

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу **Шаталина Евгения Викторовича**
«ЭМПИРИЧЕСКИЙ МОСТ И ЗАДАЧИ ТЕСТИРОВАНИЯ АДЕКВАТНО-
СТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ АНАЛИЗА ДАННЫХ»,

представленную на соискание ученой степени кандидата

физико-математических наук по специальности

05.13.17 — Теоретические основы информатики

Актуальность избранной темы.

В своей работе Е.В. Шаталин решает задачу анализа данных, а именно задачу построения результативных статистических критериев для анализа адекватности регрессионной модели рассматриваемой выборке данных. В качестве регрессора рассматриваются порядковые статистики, что само по себе является весьма необычной, хотя и довольно естественной, постановкой задачи. В самом деле, в целом ряде приложений видится логичным упорядочение данных, соответственно изучение упорядоченных данных и построение регрессионных моделей по ним является актуальным методом исследования. Как правило, эмпирические данные сами по себе не несут в себе однозначных указаний о предполагаемых моделях их описания. Даже графическое изображение исходной выборки порой может лишь намекнуть и выявить только определенный класс моделей, претендующих на способ описания указанных данных. Соответственно исследователь вынужден рассматривать несколько (иногда весьма большое количество) регрессионных моделей одновременно. Таким образом, на первый план выходит вопрос быстрого и достаточно точного алгоритма «выбраковки» неподходящих моделей. