



ОПЫТ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ЭЛЕКТОРАЛЬНЫХ ДАННЫХ

В статье обсуждаются результаты многомерного статистического анализа официальных электоральных данных, представленных в Центральную избирательную комиссию (ЦИК) Российской Федерации (РФ) избирательными комиссиями нижних уровней (избирательные комиссии субъектов РФ (ИК), территориальные избирательные комиссии (ТИК), участковые избирательные комиссии (УИК)) по итогам проведения в 2018 г. выборов президента РФ, которые в настоящее время находятся на сайте ЦИК РФ.

Описаны технология, обеспечивающая автоматическую выгрузку обсуждаемых электоральных данных с сайта ЦИК РФ в файл CVS-формата и структура выгруженных электоральных данных. Представлены случайные последовательности, составленные из значений числа зарегистрированных избирателей, числа избирателей, принявших участие в выборах, и относительной явки избирателей на выборы на уровнях УИК, ТИК и ИК, которые представляют собой выборки случайной величины с ограниченной областью рассеяния, закон распределения которых отличен от нормального закона распределения. Обоснована возможность описания связей между числом избирателей, принявших участие в выборах, от числа зарегистрированных в соответствующей избирательной комиссии данного уровня избирателей с помощью линейных моделей. Продемонстрировано отсутствие детерминированных связей между относительной явкой избирателей на выборы и числом, зарегистрированных в зарегистрированных избирателей в данных УИК, ТИК и ИК, а также количеством избирательных комиссий на каждом из уровней. На основании полученных результатов сделан вывод об отсутствии целенаправленных действий в избирательной кампании выборов Президента РФ в 2018 г., направленных на фальсификацию электоральных данных.

Предложены организационные меры, призванные упорядочить процесс публичной публикации результатов анализа электоральных данных и их интерпретации.

Ключевые слова: выборы, электоральные данные, избирательная комиссия, статистический анализ, плотность распределения, аппроксимация Розенблатта-Парзена.

THE MULTIVARIATE STATISTICAL ANALYSIS EXPERIENCE OF THE ELECTORAL DATA

The article discusses the results of a multidimensional statistical analysis of the official electoral data submitted to the Central Election Commission (CEC) of the Russian Federation (RF) by lower-level election commissions (election commissions of RF subjects (EC), territorial election commissions (TEC), precinct election commissions (PEC)) based on the results of the Russian presidential elections in 2018. This data is currently presented on the website of the CEC RF.

The technology of automatic unloading of the discussed electoral data from the website of the CEC RF into a CVS-format file and the structure of the uploaded electoral data are described. There are presented random sequences that are composed of the values of the number of registered voters, the number of voters participated in the elections, and the relative voter turnout at the PEC, TEC and EC levels, which are samples of a random variable with a limited scattering area. The distribution law of random sequences is different from the normal one. The authors substantiate the possibility to describe the relationship between the voters number participated in the elections and the voters number of registered with the relevant EC at precise level using linear models. There is demonstrated the absence of deterministic links between the relative voter turnout at the elections and the number of registered voters in the PEC, TEC and EC data, as well as the number of election commissions at each level. The obtained results prove that there are no purposeful actions in the election campaign of the Russian presidential election in 2018 year aimed at the falsification of electoral data.

There are proposed organizational measures to streamline the process of public posting of the analysis results of electoral data and their interpretations.

Keywords: elections, electoral data, election commission, statistical analysis, distribution density, Rosenblatt-Parsen approximation.

Введение

Проблемы, связанные с оценкой качества организации процесса волеизъявления избирателей и достоверности результатов выборов в те или иные органы власти и местного самоуправления, являются актуальными в каждой стране, в том числе и в РФ. Анализ практики подведения итогов проведенных выборов в РФ показывает, что, их результаты оказываются предметом острых научных и политических дискуссий. В этих дискуссиях принимают участие политики, политологи, юристы, специалисты по прикладной статистике (электоральные статистики), а также политически активные избиратели) (см., например, [1,2]). Анализ хода их проведения показывает, что в большинстве случаев, мнения участников дискуссий по обсуждаемым проблемам, в зависимости от занимаемой ими политической позиции, оказываются диаме-

трально противоположными: от полного отрицания каких-либо фальсификаций электоральных данных (например, члены ЦИК и избирательных комиссий нижних уровней), до полного отрицания возможности проведения в Российской Федерации (РФ) честных выборов (например, кандидаты, не получившие поддержки избирателей, а также члены сообщества либерально настроенных электоральных статистиков (далее для краткости – оппозиционные электоральные статистики (ОЭС)) (см., например, [3]). В этой ситуации очевидны как необходимость объективного анализа официальных электоральных данных, так и проверки, собственно, математических методов, применяемых ОЭС.

Также необходимо отметить, что электоральные данные, а также результаты их анализа и интерпретации находятся в сфере регулирования Федерального закона «Об ин-

формации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ (последняя редакция), а задача обеспечения их достоверности, как на этапе их получения, так и на этапе интерпретации, по своей сути, изоморфны задачам защиты информации, а потому для их решения необходимо применять соответствующие методы защиты информации [4]. В этой связи, разработка методов защиты электоральных данных, в том числе, и организационных, к которым, следует отнести методы оценки достоверности выводов электоральных статистиков, в особенности ОЭС, является актуальной.

Отметим, что основная проблема оценки достоверности методов, используемых ОЭС, состоит в отсутствии электоральных данных, полученных в ходе проведения «абсолютно честных» выборов», отсутствие фальсификаций в которых доказано с помощью методов, отличных от методов электоральной статистики (например, на основе данных, представленных независимыми наблюдателями, данных экзит-поллов и т.д.). Их анализ, как очевидно, мог бы позволить определить набор «эталонных» статистических характеристик, по значениям которых можно будет говорить об отсутствии или наличии фальсификаций электоральных данных.

Отметим, что в реальных условиях невозможно проведение «абсолютно честных» выборов, процедура организации и проведения которых гарантировала бы недопущение каких-либо процедурных нарушений (как преднамеренных, так и не преднамеренных, а также признание отсутствия таковых всеми участниками избирательного процесса. В этой связи для оценки адекватности известных методов выявления электоральных аномалий представляется целесообразным использовать данные, представленные избирательными комиссиями (ИК) соответствующих уровней по результатам таких выборов, в которых преимущество одного из кандидатов являлось столь значительным, что его победа на выборах ни у кого не вызывала ни малейших сомнений, а потому использование каких-либо фальсификаций не требовалось. Такими, с нашей точки зрения, были выборы Президента РФ в 2018 г., победа в которых за явным преимуществом была одержана действующим Президентом РФ В.В. Путиным, что, однако, не отменяет необходимости проведения комплексного исследования соответствующих электоральных данных с целью

подтвердить или опровергнуть данную гипотезу.

В статье обсуждаются результаты многомерного статистического анализа официальных электоральных данных, размещенных на сайте ЦИК РФ (см. [5]), по результатам проведения в 2018 г. выборов Президента РФ.

2. Структура анализируемых электоральных данных, размещенных на сайте ЦИК

Структура анализируемых электоральных данных была задана следующей древовидной структурой системы избирательных комиссий, созданных для проведения в 2018 г. выборов Президента РФ:

- вершина дерева – ЦИК,
- вершины второго уровня – 85 ИК субъектов РФ, а также две приравненные к ним ИК «г. Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ» (итого 87 ИК);
- вершины третьего уровня – 2887 территориальных избирательных комиссий (ТИК), созданных ИК субъектов РФ (ТИК ИК «г. Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ» не создавались);
- вершины четвертого уровня – 97314 участковых избирательных комиссии (УИК), созданных соответствующими ТИК субъектов РФ, а также 7 УИК, созданных ИК «г. Байконур (Республика Казахстан)» и 393 УИК, созданных в ЦИК в ИК «Территория за пределами РФ» (итого 97716 УИК).

Однако в действительности оказалось, что 21 УИК, названия которых представлены в таблице 1, не преступила к передаче данных в ЦИК РФ о результатах выборов Президента РФ в 2018 г. Таким образом, общее число УИК, представивших электоральные данные в ЦИК, составило 97695.

Таким образом, общее число УИК, проводивших в 2018 выборы Президента РФ, составило 97695.

Автоматизированная технология выгрузки первичных электоральных данных с сайта ЦИК РФ и соответствующий программный инструмент, разработанные авторами, подробно описаны в [6]. Для автоматической выгрузки обсуждаемых электоральных данных с сайта ЦИК [4] был использован обычный офисный компьютер (AMD Ryzen 5 4 ГГц, 32 Гб ОЗУ, Интернет со скоростью 50 Мб/с). На выгрузку всего объема данных и их записи на жесткий диск использованного компьютера потребовалось 8 часов, в течение которых было выполнено 201156 http-запросов и скачано около 10 Гб сырых данных. Далее из

Перечень УИК, не преступивших к передаче электоральных данных по результатам выборов Президента РФ в 2018 г.

№	Название ИК субъекта РФ	Название ТИК	УИК
1	Волгоградская область	Волгоград, Красноармейская	УИК №611
2	Город Байконур (Республика Казахстан)	Без образования ТИК	УИК №8010
3	город Москва	Нагорный район	УИК №1910
4	город Москва	район Люблино	УИК №1438
5	город Москва	район Люблино	УИК №1439
6	город Москва	район Люблино	УИК №1446
7	город Москва	район Люблино	УИК №1450
8	Калининградская область	Советская	УИК №572
9	Калининградская область	Советская	УИК №576
10	Кемеровская область	Междуреченская городская	УИК №509
11	Кемеровская область	Новокузнецк, Куйбышевская	УИК №766
12	Краснодарский край	Гулькевичская	УИК №1158
13	Московская область	Люберецкая городская	УИК №1480
14	Нижегородская область	Нижний Новгород, Канавинская	УИК №2383
15	Республика Дагестан	Дагогнинская городская	УИК №380
16	Республика Дагестан	Махачкала, Ленинская	УИК №1073
17	Республика Дагестан	Махачкала, Советская	УИК №1126
18	Республика Дагестан	Сулейман-Стальская	УИК №1284
19	Территория за пределами РФ	Без образования ТИК	УИК №8443
20	Тюменская область	Тюмень, Центральная	УИК №2239
21	Ханты-Мансийский автономный округ - Югра	Белоярская	УИК №35

первичных («сырых») данных извлечено 11Мб собственно полезных данных, которые были сохранены в CSV-файл (в формате RFC 4180) со следующей структурой (таблица 2).

Также в данный файл записывались сводные результаты по данной ТИК в строку, следующую за строкой последней УИК, созданной данной ТИК. При этом в соответствующую ячейку столбца I3 заносилось значение null-string (null-string=""). Сводные данные результаты по данной ИК, записывались в строку, следующую за строкой, содержащей сводные данные по последней ТИК, созданной данной ИК. В результате число строк файла, содержащего обсуждаемые электоральные данные ЦИК, оказалось равным 1005879. Таким образом, первичная информация представляет собой таблицу размером 1005879×23, которая упорядочена в алфавитном порядке по значениям текстовых полей I1, I2, I3, соответственно.

На первом шаге осуществлялся импорт электоральных данных ЦИК из файла «Данные ЦИК.xlsx» в рабочее пространство MATLAB. В результате импорта данных из об-

суждаемого CVS-файла в рабочем пространстве MATLAB создавалась переменная «Untitled», содержащая электоральные данные, выгруженные с сайта ЦИК РФ, которая далее переименовывалась в переменную «Data» и сохранялась в текущем каталоге в файле Data.mat. Для проведения автоматизированного статистического анализа электоральных данных, размещенных в файле Data.mat, на m-языке MATLAB была написана программная библиотека, включающая в себя соответствующий набор скриптов, обеспечивающих вычисление в выбранном разрезе статистических характеристик электоральных данных и параметров регрессионных моделей, описывающих связи между выбранными показателями результатов выборов.

