

УДК 616-006:519.237.5

DOI: [10.26102/2310-6018/2024.46.3.023](https://doi.org/10.26102/2310-6018/2024.46.3.023)

## Прогноз распространенности онкозаболеваний среди жителей Московского региона на основе модели регрессии

В.С. Степанов✉

*Центральный экономико-математический институт РАН, Москва,  
Российская Федерация*

**Резюме:** В статье сделана попытка выявить связь между частотой случаев рака на территориях и факторами окружающей среды, с учетом показателя демографии. Установлена регрессионная зависимость распространенности онкологических заболеваний на территориях городских округов Московской области и ряда округов столицы с долей пожилых жителей и рядом санитарно-гигиенических показателей территорий. Комплекс факторных объясняющих переменных включал показатель загрязненности атмосферного воздуха территории, два показателя с концентрациями озона и бенз(а)пирена, качественные переменные по уровню ее техногенного загрязнения и объемам сброса загрязненной воды, долю пожилых граждан. Учитывалось и ежедневное курение сигарет взрослыми. На этой основе построена модель регрессии с переменной структурой, которая имеет коэффициент детерминации 98,5 % и ошибку аппроксимации ниже 2 %. Параметры модели оценивались по методу наименьших квадратов по данным из 51 городского округа области и 5 округов г. Москвы. Наличие лагов у факторов позволяет сделать прогноз численности контингента лиц, страдающих опухолями любой локализации, в муниципальном разрезе и с горизонтом планирования 1 год. На основе созданной модели можно более эффективно планировать ряд мероприятий первичной профилактики и распределять медицинские ресурсы.

**Ключевые слова:** регрессионная модель, загрязнение атмосферного воздуха, сброс загрязненных сточных вод, бенз(а)пирен, приземный озон, взвешенные частицы, техногенное загрязнение, злокачественное новообразование, городской округ, муниципальный район.

**Для цитирования:** Степанов В.С. Прогноз распространенности онкозаболеваний среди жителей Московского региона на основе модели регрессии. *Моделирование, оптимизация и информационные технологии*. 2024;12(3). URL: <https://moitvvt.ru/ru/journal/pdf?id=1644> DOI: 10.26102/2310-6018/2024.46.3.023 (In Russ.).

## The forecast of the prevalence of cancer among residents of the Moscow region based on a regression model

V.S. Stepanov✉

*Central Economics and Mathematics Institute of the RAS, Moscow, the Russian Federation*

**Abstract:** The article makes an attempt to identify the relationship between cancer prevalence in urban areas and several environmental factors, taking into account a demographic indicator. The regression dependence of the prevalence of oncologic diseases in the territories of urban districts of the Moscow region and several districts of the capital with the proportion of elderly residents and a number of sanitary and hygienic indicators of the territories has been established. The complex of factor explanatory variables included the indicator of atmospheric air pollution of the territory, two variables with the concentration of surface ozone and benz(a)pyrene on it, qualitative variables in terms of the level of its man-made pollution and the volumes of polluted water discharge, the proportion of elderly population. Daily cigarette smoking by adults is also taken into account. On this basis, a regression model with a variable structure is constructed, which has a determination coefficient of 98.5% and an approximation error below 2%. The model parameters were estimated using the least squares method

based on data for 51 urban districts of the region and 5 districts of Moscow. The presence of lags in the factors makes it possible to make a forecast of the number of people suffering from tumors of any localization, in the municipal context and with a planning horizon of 1 year. Based on the created model, it is possible to plan primary prevention measures more effectively and allocate medical resources.

**Keywords:** regression model, atmospheric air pollution, discharge of polluted waste water, benz(a)pyrene, surface ozone, suspended particles, technogenic pollution, malignant neoplasm, city district, municipality.

**For citation:** Stepanov V.S. The forecast of the prevalence of cancer among residents of the Moscow region based on a regression model. *Modeling, Optimization and Information Technology*. 2024;12(3). <https://moitvivr.ru/ru/journal/pdf?id=1644> DOI: 10.26102/2310-6018/2024.46.3.023 (In Russ.).

## Введение

Развитие злокачественных новообразований (ЗНО) среди населения является важной медико-социальной проблемой, так как рак является одной из девяти социально-значимых болезней [1]; это особенно актуально для пожилых граждан. При этом если смертность россиян из-за болезней сердца и сосудов заметно снижается с 2005 г., то смертность из-за ЗНО – почти постоянна с 2000 г. [2]. Кроме прямых потерь от смертности важен высокий уровень затрат, связанных с лечением патологии и ростом инвалидности, как и потери в экономическом развитии регионов. Для решения многих из этих проблем весьма привлекательны мероприятия по первичной профилактике, уменьшающие риск развития ЗНО у населения; исследования здесь встречаются редко [3].

Этиология ЗНО имеет многофакторную природу. После серии исследований Международное агентство по изучению рака (IARC) установило с высокой степенью достоверности категории (группы) факторов риска развития ЗНО, где категория 1 – самая опасная. Она включает ряд факторов окружающей среды и образа жизни населения: курение табака, низкая физическая активность, чрезмерное потребление алкоголя, канцерогены на рабочем месте или в атмосферном воздухе, ионизирующая радиация и т. д. [4–6]. К факторам риска часто причисляют и хронический стресс, пожилой возраст, избыточную массу тела, наличие ряда хронических заболеваний или вирусных инфекций, генетические особенности и др. [1, 4]. Группа 1 включает в себя, в частности, углеводород «бенз(а)пирен» (БП), а также озон –  $O_3$ . Первое из этих веществ имеет химическую формулу  $C_{20}H_{12}$ , возникая в сильных концентрациях при сжигании угля, нефти и ее продуктов, органики, при пожарах. Всемирная организация здравоохранения (ВОЗ) рекомендует для постоянного мониторинга список из БП,  $O_3$ , взвешенных веществ ( $PM$ ) и др. [7]; о влиянии  $PM$  на риск развития ЗНО сообщается в [8]. Автор обсуждает здесь роль загрязнения атмосферного воздуха из-за  $PM$  как фактора риска канцерогенеза, в зависимости от размера, происхождения, химического состава и концентрации  $PM$  с более мелким диаметром ( $PM_{2.5}$ ,  $PM_1$ ). Эти  $PM$  – особенно дорожно-транспортного происхождения, включающие тяжелые металлы с переходной валентностью – наиболее опасны. Долговременное воздействие со стороны этих  $PM$  усиливает риск развития ЗНО у автоводителей (и у жителей вблизи загруженных дорог).

