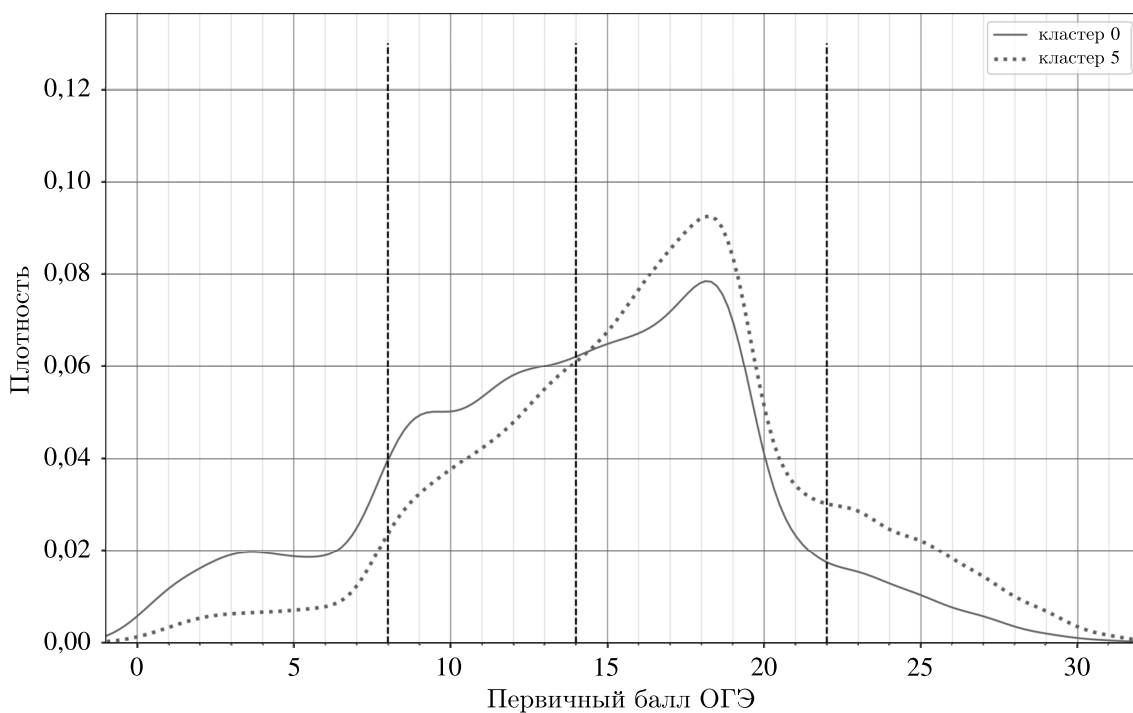


Рис. 5. Эмпирические плотности распределения первичных баллов ОГЭ по математике выделенных кластеров с подозрением на аномальное поведение

Также рассмотрим графики ЭФР, по значениям которой проводилась кластеризация. На рис. 6 построены графики ЭФР всех школ из кластеров 0 и 5 с выявленными аномалиями и кластера 2, выбранного нами в качестве эталонного для проведения визуального сравнения. На всех