В последующих разделах обсуждаются результаты статистического анализа электоральных данных, размещенных на сайте ЦИК РФ в различных измерениях, а также связей между выбранными показателями, характеризующие организацию системы избирательных комиссий различных уровней.

Структура электоральных данных, выгруженных с сайта ЦИК РФ и размещенных в CVS-файле

Наименование столбца	Тип данных	Описание
I1	Строка	наименование ИК субъекта РФ (2-й уровень)
I2	Строка	Наименование территориальной ИК субъекта РФ (3-й уровень)
I3	Строка	Наименование участковой ИК 3 уровня, сформированной территориальной ИК субъекта РФ (4-й уровень)
c1	Целое число	Число избирателей, включенных в список избирателей
c2	Целое число	Число избирательных бюллетеней, полученных участковой избирательной комиссией
c3	Целое число	Число избирательных бюллетеней, выданных избирателям, проголосовавшим досрочно
c4	Целое число	Число избирательных бюллетеней, выданных в помещении для голосования в день голосования
c5	Целое число	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования
c6	Целое число	Число погашенных избирательных бюллетеней
c7	Целое число	Число избирательных бюллетеней в переносных ящиках для голосования
c8	Целое число	Число бюллетеней в стационарных ящиках для голосования
c9	Целое число	Число недействительных избирательных бюллетеней
c10	Целое число	Число действительных избирательных бюллетеней
c11	Целое число	Число утраченных избирательных бюллетеней
c12	Целое число	Число избирательных бюллетеней, не учтенных при получении
c13	Целое число	Число избирателей, проголосовавших за Бабурина Сергея Николаевича
c14	Целое число	Число избирателей, проголосовавших за Грудинина Павла Николаевича
c15	Целое число	Число избирателей, проголосовавших за Жириновского Владимира Вольфовича
c16	Целое число	Число избирателей, проголосовавших за Путина Владимира Владимировича
c17	Целое число	Число избирателей, проголосовавших за Собчак Ксению Анатольевну
c18	Целое число	Число избирателей, проголосовавших за Сурайкина Максима Александровича
c19	Целое число	Число избирателей, проголосовавших за Титова Бориса Юрьевич
c20	Целое число	Число избирателей, проголосовавших за Явлинского Григория Алексеевича

3. Статистический анализ электоральных данных на уровне УИК

Для оценки аппроксимация плотностей распределений случайных последовательностей $\{x_i\}$ в работе были использованы ядерные аппроксимация плотности распределения (аппроксимация Розенблатта-Парзена) [7,8] данной последовательности, рассчитываемой по формуле

$$\hat{p}_h(x) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{x-x_i}{h}\right) \quad (1)$$

здесь

N – число членов случайной последовательности $\{x_i\}$;

$k(t) = \frac{d}{dt}K(t)$ – так называемая ядерная функция, где функция $K(t)$ удовлетворяет следующим условиям:

а) $K(t)$ – монотонная неубывающая функция, область значений которой принадлежит интервалу

б) $K(t) = 1 - K(t)$ – ядерная функция, симметричная относительно 0;

h – размытости параметр (называемый также полосой пропускания ядра), что $h \rightarrow 0$ при $N \rightarrow \infty$.

При использовании на практике метода Розенблатта-Парзена приходится решать следующие задачи:

1) выбора оптимальной ядерной функции $k(t)$;

2) нахождение для выбранной ядерной функции оценки оптимального значения параметра размытости h^* , обеспечивающего на выбранному критерию минимальное отклонение между аппроксимациями Розенблатта-

Парзена и соответствующими эмпирическими ФР и ПР.

Методы решения данных задач обсуждаются, например, в [9–17]. В нашей работе для оценивания оптимального значения параметра размытости был использован метод, предложенный авторами в [17].

Рассмотрим визуализацию последовательности $\{[N_{УИК}^{RE}]_i\}$, $i = 1, 97695$, составленной из значений числа избирателей, зарегистрированных соответствующей УИК, и аппроксимацию Розенблатта-Парзена ее плотности распределения, а также плотность вероятности, соответствующего нормального закона распределения с параметрами:

$$\mu_{УИК}^{RE} = \text{mean}\left(\{[N_{УИК}^{RE}]_i\}\right), \sigma_{УИК}^{RE} = \text{std}\left(\{[N_{УИК}^{RE}]_i\}\right),$$

где $\text{mean}(\square)$ – функция, возвращающая оценку среднего значения последовательности, $\text{std}(\square)$ – функция, возвращающая среднеквадратическое отклонение случайной последовательности. Также на рис. 1 представлены соответствующие плотности распределения, параметры которых вычислены в предположении о том, данная последовательность является случайной выборкой из генеральной совокупности с нормальным законом распределения, представленные на рис. 1.

Из рис. 1 видно, что последовательность $\{[N_{УИК}^{RE}]_i\}$, значения которой принадлежат отрезку $[3, 7746]$, представляет собой некоторую реализацию случайной величины с ограниченной областью рассеяния, плотность распределения которой отличается от нормального закона распределения. Отметим,

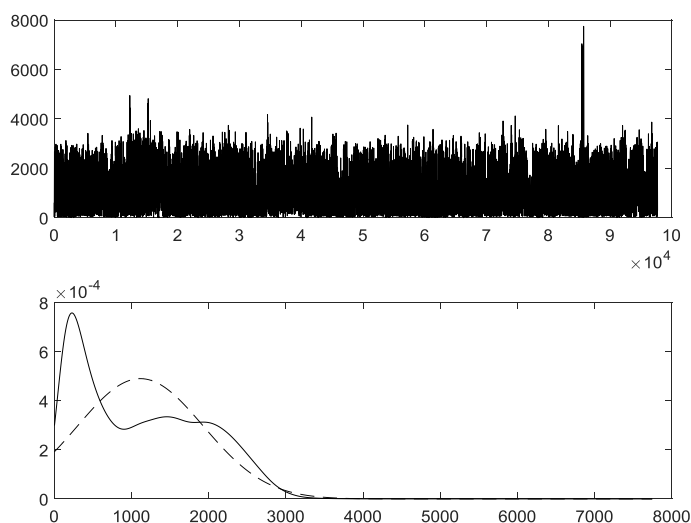


Рис. 1. Визуализация последовательности $\{[N_{УИК}^{RE}]_i\}$ (вверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^* = 110.283$) ее плотности распределения и плотности соответствующего нормального закона распределения (–) (снизу)

что наименьшее число избирателей – 3 было зарегистрировано УИК № 814, созданной ТИК «Булунская» (ИК «Республика Якутия (Саха)». Наибольшее число зарегистрированных избирателей – 7746 в УИК № 8342, созданной ИК «Территория за пределами РФ».

В связи с тем, что аппроксимация Розенблатта-Парзена $\hat{p}(\xi)$ плотности распределения последовательности $\{[N_{УИК}^{RE}]_i\}$, $i=1, 97695$ имеет несколько локальных максимумов, значения среднего числа и соответствующего среднеквадратического отклонения здесь и далее вычислялись по формулам:

$$\mu = \int_{y_{\min}}^{y_{\max}} y(\xi) \hat{p}(\xi) d\xi, \quad (2)$$

$$\sigma = \left(\int_{y_{\min}}^{y_{\max}} (y(\xi) - \mu)^2 \hat{p}(\xi) d\xi \right)^{0.5} \quad (3)$$

где $y(\xi)$ – случайная величина, $p(\xi)$ – плотность распределения $y(\xi)$, соответственно.

Значения $\mu_{УИК}^{RE}$, $\sigma_{УИК}^{RE}$, округленные до ближайших целых чисел, оказались равными 1 117 и 802, соответственно. Разделив общее число зарегистрированных избирателей

$N_{all}^{RE}=109\ 001\ 306$ на $\mu_{УИК}^{RE}=1117$, получаем оценку требуемого числа УИК $\tilde{N}_{УИК}^{RE}=97\ 584$, число которых при данной численности избирателей обеспечит среднюю численность за-регистрированных избирателей равной $\mu_{УИК}^{RE}=1117$. Сравнивая $\tilde{N}_{УИК}^{RE}=97\ 584$ и фактическое число УИК, проводивших в 2018 г. выборы Президента РФ, $N_{УИК}^{RE}=97695$, видим, что с точностью до 0,11% эти числа совпали.

Рассмотрим визуализацию последовательности $\{[N_{УИК}^V]_i\}$, $i=1, 97695$, составленной из значений числа избирателей, принявших участие в голосовании в соответствующей УИК, а также аппроксимацию Розенблатта-Парзена плотности данной последовательности и плотность вероятности, соответствующего нормального закона распределения с параметрами: соответствующего нормально-го распределения с параметрами, вычисленных численным интегрированием в соответствии с (2), (3) численным интегрированием:

$$\mu_{УИК}^V = \text{mean}\left(\{[N_{УИК}^V]_i\}\right), \sigma_{УИК}^V = \text{std}\left(\{[N_{УИК}^V]_i\}\right),$$

представленные на рис. 2.

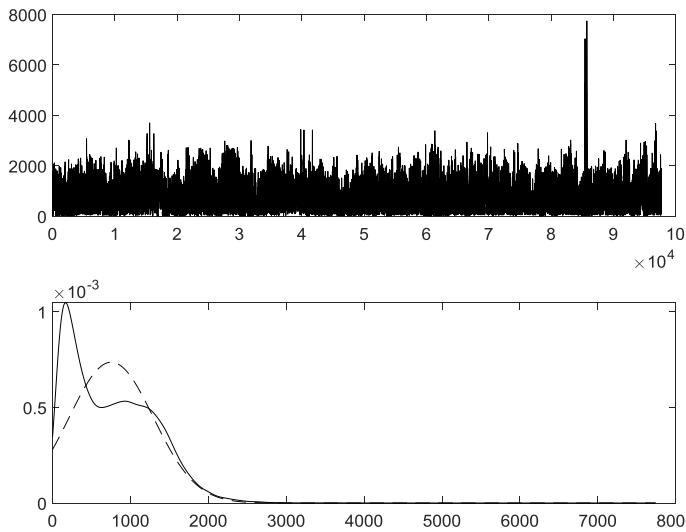


Рис. 2. Визуализация последовательности $\{[N_{УИК}^V]_i\}$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=69.104$) ее плотности распределения и плотности соответствующего нормального закона распределения (–) (снизу)

Из рис. 2 видно, что что последовательность $\{[N_{УИК}^V]_i\}$, $i=1, 97695$, значения которой принадлежат отрезку $[0, 7741]$, представляет собой некоторую реализацию случайной величины с ограниченной областью рассеяния, плотность распределения которой отличается от нормального закона распределения. Отметим, что наименьшее число избирателей, пришедших на выборы, равное 0 было зарегистрировано в УИК № 809, созданной

ТИК «Булунская» (ИК «Республика Якутия (Саха)». (При этом общее число зарегистрированных данной УИК избирателей равнялось 7), наибольшее число избирателей – 7741 в УИК № 8342, созданной ИК «Территория за пределами РФ».