Повышенные концентрации  $PM$  с разными диаметрами:  $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_1$ , – и любые концентрации  $O_3$  также неблагоприятны для жителей средних и крупных городов в плане вероятности развития ряда форм ЗНО [8, 9]; с фокусировкой на Московский регион вопрос с озоном рассмотрен в [10]. Газ  $O_3$  сравнительно редко фиксируется постами наблюдения в нашей стране; например, даже в столице концентрации  $O_3$

измеряют около 12 измерительных станций, причем даже не во всех ее округах <sup>1</sup>. Вред от концентраций  $PM_{10}$  (это крупная пыль и проч.) и  $O_3$  в воздухе городов иллюстрируется следующим примером из обзора [11]. Согласно ВОЗ, кратковременное воздействие  $PM_{10}$  увеличивает суточную смертность населения от естественных причин на 0,46–0,62 % в случае прироста концентраций  $PM_{10}$  на каждые 10 мкг/м<sup>3</sup> (в среднем, за сутки). В Москве относительный прирост этой смертности из-за такого роста концентрации  $PM_{10}$  оценивался 20 лет назад в 0,33 %. В то же время, в результате повышения концентраций  $O_3$  эта смертность выросла в среднем на 1,1 %. Иначе говоря, высокие уровни  $O_3$  приводили здесь к тройному росту риска смертности, связанного с концентрациями  $PM_{10}$ .

Кроме того, анализировались корреляционные связи между распространенностью случаев ЗНО и рядом характеристик популяционного здоровья. Для этого по данным 2014–2016 гг. рассчитывались ранговые коэффициенты корреляции Спирмена, затем после учета демографической структуры населения обнаружилось, что чем выше доля пожилых и меньше доля молодежи, тем чаще в регионах наблюдались случаи рака. Это же относится и к доле населения старше трудоспособного возраста [1, 12–14]. Аналогичные взаимосвязи наблюдаются и за рубежом: на территориях, расположенных вблизи от промышленных центров, или там, где проживает значительная доля пожилых; так у горожан выше риск развития ЗНО, чем у лиц из сельской местности [15].

Менее сильные, но значимые корреляции были выявлены с фактором «потребление алкогольных напитков» [1], что подтверждалось для варианта чрезмерного их потребления в [4, 14]. Кроме того, частота встречи случаев ЗНО в регионах была значимо выше там, где наблюдались высокие вредные выбросы в атмосферу от передвижных источников (обычно – в густонаселенных местах), а также – сброс в природу плохо очищенных сточных вод [1]. Это согласуется с выводами, полученными для модели системных взаимодействий при исследованиях распространенности ЗНО среди жителей Приморского края [16]. И, кроме того, – с тем, что факторы по загрязненности атмосферного воздуха и сбросам загрязненной воды были значимыми в моделях множественной регрессии по региональным и муниципальным данным [13, 14]. По оценке *Greenpeace*, загрязнение воздуха в городах РФ дает в среднем 5–7 % от общей смертности <sup>2</sup>.

Загрязнение воздуха определяется по значениям средних и / или максимальных разовых концентраций примесей. Степень загрязнения воздуха оценивается при сравнении фактических концентраций с предельно допустимыми концентрациями (ПДК). Основные источники загрязнения атмосферы в Московском регионе, который формируют столица РФ и ее область – это тепловые электростанции и ТЭЦ, бытовые котельные, предприятия по нефтепереработке, нефтехимии, химии. Кроме того, это предприятия автомобилестроения, металлургии, машиностроения, стройиндустрии, электротехники и различные виды транспорта (особенно – автотранспорт). Оказалось, что «около 60 % территории Московской области хронически загрязнены токсикантами промышленного происхождения» [17].

Информативность таких факторов, как возраст пациента, воздействие химических канцерогенов и ряда других, выявилась и на индивидуальном уровне при исследованиях «случай-контроль»: [4], а также в [18], где анализировалось состояние здоровья несколько сотен лиц, страдающих из-за ЗНО, и где ставилась цель по выявлению

<sup>1</sup> ГПБУ «Мосэкомониторинг». Измерительные станции. URL: <https://mosecom.mos.ru/stations> (дата обращения: 02.07.2024).

<sup>2</sup> Информационное агентство ТАСС. Карманный датчик загрязнения воздуха поступил в продажу. URL: <https://tass.ru/plus-one/4600936> (дата обращения: 02.07.2024).

значимых факторов риска развития рака. Для этого наблюдения за пациентами базовой группы, куда вошли все случаи с ЗНО, сравнивались с лицами контрольной группы (где нет ЗНО). У каждого испытуемого после анкетирования измерялись десятки показателей; около 10 из них измерялись у пациента врачом. Результаты опросов анализировались по каждому показателю через  $t$ -тест Стьюдента для сравнения средних значений в группах. В итоге среди наиболее значимых различий этих групп были: возраст человека, загрязненность воздуха на его рабочем месте и / или наличие там вредных химических веществ [18].

Сообщим ряд сведений по поводу оценки уровня  $O_3$  в городах. Инновационная фирма из Франции – *Plume Labs*, созданная около 10 лет назад, – разработала наукоемкие продукты, связанные с данными о загрязнении атмосферного воздуха в городах мира. Эти данные поставляются  $\approx 11$  тыс. станциями мониторинга. Также она создала доступное приложение для *iOS* и *Android*. Ее software ежечасно отслеживает концентрации мелкодисперсных  $PM_{2.5}$  (это – пыльца, споры и т. д.) и более крупных –  $PM_{10}$ , а также загрязняющих газов:  $NO_2$  и  $O_3$ . Эти приложения легкодоступны смартфонам из России с 2015 г.; за простым интерфейсом в них стоит мощная система обработки данных *Plume Air Cloud*. Эта фирма имеет широкую клиентскую базу в мире, а ее продукты находят применение у частных лиц, бизнеса и госструктур<sup>2</sup>. В этом году корпорация *AccuWeather*<sup>3</sup> интегрировала эти данные в ее прогнозы, поглотив и саму *Plume Labs*. В нашем случае, из *Plume Labs* были взяты ежесуточные концентрации  $O_3$  для городов РФ.

В плане выбора инструментария отметим, что в исследованиях заболеваемости и распространенности ЗНО на тех или иных территориях достаточно часто предлагаются различные модели машинного обучения: линейная регрессия и т. д. [12–14, 19–21] и др.

