



## Определение выбросов при построении многомерной регрессионной модели цен на региональном рынке жилой недвижимости

И.А. Казимиров<sup>1✉</sup>, М.В. Матвеева<sup>2</sup>, Н.В. Полякова<sup>3</sup>

<sup>1,2</sup>Иркутский национальный исследовательский технический университет, г. Иркутск, Россия

<sup>3</sup>Байкальский государственный университет, г. Иркутск, Россия

**Аннотация.** Целью проведенного исследования являлся выбор оптимального метода выявления выбросов в исходных данных при построении многомерной регрессионной модели цен на региональном рынке жилой недвижимости. Исследование выполнено на информационной базе предложений по продаже квартир в многоквартирных панельных жилых домах, расположенных в г. Иркутске. Для достижения поставленной цели было выполнено построение базовой мультипликативной многомерной регрессионной модели, описывающей зависимость стоимостных показателей от ценообразующих факторов объектов недвижимости. Построение вариативных регрессионных моделей выполнено путем итерационного удаления из базовой модели выбросов, выявленных различными методами: определение стандартного отклонения (z-оценка); определение расстояния Махаланобиса; разработанный в исследовании метод приведения цен объектов к характеристикам эталонного объекта. Выбор оптимального метода выявления выбросов в исходных данных выполнен путем сравнения характеристик итоговых вариативных многомерных регрессионных моделей, полученных после удаления из них выбросов. На основании анализа полученных результатов оптимальным методом выявления выбросов при построении многомерной регрессионной модели цен на региональном рынке жилой недвижимости признан метод приведения цен объектов к характеристикам эталонного объекта. В результате использования указанного метода существенно снижаются ошибки аппроксимации базовой многомерной регрессионной модели рынка, что повышает адекватность результатов проводимой на ее основе оценки недвижимости.

**Ключевые слова:** корреляционно-регрессионный анализ, многомерная регрессионная модель, ценообразующие факторы, выброс, z-оценка, расстояние Махаланобиса

**Для цитирования:** Казимиров И.А., Матвеева М.В., Полякова Н.В. Определение выбросов при построении многомерной регрессионной модели цен на региональном рынке жилой недвижимости // Известия вузов. Инвестиции. Строительство. Недвижимость. 2024. Т. 14. № 3. С. 462–473. <https://doi.org/10.21285/2227-2917-2024-3-462-473>. EDN: CYETFG.

### Original article

## Determination of outliers when building a multivariate regression model of prices in residential real estate market of the region

Ilya A. Kazimirov<sup>1✉</sup>, Maria V. Matveeva<sup>2</sup>, Nina V. Polyakova<sup>3</sup>

<sup>1,2</sup>Irkutsk National Research Technical University, Irkutsk, Russia

<sup>3</sup>Baikal State University, Irkutsk, Russia

**Abstract.** The paper is aimed at selecting the optimal method for identifying outliers in the initial data when building a multivariate regression model of prices in the regional residential real estate market. The study was based on offers for sale of apartments in prefab flat blocks located in Irkutsk. In this regard, a basic multiplicative multivariate regression model was built, describing the dependence of cost indicators on the pricing factors of real estate. The identified outliers were iteratively removed from the basic model. The methods for detecting outliers included calculation of standard deviation (z-score), calculation of the Mahalanobis distance, as well as a method developed in the study for bringing the prices of objects to

the characteristics of the reference object. The optimal method for detecting outliers in the initial data was selected by comparing the characteristics of the final variable-based multivariate regression models obtained after removing outliers from them. The analysis of the results proved the method of bringing the prices of objects to the characteristics of the reference object to be the optimal method of identifying outliers when building a multivariate regression model of prices in the regional residential real estate market. This method significantly reduces the approximation errors of the basic multivariate regression model of the market, thereby increasing the adequacy of the results of the real estate valuation conducted on its basis.

**Keywords:** correlation and regression analysis, multidimensional regression model, price forming factors, outlier, z-score, Mahalanobis distance

**For citation:** Kazimirov I.A., Matveeva M.V., Polyakova N.V. Determination of outliers when building a multivariate regression model of prices in residential real estate market of the region. *Proceedings of Universities. Investment. Construction. Real estate*. 2024;14(3): 462–473. (In Russ.). <https://doi.org/10.21285/2227-2917-2024-3-462-473>. EDN: CYETFG.

## ВВЕДЕНИЕ

Качество экономико-математических моделей, в частности регрессионных, описывающих зависимость стоимостных показателей объектов недвижимости от ценообразующих факторов, в большой степени зависит от качества используемых при их построении исходных данных. Наличие в исходных данных выбросов – объектов, имеющих значения цен и/или ценообразующих факторов, существенно отличающихся от остальных элементов выборки, – негативно влияет на результаты анализа, поэтому их выявление и устранение является важной частью построения регрессионной модели. Особенности региональных рынков недвижимости являются относительно небольшой объем предложений, высокая волатильность ценовых показателей и существенные отличия объектов по ценообразующим факторам на фоне низкого качества (наличия искажений, неполноты) публикуемой информации. Указанные факторы являются причиной неизбежного наличия выбросов в исследуемых выборках данных.

Анализ публикаций, связанных с разработкой и исследованием регрессионных моделей рынка недвижимости [1–12], показал, что в большинстве случаев проблема выявления выбросов в исходных данных либо не рассматривается вовсе, либо затрагивается в общих чертах, без полноценного описания используемых алгоритмов.

В целом, проблеме выявления выбросов посвящено достаточно много публикаций. Однако, вид анализируемых наборов данных чаще всего ограничен одномерными рядами.

Основная часть рассматриваемых методов поиска выбросов (в том числе используемых применительно к рынку и оценке недвижимости в работах [13]) ограничена объемом выборки до 500 элементов, что не позволяет их использовать при больших объемах данных<sup>1</sup>. Вышесказанное обосновывает целесообразность проведения дополнительных исследований в области определения выбросов в исходных данных при построении многомерных регрессионных моделей регионального рынка недвижимости.

Объектом исследования в данной работе является вторичный рынок жилой недвижимости г. Иркутска. Информационная база исследования: данные о предложениях на рынке. Выбор цен предложений обусловлен доступностью исходных данных<sup>2</sup>.

Предметом исследования являются выбросы в исходных данных, используемых при построении многомерной регрессионной модели цен на недвижимость.

Основная цель проведенного исследования заключалась в выборе оптимального метода поиска выбросов для достижения адекватных результатов оценки.