Значения $\mu_{УИК}^V$, $\sigma_{УИК}^V$, округленные до ближайших целых чисел, оказались равными 754 и 535, соответственно. Разделив $\mu_{УИК}^V$ на $\mu_{УИК}^{RE}=1117$, находим, оценку явки избирате-

лей на выборы Президента РФ в процентном отношении, которая оказалась равной 67.47%. Для сравнения явка избирателей, вычисленная как отношение официальных данных ЦИК РФ о числе зарегистрированных избирателей $N_{all}^{RE}=109\,001\,306$ и числе избирателей, принявших участие в выборах $N_{УИК}^V=73\,573\,516$, оказывается равной 67.50%. Таким образом, значения оценок явки избирателей в 2018 г. на выборы Президента РФ, оказываются отличными друг от друга не более, чем на 0.04%.

Рассмотрим визуализацию последовательности $\{[N_{УИК}^{\%}]_i\}, i=\overline{1,97695}$, составленной из отношения значений числа избирателей,

принявших участие в голосовании в соответствующей УИК, к числу зарегистрированных в данной УИК избирателей:

$$\{[N_{УИК}^{\%}]_i = 100 \cdot ([N_{УИК}^V]_i / [N_{УИК}^{ER}]_i)\}, i = \overline{1,97695},$$

а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и плотность соответствующего нормального распределения с параметрами, вычисленными в соответствие с (2), (3) численным интегрированием с помощью метода трапеций:

$$\mu_N^{\%} = \text{mean}\left(\{[N_{УИК}^{\%}]_i\}\right), \sigma_N^{\%} = \text{std}\left(\{[N_{УИК}^{\%}]_i\}\right),$$

представленные на рис. 3.

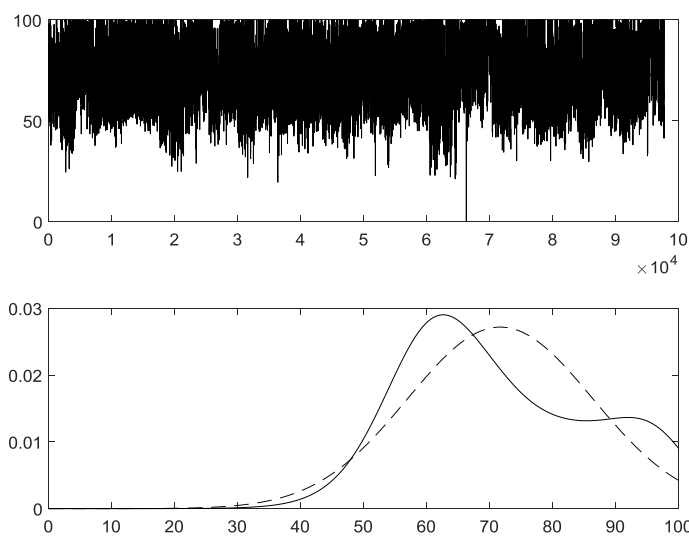


Рис. 3. Визуализация последовательности $\{[N_{УИК}^{\%}]_i\}$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=5.0$) ее плотности распределения и плотности соответствующего нормального закона распределения (снизу)

Из рис. 3 видно, что последовательность $[N_{УИК}^{\%}]_i$ представляет собой некоторую выборку случайной величины с ограниченной областью рассеяния $[0,100]\%$, плотность распределения которой отличается от нормального закона распределения. Наименьшее явка избирателей на выборы – 0% оказалась в УИК № 809, созданной ТИК «Булунская» (ИК «Республика Саха (Якутия)»), в которой были зарегистрированы 7 избирателей. Наибольшая явка избирателей на выборы – 100% оказалась в УИК № 1910, созданной ТИК «Барнаул, Октябрьская» (ИК «Алтайский край»), где был зарегистрирован 231 избиратель.

Значения $\mu_N^{\%}, \sigma_N^{\%}$ оказались равными 67.55 % и 14.35%, соответственно. Таким образом, значение $\text{mean}(\mu_{УИК}^{\%})$ согласуется с ранее полученными оценками процента средней явки

избирателей в 2018 г. на выборы Президента РФ.

Так же необходимо отметить, что, как показывает анализ аппроксимации Розенблатта-Парзена (рис. 3 (снизу)) плотности распределения случайной последовательности $[N_{УИК}^{\%}]_i$, в конечном числе УИК явка избирателей оказалась выше 90%. (Многомерный статистический анализ электоральных данных, представленных УИК, в которых явка избирателей на выборы в 2018 г. Президента РФ оказалась явка оказалось аномально высокой, являются предметом отдельной статьи.)

Рассмотрим зависимость числа избирателей, принявших участие в выборах в данной УИК, от числа зарегистрированных в ней избирателей, – $[N_{УИК}^V]_i = f([N_{УИК}^{ER}]_i)$ (рис. 4, сверху) и зависимость относительной явки

избирателей, проголосовавших в данной УИК, от числа зарегистрированных в ней избирателей (рис. 4, снизу) $[N_{УИК}^V]_i = f([N_{УИК}^{ER}]_i)$, а также графики линейных моделей, аппроксимирующих данные зависимости, которые описываются следующими статистиками, вычисленными с помощью функции fitlm.m пакета MATLAB:

Линейная модель № 1, аппроксимирующая зависимость $[N_{УИК}^V]_i = f([N_{УИК}^{ER}]_i)$

Linear regression model:
y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	44.659	0.88374	50.534	0
x1	0.63495	0.00063985	992.34	0

Number of observations: 97695, Error degrees of freedom: 97693
Root Mean Squared Error: 163
R-squared: 0.91, Adjusted R-Squared: 0.91
F-statistic vs. constant model: 9.85e+05, p-value = 0

Линейная модель № 2, аппроксимирующая зависимость $[N_{УИК}^E]_i = f([N_{УИК}^{ER}]_i)$

Linear regression model:
y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	79.511	0.073357	1083.9	0
x1	-0.0070294	5.3112e-05	-132.35	0

Number of observations: 97695, Error degrees of freedom: 97693
Root Mean Squared Error: 13.5
R-squared: 0.152, Adjusted R-Squared: 0.152
F-statistic vs. constant model: 1.75e+04, p-value = 0

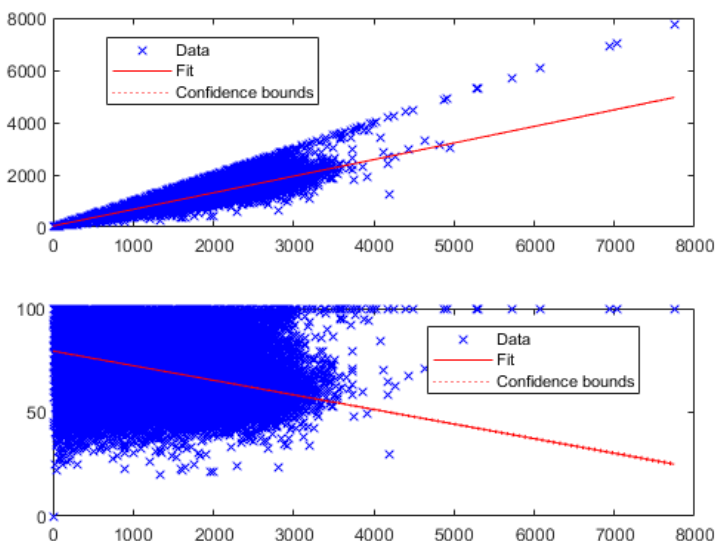


Рис. 4. Визуализация зависимостей $[N_{УИК}^V]_i = f([N_{УИК}^{ER}]_i)$, $[N_{УИК}^E]_i = f([N_{УИК}^{ER}]_i)$ и их линейных аппроксимаций

Анализ зависимостей, представленных на рис. 4, и приведенных выше статистик построенных линейных моделей позволяет сделать вывод о наличии линейной связи между числом избирателей, проголосовавших в данной УИК и числом избирателей (коэффициент детерминации модели $R^2=0.91$), зарегистрированных в данной УИК, а также наличии отрицательной линейной связи между относительной явкой избирателей и числом избирателей, зарегистрированных в данной УИК (коэффициент детерминации модели $R^2=0.152$). Линейная связь между этими показателями понятна из описанных выше результатов анализа распределений данных величин, действительно, априори, очевидно, что абсолютное число избирателей, принявших участие в выборах, N_i^V пропорционально числу проголосовавших в данной УИК избирателей, поэтому, в первом приближении можно описать связь между этими показателями с помощью линейной модели:

$$\left[N_{УИК}^V \right]_i = a_{УИК}^V \left(\left[N_{УИК}^{ER} \right]_i \right) + b_{УИК}^V,$$

где $a_{УИК}^V, b_{УИК}^V$ – коэффициенты линейной регрессии, определяемые в соответствии с методом наименьших квадратов. (Для обсуждаемой зависимости $a_{УИК}^V=0.63495, b_{УИК}^V=44.659$.) Отрицательное значение коэффициента линейной регрессии, аппроксимирующей зависимость $\left[N_{УИК}^{\%} \right]_i = \left[N_{УИК}^V \right]_i / \left[N_{УИК}^{ER} \right]_i$, $a_{УИК}^{\%} = -0.0070294$ объясняется тем, что

$$\left[N_{УИК}^{\%} \right]_i = \left[N_{УИК}^V \right]_i / \left[N_{УИК}^{ER} \right]_i \approx a_{УИК}^V + b_{УИК}^V / \left(\left[N_{УИК}^{ER} \right]_i \right) \approx a_{УИК}^V - b_{УИК}^V / \min \left(\left\{ \left[N_{УИК}^{ER} \right]_i \right\} \right) \Delta_{УИК}^{ER},$$

$$\text{где } \Delta_{УИК}^{ER} = \left[N_{УИК}^{ER} \right]_i - \min \left(\left\{ \left[N_{УИК}^{ER} \right]_i \right\} \right).$$

Таким образом, при увеличении числа избирателей, зарегистрированных в данной УИК, прямо пропорционально увеличивается число избирателей, принявших участие в выборах, и, наоборот, при увеличении числа избирателей, зарегистрированных в данной УИК, относительная явка избирателей на выборы прямо пропорционально уменьшается.

4. Анализ электоральных данных на уровне ТИК

Рассмотрим последовательность, составленную из значений числа УИК, зарегистрированных соответствующими ТИК данных ИК субъектов РФ, состоящей из 2 776 значений $\left(\left[N_{ТИК}^{УИК} \right]_i, i=1, 2776 \right)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисляемые в соот-

ветствие с (2), (3) численно с помощью метода трапеций:

$$\mu_{ТИК}^{УИК} = \text{mean} \left(\left\{ \left[N_{ТИК}^{УИК} \right]_i \right\} \right), \sigma_{ТИК}^{УИК} = \text{std} \left(\left\{ \left[N_{ТИК}^{УИК} \right]_i \right\} \right),$$

представленные на рис. 5. (При вычислении зависимостей, представленных на рис. 5 не учитывались избиратели, зарегистрированные ИК «город Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ», так как данные ИК не создавали УИК).

Из рис. 5 видно, что обсуждаемая последовательность $\left[N_{ТИК}^{УИК} \right]_i$ оказывается некоторой реализацией случайной величины с ограниченной областью рассеяния [1,304], плотность распределения которой отличается от плотности нормального закона распределения. Отметим, что наименьшее число УИК – 1 было создано ТИК «Алеутская» (ИК «Камчатский край»); наибольшее число УИК 304 – ТИК «Мурманская» (ИК «Мурманская область»).

Значения $\mu_{ТИК}^{УИК}$ и $\sigma_{ТИК}^{УИК}$ оказались равными 35 и 22, соответственно.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в данной ТИК $\left(\left[N_{ТИК}^{RE} \right]_i, i=1, 2776 \right)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленные в соответствии с (2), (3) численно с помощью метода трапеций:

$$\mu_N^{RE} = \text{mean} \left(\left\{ \left[N_{ТИК}^{RE} \right]_i \right\} \right), \sigma_N^{RE} = \text{std} \left(\left\{ \left[N_{ТИК}^{RE} \right]_i \right\} \right),$$

представленную на рис. 6.