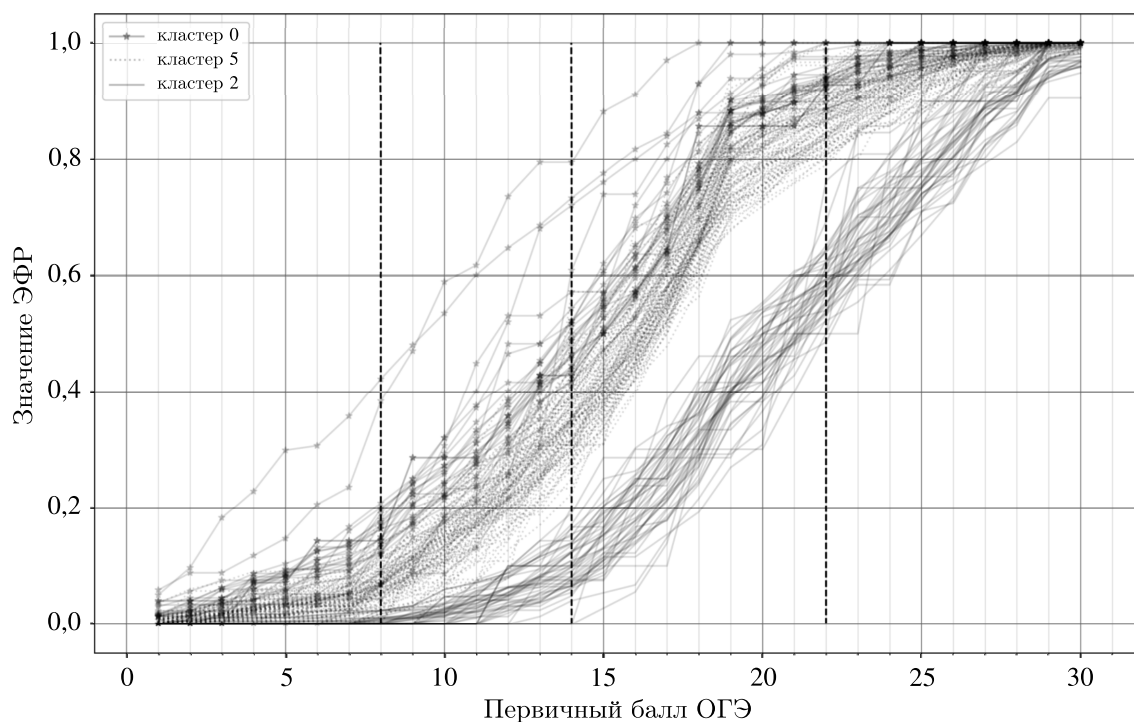


Рис. 6. Эмпирические функции распределения первичных баллов ОГЭ по математике кластеров 0, 5 и 2

трех группах ЭФР наблюдается излом между 19 и 20 баллами, что соответствует резкому снижению плотности после прохождения пика распределения. При этом в кластерах 0 и 5 наблюдается обратный излом в районе между 7 и 8 первичными баллами, что как раз и свидетельствует о наличии некоторого скачка плотности при переходе между зонами оценок 2 и 3 по пятибалльной шкале.

4. Построение простого критерия аномальности результатов экзамена

Исходя из представленных выше наблюдений и полученных выводов, в качестве метрики аномальности результатов учащихся образовательной организации может быть использована разница значений ЭФР между точками с двух сторон от границы, разделяющей оценки 2 и 3. Вычтя из этих значений общую разницу ЭФР всех учебных заведений, получим удобную для сравнения с общим распределением по региону характеристику школы $cr_1^k = F_{10}^k - F_6^k - \delta_{10,6}^{reg}$, которая будет являться статистикой первого полученного нами критерия аномальности. Здесь k — порядковый номер, отвечающий за каждую рассмотренную образовательную организацию Москвы, F_i^k — значение ЭФР в точке, соответствующей i первичным баллам, для k -го учебного заведения, а $\delta_{10,6}^{reg}$ — значение разницы ЭФР, вычисленной по результатам всех учеников региона, между 10 и 6 первичными баллами. Если у школы значение данной статистики отрицательно, это означает, что интересующий нас скачок баллов либо отсутствует, либо ниже среднего по региону. Положительное же значение cr_1^k будет свидетельствовать о более резком росте плотности на границе оценок 2 и 3, и чем выше будет это значение, тем с большей уверенностью можно будет утверждать об аномальности распределения результатов в данной организации.

Рисунок 7 содержит график распределения данной статистики по всем рассмотренным образовательным организациям. Проверка с помощью критериев согласия Колмогорова – Смирнова и Андерсона – Дарлингга отвергла гипотезы о равенстве эмпирической функции распределения данной величины с функциями распределений большинства классических непрерывных распределений. По этой причине для расчета критического значения, выше которого лежат наиболее