Целью нашей работы является построение многофакторной модели регрессии, имеющей переменную структуру. Эта модель связывает комплекс факторных переменных с зависимой переменной  $Y$ . Последней является численность  $N$  контингента лиц, которые проживают на ряде территорий г. Москвы или области и зафиксированы в учреждениях Минздрава к концу года  $t$  как страдающие от ЗНО. Значение  $N$  берется в соотношении со средней за год численностью постоянного населения территории; отсюда  $Y$  имеет размерность «число лиц со ЗНО на 100 тысяч населения». Факторные переменные включают демографический показатель территории и ряд ее санитарно-гигиенических или экологических показателей, часть из которых – качественные; их градации описывались фиктивными переменными. Также три важных фактора вошли с лагами в нашу модель.

Данная работа является продолжением исследований [13, 14], в которых решалась задача построения аналогичных моделей машинного обучения на основе анализа панельных наборов данных для вариантов «мезо-1» (муниципальный разрез) и «мезо-2» (региональный разрез). В разделе работы «Обсуждение» выполняется сравнительный анализ созданных нами моделей на нескольких примерах.

## Материалы и методы

Объектами настоящего исследования были: а) данные Росстата, полученные на территориях для городского округа (г. о.) и муниципальных районов Московского региона за 2016–2020 гг.; б) данные Росгидромета со среднегодовыми концентрациями БП в атмосферном воздухе городов на этой территории; в) данные со среднесуточными концентрациями  $O_3$  в атмосферном воздухе городских округов (или других населенных

<sup>3</sup>AccuWeather, Inc. URL: <https://accuweather.com> (дата обращения: 02.07.2024).

пунктов территории), взятые из приложения *Plume Labs*; з) данные по объемам сброса загрязненных сточных вод (Росстат, Росводресурсы); д) информация о техногенных загрязнениях (знания, данные НПО «Тайфун», картографические материалы). В таблицу с материалом обучения, на основе которой оценивались параметры модели, переменная  $Y$  обычно входила за 2020 г.; она включает значения частот выявленных случаев со ЗНО для 51 г. о. Подмосковья и 5 округов столицы с  $Y$  за 2018 год<sup>4, 5</sup>.

Во введении приводилось обоснование выбора набора факторных объясняющих переменных. Их ежегодные значения для некоторых из них были легкодоступны, а другие рассчитывались нами по простым формулам. В итоге, в роли факторов, вошедших далее в уравнение модели, причем с учетом сильных ограничений источников<sup>6, 7</sup>, были взяты:

$X_1$  – доля пожилых на территории: мужчин возраста 60+ и женщин – 55+, в %<sup>6</sup>;

$X_2$  – показатель загрязненности воздуха: доля исследованных проб атмосферного воздуха в населенных пунктах, где имело место превышение среднесуточной ПДК<sub>с.с.</sub> по любому из нормируемых вредных веществ (или доля случаев в %, где соотношение среднегодовой концентрации  $Q$  каждой из примесей в замерах было выше ПДК<sub>с.с.</sub>)<sup>7</sup>;

$X_3$  – средняя концентрация  $O_3$ , рассчитанная в промежутке 1.VII–5.IX.2022 г., в атмосферном воздухе территории: городе или другом ее населенном пункте, в мг/м<sup>3</sup>;

$X_4$  – переменная со среднегодовым уровнем БП, который воздействовал на организм среднестатистического жителя территории в результате вдыхания им атмосферного воздуха и из-за курения сигарет; расчет значений фактора  $X_4$  сделан по формуле из [13, с. 5] и данным<sup>8, 9</sup>;

$d_1, d_2$  – фиктивные переменные (dummy), которые описывают уровень техногенного загрязнения территории из-за различных причин: уровень радиационного фона, наличие неподалеку мусоросжигающих заводов, полигонов с бытовыми отходами и т. д.; при этом  $d_1, d_2$  брались соответственно равными (0,0), если он был «низкий»; или (1,0) – «умеренный» и (0,1) – «повышенный»;

$d_3$  – dummy-переменная, кодирующая уровень еще одной переменной  $X_5$ :  $d_3 = 1$ , если  $X_5 \geq 5$  или  $d_3 = 0$ , если  $X_5 < 5$ ; здесь  $X_5 = (V_{св}/S)^{0,5}$ , где  $V_{св}$  – объем сброса загрязненных стоков в году  $t$  на территории в поверхностные водные объекты, [млн. м<sup>3</sup>], а  $S$  – ее площадь, [тыс. км<sup>2</sup>]<sup>10, 6</sup>. При этом в числителе у переменной  $X_5$  взята взвешенная

<sup>4</sup> Распоряжение Правительства Московской области от 25 июня 2019 г. N 521-ПП «Об утверждении региональной программы «Борьба с онкологическими заболеваниями» на период 2019-2024 годов (в ред. распоряжений Правительства Московской области от 29 октября 2019 г. N 890-ПП; от 21 июня 2021 г. N 485-ПП): Таблица 21. Показатель распространенности ЗНО на 100 тыс. населения; Региональная программа «Борьба с онкологическими заболеваниями» для г. Москвы от 17 июня 2019 г. (утверждена Заместителем Мэра Москвы в Правительстве Москвы по вопросам социального развития А.В. Раковой).

<sup>5</sup> г. Зеленоград, г.о.Троицк и округ Новая Москва и территориальные округа: ВАО, СВАО и ЮЗАО.

<sup>6</sup> Управление Федеральной службы государственной статистики по г. Москве и Московской области (Мосстат). База данных «Показатели муниципальных образований (БД ПМО)». URL: <https://77.rosstat.gov.ru/folder/68025> (дата обращения: 09.07.2024).

<sup>7</sup> Государственный доклад «О состоянии санитарно-эпидемиологического благополучия населения в Московской области в 2018 году». URL: <https://50.rospotrebnadzor.ru/documents/10156/eb48963a-56ad-41b8-be43-ce48ea5c9e45> (дата обращения: 09.07.2024).

<sup>8</sup> НПО «Тайфун». Справки, ежегодники по загрязнению ОС: Справка о среднемесячных концентрациях бенз(а)пирена в воздухе городов на территории России за 2010-2019 гг. URL: <https://www.rpatyphoon.ru/products/pollution-media.php> (дата обращения: 09.07.2024).

<sup>9</sup> Федеральная служба государственной статистики (Росстат). Итоги комплексного наблюдения условий жизни населения в 2018 году: Таблица 39. Распространенность курения среди лиц в возрасте 15 лет и более. URL: [https://rosstat.gov.ru/free\\_doc/new\\_site/KOUZ18/index.html](https://rosstat.gov.ru/free_doc/new_site/KOUZ18/index.html) (дата обращения: 09.07.2024).

<sup>10</sup> Федеральная служба государственной статистики (Росстат). Охрана окружающей среды в России – 2012 г.: Городская среда: Таблица 11.16. Сброс загрязненных сточных вод в поверхностные водные объекты по

сумма объемов загрязненной воды, которая сброшена без очистки, вместе со сбросами ненормативно очищенной воды; объем последних был уменьшен в 4,5 раза.

