

В диссертационный совет Д 01.024.04
при ГОУВПО «ДОНЕЦКИЙ
НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»
и ГОУВПО «ДОНЕЦКИЙ
НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»
ученому секретарю
Завадской Татьяне Владимировне
283001, г. Донецк, ул. Артема, 58

ОТЗЫВ

Официального оппонента на диссертацию Рычки Ольги Валентиновны на тему: «Совершенствование методов выявления и корректировки аномальных измерений для повышения качества линейных регрессионных моделей», на соискание ученой степени кандидата технических наук по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ (технические науки)

1 Актуальность темы диссертационной работы

Одним из основных этапов анализа системы является построение соответствующей математической модели. От точности построенной модели зависит эффективность принимаемых решений.

Одним из инструментов анализа данных и обнаружения в них закономерностей выступает регрессионный анализ. Частным, но важным примером применения регрессионного анализа являются парные линейные модели, которые имеют широкую область практического применения в различных областях науки и техники. Построенное таким образом линейное уравнение может быть первым шагом для построения более сложных моделей.

Одним из преимуществ парных линейных регрессионных моделей является возможность приведения большинства нелинейных моделей к кусочно-линейному виду.

Для повышения качества принимаемых решений необходима предварительная обработка данных, включающая в себя выявление аномальных измерений (выбросов) и их последующую обработку.

Вместе с тем, в настоящее время отсутствуют методы поиска и корректировки аномальных измерений, способные полноценно решить данную задачу в случае использования линейных регрессионных моделей.

Диссертационное исследование Рычки О.В. затрагивает одну из главных проблем анализа пар данных, имеющих взаимозависимость, а именно, отсутствие возможности одновременного обнаружения и корректировки аномальных измерений по двум переменным (зависимой и независимой) для построения адекватной регрессионной модели.

Несмотря на многообразие существующих подходов к обнаружению выбросов в исходных статистических данных, большинство из них обладают рядом недостатков:

- методы применимы только для одномерных выборок (чаще всего временных рядов);
- на каждой итерации метода осуществляется анализ лишь одного подозрительного значения;
- описанные в литературе методы являются чувствительными к объёму исходной выборки;
- методы, основанные на поиске отклонения от среднего или предыдущего (последующего) значения не дают положительных результатов при наличии зависимости между несколькими переменными.

Следовательно, совершенствование методов выявления и корректировки аномальных измерений с целью повышения качества парных регрессионных моделей для дальнейшего их использования при принятии решений является актуальной научно-прикладной задачей.

2 Обоснованность научных положений, выводов и рекомендаций, сформулированных в диссертации

Обоснованность научных положений, выводов и рекомендаций обеспечивается достаточным количеством проведенных экспериментальных исследований с использованием реальных и модельных данных. Научные положения, выводы и рекомендации, сформулированные в работе, подкреплены реальными данными, наглядно представленными в приведенных таблицах и рисунках. Подготовка, анализ исходных данных и интерпретация итоговых результатов базируются на современных методах обработки информации и статистического анализа.

3 Достоверность и новизна научных положений, выводов и рекомендаций

Достоверность полученных в работе результатов обеспечивается корректным использованием математического аппарата для предложенных методов и проверкой результатов моделирования на реальных данных.

Научная новизна полученных результатов заключается в следующем:

1. Разработан новый метод поиска аномалий, основанный на построении области надёжности, которая зависит от угла наклона функции регрессии, доверительной вероятности и соответствующего коэффициента, что позволяет одновременно обнаруживать аномальные измерения как по независимой переменной (X), так и по зависимой переменной (Y). Это приводит к повышению качества прогнозов (от 10%), полученных по линейным регрессионным моделям, возрастанию коэффициента детерминации (от 10% до 30%) и уменьшению трудоёмкости (количество элементарных операций) – до $2 \cdot 10^n$ раз, по сравнению с существующими методами.

2. Предложены метод и алгоритм корректировки аномалий, отличием которых от известных является изменение значений аномальных статистических данных на значения, соответствующие рассчитанной области надёжности, а также отсутствием сокращения объема исходных статистических данных из-за отбрасывания, что особенно важно при моделировании на выборках малого объёма.

3. Предложены два упрощения алгоритма поиска аномальных данных, позволяющие сократить трудоёмкость анализа, выбор которых зависит от надёжности исходных значений X и Y, что приводит к обнаружению аномальных данных по одной из соответствующих переменных. Спецификой первого упрощения является то, что оно может быть использовано для поиска аномальных данных при многомерной линейной зависимости.