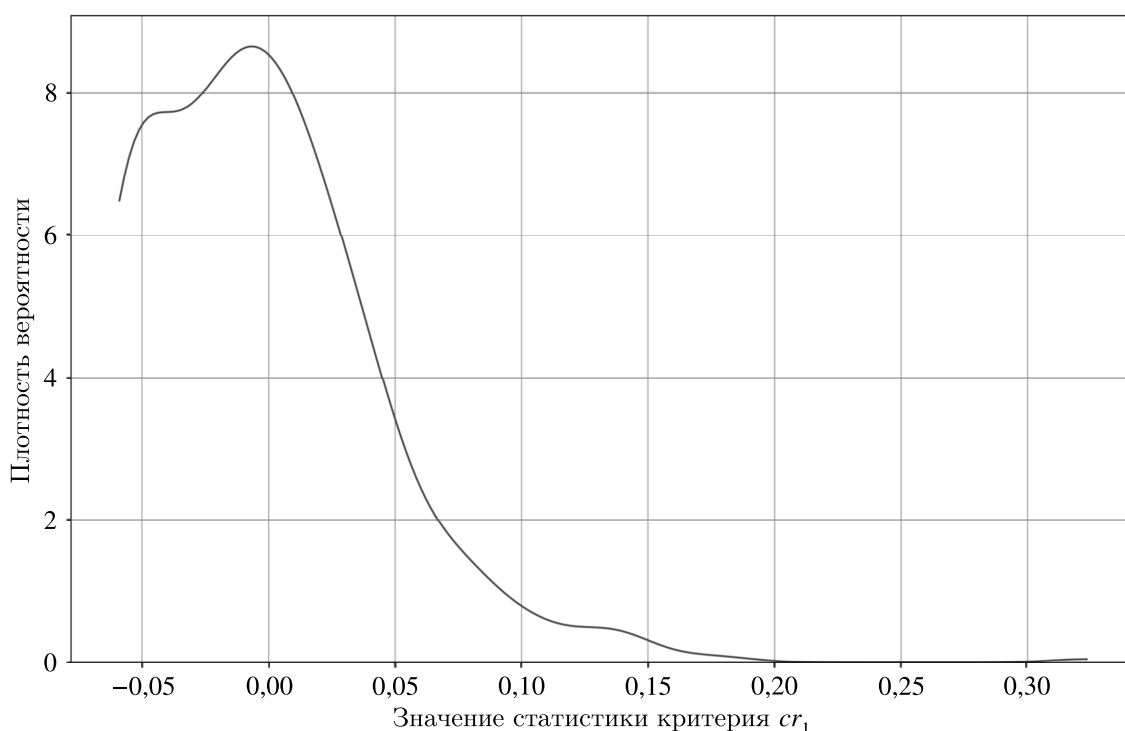


Рис. 7. Плотность распределения статистики критерия cr_1

экстремальные значения выборки, используется выборочный квантиль уровня 0,9. При условии непрерывной дифференцируемости функции распределения в окрестности квантиля u_p уровня p и не равенства нулю плотности распределения f в точке u_p можно утверждать о том, что выборочный квантиль \hat{u}_p является асимптотически нормальной оценкой теоретического квантиля u_p с асимптотической дисперсией $\frac{p(1-p)}{f(u_p)^2}$ [Van der Vaart, 1998]. Значение выборочной квантили уровня 0,9 распределения величины cr_1 равно примерно 0,053. Таким образом, построенный критерий принимает следующий вид:

$$S_1 = \{cr_1^k \geq 0,053\}. \tag{1}$$

5. Построение общего критерия аномальности

Как уже было отмечено в § 2, утверждать о наличии аномалий в результатах ОГЭ образовательной организации только по построенному выше критерию S_1 некорректно. Поэтому нами было решено рассмотреть у данного набора учеников результаты сдачи диагностической работы по математике в 8 классе с организованным видеоконтролем. Это необходимо для проведения сравнения характера распределений результатов двух схожих работ: если в учебном заведении, попадающем в простой критерий аномальности по результатам ОГЭ, наблюдаются такие же паттерны поведения на диагностической работе в 8 классе, то это будет считаться не аномалией, а особенностью данного заведения. Таким образом, нашей целью является получение второй статистики критерия аномальности, которая будет отвечать за сравнение результатов экзамена и диагностической работы, а затем формирование с ее помощью общего критерия аномальности.