Основными математическими методами были корреляционный анализ, множественный регрессионный анализ, методы теории проверки гипотез. Качество построенной модели оценивалось по  $F$ -критерию Фишера, подправленному коэффициенту детерминации  $R_{adj}^2$  (т. е. доле дисперсии величины  $Y$ , которая объясняется моделью) и ошибке аппроксимации MAPE («средняя процентная ошибка») [13, 22]. Кроме того, использовались статистические тесты: критерий Стьюдента для  $H_0$ : « $\beta$ -коэффициент в модели при данном факторе равен нулю» и тест Дарбина-Уотсона на наличие в остатках регрессии автокорреляции через статистику  $DW$  [22]. Для правильной интерпретации этих тестов еще применялись два различных критерия согласия распределения остатков регрессии с нормальным законом распределения. Дополнительно проверялось условие гомоскедастичности после рассмотрения особенностей диаграммы с этими остатками.

Обработка данных и их статистический анализ выполнены средствами *MS-Excel*.

### Результаты

После работы с данными, приближенно аналогичной [13, 14], была создана модель:

$$Y^{(t)} = b_0 + b_1 \cdot X_1^{(t-1)} + b_2 \cdot X_2^{(t-2)} + b_3 \cdot X_3 + b_4 \cdot \ln(X_4^{(t-1)}) + b_5 \cdot d_1 + b_6 \cdot d_2 + b_7 \cdot d_3, \quad (1)$$

где  $Y^{(t)}$  (или, для краткости,  $Y$ ) – контингент лиц, страдающих из-за ЗНО, зафиксированных к концу года  $t$  для городского округа или муниципального района в расчете на  $10^5$  его жителей. На соответствующей территории также наблюдались значения факторов  $X_1, \dots, X_4, d_1, d_2, d_3$ ; часть из них взята с лагами относительно текущего года  $t$ . Постоянная составляющая в (1) – сумма первого и трех последних ее членов, – изменяется по территориям в зависимости от сочетания значений указанных факторов  $d_1, \dots, d_3$ .

В Таблице 1 приводятся значения оценок  $b_0, \dots, b_7$  для  $\beta$ -коэффициентов, которые найдены по методу наименьших квадратов (МНК), а также –  $t$ -статистик Стьюдента. Последние нужны для проверки статистической гипотезы  $H_0$ : « $\beta_j = 0$ » о значимости факторов (при  $j = 1, \dots, 7$ ). Далее даны достигнутые здесь уровни значимости ( $p$ -значения).

Таблица 1 – Результаты оценивания  $\beta$ -коэффициентов, значения лагов и статистик  
Table 1 – Results of  $\beta$ -coefficients estimation, lag values and statistics

Факторная объясняющая переменная		Лag $\tau$	Оценка $\beta$	Значение $t$
1	2	3	4	5
$b_0$ – оценка константы для параметра $\beta_0$			708,93	(9,3)
$X_1$	Доля жителей старше трудоспособного возраста, в %	1	79,51	(27,1)
$X_2$	Частота загрязнений атмосферного воздуха, в %	2	93,96	(4,5)
$X_3$	Среднесуточная концентрация озона – $O_3$ , в (мг/м <sup>3</sup> )		11,65	(2,3)

отдельным городам с численностью населения 100 и более тысяч человек. URL: [https://rosstat.gov.ru/bgd/regl/b12\\_54/Main.htm](https://rosstat.gov.ru/bgd/regl/b12_54/Main.htm) (дата обращения: 09.07.2024).

Таблица 1 (продолжение)  
Table 1 (continued)

$\ln(X_4)$	Концентрация бенз(а)пирена (БП), в долях ПДК, или $X_4$	1	358,7	(16,2)
$d_1$	Уровень загрязнения территории – «средний»		305,12	(14,7)
$d_2$	Уровень загрязнения территории – «более высокий»		842,78	(18,9)
$d_3$	Есть ли высокие уровни сброса загрязненной воды?	3	143,68	(6,3)
$n = 56$	$R^2_{adj} = 98,5\%$ , $RMSE = 60,7$ ; $DW = 1,95$ ; $F(48,7) = 510,6$		$D_n = 0,086$	$\chi^2 = 2,36$

Приложение формулы (1) ко всем  $n$  наблюдениям из Московского региона, образовавшим строки таблицы обучения, приводит к средней процентной ошибке  $MAPE = 1,8(\pm 1,23)\%$ . Оценка коэффициента детерминации  $R^2$  получилась равной  $98,67\%$ , а ее скорректированное с целью уменьшения смещения значение  $R^2_{adj} = 98,5\%$  [22]. В столбце 4 приводятся численные значения  $b_0, \dots, b_7$ , а в столбце 5 – значения  $t$ -статистик. Для константы и каждого из шести факторов, кроме переменной  $O_3$ ,  $p$ -значение было меньше, чем  $0,0001$ , тогда как для фактора  $O_3$  оно было равным  $0,0256$ . В нижней строке таблицы также приводится среднеквадратичная ошибка  $RMSE$ , статистика  $DW \approx 2$  и  $F$ -статистика Р. Фишера, которая является высокой, и поэтому гипотеза  $H_0: \langle \beta_1 = \dots = \beta_7 = 0 \rangle$  отвергается с  $p$ -значением явно меньшим, чем  $0,0001$ . Последние две статистики относятся к критериям согласия. В правом нижнем углу здесь приводится значение статистики  $\chi^2$ . Верхняя  $5\%$ -ная точка  $Q$  распределения  $\chi^2$  с параметром  $\nu = 6$  равна  $12,6$ ; поэтому наблюдаемое значение  $\chi^2$  не попадает в критическую область. В связи с этим сложная гипотеза о том, что остатки модели распределены нормально (со средним  $0$  и неизвестной дисперсией  $d^2$ ) по критерию  $\chi^2$  не отвергается. Также максимальное расхождение  $D_n$  между эмпирической функцией распределения и функцией Гаусса – малое (Рисунок 1 б) и гипотеза о нормальности не отвергается и по тесту Лиллиефорса.

Свойство гомоскедастичности остатков регрессии проверялось на диаграмме их рассеяния (Рисунок 1 а), на которой по оси абсцисс отложены оценки  $Y$  через (1), а по оси ординат – остатки для  $n$  наблюдений из таблицы с материалом обучения, по которым оценивались параметры. Заметного нарушения условия гомоскедастичности здесь не наблюдается [22]. Также на качественном уровне был исследован вопрос о мультиколлинеарности переменных. Этот термин обозначает сильную корреляцию между факторами; ее наличие ухудшает качество построенных оценок параметров. Оказалось, что в матрице коэффициентов корреляции между парами переменных нет таких, которые по модулю выше  $0,6$ ; причем коэффициент корреляции  $X_j$  с  $Y$  обычно был выше, чем коэффициент между переменной  $X_j$  и каждым из прочих шести факторов.