Из рис. 6 видно, что обсуждаемая последовательность $\left[N_{ТИК}^{RE} \right]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [131,335 978], закон распределения которой отличается от нормального закона распределения. Отметим, что наименьшее число избирателей – 131 было зарегистрировано в ТИК «Новороссийск, Судовая», созданной ИК «Краснодарский край». При этом общее число УИК, созданных ТИК «Новороссийск, Судовая» равнялось 5 (УИК № 3695–3699). Наибольшее число избирателей – 335 978 было зарегистрировано в ТИК «Тольятти, Автозаводская» (ИК «Самарская область»). При этом общее число УИК, созданных ТИК «Тольятти, Автозаводская» равнялось 147.

Значения $\mu_{ТИК}^{RE}$ и $\sigma_{ТИК}^{RE}$ оказались равными 39 053 и 40 891, соответственно. Сравнивая отношение $\mu_{ТИК}^{RE} / \mu_{ТИК}^{УИК} = 1116$ с оценкой сред-

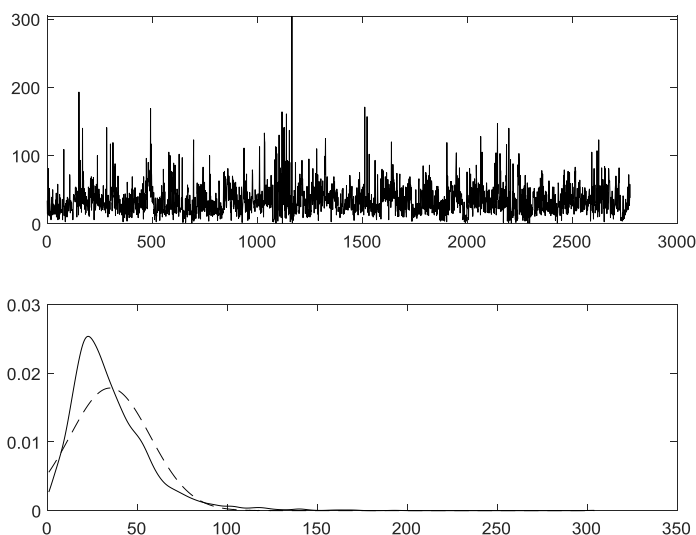


Рис. 5. Визуализация последовательности $[N_{TIK}^{YIK}]$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=3.5376$) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (снизу)

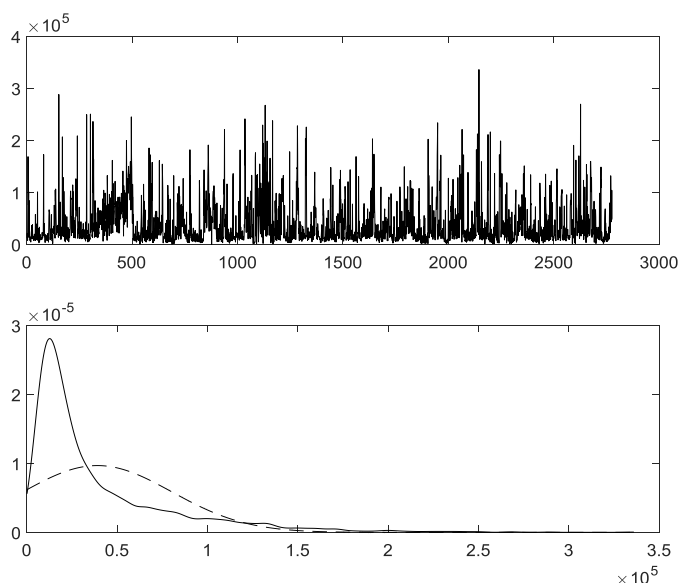


Рис. 6. Визуализация последовательности $[N_{TIK}^{RE}]$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=4165.74$) ее плотности распределения и плотность распределения соответствующего нормального закона распределения (--) (снизу)

него числа избирателей, зарегистрированных в УИК, с соответствующим значением $\mu_{YIK}^{RE}=1117$, вычисленным на основе анализа последовательности $[N_{YIK}^{RE}]$, приходим к выводу, что значения данных величин с точностью 0.09% совпадают. Следовательно, оценки среднего числа избирателей, зарегистрированных в данной УИК, вычисленные на уровне ТИК и на уровне УИК согласуются друг с другом.

Рассмотрим последовательность, составленную из значений числа избирателей, заре-

гистрированных в данной ТИК и принявших участие в выборах $([N_{TIK}^V]_i, i=1, 2776)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленными в соответствии с (2), (3) численно с помощью метода трапеций:

$$\mu_{TIK}^V = \text{mean}(\{[N_{TIK}^V]_i\}), \sigma_{TIK}^V = \text{std}(\{[N_{TIK}^V]_i\}),$$

представленную на рис. 7.

Из рис. 7 видно, что видно, что обсуждае-

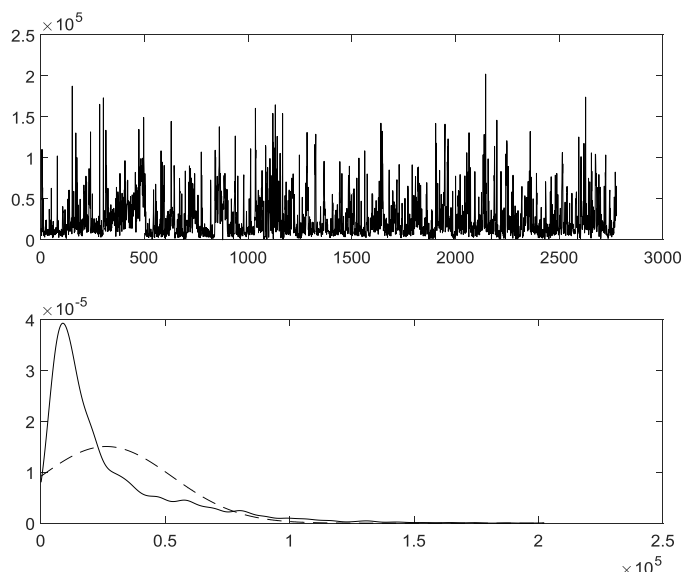


Рис. 7. Визуализация последовательности $[N^V_{ТИК}]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=2924.2$) ее плотности распределения и плотности распределения соответствующего нормального закона распределения

мая последовательность $[N^V_{ТИК}]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [126,202 275], закон распределения которой отличается от нормального закона распределения. Наименьшее число избирателей, принявших участие в выборах, – 126 было зарегистрировано в ТИК «Новороссийск, Судовая», созданной ИК «Краснодарский край». Наибольшее число избирателей – 202 275 было зарегистрировано в ТИК «Тольятти, Автозаводская» (ИК «Самарская область»).

Значения $\mu^V_{ТИК}$ и $\sigma^V_{ТИК}$ значения оказались равными 26 312 и 26 314, соответственно. Следовательно, оценки среднего числа избирателей, зарегистрированных в данной УИК, вычисленные на уровне ТИК и на уровне УИК согласуются друг с другом. Разделив $\mu^V_{ТИК}=26\,312$ на $\mu^{RE}_{all}=39\,053$, находим, оценку явки избирателей на выборы Президента РФ в процентном отношении, которая оказалось равной 67.38%. Для сравнения явка избирателей, вычисленная как отношение официальных данных ЦИК РФ о числе зарегистрированных избирателей $N^{RE}_{all}=109\,001\,306$ и числе избирателей, принявших участие в выборах $N^{V}_{УИК}=73\,573\,516$, оказывается равной 67.50%. Таким образом, значения оценок явки избирателей в 2018 г. на выборы Президента РФ, отказываются отличными друг от друга не более, чем на 0.18%.

Рассмотрим визуализацию последовательности $\{[N^{\%}_{ТИК}]_i\}, i = \overline{1, 2776}$, составленной

из отношения значений числа избирателей, принявших участие в голосовании в соответствующей УИК, к числу зарегистрированных в данной УИК избирателей:

$$\{[N^{\%}_{ТИК}]_i = 100([N^V_{ТИК}]_i / [N^{ER}_{ТИК}]_i)\}, i = \overline{1, 2776},$$

а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленным в соответствие с (2), (3) численно с помощью метода трапеций:

$$\mu^{\%}_{ТИК} = \text{mean}(\{[N^{\%}_{ТИК}]_i\}), \sigma^{\%}_{ТИК} = \text{std}(\{[N^{\%}_{ТИК}]_i\}),$$

представленные на рис. 8.

Из рис. 8 видно, что видно, что обсуждаемая последовательность $[N^{\%}_{ТИК}]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [44.22,99.80]%, закон распределения которой отличается от нормального закона распределения. Наименьшая явка избирателей – 44.22% оказалось в ТИК «Власихинская городская», созданная ИК «Московская область» (число УИК – 9, число зарегистрированных избирателей – 18 255). Наибольшая явка избирателей – 99.80% оказалась в ТИК «Новая земля», созданной ИК «Архангельская область» (число УИК – 6, число зарегистрированных избирателей – 1 957).

Значения $\mu^{\%}_{ТИК}$ и $\sigma^{\%}_{ТИК}$ оказались равными 65.59% и 13.20%, соответственно. Таким образом, значения оценок явки избирателей на

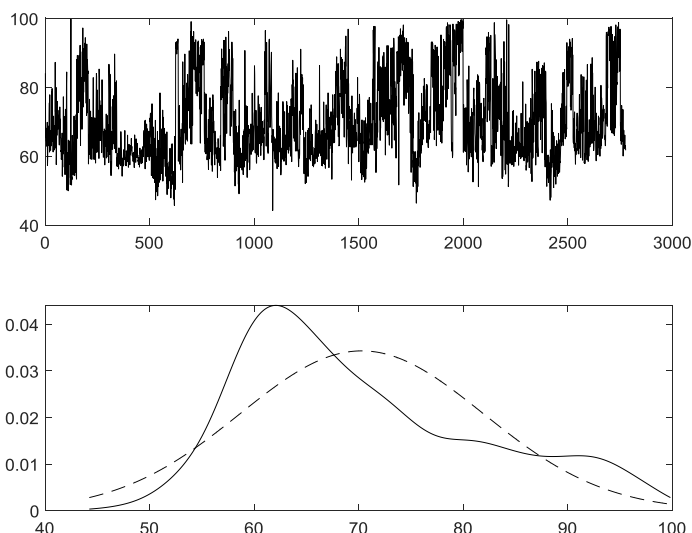


Рис. 8. Визуализация последовательности $[N_{ТИК}^{\%}]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=2.293$) ее плотностей распределения и плотностей распределения соответствующего нормального закона распределения (–) (снизу)

выборы Президента РФ в 2018 г. на уровне ТИК, отказались отличными друг от друга не более, чем на 2.4%.

Так же необходимо отметить, что, как показывает анализ аппроксимации Розенблатта-Парзена (рис. 8 (снизу)) плотности распределения случайной последовательности $[N_{ТИК}^{\%}]_i$ в конечном числе избирательных комиссий явка избирателей оказалась выше 90%. (Многомерный статистический анализ электоральных данных на уровне ТИК, в которых явка избирателей на выборы в 2018 г. Президента РФ оказалась явкой оказалось аномально высокой, являются предметом последующей отдельной статьи.)