Исходя из поставленной цели, задачами проведенного исследования являлись:

– построение базовой многомерной регрессионной модели, описывающей зависимость стоимостных показателей от ценообразующих факторов объектов недвижимости;

– построение вариативных регрессионных моделей путем итерационного удаления из базовой модели выбросов, выявленных различными методами;

<sup>1</sup>Грибовский С.В., Сивец С.А., Левыкина И.А. Математические методы оценки стоимости недвижимого имущества. М.: Маросейка, Книжная линия, 2014. 352 с.

<sup>2</sup>Недвижимость Иркутска // Poni38.ru. Режим доступа: <http://poni38.ru/> (дата обращения: 04.01.2023).

– выбор оптимального метода выявления выбросов в исходных данных путем сравнения характеристик итоговых вариативных многомерных регрессионных моделей.

### МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

*Построение базовой многомерной регрессионной модели, описывающей зависимость стоимостных показателей от ценообразующих факторов объектов недвижимости*

Выполнено в порядке, изложенном в [14, 15].

Использовалась мультипликативная многомерная регрессионная модель вида:

$$y = \beta_0 \times x_1^{\beta_1} \times x_2^{\beta_2} \times \dots \times x_n^{\beta_n} \times \varepsilon \quad (1)$$

где

$y$  – модельное значение результирующей переменной;

$x_1, x_2, \dots, x_n$  – факторные переменные;

$\beta_0, \beta_1, \dots, \beta_n$  – коэффициенты уравнения регрессионной модели;

$n$  – число факторных переменных, используемых в модели;

$\varepsilon$  – ошибка аппроксимации.

В качестве результирующей переменной ( $y$ ) в модели использовалась начальная удельная цена предложения общей площади квартир (руб./м<sup>2</sup>) на дату публикации предложения о продаже объекта.

В качестве ценообразующих факторов (факторных переменных  $x_1, x_2, \dots, x_n$ ) в модели использовались:

1) Характеристики квартиры:

– общая площадь помещений (без учета лоджий и балконов, округленно до 10 м<sup>2</sup>, округление площади помещений позволяет рассматривать указанную факторную переменную как дискретную, что дает возможность выполнить для нее подбор аппроксимирующей функции методом оптимизации модели [17], не привязываясь к определенным, субъективно заданным зависимостям);

– этаж;

– внутренняя отделка и ее состояние;

– наличие балкона и его остекление;

– наличие мебели.

2) Характеристики здания и местоположения:

– комплексный ценообразующий фактор с условным названием «Адрес (дом)», агрегирующий влияние на стоимость объекта характеристик, общих для всех квартир в здании.

3) Характеристики рынка:

– индекс цен на дату публикации предложения о продаже объекта.

При формировании выборки объектов, в целях повышения ее однородности, учитывались следующие ограничения:

1) Исследование выполнено на информационной базе предложений по продаже квартир в многоквартирных панельных жилых домах до 2004 г. постройки.

2) Из выборки удалены объекты:

– площадью менее 15 м<sup>2</sup> и/или без санузла, указанные объекты по своим характеристикам ближе к сегменту рынка по продаже комнат;

– многоуровневые квартиры;

– квартиры с отдельным входом и квартиры, явно позиционируемые продавцом как подходящие под коммерческое использование;

– квартиры с черновой и получистовой отделкой;

– квартиры с отделкой в «хорошем/неудовлетворительном состоянии» (объекты, включающие помещения, существенно отличающиеся по состоянию отделки).

В выборку включены предложения с 2018 по 2022 гг. Рассматриваемый период обусловлен имеющейся информацией о предложениях, а также гипотезой об изменении степени влияния ценообразующих факторов на стоимость недвижимости с течением времени.

Ограничение длительности исследуемого периода (пять лет) принято по результатам сравнительного анализа моделей, построенных по предложениям с периодами различной продолжительности.

*Построение вариативных регрессионных моделей*

Выполнено путем удаления из базовой модели выбросов, выявленных различными методами.

Алгоритм поиска и удаления выбросов в каждой вариативной модели – итерационный.

На первой итерации выполнен поиск выбросов в исходных данных базовой регрессионной модели. Последующие итерации включали:

– удаление из выборки выявленных на предыдущей итерации выбросов;

– расчет регрессионной модели;

– поиск выбросов в данных полученной регрессионной модели.

Итоговые вариативные регрессионные модели фиксировались при условии отсутствия в итерации выявленных выбросов.

Определение выбросов в исходных данных при построении вариативных регрессионных моделей производилось:

1) Методом определения стандартного отклонения ( $z$ -оценка).

2) Методом определения расстояния Махаланобиса.

3) Методом приведения цен объектов к характеристикам эталонного объекта.

*Метод определения стандартного отклонения (z-оценка)*

Является распространенным методом выявления выбросов в одномерных рядах данных, в т. ч. в задачах построения эконометрических моделей рынка недвижимости [16].

Основан на определении для каждого элемента одномерного набора данных стандартизированной оценки (z-оценки), показывающей, сколько стандартных отклонений составляет его отличие от среднего значения по выборке.

Метод определяет выброс как элемент со значением z-оценки за пределами (выше или ниже) заданного числа стандартных отклонений от среднего значения по выборке. Чаще всего – 3.

Ограничение метода: выборка должна иметь нормальное распределение.

Поиск выбросов выполнен в свободно распространяемой программной среде вычислений «Loginom Community» (версия программы 7.0.2)<sup>3</sup>. Порядок выявления выбросов изложен в материалах по теме<sup>4,5</sup>.

Выбросы выявлялись как по результирующей переменной (y), так и по всем факторным переменным  $x_1, x_2, \dots, x_n$ . В качестве проверяемых данных для факторных переменных использовались итоговые оптимизированные коды их градаций, полученные с помощью инструмента Microsoft Excel «Поиск решения» [17]. Выбросы определялись при значении z-оценки для любой переменной объекта за пределами трех стандартных отклонений от среднего значения по выборке. Учитывая то, что распределение оптимизированных кодов части ценообразующих факторов не является нормальным, использование данного метода поиска выбросов в настоящем исследовании носит факультативный характер и выполнено в целях проверки возможности его практического использования для решения поставленной задачи.

*Метод определения расстояния Махаланобиса*

Является одним из наиболее распространенных в зарубежной практике [18–27] и единственным рекомендуемым нормативными документами Российской Федерации методом выявления выбросов в многомерных рядах данных<sup>6</sup>.

Основан на преобразовании многомерных данных к одномерной статистике, в качестве которой используется расстояние Махаланобиса, являющееся мерой расстояния от многомерного наблюдения до выборочного среднего набора данных, нормированного при помощи выборочной ковариационной матрицы.