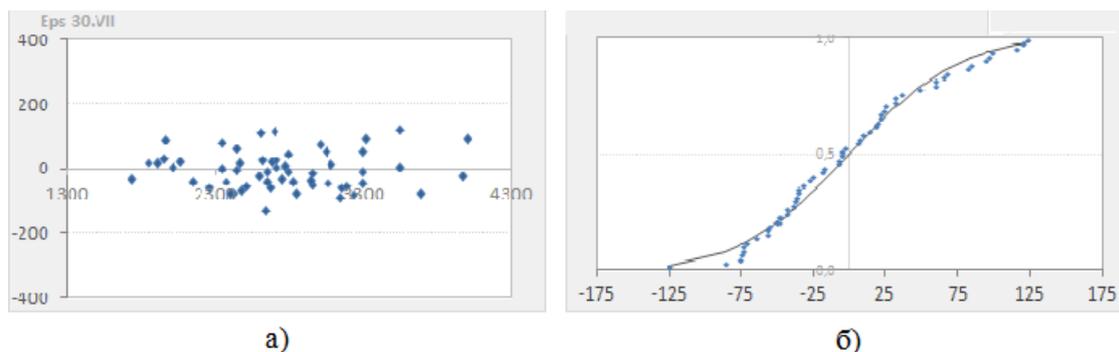


Рисунок 1 – Остатки в модели (1): а) их разброс; б) две функции распределения  
Figure 1 – Residuals in the model (1): a) their spread; b) two distribution functions

Из изложенного выше следует, что выполняются условия теоремы Гаусса-Маркова, в которой обоснована корректность применения в модели регрессии оценок по МНК [22].

### Обсуждение

После изучения матрицы корреляций с переменными, значений  $t$ -статистик, диаграммы рассеяния  $Y$  с  $X_i$  и средних коэффициентов эластичности следует, что основным фактором, определяющим распространенность ЗНО на территории, была доля лиц старшего возраста среди жителей. Следующий по важности фактор – концентрация БП, обусловленная и загрязнениями атмосферы, и ежесуточным курением сигарет. Потом идет концентрация  $O_3$ . Это соответствует тому, что БП и  $O_3$ , согласно критериям IARC, очень опасны, так как относятся к группе 1. А затем идет техногенное загрязнение территории. Про важную роль учета тропосферного  $O_3$  писали и наши видные ученые<sup>11</sup>. Также известно, что концентрации  $O_3$  растут от года к году в среднем на  $\approx 1-2\%$ .

Далее приведем примеры приложений формулы (1) к прогнозированию  $Y$ , в ее сопоставлении с  $Y$  по факту, согласно сведениям региональных программ. Так, для нашей северной столицы и ее области опубликованы значения  $Y$  в 2017–2018 гг. Если нашу модель применить к 11 административным районам г. о. Санкт-Петербурга с  $Y$  за 2017 г.<sup>12</sup> и одному району Ленобласти за 2018 г., то это приведет к значению  $MAPE = 1,34(\pm 1,6)\%$ .

На Рисунке 2 показаны другие примеры, где модель (1) обозначена через mod.3. Здесь взяты данные для г. о. из Ростовской, Рязанской областей, Республики Башкортостан и Краснодарского края в сравнении с фактами, взятыми из их программ борьбы с болезнями из онкологии. Кроме того, на графиках приводятся аналогичные прогнозы из [23, 24], полученные по моделям регрессии: mod.2 – из [13] и mod.1 – [14].

В работе [25] приводится карта Ростовской области, где имеется и г. Азов; в ее основе лежит комплексный показатель «медико-экологической безопасности». Последний построен по комплексу переменных: частоте случаев 4-х конкретных локализаций ЗНО, значению комплексной «антропо- и техногенной нагрузки» (уровень загрязнения воздуха, воды, почвы, рост шумовой нагрузки), а также – оригинальному индексу природно-ресурсного потенциала. Небольшой г. о. Донецк лежит на границе области: в 26 км западнее от г. Каменск-Шахтинский (Рисунок 2). Также здесь на основе

<sup>11</sup> ИА REGNUM. Котельников С.Н., Степанов Е.В., Щербаков И.А., Ивашкин В.Т. Приземный озон в России: чем он опасен. URL: <https://regnum.ru/article/2099148> (дата обращения: 10.07.2024).

<sup>12</sup> Административные районы г.о. Санкт-Петербурга: Василеостровский, Выборгский, Кировский, Колпинский, Красносельский, Кронштадский, Московский, Невский, Петродворцовый, Пушкинский, Фрунзенский; Киришский район Ленинградской области.

software *ArcGIS-10* анализируется заболеваемость раком предстательной железы за 15 лет. Здесь наглядно показана связь показателей экологии с рисками развития этой локализации ЗНО.

Следует также обратить внимание на весьма важную для Московского региона проблему утилизации твердых бытовых отходов (ТБО), в которых почти 1/3 составляют неперерабатываемые отходы [26]. Только 10 % из них подвергались в 2017 г. промышленной переработке на объектах г. Москвы, а очень высокие 62 % вывозились на полигоны в Подмоскowie. Наиболее активно использовались в нем два полигона: ТБО «Тимохово» (Богородский район) и «Сабурово» (Щелковский район). Поэтому мы приписывали здесь значения для  $d_1$ ,  $d_2$  в первом случае – (1, 0) и (0, 1) – во втором. Они связаны с объектами, негативно влияющими на состояние окружающей среды<sup>13</sup>.