Рассмотрим зависимость числа избирателей, принявших участие в выборах в данной УИК, от числа зарегистрированных в ней избирателей, $[N_{ТИК}^V]_i = f([N_{ТИК}^{ER}]_i)$ (рис. 9, сверху) и зависимость относительной явки избирателей, проголосовавших в данной УИК

от числа зарегистрированных в ней избирателей (рис. 9, снизу) $[N_{ТИК}^{\%}]_i = f([N_{ТИК}^{ER}]_i)$, а также графики линейных моделей, аппроксимирующих данные зависимости, которые описываются следующими статистиками, вычисленными с помощью функции fitlm.m пакета MATLAB:

Из рис. 9 видно, что зависимость $[N_{ТИК}^V]_i = f([N_{ТИК}^{ER}]_i)$ аппроксимируется с помощью линейной функции. При этом, как видно, из вычисленных статистик данной аппроксимации коэффициент $R^2=0.97$, что свидетельствует о высокой линейной связи между переменными $N_{ТИК}^V$ и $N_{ТИК}^{ER}$ которая обусловлена тем, что при отсутствии фальсификаций электоральных данных $N_{ТИК}^V \approx \mu_{ТИК}^{\%} N_{ТИК}^{ER}$. При этом, одновременно, относительная явка избирателей на выборы $N_{ТИК}^{\%}$ в данной ТИК оказывается независимой от числа зарегистрированных в ней избирателей $N_{ТИК}^{ER}$ ($R^2=0.0566$), что, с нашей точки зрения, также является кос-

Статистики линейной модели № 1, аппроксимирующей зависимость

$$[N_{ТИК}^V]_i = f([N_{ТИК}^{ER}]_i)$$

Linear regression model:

$$y \sim 1 + x1$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1575.1	119.17	13.217	1.0113e-38
x1	0.63332	0.0020989	301.74	0

Number of observations: 2776, Error degrees of freedom: 2774

Root Mean Squared Error: 4.55e+03

R-squared: 0.97, Adjusted R-Squared: 0.97

F-statistic vs. constant model: 9.1e+04, p-value = 0

Статистики линейной модели № 2, аппроксимирующей зависимость

$$\left[N_{ТИК}^{%}\right]_i = f\left(\left[N_{ТИК}^{ER}\right]_i\right)$$

Linear regression model:

$$y \sim 1 + x1$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	72.905	0.29584	246.43	0
x1	-6.721e-05	5.2104e-06	-12.899	5.1648e-37

Number of observations: 2776, Error degrees of freedom: 2774

Root Mean Squared Error: 11.3

R-squared: 0.0566, Adjusted R-Squared: 0.0562

F-statistic vs. constant model: 166, p-value = 5.16e-37

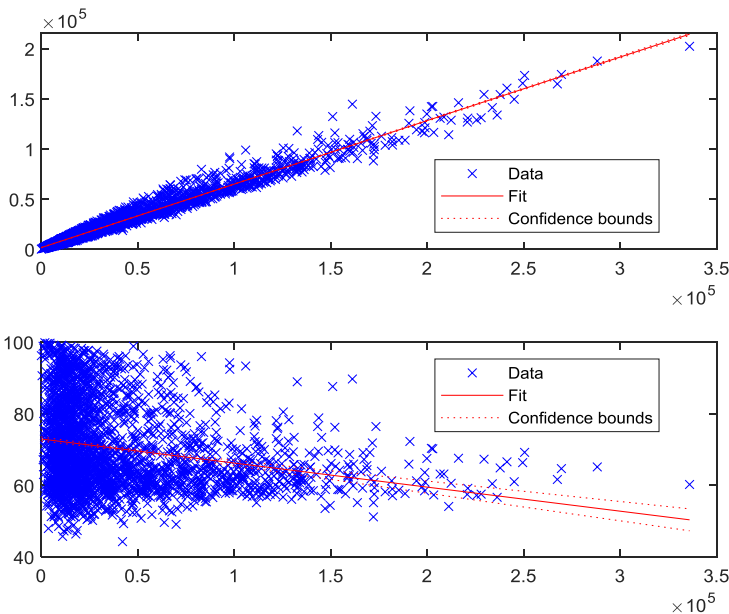


Рис. 9. Визуализация зависимостей $\left[N_{ТИК}^V\right]_i = f\left(\left[N_{ТИК}^{ER}\right]_i\right)$, $N_i^{%} = f\left(\left[N_{ТИК}^{ER}\right]_i\right)$ и их линейных аппроксимаций

венным подтверждением честного проведения в 2018 г. выборов Президента РФ.

На уровне ТИК помимо статистических характеристик электоральных данных на уровне УИК, обсуждавшиеся в предыдущем разделе, появляется дополнительный количественный показатель – число УИК, зарегистрированных данной ТИК. В этой связи рассмотрим зависимости числа УИК от числа зарегистрированных в соответствующих ТИК $N_{ТИК}^{УИК} = f\left(N_{ТИК}^{RE}\right)$, а также относительной явки от числа УИК, зарегистрированных данной ТИК, и их линейные аппроксимацию, статистики которых приведены ниже, представленные на рис. 10.

Из рис. 10 и анализа статистик аппроксимации видно:

1. Зависимость $N_{ТИК}^{УИК} = f\left(N_{ТИК}^{RE}\right)$ с высоким значением коэффициента $R^2=0.657$ может быть

аппроксимирована линейной зависимостью. Данный результат позволяет сделать вывод о том, что при формировании УИК их число выбиралось прямо пропорционально числу избирателей, зарегистрированных на территории, данной ТИК. Меньшее значение коэффициента $R^2=0.657$ в рассматриваемом случае, чем у аналогичной линейной модели, аппроксимирующей зависимость $N_{ТИК}^V = f\left(N_{ТИК}^{ER}\right)$ ($R^2=0.97$), по-видимому, объясняется тем, что число ТИК, создаваемых данной ИК субъекта РФ, зависит не только от абсолютного числа избирателей, зарегистрированных на территории данных ТИК, но и от его распределения по ее территории.

2. Коэффициент линейной модели, аппроксимирующей зависимость $\left[N_{ТИК}^{%}\right]_i = f\left(\left[N_{ТИК}^{УИК}\right]_i\right)$, $R^2=0.0138$, что позволяет отвергнуть гипотезу о линейной связи между

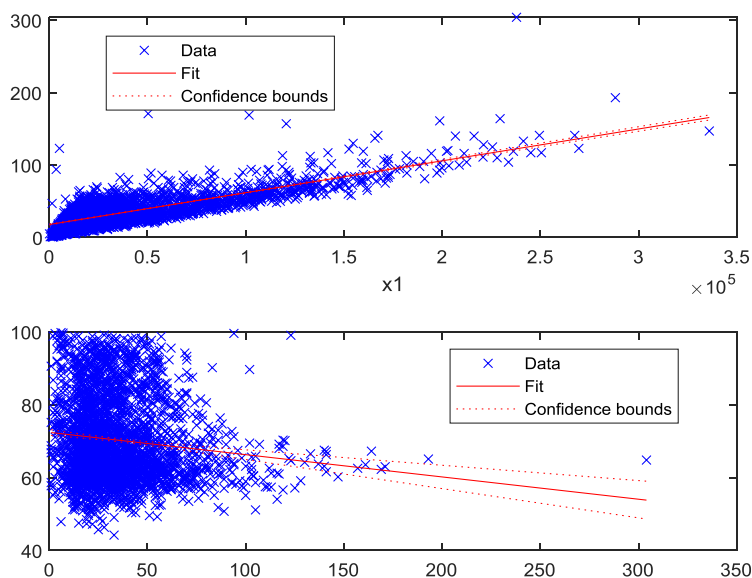


Рис. 10. Визуализация зависимости $[N_{TIK}^{YIK}]_i = f([N_{TIK}^{RE}]_i)$ и ее линейной аппроксимации (сверху), зависимости $[N_{YIK}^%]_i = f([N_{TIK}^{YIK}]_i)$ и ее линейной аппроксимации (снизу)

Статистики линейной модели, аппроксимирующей зависимость $[N_{TIK}^{YIK}]_i = f([N_{TIK}^{RE}]_i)$

Linear regression model:

$y \sim 1 + x1$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	17.858	0.34257	52.129	0
x1	0.00043982	6.0334e-06	72.898	0

Number of observations: 2776, Error degrees of freedom: 2774

Root Mean Squared Error: 13.1

R-squared: 0.657, Adjusted R-Squared: 0.657

F-statistic vs. constant model: 5.31e+03, p-value = 0

subplot(2,1,2);plot mdl)

Статистики линейной зависимости, аппроксимирующей зависимость

$$[N_{TIK}^{YIK}]_i = f([N_{TIK}^{RE}]_i)$$

Linear regression model:

$y \sim 1 + x1$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	72.423	0.40809	177.46	0
x1	-0.06118	0.0098179	-6.2315	5.3221e-10

Number of observations: 2776, Error degrees of freedom: 2774

Root Mean Squared Error: 11.6

R-squared: 0.0138, Adjusted R-Squared: 0.0134

F-statistic vs. constant model: 38.8, p-value = 5.32e-10

переменными N_{TIK}^{YIK} , N_{TIK}^{RE} и принять альтернативную гипотезу о независимости данных переменных друг от друга.

4. Анализ электоральных данных на уровне ИК

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и

приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $[N_{IK}^{RE}]_i, i=1,87$, аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность соответствующего нормального закона распределения с параметрами, вычисленными в соответствие с (2), (3) численно с помощью метода трапеций:

$\mu_{ИК}^{RE} = \text{mean}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right), \sigma_{ИК}^{RE} = \text{std}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right),$
представленные на рис. 11.

Из рис. 11 видно, что случайная последовательность $\left[N_{ИК}^{RE}\right]_i, i = \overline{1, 87}$ является некоторой реализацией случайной величиной с

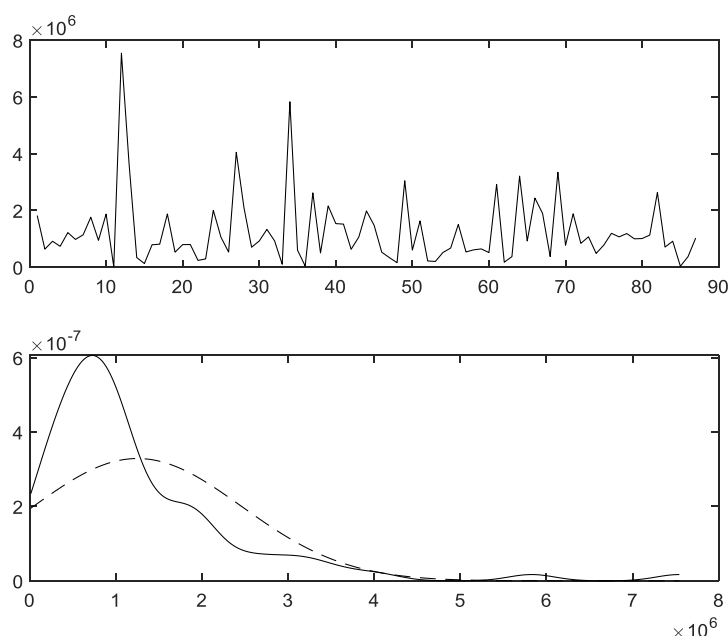


Рис. 11. Визуализация последовательности $\left[N_{ИК}^{RE}\right]$ (снизу), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=300.21$) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (---) (снизу)

ограниченной областью рассеяния [14 575,7 543 682], закон распределения которой отличается от нормального закона. Наименьшее число избирателей – 14 575 было зарегистрировано в ИК «город Байконур (Республика Казахстан)» (Республика Казахстан) (число созданных УИК 7). Наибольшее число зарегистрированных избирателей – 7 543 682 было зарегистрировано в ИК «Московская область» (число созданных ТИК – 73, число созданных УИК 4238).