Метод определяет выбросом наблюдение с расстоянием Махаланобиса, чей уровень значимости (p-значение) превышает заданное критическое значение.

Поиск выбросов выполнен в свободно распространяемой программной среде вычислений «R» (версия программы 4.3.0)<sup>7</sup>.

Порядок выявления выбросов изложен в материалах по теме<sup>8,9</sup>. При определении выбросов в исследуемый многомерный набор данных включались результирующая переменная (y) и все факторные переменные  $x_1, x_2, \dots, x_n$ . В качестве проверяемых данных для факторных переменных использовались итоговые оптимизированные коды их градаций, полученные с помощью инструмента Microsoft Excel «Поиск решения» [17]. Выбросы определялись при уровне значимости (p-значении) расстояния Махаланобиса у объекта меньше 0,001.

*Метод приведения цен объектов к характеристикам эталонного объекта*

Отсутствие существенных улучшений качества базовой регрессионной модели в результате удаления из нее выбросов, определенных при помощи вышеуказанных методов (таблица), привело к необходимости разработки собственного метода их выявления. В основе разработанного метода лежит допущение о том, что при корректно подобранных значениях коэффициентов регрессионной модели и оптимизированных кодов градаций ценообразую-

<sup>3</sup>Loginom. Режим доступа: <https://loginom.ru/download> (дата обращения: 23.11.2023).

<sup>4</sup>Loginom. Режим доступа: <https://loginom.ru/download> (дата обращения: 23.11.2023).

<sup>5</sup>Обнаружение и коррекция одномерных выбросов в данных // Loginom. Режим доступа: <https://loginom.ru/blog/outliers> (дата обращения: 23.11.2023).

<sup>6</sup>ГОСТ Р ИСО 16269-4-2017. Статистические методы. Статистическое представление данных. Ч. 4. Выявление и обработка выбросов.

<sup>7</sup>The R Project for Statistical Computing. Режим доступа: <https://www.r-project.org/> (дата обращения: 23.11.2023).

<sup>8</sup>Как рассчитать расстояние Махаланобиса в SPSS // Кодкамп. Режим доступа: <https://www.codecamp.ru/blog/mahalanobis-distance-spss/> (дата обращения: 23.11.2023).

<sup>9</sup>Как рассчитать расстояние Махаланобиса в R // Кодкамп. Режим доступа: <https://www.codecamp.ru/blog/mahalanobis-distance-r/> (дата обращения: 23.11.2023).

щих факторов цены всех объектов, приведенные к некоторым заданным эталонным характеристикам (скорректированные с учетом отличий по ценообразующим факторам от эталонного объекта), должны быть равны между собой, либо иметь случайное отклонение от средней величины по выборке. Исходя из данного допущения, цены всех объектов, имеющих одинаковое значение (градацию) одного из ценообразующих факторов, и скорректированные с учетом отличий от эталонного объекта по остальным ценообразующим факторам, для однородной выборки также должны быть равны между собой, либо иметь случайное отклонение от средней величины. Все объекты, скорректированные указанным образом цены которых существенно отличаются от средней величины по выборке, являются выбросами.

Алгоритм предлагаемого метода предусматривает следующие шаги:

1) Формирование из «базовой» (полной) выборки объектов т. н. «микровыборок», включающих объекты, имеющие одинаковое значение (градацию) одного из ценообразующих факторов. «Микровыборки» формируются для каждой градации каждого используемого в модели ценообразующего фактора.

2) Корректировка цен объектов в целях их приведения к заданным характеристикам эталонного объекта. Проводится по всем ценообразующим факторам, кроме фактора, по которому сформирована «микровыборка». Характеристики эталонного объекта, как правило, определяются по критерию максимального присутствия объектов с данными параметрами в «базовой» выборке. Поправочные коэффициенты рассчитываются по коэффициентам регрессионной модели и оптимизированным кодам градаций ценообразующих факторов.

3) Проверка скорректированных цен объектов в каждой «микровыборке» на наличие выбросов. Объекты, скорректированные цены которых определены внутри «микровыборки» как выбросы, считаются таковыми и для «базовой» выборки объектов.

Приведение цен объектов к характеристикам эталонного объекта позволяет рассматривать «микровыборки» как одномерные ряды данных с одним проверяемым на наличие выбросов фактором – скорректированной ценой объекта.

Поскольку количество уникальных градаций ценообразующих факторов (и, соответ-

ственно, формируемых «микровыборок») достаточно большое, размер каждой «микровыборки» в большинстве случаев существенно меньше «базовой» выборки, что позволяет использовать для поиска выбросов стандартные, хорошо проработанные критерии их выявления в однородных рядах данных, в т. ч. малого объема.

В проведенном исследовании поиск выбросов в «микровыборках» выполнен по критериям:

– Смирнова-Граббса (определен по [13] и специализированной литературе)<sup>10</sup>;

– Граббса (определен по [28]);

– Шовене (определен по [28]).

Проверка гипотезы о нормальном распределении (для «микровыборок» размером более 25 элементов) выполнена:

– по коэффициенту вариации (определен по [13] и специализированной литературе);

– по критерию среднего абсолютного отклонения (определен по [13] и специализированной литературе);

– по размаху варьирования (определен по [13] и специализированной литературе);

– с помощью показателей асимметрии и эксцесса (определены по [13] и специализированной литературе);

– по критерию Колмогорова-Смирнова (определен по [28]);

– по критерию Смирнова-Крамера фон Мизеса ( $n\omega^2$ ) (определен по [28]).

Поиск выбросов и проверка гипотезы о нормальном распределении выполнены при уровне значимости  $\alpha=0,05$ .

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Выбор оптимального метода выявления выбросов в исходных данных выполнен путем сравнения характеристик итоговых вариативных многомерных регрессионных моделей, полученных после удаления из них выбросов, выявленных различными методами (таблица).

Сравнение качества регрессионных моделей выполнено по следующим показателям:

1. Скорректированный на число степеней свободы коэффициент детерминации  $R^2$  скор (определен по специализированной литературе).

2. Средняя и максимальная (по модулю) ошибка аппроксимации (определены по специализированной литературе).

3. Количество выявленных и удаленных выбросов (в % к объему выборки базовой модели).

<sup>10</sup>Грибовский С.В., Сивец С.А., Левыкина И.А. Математические методы оценки стоимости недвижимого имущества. М.: Маросейка, Книжная линия, 2014. 352 с.