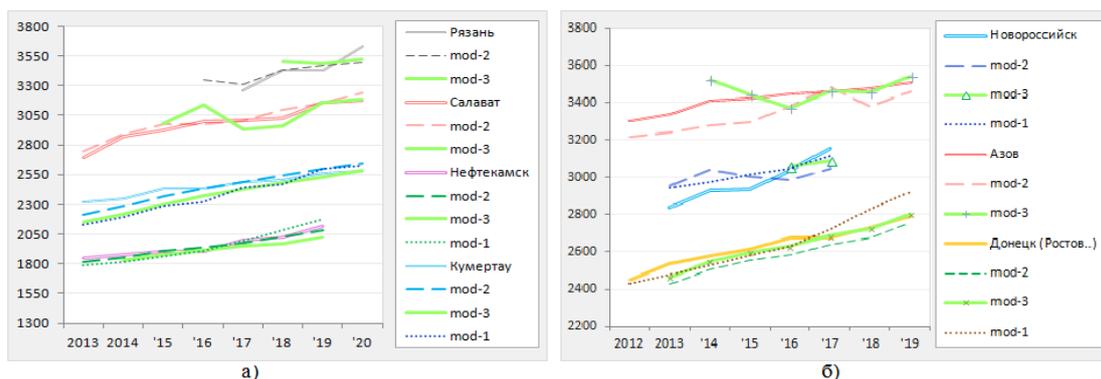


Рисунок 2 – Распространенность ЗНО в городах (г.о.): а) – из ЦФО, ПФО; б) – из ЮФО  
Figure 2 – Cancer prevalence in a few towns: a) by Central part of Russia; b) by the South part

В 2016 г. в нашей столице было три мусоросжигательных завода «Спецзаводы № 2, 3, 4». В результате работы таких заводов возрастает риск развития ЗНО. Переработке в столице тогда подвергались всего 10 % ТБО и около 60 % промышленных отходов; важно, что последние отходы, для 3–4-го классов их опасности, размещались на полигонах ТБО в области. Среди населения, что проживает в непосредственной близости от ряда полигонов ТБО, наблюдается высокий уровень заболеваемости ЗНО. Количество действующих полигонов ТБО в этой области в 2017 г. было 18, причем 12 расположены близко к жителям; также здесь приводятся весьма полезные карты по этой проблематике [26]. К 2017 году было одобрено строительство еще пяти новых полигонов и четырех мусоросжигающих заводов ТБО в Новой Москве и области: вблизи г. о. Воскресенск, Солнечногорск, Наро-Фоминск и Электросталь. Поэтому для г. о. Воскресенск и Наро-Фоминский переменным  $d_1$ ,  $d_2$  были приписаны значения (0,1); кроме того, был взят только сам г. Солнечногорск. Пара  $d_1$ ,  $d_2$  имеет значения (0,1) и для г. о. Волоколамский, где наблюдается повышенный радиационный фон<sup>14</sup> и имеется полигон ТБО «Ярцево».

<sup>13</sup> Филиал «Газпром ПХГ» (подземное хранение газа); фирма «ЭНА» (производство агрегатов для электронасосов); «Опытный завод № 31-ГА» (производство аэродромной техники); завод «Спецмонтажизделие» (метизное производство); «ОМЗ НИИХИММАШ» (производство машин и оборудования для изготовления пищевых продуктов); «Щелковский завод вторичных драгоценных металлов» (переработка сырья с драгметаллами); «Щёлковский металлургический завод» (производство лент из холоднокатаных сталей и ряда сплавов); «Щёлковское рудоуправление» (добыча и обработка известняка, гипсового камня); «Щёлково Агрохим» (производство пестицидов для сельского хозяйства).

<sup>14</sup> НПО «Тайфун». Справки, ежегодники по загрязнению ОС: Данные по радиоактивному загрязнению территории населенных пунктов РФ цезием-137, стронцием-90 и плутонием-239+240. URL: <https://www.rpatyphoon.ru/products/pollution-media.php> (дата обращения: 14.07.2024).

## Заключение

Построена и оценена по точности регрессионная модель с переменной структурой. Она связывает численность контингента лиц, страдающих от ЗНО и проживающих на административных территориях Московского региона с рядом переменных: по возрастному составу жителей, ежедневному курению сигарет взрослыми, состоянию окружающей среды (загрязнения атмосферного воздуха и сбросы загрязненной сточной воды) и другим техногенным загрязнениям. После обработки 56 наблюдений, собранных в муниципальном разрезе, модель объяснила 98,5 % дисперсии зависимой переменной и привела к средней процентной ошибке MAPE = 1,8(±1,23) %. Созданная модель сопоставлялась по точности с двумя примерно аналогичными моделями для 7 городов центральной и южной частей РФ за ряд лет; результаты наглядно показаны на графиках.

На основе этих моделей можно сделать прогноз численности пациентов, страдающих из-за ЗНО, в значительном числе муниципальных территорий РФ с горизонтом в 1 год. Такие прогнозы полезны при разработке мероприятий по медико-демографической и экологической политике, проводимой на территориях городских округов или муниципальных районов Московского региона и ряда других субъектов РФ.

## СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Будилова Е.В., Лагутин М.Б. Социально значимые заболевания населения России и факторы среды (по 84 субъектам РФ за 2014–2016 гг.). *Вестник Московского университета. Серия XXIII: Антропология*. 2019;(4):87–104. <https://doi.org/10.32521/2074-8132.2019.4.087-104>  
Budilova E.V., Lagutin M.B. Socially significant diseases of the Russian population and environmental factors (84 regions of the Russian Federation for 2014–2016). *Vestnik Moskovskogo universiteta. Seriya XXIII: Antropologiya = Moscow University Anthropology Bulletin*. 2019;(4):87–104. (In Russ.). <https://doi.org/10.32521/2074-8132.2019.4.087-104>
2. Kolpak E.P., Frantsuzova I.S., Kuvshinova K.V., Senkov R.E. Neoplasm Morbidity among the Population of Russia. *International Journal of Advanced Biotechnology and Research*. 2017;8(3):2315–2322.
3. Доможирова А.С. Комплексное медико-статистическое прогнозирование как основа перспективного планирования специализированной онкологической помощи в регионах. *Онкология. Журнал им. П.А. Герцена*. 2016;5(1):47–50. <https://doi.org/10.17116/onkolog20165147-50>  
Domozhirova A.S. Complex health statistical prediction as a basis for the long-term planning of specialized cancer care in the areas of the Chelyabinsk Region. *Onkologiya. Zhurnal im. P.A. Gertsena = P.A. Herzen Journal of Oncology*. 2016;5(1):47–50. (In Russ.). <https://doi.org/10.17116/onkolog20165147-50>.
4. Ефимова Н.В., Мыльникова И.В. Оценка вклада онкогенных факторов в риск развития злокачественных новообразований у городского населения трудоспособного возраста. *Анализ риска здоровью*. 2021;(3):99–107. <https://doi.org/10.21668/health.risk/2021.3.09>  
Efimova N.V., Myl'nikova I.V. Assessment of the contribution made by oncogenic factors to the risk of malignant neo-plasms development for the urban population of working age. *Health Risk Analysis*. 2021;(3):99–106. <https://doi.org/10.21668/health.risk/2021.3.09.eng>