Значения $\mu_{ИК}^{RE}$ и $\sigma_{ИК}^{RE}$, округленные до ближайшего целого числа, оказались равными 1 216 166 и 1 089 289, соответственно. Сравнивая значение $\mu_{ИК}^{RE}$ с отношением числа всех зарегистрированных избирателей $N_{all}^{RE} = 109\,001\,306$ к числу ИК субъектов РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)» и «Территория за пределами РФ», $N_{ИК} = 89$:

$$N_{all}^{RE} / N_{ИК} = 1\,224\,733,$$

видим, что с точностью до 0.7% эти значения оказываются одинаковыми.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Ре-

спублика Казахстан)», «Территория за пределами РФ», $\left[N_{ИК}^{RE}\right]_i, i = \overline{1, 87}$, которые приняли участие в выборах Президента РФ в 2018 г., аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность распределения соответствующего нормального закона распределения с параметрами, вычисляемыми в соответствии с (2), (3) численно с помощью метода трапеций

$$\mu_{ИК}^{RE} = \text{mean}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right), \sigma_{ИК}^{RE} = \text{std}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right),$$

представленные на рис. 12.

Из рис. 12 видно, что случайная последовательность $\left[N_{ИК}^{RE}\right]_i, i = \overline{1, 87}$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [14 575,7 543 682]. Наименьшее число избирателей, принявших участие в выборах, – 96 595 было зарегистрировано в ИК «город Байконур (Республика Казахстан)» (Республика Казахстан) (число созданных УИК 7). Наибольшее число избирателей, принявших участие в выборах, – 7 543 682 было зарегистрировано в ИК «город Москва» (число созданных ТИК – 128, число созданных УИК – 3 615).

Значения $\mu_{ИК}^{V}$ и $\sigma_{ИК}^{V}$, округленные до ближайшего целого числа, оказались равными

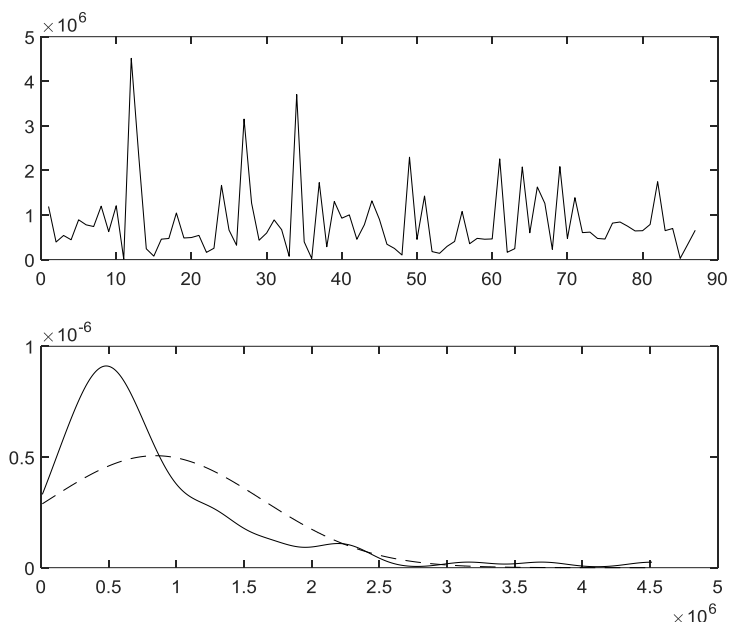


Рис. 12. Визуализация последовательности $[N_{ИК}^V]_i$ (снизу), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=184576$) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (---) (снизу)

824028 и 722799, соответственно. Сравнивая отношение $\mu_{ИК}^V$ к $\mu_{ИК}^{RE}$:

$$(\mu_{ИК}^V / \mu_{ИК}^{RE}) \cdot 100 = 67.76\%,$$

с официальным значением явки избирателей в 2018 г. на выборы Президента РФ (67.5%), видим, что найденная оценка явки избирателей в 2018 г. на выборы Президента РФ отличается от соответствующего значения, вычисленного по данным сайта ЦИК, не более, чем на 0.39%.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $[N_{ИК}^{\%}]_i, i=1,87$, которые приняли участие в выборах Президента РФ в 2018 г., аппроксимацию Розенблатта-Парзена ее плотности вероятности соответствующего нормального закона распределения с параметрами, вычисляемыми в соответствие с (2), (3) численно с помощью метод трапеций:

$$\mu_{ИК}^{\%} = \text{mean}(\{[N_{ИК}^{\%}]_i\}), \sigma_{ИК}^{\%} = \text{std}(\{[N_{ИК}^{\%}]_i\}),$$

представленные на рис. 13.

Из рис. 13 видно, что обсуждаемая последовательность $[N_{ИК}^{\%}]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [50.68,98.01]%, закон распределения которой отличается от

нормального закона распределения. Наименьшая явка избирателей – 50.68% оказалась в ИК «Иркутская область» (число ТИК – 45, число УИК – 1 916, число зарегистрированных избирателей – 1 877 547). Наибольшая явка избирателей – 98.01% оказалась в ИК «Территория за пределами РФ», (ТИК не создавались, число УИК – 393, число зарегистрированных избирателей – 483 957).

Значения $\mu_{ИК}^{\%}$ и $\sigma_{ИК}^{\%}$ оказались равными 66.22% и 9.68%, соответственно. Таким образом, значения оценки явки избирателей в 2018 г. на выборы Президента РФ на уровне ИК, отказались отличными от официальных данных не более, чем на 1.8%.

Рассмотрим последовательность, составленную из значений числа УИК, созданных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $[N_{ИК}^{УИК}]_i, i=1,87$, аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность соответствующего нормального закона распределения с параметрами, вычисляемые в соответствие с (2), (3) численно:

$$\mu_{ИК}^{УИК} = \text{mean}(\{[N_{ИК}^{УИК}]_i\}), \sigma_{ИК}^{УИК} = \text{std}(\{[N_{ИК}^{УИК}]_i\}),$$

представленные на рис. 14.

Из рис. 14 видно, что случайная последовательность $[N_{ИК}^{УИК}]_i, i=1,87$ является некоторой реализацией случайной величиной с

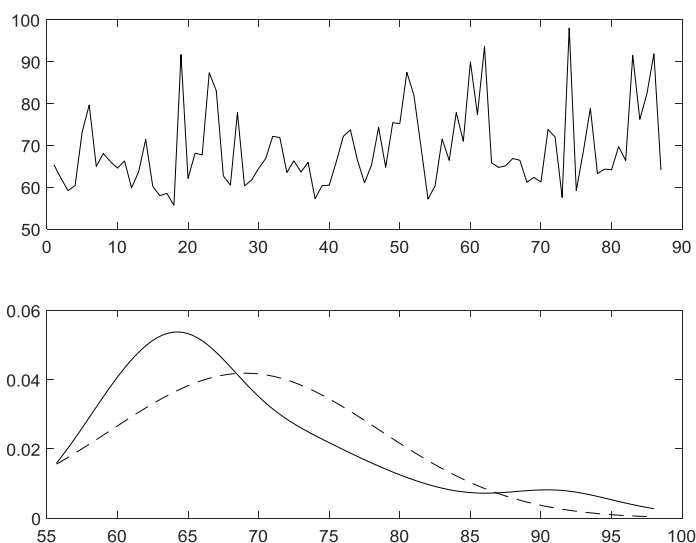


Рис. 13. Визуализация последовательности $[N_{ИК}^{\%}]_i$ и аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=3.3813$) (сверху), ее плотности распределения и плотности соответствующего нормального закона распределения (--) (снизу)

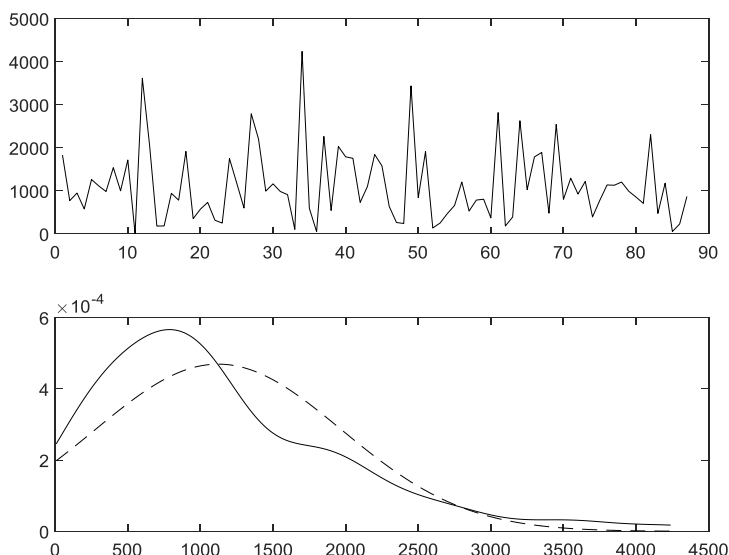


Рис. 14. Визуализация последовательности $[N_{ИК}^{УИК}]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=300.21$) ее плотности распределения и плотность распределения соответствующего нормального закона распределения (--) (снизу)

ограниченной областью рассеяния [4,4 238], закон распределения которой отличается от нормального закона распределения. Наименьшее число УИК – 7 было создано в ИК «город Байконур (Республика Казахстан)» (число зарегистрированных избирателей – 14 575). Наибольшее число УИК – 4 238 было создано в ИК «Московская область» (число зарегистрированных избирателей – 5829578).

Значения $\mu_{ИК}^{УИК}$ и $\sigma_{ИК}^{УИК}$, округленные до ближайшего целого числа, оказались равными 1 106 и 800, соответственно. Отметим, что

значение $\mu_{ИК}^{УИК}$ согласуется со значением $\mu_{УИК}^{RE}=1117$, ранее вычисленным на основе анализа последовательности $\{[N_{УИК}^{RE}]_i\}$, $i = \overline{1, 97695}$.

Рассмотрим последовательность, составленную из значений числа ТИК, созданных в ИК данного субъекта РФ, $[N_{ИК}^{ТИК}]_i, i = \overline{1, 85}$ (напомним, что в ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ» ТИК не создавались), аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность распределения соответствующе-

го нормального закона распределения с параметрами, вычисляемыми в соответствие с (2), (3) численно с помощью методатрапеций

$$\mu_{ИК}^{ТИК} = \text{mean}\left(\left\{\left[N_{ИК}^{ТИК}\right]_i\right\}\right), \sigma_{ИК}^{ТИК} = \text{std}\left(\left\{\left[N_{ИК}^{ТИК}\right]_i\right\}\right),$$

представленные на рис. 15.

Из рис. 15 видно, что случайная последовательность $\left[N_{ИК}^{ТИК}\right]_i, i=1,85$ является некоторой реализацией случайной величины с ограниченной областью рассеяния $[2,128]$, закон распределения которой отличается от нормального закона. Наименьшее число ТИК – 2 было создано в ИК «Ненецкий автономный округ» (число УИК – 51, число зарегистрированных избирателей – 39470).

Наибольшее число ТИК – 128 было создано в ИК «город Москва» (число УИК – 3 615, число зарегистрированных избирателей – 7 543 682).

Значения $\mu_{ИК}^{ТИК}$ и $\sigma_{ИК}^{ТИК}$, округленные до ближайшего целого числа, оказались равными 32 и 19, соответственно. Вычисляя произведение $\mu_{ИК}^{ТИК}$ и $\mu_{ИК}^{РЕ}$, получаем:

$$\mu_{ИК}^{ТИК} \cdot \mu_{ИК}^{РЕ} = 1249696,$$

что согласуется с соответствующим значением $\mu_{ИК}^{РЕ} = 1216166$.