Таким образом, общий критерий аномальности, по сравнению с простым, должен дополнительно проводить фильтрацию образовательных организаций, отличительной особенностью которых является большой скачок плотности распределения результатов в области перехода оценок 2 и 3 в обеих работах — диагностической работе за 8 класс и ОГЭ по математике. Соответственно, вторую статистику критерия необходимо построить в полной аналогии с cr_1 , как

разницу скачков ЭФР результатов диагностической работы отдельной школы и всего региона в целом. Основываясь на графике распределения результатов диагностики, представленном на рис. 2, были выбраны граничные значения для данного скачка, равные 2 и 3 первичным баллам. В итоге вторая статистика критерия аномальности имеет следующий вид: $cr_2^k = G_3^k - G_2^k - \delta_{3,2}^{reg}$, где G_i^k — значение ЭФР результатов диагностики в точке, соответствующей i первичным баллам, для k -го учебного заведения, а $\delta_{3,2}^{reg}$ — общее значение разницы ЭФР, посчитанной по результатам всех учащихся Москвы, между 3 и 2 первичными баллами.

На рис. 8 содержится график совместного распределения статистик cr_1 и cr_2 по всем рассмотренным учебным заведениям Москвы. Заметим, что большая часть точек на графике образуют скопление в виде клина под некоторой линейной функцией с положительным коэффициентом наклона. Это свидетельствует о том, что во многих школах рассмотренный нами скачок ЭФР в зоне оценок 2 и 3 за диагностику больше скачка в результатах ОГЭ или примерно равен ему. Однако важно отметить наличие нескольких выбросов в зоне над основным скоплением. Эти точки соответствуют учебным заведениям, которые в рамках нашей работы считаются подозрительными. Результаты диагностики их учеников не сильно отличаются от большинства школ и имеют низкий скачок в зоне оценок 2 и 3, но при этом в результатах ОГЭ это изменение ЭФР сильно выше, чем у остальных. Вторая часть общего критерия аномальности должна выявлять именно такие случаи.