5. Заридзе Д.Г., Максимович Д.М. Профилактика злокачественных новообразований. *Успехи молекулярной онкологии*. 2017;4(2):8–25. <https://doi.org/10.17650/2313-805X-2017-4-2-8-25>  
Zaridze D.G., Maksimovich D.M. Prevention of malignant neoplasms. *Uspekhi molekulyarnoy onkologii = Advances in Molecular Oncology*. 2017;4(2):8–25. (In Russ.). <https://doi.org/10.17650/2313-805X-2017-4-2-8-25>
6. Соленова Л.Г. Современные подходы к оценке влияния загрязнения окружающей среды на онкологический риск. *Успехи молекулярной онкологии*. 2020;7(1):17–22. <https://doi.org/10.17650/2313-805X-2020-7-1-17-22>  
Solénova L.G. Current approaches to assessment of the impact of the environmental contamination on cancer risk. *Uspekhi molekulyarnoy onkologii = Advances in Molecular Oncology*. 2020;7(1):17–22. (In Russ.). <https://doi.org/10.17650/2313-805X-2020-7-1-17-22>
7. Салтыкова М.М., Балакаева А.В., Федичкина Т.П., Бобровницкий И.П. Основные причины смертности, обусловленной загрязнением воздуха. *Гигиена и санитария*. 2020;99(4):337–343. <https://doi.org/10.47470/0016-9900-2020-99-4-337-343>  
Saltykova M.M., Balakaeva A.V., Fedichkina T.P., Bobrovnikskii I.P. Leading air pollution related reasons of death. *Gigiena i sanitaria = Hygiene and Sanitation*. 2020;99(4):337–343. (In Russ.). <https://doi.org/10.47470/0016-9900-2020-99-4-337-343>
8. Колпакова А.Ф. О связи антропогенного загрязнения воздуха взвешенными частицами с риском развития онкологических заболеваний (обзор литературы). *Гигиена и санитария*. 2020;99(3):298–302. <https://doi.org/10.47470/0016-9900-2020-99-3-298-302>  
Kolpakova A.F. On the relationship of anthropogenic air pollution by particulate matter with cancer risk. *Gigiena i sanitaria = Hygiene and Sanitation*. 2020;99(3):298–302. (In Russ.). <https://doi.org/10.47470/0016-9900-2020-99-3-298-302>
9. Звягинцев А.М., Кузнецова И.Н., Шалыгина И.Ю., Лезина Е.А., Лапченко В.А., Никифорова М.П., Демин В.И. Исследования и мониторинг приземного озона в России. *Труды Гидрометцентра России*. 2017;(365):56–70.  
Zvyagintsev A.M., Kuznetsova I.N., Shalygina I.Yu., Lezina E.A., Lapchenko V.A., Nikiforova M.P., Demin V.I. Investigation and monitoring of surface ozone in Russia. *Trudy Gidrometcentra Rossii = Proceedings of the Hydrometcentre of Russia*. 2017;(365):56–70. (In Russ.).
10. Звягинцев А.М. Статистическое прогнозирование концентрации приземного озона в г. Москва. *Метеорология и гидрология*. 2008;(8):49–59.  
Zvyagintsev A.M. Statistical forecast of surface ozone concentrations in Moscow. *Meteorologiya i gidrologiya*. 2008;(8):49–59. (In Russ.).
11. Самородская И.В., Семенов В.Ю., Бойцов С.А. Влияние медицинских и немедицинских факторов на смертность населения: факторы окружающей среды. *Проблемы социальной гигиены, здравоохранения и истории медицины*. 2017;25(5):260–265.  
Samorodskaya I.V., Semenov V.Yu., Boitsov S.A. The impact of medical and non-medical factors on population mortality: environment factors. *Problemi socialnoi gigieni, zdravookhraneniya i istorii meditsini = Problems of Social Hygiene, Public Health and History of Medicine*. 2017;25(5):260–265. (In Russ.).
12. Тарасова С.А. Моделирование онкологической заболеваемости населения Курской области. *System Analysis & Mathematical Modeling*. 2023;5(2):142–152. [https://doi.org/10.17150/2713-1734.2023.5\(2\).142-152](https://doi.org/10.17150/2713-1734.2023.5(2).142-152)

- Tarasova S.A. Modeling of Oncological Morbidity of Kursk Region's Population. *System Analysis & Mathematical Modeling*. 2023;5(2):142–152. (In Russ.). [https://doi.org/10.17150/2713-1734.2023.5\(2\).142-152](https://doi.org/10.17150/2713-1734.2023.5(2).142-152)
13. Степанов В.С. Прогноз общей онкологической заболеваемости в регионах и муниципалитетах России на основе многофакторной модели. *Моделирование, оптимизация и информационные технологии*. 2023;11(1). <https://doi.org/10.26102/2310-6018/2023.40.1.022>  
Stepanov V.S. The forecast of cancer prevalence in the regions and municipalities of Russia based on a multivariate model. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii = Modeling, Optimization and Information Technology*. 2023;11(1). (In Russ.). <https://doi.org/10.26102/2310-6018/2023.40.1.022>
14. Степанов В.С., Рыбкина И.Д., Орлова Е.С. Факторы окружающей среды и вредных привычек в моделях распространенности онкопатологии в муниципальных образованиях Алтайского края и других регионов. *Моделирование, оптимизация и информационные технологии*. 2023;11(4). <https://doi.org/10.26102/2310-6018/2023.43.4.022>  
Stepanov V.S., Rybkina I.D., Orlova E.S. Environmental factors and bad habits in models of the malignancy prevalence in the municipalities of Altai Krai and other regions. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii = Modeling, Optimization and Information Technology*. 2023;11(4). (In Russ.). <https://doi.org/10.26102/2310-6018/2023.43.4.022>
15. Radespiel-Tröger M., Geiss K., Twardella D., Maier W., Meyer M. Cancer incidence in urban, rural, and densely populated districts close to core cities in Bavaria, Germany. *International Archives of Occupational and Environmental Health*. 2018;91(2):155–174. <https://doi.org/10.1007/s00420-017-1266-3>
16. Веремчук Л.В., Жерновой М.В., Кику П.Ф. Экологическая зависимость распространения онкологических заболеваний в Приморском крае. *Успехи наук о жизни*. 2010;(2):107–114. <https://elibrary.ru/item.asp?id=16551607>  
Veremchuk L.V., Zhernovoi M.V., Kiku P.F. Environmental dependence of oncological diseases in the Primorye territory. *Uspexhi nauk o zhizni*. 2010;(2):107–114. (In Russ.). <https://elibrary.ru/item.asp?id=16551607>
17. Черногаева Г.М., Жадановская Е.А., Малеванов Ю.А. Источники загрязнения и качество атмосферного воздуха Московского региона. *Известия Российской академии наук. Серия географическая*. 2019;(2):109–116. <https://doi.org/10.31857/S2587-556620192109-116>  
Chernogaeva G.M., Zhadanovskaya E.A., Malevanov Yu.A. Pollution sources and air quality in the Moscow Region. *Izvestiya Rossiiskoi akademii nauk. Seriya geograficheskaya*. 2019;(2):109–116. (In Russ.). <https://doi.org/10.31857/S2587-556620192109-116>
18. Агарков А.И. Анализ взаимосвязи и значимости медико-социальных факторов риска развития злокачественных новообразований у лиц, проживающих в промышленно развитом регионе. *Системный анализ и управление в биомедицинских системах*. 2014;13(1):240–244.  
Agarkov A.I. Analysis of interrelation and importance of medical and social risk factors of malignant new growths development in industrially developed region. *Sistemnyj analiz i upravlenie v biomedicinskih sistemah = System Analysis and Management in Biomedical Systems*. 2014;13(1):240–244. (In Russ.).
19. Ермолицкая М.З., Кику П.Ф., Абакумов А.И. Модель оценки влияния показателей качества жизни на заболеваемость раком молочной железы. *Здравоохранение*