Рассмотрим графики зависимостей числа УИК и числа ТИК, зарегистрированных в дан-

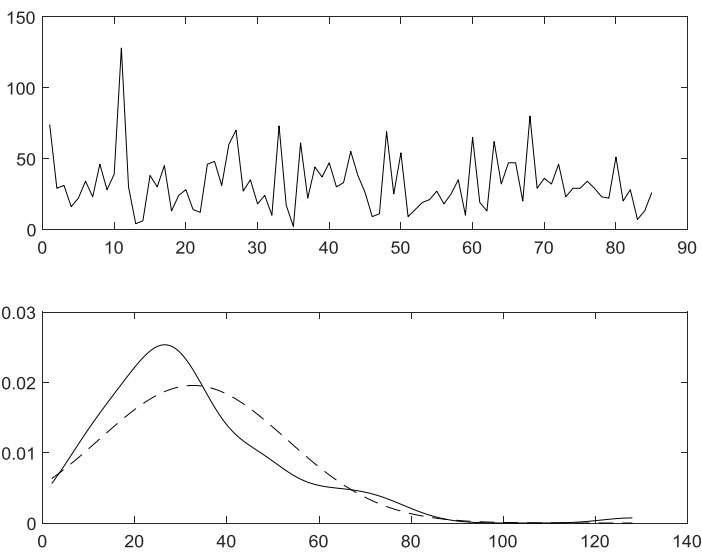


Рис. 15. Визуализация последовательности $\left[N_{ИК}^{ТИК}\right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, $h^*=6.489$) ее плотности распределения и плотности распределения соответствующего нормального закона (–) (снизу)

ной ИК, от числа избирателей, зарегистрированных в данной ИК – $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{РЕ}\right]_i\right)$, $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{РЕ}\right]_i\right)$, соответственно, их линейные аппроксимации, представленные на рис. 16, а также соответствующие статистики, представленные ниже.

Из рис. 16 видно, что зависимости $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{РЕ}\right]_i\right)$, $\left[N_{ИК}^{ТИК}\right]_i = f\left(\left[N_{ИК}^{РЕ}\right]_i\right)$ могут быть аппроксимированы с помощью линейной функции. При этом, как видно, из приведенных выше статистик данных аппроксимаций у линейной модели № 1 коэффициент $R^2=0.667$, у линейной модели № 2 – $R^2=0.759$, что, с нашей точки зрения, свидетельствует о достаточно высокой линейной связи между переменными $N_{ИК}^{УИК}$ и $N_{ИК}^{РЕ}$, а также $N_{ИК}^{УИК}$ и $N_{ИК}^{РЕ}$. Более высокое значение коэффициента R^2 у второй линейной модели объ-

ясняется более высоким влиянием распределения избирателей по УИК, чем распределения избирателей по ТИК. Таким образом, результаты статистического анализа электоральных данных на уровне ИК, также подтверждают сделанные выше выводы на уровнях УИК и ТИК об отсутствии фальсификаций электоральных данных в ходе проведения в 2018 г. выборов Президента РФ.

Рассмотрим графики зависимостей относительной явки избирателей от числа УИК и числа ТИК, зарегистрированных в данной ИК, – $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$, $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$, соответственно, их линейные аппроксимации, представленные на рис. 17, а также соответствующие статистики, представленные ниже.

Анализ зависимостей $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$, $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$ представленных на рис.

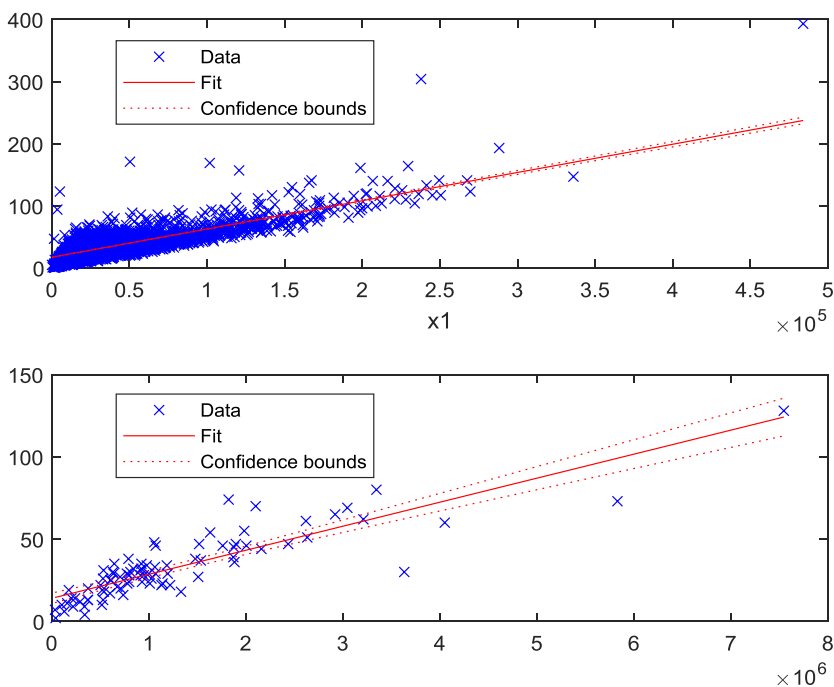


Рис. 16. Визуализация зависимостей $[N_{IK}^{YIK}]_i = f([N_{IK}^{RE}]_i)$ (сверху), $[N_{IK}^{TIK}]_i = f([N_{IK}^{RE}]_i)$ (снизу) и их линейных аппроксимаций

Линейная модель № 1. Статистики линейной модели, аппроксимирующей зависимость

$$[N_{IK}^{YIK}]_i = f([N_{IK}^{RE}]_i)$$

Linear regression model:

$$y \sim 1 + x1$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	17.329	0.3487	49.695	0
x1	0.00045462	6.0647e-06	74.962	0

Number of observations: 2778, Error degrees of freedom: 2776

Root Mean Squared Error: 13.4

R-squared: 0.669, Adjusted R-Squared: 0.669

F-statistic vs. constant model: 5.62e+03, p-value = 0

Linear regression model:

$$y \sim 1 + x1$$

Линейная модель № 2. Статистики линейной модели, аппроксимирующей зависимость

$$[N_{IK}^{TIK}]_i = f([N_{IK}^{RE}]_i)$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	14.02	1.5883	8.8269	1.4501e-13
x1	1.4602e-05	9.0308e-07	16.169	2.2484e-27

Number of observations: 85, Error degrees of freedom: 83

Root Mean Squared Error: 10.1

R-squared: 0.759, Adjusted R-Squared: 0.756

F-statistic vs. constant model: 261, p-value = 2.25e-27

18 и статистик из линейных аппроксимаций, представленных выше позволяет сделать вывод об отсутствии детерминированной зависимости между переменными $N_{ИК}^{\%}$ и $N_{ИК}^{УИК}$ а также $N_{ИК}^{\%}$ и $N_{ИК}^{ТИК}$, то есть на уровне ИК явка избирателей, выраженная в процентах, не зависит от числа УИК и ТИК, зарегистрированных данной ИК субъекта РФ.

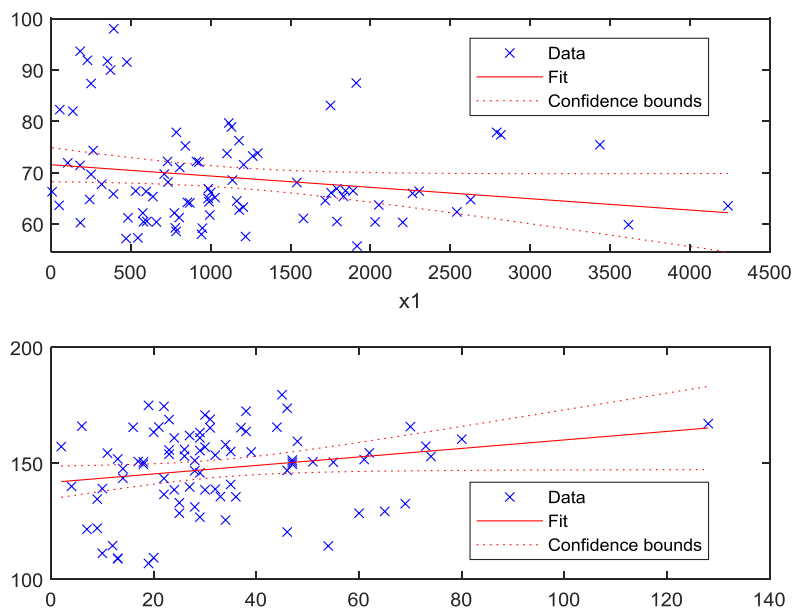


Рис. 17. Визуализация зависимостей $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$ (сверху), $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$ (снизу) и их линейных аппроксимаций

Линейная модель № 1. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$

Linear regression model:
y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	71.529	1.6751	42.702	3.2903e-59
x1	-0.0022081	0.0011919	-1.8526	0.067406

Number of observations: 87, Error degrees of freedom: 85
Root Mean Squared Error: 9.4
R-squared: 0.0388, Adjusted R-Squared: 0.0275
F-statistic vs. constant model: 3.43, p-value = 0.0674

Линейная модель № 2. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$

Linear regression model:
y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	141.75	3.5616	39.799	7.6562e-56
x1	0.18349	0.092648	1.9805	0.050964

Number of observations: 85, Error degrees of freedom: 83
Root Mean Squared Error: 17.3
R-squared: 0.0451, Adjusted R-Squared: 0.0336
F-statistic vs. constant model: 3.92, p-value = 0.051

Заключение

Проведенный многомерный статистический анализ электоральных данных, представленных в ЦИК РФ по результатам выборов во 2018 г. Президента РФ, на уровнях участковых избирательных комиссий (УИК), территориальных избирательных комиссий (ТИК) и избирательных комиссий субъектов РФ (ИК) и приравненных к ним ИК «город Байконур» (Республика Казахстан)», «Территория за пределами РФ» позволяют сделать следующие выводы:

1. случайные последовательности, составленные из значений числа зарегистрированных избирателей, числа избирателей, принявших участие в выборах, и относительной явки избирателей на выборы на уровнях УИК, ТИК и ИК, представляют собой некоторые выборки случайной величины с ограниченной областью рассеяния;

2. плотности распределений изученных случайных последовательностей не описываются нормальным законом распределения;

3. число УИК, ТИК, проводивших выборы, линейно зависит от числа зарегистрированных в данных комиссиях избирателей;

4. число избирателей, принявших участие в выборах, на уровне УИК, ТИК и ИК, не зависит от числа избирателей, зарегистрированных в данных избирательных комиссиях;

5. относительная явка избирателей на выборы, выраженная в процентах, на уровне УИК, ТИК и ИК не зависит от числа избирателей, зарегистрированных в данных избирательных комиссиях, а также числа зарегистрированных в них УИК и ТИК;

6. результаты статистического анализа электоральных данных на уровнях УИК, ТИК и ИК согласуются друг с другом.

Таким образом, приведенные выше выводы, с нашей точки зрения свидетельствуют об отсутствии признаков фальсификации электоральных данных, как на этапе формирования УИК, ТИК и ИК и регистрации избирателей в данных избирательных комиссиях, так и этапе проведения выборов. Для окончательного вынесения вердикта о возможности использования изученных электоральных

данных в качестве «эталонных» для объективной оценки работоспособности математических методов, которые, по заявлениям ОЭС, обеспечивают выявление фальсификаций результатов выборов, далее будут изучены результаты подсчета голосов, поданных за каждого из кандидатов на пост Президента РФ, а также проведен многомерный анализ электоральных данных, представленных УИК с аномально высокой относительной явкой.