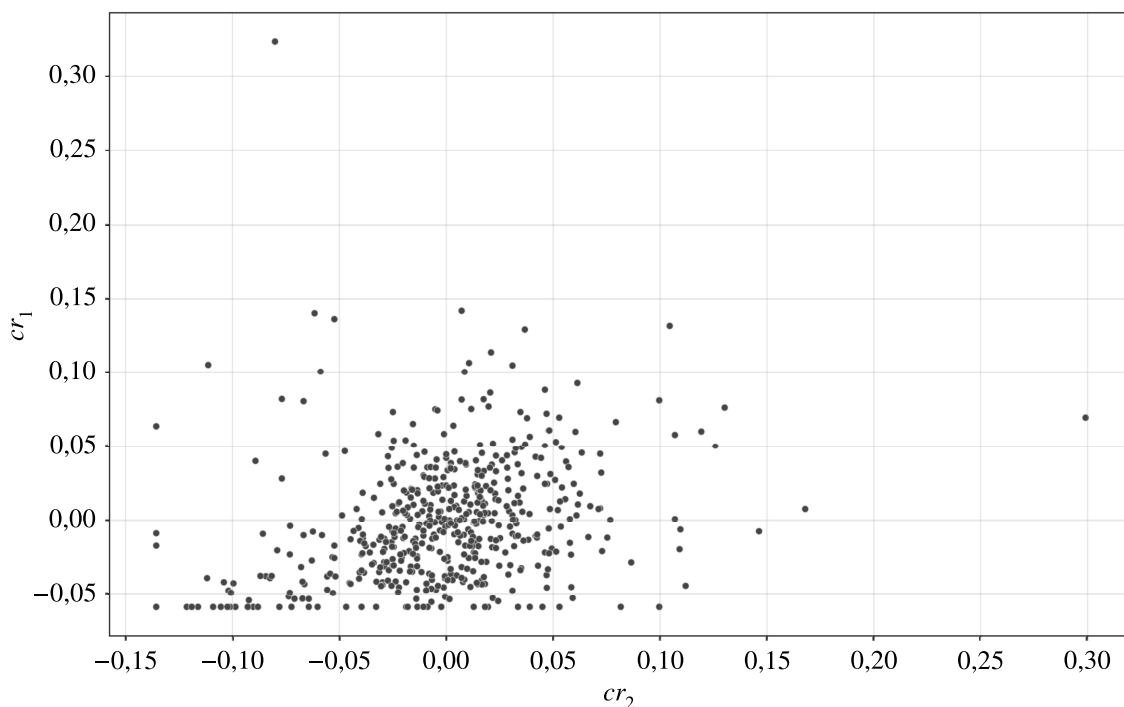


Рис. 8. Взаимное распределение статистик cr_1 и cr_2 критерия аномальности

Как было отмечено выше, основное скопление точек похоже на клин под некоторой прямой. По этой причине логичным решением данной задачи будет нахождение данной линейной функции, которая отделит точки с аномальным поведением от остальных. Для этого воспользуемся одним из наиболее популярных методов выявления аномалий, упомянутых выше, — одноклассовым методом опорных векторов, который строит линейную гиперплоскость в пространстве признаков, наилучшим образом отделяющую выбросы от основной массы объектов [Vapnik et al., 1996; Scholkopf, Smola, 2001; Muller et al., 2001]. Обучив модель одноклассового метода

опорных векторов с линейным ядром и нижней границей доли опорных векторов $\nu = 0,5$ на имеющихся данных, мы получили следующее уравнение разделяющей прямой — $y = 0,835x + 0,096$. На рис. 9 содержится график совместного распределения статистик критерия с двумя разделяющими прямыми. Сплошная линия отвечает за разделяющую гиперплоскость, построенную с помощью одноклассового метода опорных векторов. Пунктирная линия соответствует простому критерию аномальности результатов ОГЭ.