4. Обоснована возможность применения предлагаемого в работе метода обнаружения выбросов не только для линейных регрессионных прогнозных уравнений, но и для нелинейных моделей с внутренней линейностью.

Результаты работы могут быть использованы как в дальнейших теоретических исследованиях, так и в их реализациях. Практическое значение полученных результатов состоит в возможности использования предложенного алгоритма поиска аномальных данных и методов их последующей обработки, а также разработанного комплекса программ на практике в различных предметных областях, таких как здравоохранение, экономика и других, при решении задач прогнозирования, проектирования, оптимизации и т.д.

Практическое значение диссертационной работы подтверждено внедрением результатов исследования в ООО НПО «Интермет» (справка о внедрении от 23 июня 2021 г.), в научно-исследовательские госбюджетные работы ГОУВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» (справка о внедрении № 29-13/15 от 05 июля 2021 г.), в учебный процесс ГОУВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» при чтении лекций и проведении лабораторных работ на кафедре «Программная инженерия» им. Л.П. Фельдмана по дисциплинам: «Эмпирические методы программной

инженерии», «Численные методы в информатике» (справка о внедрении № 29-12/15 от 05 июля 2021 г.).

Основные научные и практические результаты диссертации достаточно полно опубликованы в 17 научных работах, в том числе: 2 работы в изданиях, входящих в перечень специализированных научных изданий, утвержденный ВАК ДНР; 4 работы в изданиях, входящих в перечень специализированных научных изданий, утвержденных ВАК Украины; 2 – в других научных изданиях (в том числе 1 монография); 9 – по материалам конференций.

4 Анализ содержания диссертации

Диссертация является законченным научным трудом, изложена на 162 страницах машинописного текста и состоит из введения, четырех разделов, заключения, списка литературы и 4 приложений, включая 44 рисунка, 36 таблиц, список литературы из 98 наименований.

Во введении обоснована актуальность, сформулированы цели и задачи исследования, изложены научные положения, выносимые на защиту, определены практическая и теоретическая значимость результатов работы, представлены сведения об их апробации.

В первом разделе проведён анализ существующих методов обнаружения и последующей корректировки аномальных измерений в исходных статистических данных, выявлены имеющиеся недостатки данных методов и определены перспективы их устранения.

На основе проведённого анализа автором обоснована необходимость в разработке нового алгоритма поиска аномальных наблюдений в исходных данных, которые используются для построения регрессионной модели и методов их последующей корректировки, а также комплекса программ для автоматизированного анализа данных.

Во втором разделе приводится описание сущности и основных параметров предлагаемого автором подхода для обнаружения аномальных данных и методов их последующей корректировки, а также модификаций разработанных методов.

Выбраны рациональные критерии оценки эффективности предложенных методов. Обосновывается возможность применения методов для анализа данных, имеющих нелинейную зависимость с внутренней линейностью, а также использование одной из модификаций для данных с многомерной линейной зависимостью. Доказана эффективность предложенных методов.

В третьем разделе описан разработанный автором комплекс программ для автоматизации поиска аномальных данных и их последующей модификации, состоящий из главного интерфейса, написанного на языке C# в среде Microsoft Visual Studio 2017 и макросов, разработанных с

применением языка программирования Visual Basic for Applications для Microsoft Excel.

Отличительными особенностями данного комплекса программ является простота применения, высокая точность определения аномальных данных, возможность графического отображения области надёжности, а также умеренные потребности в вычислительных ресурсах.

В четвертом разделе проведён сравнительный анализ предложенных автором методов между собой и с наиболее известными методами, доказана эффективность разработанных методов (нахождение всех аномальных данных, повышение значения коэффициента детерминации более, чем на 30%) на различных наборах экспериментальных данных, а также для данных, имеющих нелинейную зависимость с внутренней линейностью, даны рекомендации по выбору одного из методов или их модификаций.

Все эксперименты проводились с применением разработанного автором комплекса программ.

5 Основные результаты диссертации

Основными результатами диссертации являются:

1. Предложен и обоснован новый алгоритм для поиска выбросов в линейной регрессионной модели, а на его основе построены два метода обработки исходных данных.

2. Продемонстрирована применимость предложенных методов к реальным задачам анализа данных.

3. Проведено сравнение предложенных методов с наиболее известными, которое демонстрирует эффективность и предпочтительность разработанных в диссертации методов.

4. Сформулированы рекомендации к практическому применению предложенных методов в задачах анализа данных.