Статистические характеристики регрессионных моделей  
 Statistical characteristics of the regression models

Характеристика	Значение			
	Базовая модель	Модель без выбросов, выявленных нижеуказанным методом		
		Метод определения стандартного отклонения (z-оценка)	Метод определения расстояния Махаланобиса	Метод приведения цен объектов к характеристикам эталонного объекта
Объем выборки (количество предложений по продаже) после удаления выявленных выбросов	2 253	2 192	2 187	1 732
Количество выявленных и удаленных выбросов (в % к объему выборки базовой модели)	–	2,7 %	2,9 %	23,1 %
Нормализованный коэффициент детерминации ( $R^2_{скор}$ )	0,897	0,893	0,901	0,946
Средняя ошибка аппроксимации, %	-0,27 %	-0,27 %	-0,24 %	-0,13 %
Максимальная (по модулю) ошибка аппроксимации, %	31,30 %	30,90 %	26,63 %	12,82 %

Указанный критерий носит индикативный характер, поскольку количество обнаруженных выбросов само по себе не свидетельствует о качестве метода их выявления.

Распределение остатков регрессионных моделей приведено на рис. 1–4.

На основании анализа полученных результатов наилучшими характеристиками обладает регрессионная модель, полученная после удаления выбросов, выявленных методом приведения цен объектов к характеристикам эталонного объекта.

Остатки модели

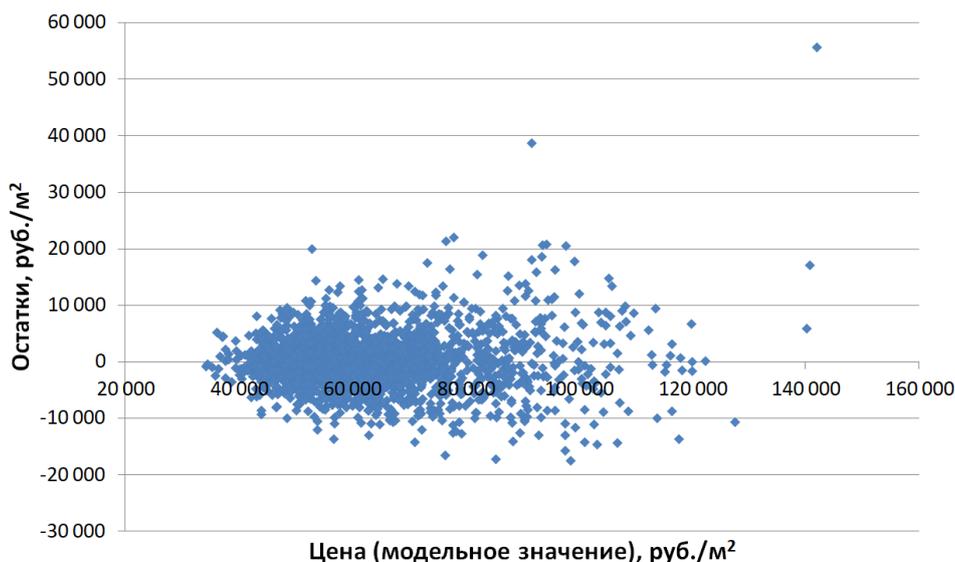
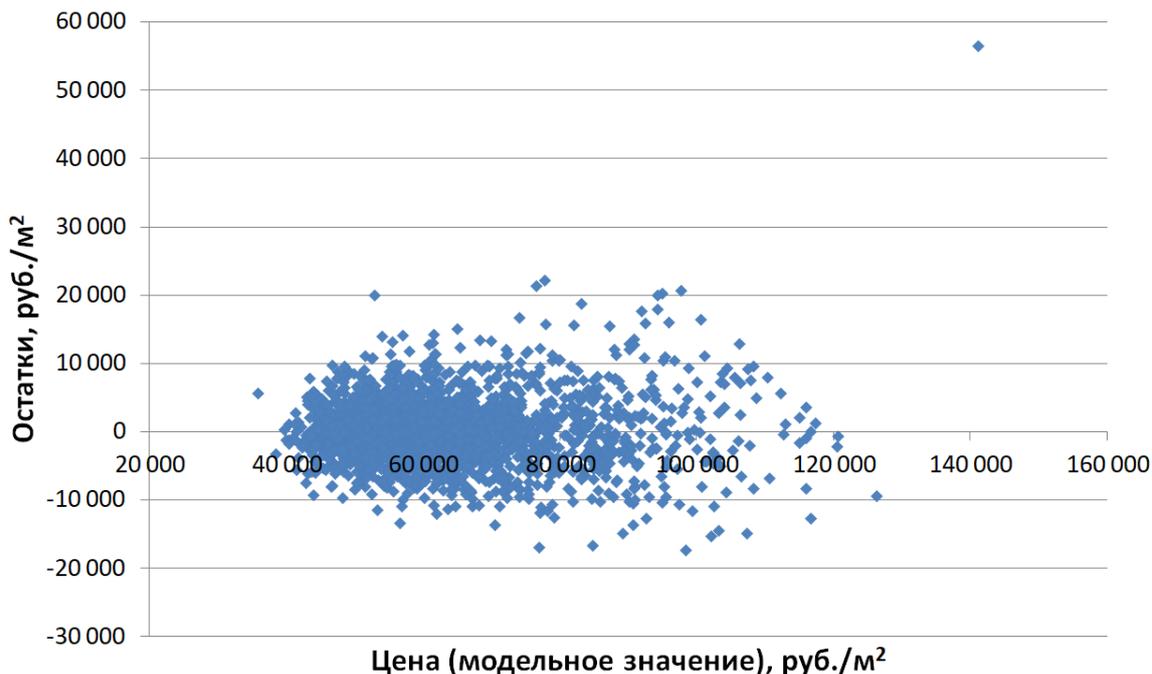