- Российской Федерации*. 2024;68(2):116–122. <https://doi.org/10.47470/0044-197X-2024-68-2-116-122>
- Ermolitskaya M.Z., Kiku P.F., Abakumov A.I. Model for assessing the influence of quality of life indicators on the breast cancer incidence. *Zdravoohranenie Rossijskoj Federacii = Health care of the Russian Federation*. 2024;68(2):116–122. (In Russ.). <https://doi.org/10.47470/0044-197X-2024-68-2-116-122>
20. Мешков Н.А., Вальцева Е.А., Баева Ю.И., Крылицына Е.А. Оценка обусловленности заболеваемости населения города Самары воздействием факторов среды обитания. *Известия Самарского научного центра Российской академии наук*. 2017;19(2-2):300–306.  
Meshkov N.A., Valtseva E.A., Baeva Yu.I., Krylitsyna E.A. Assessment the conditionality of Samara city population incidence under the influence of environmental factors. *Izvestiya Samarskogo nauchnogo centra Rossijskoj akademii nauk = Izvestia of Samara Scientific Center of the Russian Academy of Sciences*. 2017;19(2-2):300–306. (In Russ.).
21. Sekeroglu B., Tuncal K. Prediction of cancer incidence rates for the European continent using machine learning models. *Health Informatics Journal*. 2021;27(1). <https://doi.org/10.1177/1460458220983878>
22. Карлберг К. *Регрессионный анализ в Microsoft Excel*. Москва: Вильямс; 2017. 400 с.  
Carlberg C. *Regression Analysis. Microsoft Excel*. Moscow: Vil'yams; 2017. 400 p. (In Russ.).
23. Степанов В.С. О применении моделей регрессии к статистике территорий с целью улучшения средней продолжительности жизни. В сборнике: *III-й Международный демографический форум «Демография и глобальные вызовы»: Том 2: Секция 2–5, 23–25 мая 2024 года, Воронеж, Россия*. Воронеж: Издательство «Цифровая полиграфия»; 2024. С. 765–776. <https://elibrary.ru/item.asp?id=69213259&pff=1>  
Stepanov V.S. On the application of regression models to statistics of territories in order to improve the average life expectancy. In: *III International Demographic Forum "Demography and Global Challenges": Volume 2, Section 2–5, 23–25 May 2024, Voronezh, Russia*. Voronezh: Izdatel'stvo "Tsifrovaya poligrafija"; 2024. pp. 765–776. (In Russ.). <https://elibrary.ru/item.asp?id=69213259&pff=1>
24. Степанов В.С. О приложении моделей регрессии: оценка уровня благосостояния и продолжительности жизни в ПФО. В сборнике: *Инновационные технологии управления социально-экономическим развитием регионов России: Материалы XVI Международной научно-практической конференции, 23–24 мая 2024 года, Уфа, Россия*. Уфа: ИСЭИ УФИЦ РАН; 2024. С. 186–192. <https://doi.org/10.31040/978-5-6051263-4-8>  
Stepanov V.S. On application of regression models: assessment of well-being and life expectancy in the Volga Federal District. In: *Innovatsionnye tekhnologii upravleniya sotsial'no-ekonomicheskim razvitiem regionov Rossii: Materialy XVI Mezhdunarodnoi nauchno-prakticheskoi konferentsii, 23–24 May 2024, Ufa, Russia*. Ufa: ISEI UFITs RAN; 2024. pp. 186–192. (In Russ.). <https://doi.org/10.31040/978-5-6051263-4-8>
25. Архипова О.Е., Черногубова Е.А., Чибичян М.Б., Коган М.И. Эпидемиология рака предстательной железы в Ростовской области. Пространственно-временная статистика. *Онкоурология*. 2016;12(4):52–59. <https://doi.org/10.17650/1726-9776-2016-12-4-52-59>  
Arkhipova O.E., Chernogubova E.A., Chibichyan M.B., Kogan M.I. Epidemiology of prostate cancer in the Rostov region. Spatio-temporal statistics. *Onkourologiya = Cancer Urology*. 2016;12(4):52–59. (In Russ.). <https://doi.org/10.17650/1726-9776-2016-12-4-52-59>

26. Мелконян Р.Г. Мусорный кризис Москвы и Московской области. *Качество и жизнь*. 2017;(4):71–77.  
Melkonyan R.G. Garbage Crisis of Moscow and the Moscow Region. *Kachestvo i zhizn'* = *Quality and Life*. 2017;(4):71–77. (In Russ.).

#### ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

**Степанов Владимир Сергеевич**, кандидат физико-математических наук, ведущий научный сотрудник Центрального экономико-математического института РАН, Москва, Российская Федерация.

*e-mail*: [vladstep0355@gmail.com](mailto:vladstep0355@gmail.com),

[stepanov@cemi.rssi.ru](mailto:stepanov@cemi.rssi.ru)

ORCID: [0000-0002-4478-376X](https://orcid.org/0000-0002-4478-376X)

**Vladimir S. Stepanov**, Candidate of Physical and Mathematical Sciences, Leading Researcher at the Central Economics and Mathematics Institute of the RAS, Moscow, the Russian Federation.

*Статья поступила в редакцию 09.08.2024; одобрена после рецензирования 04.09.2024; принята к публикации 12.09.2024.*

*The article was submitted 09.08.2024; approved after reviewing 04.09.2024; accepted for publication 12.09.2024.*