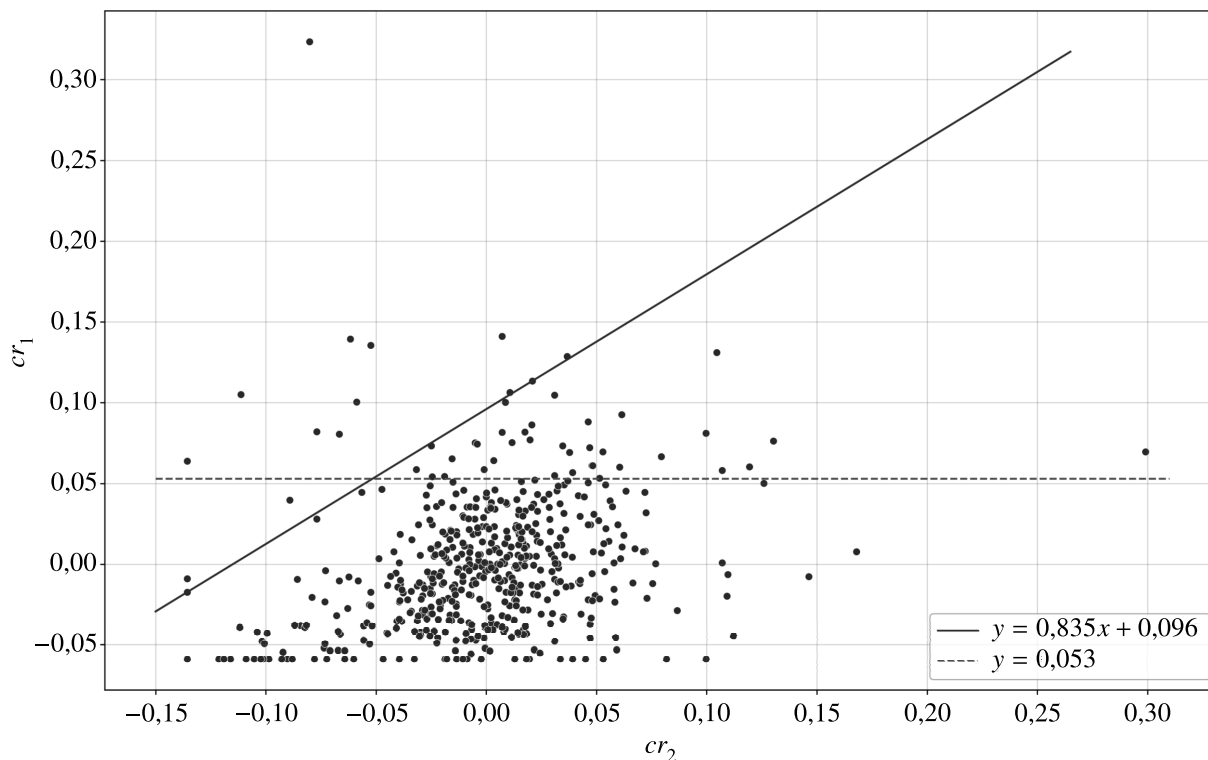


Рис. 9. Взаимное распределение статистик cr_1 и cr_2 с разделяющими прямыми общего критерия аномальности

По логике построения, в общий критерий аномальности должны попадать точки, которые лежат над обеими прямыми на рис. 9. Таким образом, он будет принимать следующий вид:

$$S_2 = \{cr_1^k \geq 0,053\} \cup \{cr_1^k \geq 0,835cr_2^k + 0,096\}. \tag{2}$$

В рассматриваемом нами случае в данный критерий попадает 12 точек, обозначающих учебные заведения с аномальными результатами ОГЭ по математике, которые не соответствуют паттерну их результатов на диагностической работе в 8 классе и поэтому считаются подозрительными. При этом отметим, что 10 из них в результате кластеризации ЭФР (рис. 4) попали в кластеры 0 и 5, которые имели самые заметные резкие скачки в зоне перехода между оценками 2 и 3.

Как и в работе [Кузнецов, Семенов, 2017], для оценки качества выявления аномалий в данных были посчитаны среднеквадратичные отклонения:

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

Для всей выборки оно составило 0,063, а для выявленных общим критерием аномальных школ — 0,165. При этом если убрать из рассмотрения все выбросы, то среднеквадратичное от-

клонение для оставшихся объектов становится 0,06. Таким образом, мы действительно удалили точки, которые в среднем находятся примерно в 2,6 раза дальше от общего центра. Если же рассмотреть составные части общего критерия по отдельности, то точки, попадающие в простой критерий S_1 , имеют среднеквадратичное отклонение 0,128, а точки, лежащие над разделяющей гиперплоскостью метода опорных векторов, — 0,157. Из этого можно также сделать вывод, об осмысленности их объединения в общий критерий.

6. Заключение

В рамках проведенного исследования были построены два критерия для выявления аномалий в результатах ОГЭ по математике на основе общего распределения первичных баллов в образовательной организации в частности и регионе в целом. Было зафиксировано, что для интересующих нас подозрительных результатов характерен большой скачок эмпирической функции распределения в зоне первичных баллов, соответствующей переходу между итоговыми оценками 2 и 3.

Первый простой критерий S_1 основан на сравнении величины данного скачка с общерегиональным уровнем. Аппроксимировав распределение построенной статистики cr_1 , описывающей изменение ЭФР в рассматриваемой области, с помощью нормального распределения, мы определили ее критическое значение, соответствующее 90-й процентилю данного распределения, и составили формулу критерия (1). В результате данный критерий S_1 выделил 47 школ, которые имеют аномально высокое значение статистики, вероятность получения которого составляет менее 0,1.

Однако подобные результаты образовательной организации могут не являться сигналом о нечестности при проведении экзамена, а свидетельствовать о каких-то особенностях контингента или программы подготовки. По этой причине было проведено сравнение с объективными результатами другой работы по математике, в качестве которой была выбрана диагностика в 8 классе, проводящаяся с использованием видеоконтроля. Для сравнения двух работ была построена статистика cr_2 , характеризующая подобный скачок в зоне первичных баллов около границы оценок 2 и 3 за диагностическую работу. С помощью одноклассового метода опорных векторов была построена прямая, наилучшим образом выделяющая выбросы в совместном распределении статистик cr_1 и cr_2 .

Формула (2) общего критерия аномальности S_2 была получена путем объединения простого критерия S_1 и построенной разделяющей гиперплоскости. В результате выделены 12 образовательных учреждений, имеющих аномальные результаты ОГЭ по математике, которые не характерны для школ Москвы и для рассмотренного набора учеников. При среднеквадратичном отклонении всей выборки, равном 0,063, выделенные выбросы расположены на расстоянии 0,165 от среднего (то есть примерно в 2,6 раза дальше).