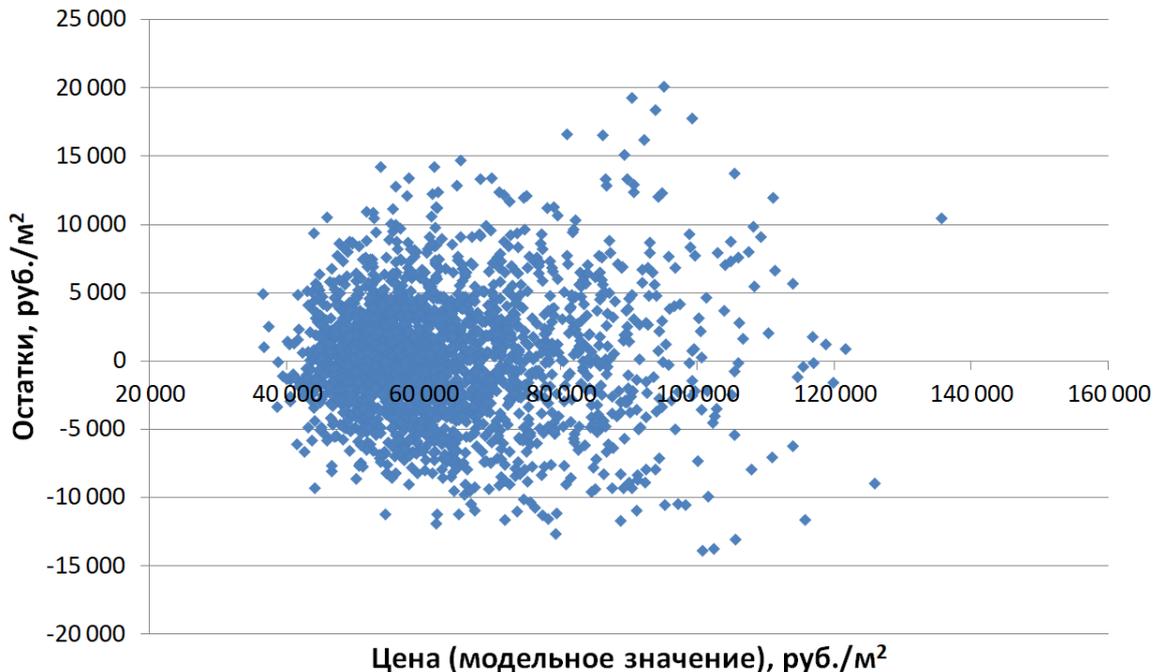
Рис. 1. Распределение остатков базовой регрессионной модели  
 Fig. 1. Distribution of residuals of the basic regression model

### Остатки модели



**Рис. 2. Распределение остатков регрессионной модели без выбросов, выявленных методом определения стандартного отклонения (z-оценка)**  
**Fig. 2. Distribution of regression model residuals without outliers identified by the standard deviation method (z-score)**

### Остатки модели

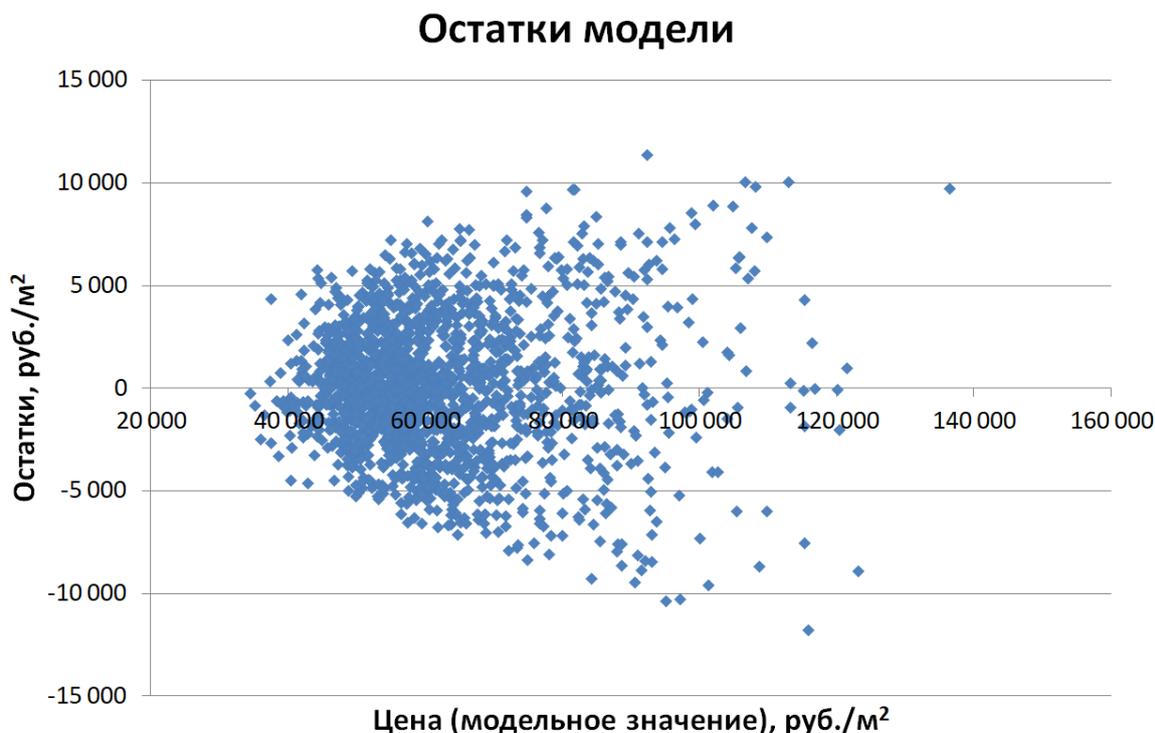


**Рис. 3. Распределение остатков регрессионной модели без выбросов, выявленных методом определения расстояния Махаланобиса**  
**Fig. 3. Distribution of regression model residuals without outliers identified by the Mahalanobis distance method**

Указанная модель обладает максимальным нормированным коэффициентом детерминации, минимальными средней и максимальной (по модулю) ошибками аппроксимации.

Распределение остатков модели (рис. 4) не имеет выраженной зависимости, что позволяет

рассматривать их как случайные. Вместе с тем, количество выявленных и удаленных выбросов в указанной модели максимальное из всех и составляет 23,1 % от объема выборки базовой модели, что несколько выше допустимого числа выбросов (20 % от общей численности объектов)<sup>11</sup>.



**Рис. 4. Распределение остатков регрессионной модели без выбросов, выявленных методом приведения цен объектов к характеристикам эталонного объекта**  
**Fig. 4. Distribution of regression model residuals without outliers identified by the method of bringing object prices to the characteristics of a reference object**

## ЗАКЛЮЧЕНИЕ

На основании вышеизложенного материала, можно сделать следующий вывод: оптимальным методом выявления выбросов при построении многомерной регрессионной модели цен на региональном рынке жилой недвижимости является разработанный в данном исследовании и описанный выше метод приведения цен объектов к характеристикам эталонного объекта. Не принимая во внимание на самую высокую трудоемкость среди рассмотренных методов.

В результате использования указанного метода существенно снижаются ошибки аппроксимации базовой многомерной регрессионной модели, что повышает адекватность результатов проводимой на ее основе оценки недвижимости.

Следует отметить, что при практическом построении регрессионных моделей рынка недвижимости, до удаления из выборки выявленных выбросов, следует в индивидуальном порядке проверять адекватность первичных данных по каждому объекту-выбросу. Необходимо учитывать наличие опечаток в тексте объявлений, отсутствие противоречий между текстом и данными фотофиксации объекта, правильность указания площади помещений (с балконом/без балкона) и ее соответствие данным Росреестра, и т. д.

Указанная мера позволяет существенно снизить количество выбросов и повысить качество регрессионной модели. Вместе с тем, указанная процедура ни в коем случае не должна «подгонять» исходные данные под достижение наилучших результатов.

<sup>11</sup>Чуракова И.Я. Приблизительная оценка. Направления использования методик выявления аномальных наблюдений при решении задач операционного менеджмента. Научный доклад № 13. СПб.: ВШМ СПбГУ, 2010. 10 с.

СПИСОК ИСТОЧНИКОВ

1. Алексеев А.О., Харитонов В.А., Ясницкий В.Л. Разработка концепции комплексного нейросетевого моделирования процессов массовой оценки и сценарного прогнозирования рыночной стоимости жилой недвижимости // Известия вузов. Инвестиции. Строительство. Недвижимость. 2018. Т. 8. № 1. С. 11–22. EDN: YUVVIC.
2. Баринов Н.П. Применение регрессионного анализа в задачах индивидуальной и массовой оценки объектов недвижимости // Вопросы оценки. 2022. № 1 (106). С. 34–46. EDN: PCGUXU.
3. Баринов Н.П. Сравнительный подход к оценке недвижимости. Современный взгляд // Вопросы оценки. 2019. № 1 (95). С. 2–16. EDN: SAEPVN.
4. Лейфер Л.А., Акобян А.А. О точности методов оценки недвижимости, основанных на машинном обучении // Вопросы оценки. 2021. № 1 (103). С. 26–32. EDN: RFGGBM.
5. Каммероу М. Статистическое определение стоимости // Вопросы оценки. 2019. № 3 (97). С. 2–17. EDN: MSPQKH.
6. Махт В.А., Карпова О.А., Артамонова С.Ф. Моделирование кадастровой стоимости земель садоводческих объединений Республики Крым // Вестник Сибирского государственного университета геосистем и технологий. 2020. Т. 25. № 2. С. 209–221. <https://doi.org/10.33764/2411-1759-2020-25-2-209-221>. EDN: KDMUNB.
7. Пузыня Н.Ю., Мягков В.Н., Грибовский С.В., Цымбалов В.В., Петухова Р.А., Шведова Н.Ю. [и др.] Современные тенденции рынка и оценка рыночной стоимости. СПб: Санкт-Петербургский государственный экономический университет, 2020. 250 с. EDN: MPDVZF.
8. Пасимовски Ю. Как дискредитировать большинство оценок недвижимости за одну минуту // Вопросы оценки. 2021. № 4 (105). С. 39–45. EDN: EFYJDD.
9. Петров Ю.Ю. Особенности оценки земель недропользования // Вопросы оценки. 2019. № 2 (96). С. 47–50. EDN: JDXVDC.
10. Пузыня Н.Ю., Мягков В.Н., Нейман Е.И., Слуцкий А.А., Козин П.А., Кузнецов Д.Д. [и др.] Проблемы, современные реалии и новые вызовы в оценочной деятельности: монография. СПб: Санкт-Петербургский государственный экономический университет, 2022. 148 с. EDN: EKLVMX.
11. Слуцкий А.А. «На пороге» регрессионного анализа в оценке // Вопросы оценки. 2022. № 1 (106). С. 22–33. EDN: MPFWGU.
12. Щеглов Е.В. Прогнозирование рыночной стоимости коммерческой недвижимости на основе показателей экономического развития территории // Российское предпринимательство. 2016. Т. 17. № 19. С. 2539–2554. <https://doi.org/10.18334/rp.17.19.36621>. EDN: WWIDHR.
13. Анисимова И.Н., Баринов Н.П., Грибовский С.В. О повышении достоверности оценки рыночной стоимости методом сравнительного анализа // Вопросы оценки. 2002. № 1. С. 2–10. EDN: OOKJLV.
14. Казимиров И.А., Ощерин Л.А., Сахилтарова С.В. Разработка и исследование многомерных регрессионных моделей вторичного рынка жилой недвижимости // Известия вузов. Инвестиции. Строительство. Недвижимость. 2017. Т. 7. № 4. С. 87–107. EDN: YLNHAL.
15. Казимиров И.А., Пешков В.В. Определение динамики цен на вторичном рынке жилой недвижимости с использованием многомерной регрессионной модели // Известия вузов. Инвестиции. Строительство. Недвижимость. 2019. Т. 9. № 3. С. 476–487. <https://doi.org/10.21285/2227-2917-2019-3-476-487>. EDN: WYQIMZ.
16. Куликов Д.Л. Оценка стоимости объектов недвижимости при помощи алгоритмов машинного обучения: изучение рынка недвижимости города Дубны // Системный анализ в науке и образовании. 2019. № 3. С. 32–39. EDN: THLPSB.
17. Баринов Н.П. Оценка рыночной стоимости земельного участка методом многомерного регрессионного анализа // Бюллетень RWAY. 2014. № 232. С. 24–32. Режим доступа: [https://www.avg.ru/wp-content/uploads/2023/01/2014\\_RWAY\\_232\\_OcenkaZU\\_Regress\\_24-32.pdf](https://www.avg.ru/wp-content/uploads/2023/01/2014_RWAY_232_OcenkaZU_Regress_24-32.pdf) (дата обращения: 19.01.2024).
18. Cabana E., Lillo R.E., Laniado H. Multivariate Outlier Detection Based on a Robust Mahalanobis Distance with Shrinkage Estimators // Statistical Papers. 2021. Vol. 62. P. 1583–1609. <https://doi.org/10.1007/s00362-019-01148-1>.
19. Leys Ch., Klein O., Dominicy Y., Ley Ch. Detecting Multivariate Outliers: Use A Robust Variant of the Mahalanobis Distance // Journal of Experimental Social Psychology. 2018. Vol. 74. P. 150–156. <https://doi.org/10.1016/j.jesp.2017.09.011>.
20. Leys Ch., Delacre M., Mora Y.L., Lakens D., Ley Ch. How to Classify, Detect, and Manage Univariate and Multivariate Outliers, with Emphasis on Pre-Registration // International Review of Social Psychology. 2019. Vol. 32. Iss. 1. P. 1–5. <https://doi.org/10.5334/irsp.289>.
21. Magyar B., Kenyeres A., Toth S., Hajdu I., Horvath R. Spatial Outlier Detection on Discrete GNSS Velocity Fields Using Robust Mahalanobis-Distance-Based Unsupervised Classification // GPS Solutions. 2022. Vol. 26. P. 1–11. <https://doi.org/10.1007/s10291-022-01323-2>.