Опыт, полученный авторами в ходе проведенного исследования, свидетельствует о необходимости правового регулирования в области анализа электоральных данных и публикации результатов. В том числе, следует на законодательном уровне закрепить обязательную публикацию описаний:

- первичных электоральных данных и использованных технологий их выгрузки из соответствующих хранилищ данных;

- использованных методик трансформации первичных электоральных в форму, пригодную для проведения их статистического анализа, и обеспечение публичного доступа к этим данным;

- использованных методов и алгоритмов анализа электоральных данных и полученных результатов, а также собственно электоральных данных.

Это позволит проводить независимую верификацию результатов выборов, проводимых РФ в органы государственной власти РФ различных уровней, на основании которых в настоящее время ОЭС предпринимают многочисленные попытки делигитимации результатов выборов с целью, в том числе, обострения общественно-политической ситуации в стране, и переводить послевыборные политические дискуссии (в большинстве своем политизированные) в дискуссии научные.

Статистический анализ результатов голосования за конкретных кандидатов в Президенты РФ в 2018 г., а также электоральных данных на участках с высокой явкой избирателей являются предметом последующих публикаций.

Литература

1. Левин Б. Р. Вероятностные модели и методы в системах связи и управления/Б. Р. Левин, В. Шварц// – М.: Радио и связь, 1985. – 312 с.
2. Тихонов В. И. Статистическая радиотехника/В.И. Тихонов// –М.: Радио и связь, 1982. – 624 с.
3. Кропотов Ю. А Методы оценивания моделей плотности вероятностей акустических сигналов в

телекоммуникациях аудиообмена/ Ю.А. Кропотов// Системы управления, связи и безопасности, 2017. № 1. С. 26–39.

4. Кропотов Ю.А. Моделирование и методы исследования акустических сигналов, шумов и помех в системах телекоммуникаций/ Ю.А. Кропотов, В.А. Ермолаев// – Берлин: Директ-Медиа, 2016. – 251 с.

5. Сушко Д.В. Оптимальная аппроксимация частотных вероятностей/Д.В. Сушко//Информационные процессы, 2018. Том 18. № 1. С. 40–54.

6. Новоселов А.А. Параметризация моделей управляемых систем // Вестник государственного аэрокосмического университета им. академика М.Ф. Решетнева, 2010. № 5 (31). С. 52–56.

7. Поршнев С.В., Копосов А.С. Случайные величины с ограниченной областью рассеяния: математическое и алгоритмическое обеспечение для оценивания плотностей вероятностей и функций распределений/С.В. Поршнев, А.С. Копосов// –М.: Горячая линия-Телеком, 2018. –184 с.

8. Киреева Н.В., Чупахина Л.Р. Сравнение возможностей использования различных методов аппроксимации для анализа трафика с самоподобным распределением/Н.В. Киреев, Л.Н. Чупахина// Международный журнал прикладных и фундаментальных исследований, 2016. № 12. С. 1287–1289.

9. Поршнев С.В., Божалкин Д.А. Математическое и алгоритмическое обеспечение для анализа характеристик информационных потоков в магистральных интернет-каналах/С.В. Поршнев, Д.А. Божалкин// –М.: Горячая линия-Телеком, 2021. –214 с.

10. Сызранцев В.Н. Расчет прочностной надежности изделий на основе методов непараметрической статистики/В.Н. Сызранцев, Я.Н. Невелев, С.Л. Голофаст// –Новосибирск: Наука, 2008. –128 с.

11. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников/ А.И. Кобзарь// –М.: ФИЗМАТЛИТ, 2006. –816 с.

12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // The Annals of Mathematical Statistics., 1956. – Т. 27. Vol. 3. –P. 832–837.

13. Боровков А.А. Математическая статистика. – М.: Наука, 1984, 472 с.

14. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P. 1065–1076.

15. Суетин П.К. Классические ортогональные многочлены. – М.: Наука, 1976. – 328 с.

16. Димаки А.В., Светлаков А.А. Аппроксимация плотностей распределений случайных величин с применением ортогональных полиномов Чебышева–Эрмита/А.В. Димаки, А.А. Светлаков// Известия Томского политехнического университета, 2006. Т. № 8. С. 6–11.

17. Голик Ф. В. Аппроксимация эмпирических распределений вероятностей полиномами Бернштейна/ Ф.В. Голик// Журнал радиоэлектроники [электронный журнал]. 2018. № 7. Режим доступа: <http://jre.cplire.ru/jre/jul18/5/text.pdf> (дата обращения 16 октября 2021 г.) DOI 10.30898/1684-1719.2018.7.5.

18. Лемешко Б.Ю. О нахождении параметра размытости непараметрических оценок функции плотности/ Б.Ю. Лемешко, С.Н. Постовалов, А.В. Французов// Труды V международной конференции «Актуальные проблемы электронного приборостроения», АПЭП-2000. Новосибирск 26–29 сентября, 2000. В 7 тт. –Новосибирск: Новосибирский государственный технический университет, 2000. –Т. 6. – С. 17–20.

19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, Annal Statistic, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: <https://doi.org/10.1214/10-AOS799>.

20. Silverman B.W. Density Estimation for Statistics and Data Analysis. –Chapman and Hall/CRC, 1986. –175 p.

21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. Pattern Recognition Letters, Volume 33, Issue 13, 2012, P. 1717–1724.

22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, Electric Power Systems Research, Volume 121, April 2015, P. 28–37.

23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, Energy Conversion and Management, Volume 93, 2015, P. 414-434.

24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association, Volume 91, № 433, March 1996, P. 401–4071.

25. Симахин В.А. Робастные непараметрические оценки: адаптивные оценки взвешенного максимального правдоподобия в условиях статистической априорной неопределенности/ В.А. Симахин // –Saarbrücken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –292 с.

References

1. Levin B. R. Veroyatnostnyye modeli i metody v sistemakh svyazi i upravleniya/B. R. Levin, V. Shvarts// – M.: Radio i svyaz', 1985. – 312 s.
2. Tikhonov V. I. Statisticheskaya radiotekhnika/V.I. Tikhonov// –M.: Radio i svyaz', 1982. – 624 s.
3. Kropotov YU. A Metody otsenivaniya modeley plotnosti veroyatnostey akusticheskikh signalov v telekommunikatsiyakh audioobmena/ YU.A. Kropotov// Sistemy upravleniya, svyazi i bezopasnosti, 2017. № 1. S. 26–39.
4. Kropotov YU.A. Modelirovaniye i metody issledovaniya akusticheskikh signalov, шумов i pomekh v sistemakh telekommunikatsiy/ YU.A. Kropotov, V.A. Yermolayev// – Berlin: Direkt-Media, 2016. – 251 c.
5. Sushko D.V. Optimal'naya approksimatsiya chastotnykh veroyatnostey/D.V. Sushko// Informatsionnyye protsessy, 2018. Tom 18. № 1. S. 40–54.
6. Novoselov A.A. Parametrizatsiya modeley upravlyayemykh sistem // Vestnik gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva, 2010. № 5 (31). S. 52–56.
7. Porshnev S.V., Koposov A.S. Sluchaynyye velichiny s ogranichennoy oblast'yu rasseyaniya: matematicheskoye i algoritmicheskoye obespecheniye dlya otsenivaniya plotnostey veroyatnostey i funktsiy raspredeleniy/S.V. Porshnev, A.S. Koposov// –M.: Goryachaya liniya-Telekom, 2018. –184 s.
8. Kireyeva N.V., Chupakhina L.R. Sravneniye vozmozhnostey ispol'zovaniya razlichnykh metodov approksimatsii dlya analiza trafika s samopodobnym raspredeleniyem/N.V. Kireyev, L.N. Chupakhina// Mezhdunarodnyy zhurnal prikladnykh i fundamental'nykh issledovaniy, 2016. № 12. S. 1287–1289.
9. Porshnev S.V., Bozhalkin D.A. Matematicheskoye i algoritmicheskoye obespecheniye dlya analiza kharakteristik informatsionnykh potokov v magistral'nykh internet-kanalakh/S.V. Porshnev, D.A. Bozhalkin// –M.: Goryachaya liniya-Telekom, 2021. –214 s.
10. Syzrantsev V.N. Raschet prochnostnoy nadezhnosti izdeliy na osnove metodov neparametricheskoy statistiki/V.N. Syzrantsev, YA.N. Nevelev, S.L. Golofast// –Novosibirsk: Nauka, 2008. –128 s.
11. Kobzar' A.I. Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov/A.I. Kobzar'// –M.: FIZMATLIT, 2006. –816 s.
12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // The Annals of Mathematical Statistics., 1956. – T. 27. Vol. 3. –P. 832–837.
13. Borovkov A.A. Matematicheskaya statistika. – M.: Nauka, 1984, 472 s.
14. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P. 1065–1076.
15. Suyetin P.K. Klassicheskiye ortogonal'nyye mnogochleny. – M.: Nauka, 1976. – 328 s.
16. Dimaki A.V., Svetlakov A.A. Approksimatsiya plotnostey raspredeleniy sluchaynykh velichin s primeneniye ortogonal'nykh polinomov Chebysheva–Ermita/A.V. Dimaki, A.A. Svetlakov// Izvestiya Tomskego politekhnicheskogo universiteta, 2006. T. № 8. S. 6–11.
17. Golik F. V. Approksimatsiya empiricheskikh raspredeleniy veroyatnostey polinomami Bernshteyna/ F.V. Golik// Zhurnal radioelektroniki [elektronnyy zhurnal]. 2018. № 7. Rezhim dostupa: <http://jre.cplire.ru/jre/jul18/5/text.pdf> (data obrashcheniya 16 oktyabrya 2021 g.) DOI 10.30898/1684-1719.2018.7.5.
18. Lemesheko B.YU. O nakhozhenii parametra razmytosti neparametricheskikh otsenok funktsii plotnosti/ B.YU. Lemesheko, S.N. Postovalov, A.V. Frantsuzov// Trudy V mezhdunarodnoy konferentsii «Aktual'nyye problemy elektronnoy priborostroyeniya», APEP-2000. Novosibirsk 26–29 sentyabrya, 2000. V 7 tt. –Novosibirsk: Novosibirskiy gosudarstvennyy tekhnicheskii universitet, 2000. –T. 6. –S. 17–20.
19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, Annal Statistic, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: <https://doi.org/10.1214/10-AOS799>.
20. Silverman B.W. Density Estimation for Statistics and Data Analysis. –Chapman and Hall/CRC, 1986. –175 p.
21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. Pattern Recognition Letters, Volume 33, Issue 13, 2012, P. 1717–1724.
22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, Electric Power Systems Research, Volume 121, April 2015, P. 28–37.
23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, Energy Conversion and Management, Volume 93, 2015, P. 414–434.
24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association, Volume91, № 433, March 1996, P. 401–4071.

ПОРШНЕВ Сергей Владимирович, доктор технических наук, профессор, директор Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: s.v.porshnev@urfu.ru

РЯБКО Николай Юрьевич, аспирант Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: N.Yu.Ryabko@urfu.ru

PORSHNEV Sergey Vladimirovich, Doctor of Technical Sciences, Professor, Director of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: s.v.porshnev@urfu.ru

RYABKO Nikolay Yurievich, post-graduate student of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education "Ural Federal University named after the first President of Russia B.N. Yeltsin". Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: N.Yu.Ryabko@urfu.ru