Таким образом, в рамках работы была разработана методика по выявлению наиболее подозрительных аномальных результатов экзаменов, которые могут свидетельствовать о нечестности при его проведении. Построенные критерии могут быть использованы органами управления образования для обнаружения массовых нарушений процедуры проведения экзамена и разработки мер по предотвращению подобных случаев в будущем. А так как число выявленных аномалий достаточно небольшое, то это позволит работать с данными образовательными организациями адресно.

Список литературы (References)

Кузнецов А. С., Семенов Е. Ю. Выбор метода обнаружения аномалий в образовательных данных // International Journal of Open Information Technologies. — 2017. — Т. 5, № 4. — С. 1–6.

- Kuznetsov A. S., Semenov E. Yu.* Vybor metoda obnaruzheniya anomalij v obrazovatel'nykh dannykh [The choice of anomaly detection method for educational data] // *International Journal of Open Information Technologies*. — 2017. — Vol. 7, No. 4. — P. 1–6 (in Russian).
- Ладанова С. В., Юдин Д. А.* Детекция аномалий для комплексирования данных лидарной и спутниковой локализации транспортного средства // *Труды Московского физико-технического института*. — 2023. — Т. 15, № 3. — С. 59–72.
- Ladanova S. V., Yudin D. A.* Detektsiya anomalij dlya kompleksirovaniya dannykh lidarnoj i sputnikovoj lokalizatsii transportnogo sredstva [Anomaly detection for integrating lidar and satellite vehicle localization data] // *Proceedings of MIPT*. — 2023. — Vol. 15, No. 3. — P. 59–72 (in Russian).
- Шмелева Е. Д.* Плагиат и списывание в российских вузах: роль образовательной среды и индивидуальные характеристики студента // *Вопросы образования*. — 2016. — № 1. — С. 84–109.
- Shmeleva E. D.* Plagiat i spisyvanie v rossiiskikh vuzakh: rol' obrazovatel'noy sredy i individualnykh kharakteristik studenta [Plagiarism and cheating in Russian universities: the role of the learning environment and personal characteristics of students] // *Educational Studies*. — 2016. — No. 1. — P. 84–109 (in Russian).
- Шолохова А. А.* Поиск аномалий в сенсорных данных на примере анализа движения морского судна // *Моделирование, оптимизация и информационные технологии*. — 2017. — № 3 (18). — С. 19.
- Sholokhova A. A.* Poisk anomalij v sensorykh dannykh na primere analiza dvizheniya morskogo sudna [Anomaly detection in sensor data in application to the analysis of maritime vessel motion] // *Modeling, optimization and information technology*. — 2017. — No. 3 (18). — P. 19 (in Russian).
- Alexandron G., Yoo L. Y., Ruiperez-Valiente J. A., Pritchard D. E.* Towards a general purpose anomaly detection method to identify cheaters in massive open online courses // *Proceedings of The 12th International Conference on Educational Data Mining*. — 2019. — P. 480–483.
- Alsabhan W.* Student cheating detection in higher education by implementing machine learning and LSTM techniques // *Sensors*. — 2023. — Vol. 23, No. 8. — P. 4149.
- Belov D.* Detection of answer copying based on the structure of a high-stakes test // *Applied Psychological Measurement*. — 2011. — Vol. 35, No. 7. — P. 495–517.
- Belov D., Wollack J.* Graph theory approach to detect examinees involved in test collusion // *Applied Psychological Measurement*. — 2021. — Vol. 45, No. 4.
- Borge S.* *Academic cheating and stressors at the university level*. — Worcester: Assumption University, 2004.
- Borisova E., Peresetsky A.* Do secrets come out? Statistical evaluation of student cheating // *Applied Econometrics*. — 2016. — Vol. 44, No. 4. — P. 119–130.
- Bulut O., Gorgun G., He S.* Unsupervised anomaly detection in sequential process data // *Zeitschrift für Psychologie*. — 2024. — Vol. 232, No. 2. — P. 74–94.
- Cavalcanti E. R., Pires C. E., Cavalcanti E. P., Pires V. F.* Detection and evaluation of cheating on college exams using supervised classification // *Informatics in Education*. — 2012. — Vol. 11, No. 2. — P. 169–190.
- Crispim Romao M., Castro N. F., Pedro R.* Finding new physics without learning about it: anomaly detection as a tool for searches at colliders // *The European Physical Journal C – Particles and Fields*. — 2021. — Vol. 81, No. 1. — P. 27.
- Dee T. S., Dobbie W., Jacob B. A., Rockoff J.* The causes and consequences of test score manipulation: evidence from the New York regents examinations // *American Economic Journal: Applied Economics*. — 2019. — Vol. 11, No. 3. — P. 382–423.
- Dillon R., De Marsico M.* Behavioral biometrics for remote exam integrity: continuous authenticity assessment via keystroke dynamics // *Procedia Computer Science*. — 2025. — Vol. 274. — P. 402–411.
- Erdem B., Karabatak M.* Cheating detection in online exams using deep learning and machine learning // *Applied Sciences (Switzerland)*. — 2025. — Vol. 15, No. 1. — P. 400.
- Feng B., Zhang L.* Optimizing the isolation forest algorithm for identifying abnormal behaviors of students in education management big data // *Journal of Artificial Intelligence and Technology*. — 2024. — Vol. 4, No. 1. — P. 31–39.