22. Mayrhofer M., Filzmoser P. Multivariate Outlier Explanations Using Shapley Values and Mahalanobis Distances // *Econometrics and Statistics*. 2023. P. 1–24.
23. Afzal S., Afzal A., Amin M., Saleem S., Ali N., Sajid M. A Novel Approach for Outlier Detection in Multivariate Data // *Mathematical Problems in Engineering*. 2021. Vol. 2021. P. 1–12. <https://doi.org/10.1155/2021/1899225>.
24. Santos-Pereira C.M., Pires A.M. Detection of Outliers in Multivariate Data: A Method Based on Clustering and Robust Estimators. Heidelberg: Physica, 2002. [https://doi.org/10.1007/978-3-642-57489-4\\_41](https://doi.org/10.1007/978-3-642-57489-4_41).
25. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff a New Robust Estimator to Detect Outliers for Multivariate Data // *Journal of Physics: Conference Series*. 2019. P. 1–10. <https://doi.org/10.1088/1742-6596/1366/1/012104>.
26. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff Comparison of Robust Estimators' Performance for Detecting Outliers in Multivariate Data // *Journal of Statistical Modeling & Analytics*. 2021. Vol. 3. Iss. 2. P. 36–64. <http://doi.org/10.22452/josma.vol3no2.3>.
27. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff a Review on Outliers-Detection Methods for Multivariate Data // *Journal of Statistical Modeling & Analytics*. 2021. Vol. 3. Iss. 1. P. 1–15. <https://doi.org/10.22452/josma.vol3no1.1>.
28. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.

## REFERENCES

1. Alekseev A.O., Kharitonov V.A., Yasnitsky V.L. Development of The Concept of Complex Neural Simulation of Processes of Mass Estimation and Scenary Forecasting of Market Cost of Housing Real Estate. *Proceedings of Universities. Investment. Construction. Real estate*. 2018;8(1):11-22. (In Russ.). EDN: YUVVIC.
2. Barinov N.P. Application of Regression Analysis in Tasks of Individual and Visual Assessment of Real Estate Objects. *The Appraisal Issues*. 2022;1(106):34-46. (In Russ.). EDN: PCGUXU.
3. Barinov N.P. Comparative Approach to Real Estate Valuation. Modern View. *The Appraisal Issues*. 2019;1(95):2-16. (In Russ.). EDN: SAEPVN.
4. Leifer L.A., Akobyan A.A. On The Accuracy of Real Estate Valuation Methods Based on Machine Learning. *The Appraisal Issues*. 2021;1(103):26-32. (In Russ.). EDN: RFGGBM.
5. Kammerou M. Statistical Determination of Cost. *The Appraisal Issues*. 2019;3(97):2-17. (In Russ.). EDN: MSPQKH.
6. Makht V.A., Karpova O.A., Artamonova S.F. Modeling Cadastral Value of Associations of Private Gardeners' Lands of the Republic of Crimea. *Vestnik of the Siberian State University of Geosystems and Technologies*. 2020;25(2):209-221. (In Russ.). <https://doi.org/10.33764/2411-1759-2020-25-2-209-221>. EDN: KDMUNB.
7. Puzynya N.Yu., Myagkov V.N., Gribovskii S.V., Tsymbalov V.V., Petukhova R.A., Shvedova N.Yu. [et al.] *Current Market Trends and Market Value Assessment*. St. Petersburg: St. Petersburg State University of Economics, 2020. 250 p. (In Russ.). EDN: MPDVZF.
8. Pasimovski Yu. How to Discredit Most Real Estate Appraisals in One Minute. *The Appraisal Issues*. 2021;4(105):39-45. (In Russ.). EDN: EFYJDD.
9. Petrov Yu.Yu. Features of The Assessment of Subsoil Use Lands. *The Appraisal Issues*. 2019;2(96):47-50. (In Russ.). EDN: JDXVDC.
10. Puzynya N.Yu., Myagkov V.N., Neiman E.I., Slutskii A.A., Kozin P.A., Kuznetsov D.D. [et al.] *Problems, Modern Realities and New Challenges in Valuation Activities: Monograph*. St. Petersburg: St. Petersburg State Economic University, 2022. 148 p. (In Russ.). EDN: EKLVMX.
11. Slutskii A.A. "On The Threshold" of Regression Analysis in Assessment. *The Appraisal Issues*. 2022;1(106):22-33. (In Russ.). EDN: MPFWGU.
12. Tscheglov E.V. Forecasting of The Market Value of the Commercial Real Estate Based On Indices of the Territorial Economic Development. *Russian Journal of Entrepreneurship*. 2016;17(19):2539-2554. (In Russ.). <https://doi.org/10.18334/rp.17.19.36621>. EDN: WWIDHR.
13. Anisimova I.N., Barinov N.P., Gribovskii S.V. On Increasing the Reliability of Market Value Assessment Using the Method of Comparative Analysis. *The Appraisal Issues*. 2002;1:2-10. (In Russ.). EDN: OOKJLV.
14. Kazimirov I.A., Oscherin L.A., Sakhiltarova S.V. Development and Research of Multisized Regressive Models of Second Hand Market of Residential Real Estate. *Proceedings of Universities. Investment. Construction. Real estate*. 2017;7(4):87-107. (In Russ.). EDN: YLNXAL.
15. Kazimirov I.A., Peshkov V.V. Determination of Price Behaviour in The Secondary Residential Real Estate Market Using a Multidimensional Regression Model. *Proceedings of Universities. Investment. Construction. Real estate*. 2019;9(3):476-487. (In Russ.). <https://doi.org/10.21285/2227-2917-2019-3-476-487>. EDN: WYQIMZ.