5. Показано, что предложенный подход позволяет обнаружить выбросы и скорректировать вид модели без дополнительного графического отображения (на примере трёх наборов данных из «квартета Энскомба»).

6 Замечания и вопросы по диссертации

1. Иногда автор использует разные обозначения для одних и тех же понятий в различных частях текста диссертации (например: аномалии, выбросы, ненадёжные измерения). Это усложняет чтение работы.

2. Во введении задача построения «...качественной модели ...» сформулирована как многокритериальная задача. Решена ли эта задача в данной работе в такой постановке (стр.5, второй абзац; стр.10, абзац 1)?

3. В диссертации присутствует ряд опечаток и пунктуационных неточностей (стр.5, 4-й абзац; стр.67, абзац 4 и др).

4. В первом пункте научной новизны сформулировано, что область надёжности «...зависит от наклона уравнения регрессии...». Точнее здесь

можно сказать о зависимости от угла наклона графика функции регрессии (стр.8, абзац 5 и стр.21 формула (1.2)).

5. Во введении сказано «...В работе определены *оптимальные* параметры использования предложенных методов...». Как доказана их оптимальность и какому критерию были получены эти параметры (стр.9, абзац 7; стр.10 абзац 1)?

6. Не указаны возможные значения индексов переменных T_n и x_{n-1} (стр.24-25, после формулы (1.6)).

7. В пояснении к формуле (2.9) на стр.49 указано, что $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ является математическим ожиданием. Точнее сказать, что это оценка математического ожидания. Замечание относится ко всем формулам расчета СКО s и дисперсии s^2 .

8. В описании сущности метода повышения качества модели, основанного на переносе ненадёжных данных, не приводится обоснования их переноса на границы области надёжности (стр.62, п.(2.2.1)).

9. В работе не приводятся полные алгоритмы предложенных методов, а их приведенные схемы представляют лишь одну итерацию метода. Т.о., не представляется возможным установить: проверяется ли в программном комплексе условие нормальности распределения невязок, каким образом реализуется принятие решение об окончании построения модели, как реализуется многокритериальная оптимизация, и т.д.

10. Отсутствует руководство пользователя для разработанного комплекса программ.

Заключение

Отмеченные в отзыве замечания не исключают общей положительной оценки диссертации. Диссертационная работа Рычки Ольги Валентиновны «Совершенствование методов выявления и корректировки аномальных измерений для повышения качества линейных регрессионных моделей» является законченной научно-квалификационной работой, в которой дано теоретическое обоснование и приведено решение важной научно-практической задачи совершенствования методов поиска и обработки аномальных данных с целью повышения качества парных линейных регрессионных моделей.


В целом, работа выполнена на высоком научном уровне, содержит новые теоретические и практические результаты. Полученные результаты, положения и выводы соответствует паспорту специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ (технические науки), в частности: п.3 «Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий»; п.5 «Комплексные исследования научных и технических проблем с применением современных технологий»

математического моделирования и вычислительного эксперимента»; п.6 «Разработка новых математических методов и алгоритмов проверки адекватности математических моделей объектов на основе данных натурального эксперимента».

Диссертационная работа удовлетворяет требованиям п. 2.2. «Положения о присуждении ученых степеней» Донецкой Народной Республики, а ее автор, Рычка Ольга Валентиновна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ (технические науки).

Официальный оппонент:


Доктор технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации» и специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ», доцент, профессор кафедры «Программное обеспечение вычислительной техники и автоматизированных систем», профессор кафедры «Вычислительные системы и информационная безопасность» ФГБОУ ВО «Донской государственный технический университет»


(подпись)

В.Г. Кобак

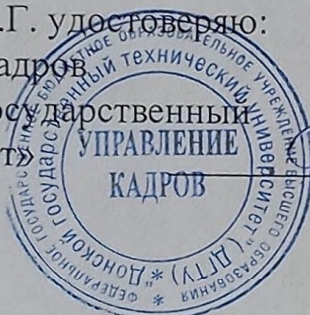
Адрес: 344000, Российская Федерация,
ЮФО, Ростовская область, г. Ростов-на-Дону,
пл. Гагарина, 1,
тел.: 8-800-100-1930
факс: (863)232-7953
e-mail: reception@donstu.ru

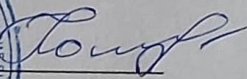
Я, Кобак Валерий Григорьевич, даю согласие на автоматизированную обработку моих персональных данных.



В.Г. Кобак

Подпись д.т.н. Кобака В.Г. удостоверяю:
начальник управления кадров
ФГБОУ ВО «Донской государственный
технический университет»





О.И. Костина