- Grekov A. N., Kabanov A. A., Vyshkvarkova E. V., Trusevich V. V.* Anomaly detection in biological early warning systems using unsupervised machine learning // *Sensors*. — 2023. — Vol. 23, No. 5. — P. 2687.
- Ishida E. E. O., Mondon F., Sreejith S., Kornilov M. V., Malanchev K. L., Pruzhinskaya M. V., Volnova A. A., Korolev V. S., Malancheva A. A., Das S.* Active Anomaly Detection for time-domain discoveries // *Astronomy and Astrophysics*. — 2021. — Vol. 650. — A195.
- Jacob B. A., Levitt S. D.* Rotten apples: an investigation of the prevalence and predictors of teacher cheating // *Quarterly Journal of Economics*. — 2003. — Vol. 118, No. 3. — P. 843–877.
- Kamalov F., Sulieman H., Santandreu Calonge D.* Machine learning based approach to exam cheating detection // *PLoS ONE*. — 2021. — Vol. 16, No. 8. — e0254340.
- Muller K. R., Mika S., Ratsch G., Tsuda K., Scholkopf B.* An introduction to kernel-based learning algorithms // *IEEE Transactions on Neural Networks*. — 2001. — Vol. 12, No. 2. — P. 181–201.
- Pan Y., Sinharay S., Livne O., Wollack J. A.* A machine learning approach for detecting item compromise and preknowledge in computerized adaptive testing // *Psychological Test and Assessment Modeling*. — 2022. — Vol. 64, No. 4. — P. 385–424.
- Peleg O., Raytan M., Borenstein E.* KADAIF: an anomaly detection method for complex microbiome data // *Bioinformatics*. — 2025. — Vol. 41, No. 10. — btaf520.
- Ranger J., Schmidt N., Wolgast A.* The detection of cheating on E-exams in higher education — the performance of several old and some new indicators // *Frontiers in Psychology*. — 2020. — Vol. 11. — 568825.
- Scholkopf B., Smola A.* Learning with kernels: support vector machines, regularization, optimization, and beyond. — Cambridge, MA, USA: MIT Press, 2001.
- Shumilo L.* Automatic anomaly detection methodology for crop classification data using morphological features // 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS). — 2020. — P. 1–5.
- Smith N. R., Keller L. A., Feinberg R. A., Liu C.* Few and different: detecting examinees with preknowledge using extended isolation forests // *Applied psychological measurement*. — 2025.
- Van der Vaart A. W.* Asymptotic statistics. — Cambridge, MA, USA: Cambridge University Press, 1998.
- Vapnik V., Golowich S., Smola A.* Support vector method for function approximation, regression estimation and signal processing // *Proceedings of the 10th International Conference on Neural Information Processing Systems (NIPS'96)*. — 1996. — P. 281–287.