16. Kulikov D.L. Real Estate Valuation with The Help of Machine Learning Algorithms: A Study of the Real Estate Market of the City of Dubna. *System Analysis in Science and Education*. 2019;3:32-39. (In Russ.). EDN: THLPSB.
17. Barinov N.P. Assessment of The Market Value of a Land Plot Using Multivariate Regression Analysis. *Byulleten' RWAY*. 2014;232:24-32. Available from: [https://www.avg.ru/wp-content/uploads/2023/01/2014\\_RWAY\\_232\\_OcenkaZU\\_Regress\\_24-32.pdf](https://www.avg.ru/wp-content/uploads/2023/01/2014_RWAY_232_OcenkaZU_Regress_24-32.pdf) [Accessed 19 January 2024]. (In Russ.).
18. Cabana E., Lillo R.E., Laniado H. Multivariate Outlier Detection Based on a Robust Mahalanobis Distance with Shrinkage Estimators. *Statistical Papers*. 2021;62:1583-1609. <https://doi.org/10.1007/s00362-019-01148-1>.
19. Leys Ch., Klein O., Dominicy Y., Ley Ch. Detecting Multivariate Outliers: Use A Robust Variant of the Mahalanobis Distance. *Journal of Experimental Social Psychology*. 2018; 74:150-156. <https://doi.org/10.1016/j.jesp.2017.09.011>.
20. Leys Ch., Delacre M., Mora Y.L., Lakens D., Ley Ch. How to Classify, Detect, and Manage Univariate and Multivariate Outliers, with Emphasis on Pre-Registration. *International Review of Social Psychology*. 2019;32(1):1-5. <https://doi.org/10.5334/irsp.289>.
21. Magyar B., Kenyeres A., Toth S., Hajdu I., Horvath R. Spatial Outlier Detection on Discrete GNSS Velocity Fields Using Robust Mahalanobis-Distance-Based Unsupervised Classification. *GPS Solutions*. 2022;26:1-11. <https://doi.org/10.1007/s10291-022-01323-2>.
22. Mayrhofer M., Filzmoser P. Multivariate Outlier Explanations Using Shapley Values and Mahalanobis Distances. *Econometrics and Statistics*. 2023:1-24.
23. Afzal S., Afzal A., Amin M., Saleem S., Ali N., Sajid M. A Novel Approach for Outlier Detection in Multivariate Data. *Mathematical Problems in Engineering*. 2021;2021:1-12. <https://doi.org/10.1155/2021/1899225>.
24. Santos-Pereira C.M., Pires A.M. *Detection of Outliers in Multivariate Data: A Method Based on Clustering and Robust Estimators*. Heidelberg: Physica, 2002. [https://doi.org/10.1007/978-3-642-57489-4\\_41](https://doi.org/10.1007/978-3-642-57489-4_41).
25. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff a New Robust Estimator to Detect Outliers for Multivariate Data. *Journal of Physics: Conference Series*. 2019:1-10. <https://doi.org/10.1088/1742-6596/1366/1/012104>.
26. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff Comparison of Robust Estimators' Performance for Detecting Outliers in Multivariate Data. *Journal of Statistical Modeling & Analytics*. 2021;3(2):36-64. <http://doi.org/10.22452/josma.vol3no2.3>.
27. Sharifah Sakinah Syed Abd Mutalib, Siti Zanariah Satari, Wan Nur Syahidah Wan Yusoff A Review on Outliers-Detection Methods for Multivariate Data. *Journal of Statistical Modeling & Analytics*. 2021;3(1):1-15. <https://doi.org/10.22452/josma.vol3no1.1>.
28. Kobzar' A.I. *Applied Mathematical Statistics. For Engineers and Scientists*. Moscow: FIZMATLIT, 2006. 816 p. (In Russ.).

#### Информация об авторах

**Казимиров Илья Александрович**,  
старший преподаватель кафедры экспертизы  
и управление недвижимостью,  
Иркутский национальный исследовательский  
технический университет,  
664074, г. Иркутск, ул. Лермонтова, 83, Россия,  
✉e-mail: [ilya@appfordepo.nichost.ru](mailto:ilya@appfordepo.nichost.ru)  
<https://orcid.org/0000-0002-3708-256X>  
Author ID: 953838

**Матвеева Мария Витальевна**,  
д.э.н., профессор,  
профессор кафедры экспертиза и управление  
недвижимостью,  
Иркутский национальный исследовательский  
технический университет,  
664074, г. Иркутск, ул. Лермонтова, 83, Россия,  
e-mail: [expertiza@istu.edu](mailto:expertiza@istu.edu)  
<https://orcid.org/0000-0002-9390-5444>  
Author ID: 505831

#### Information about the authors

**Ilya A. Kazimirov**,  
Senior Lecturer of the Department of Real Estate  
Expertise and Management,  
Irkutsk National Research Technical University,  
83 Lermontov St., Irkutsk 664074,  
Russia,  
✉e-mail: [ilya@appfordepo.nichost.ru](mailto:ilya@appfordepo.nichost.ru)  
<https://orcid.org/0000-0002-3708-256X>  
Author ID: 953838

**Maria V. Matveeva**,  
Dr. Sci. (Econ.), Professor,  
Professor of the Department of Real Estate  
Expertise and Management,  
Irkutsk National Research Technical University,  
83 Lermontov St., Irkutsk 664074,  
Russia,  
e-mail: [expertiza@istu.edu](mailto:expertiza@istu.edu)  
<https://orcid.org/0000-0002-9390-5444>  
Author ID: 505831

**Полякова Нина Владимировна,**  
д.э.н, профессор, профессор кафедры  
журналистики и маркетинговых технологий,  
Байкальский государственный университет,  
664003, г. Иркутск, ул. Ленина, 11,  
Россия,  
e-mail: Polyakova.ninotchka@yandex.ru  
<https://orcid.org/0009-0001-5785-3616>  
Author ID: 412361

**Nina V. Polyakova,**  
Dr. Sci. (Econ.), Professor, Professor  
of the Department of Journalism  
and Marketing Technologies,  
Baikal State University,  
11 Lenin St., Irkutsk, 664003, Russia,  
e-mail: Polyakova.ninotchka@yandex.ru  
<https://orcid.org/0009-0001-5785-3616>  
Author ID: 412361

#### **Вклад авторов**

Все авторы сделали эквивалентный вклад в подготовку публикации.

#### **Contribution of the authors**

The authors contributed equally to this article.

#### **Конфликт интересов**

Авторы заявляют об отсутствии конфликта интересов.

#### **Conflict of interests**

The authors declare no conflict of interests regarding the publication of this article.

Все авторы прочитали и одобрили окончательный вариант рукописи.

The final manuscript has been read and approved by all the co-authors.

#### **Информация о статье**

Статья поступила в редакцию 28.02.2024.  
Одобрена после рецензирования 26.03.2024.  
Принята к публикации 01.04.2024.

#### **Information about the article**

The article was submitted 28.02.2024.  
Approved after reviewing 26.03.2024.  
Accepted for publication 01.04.2024.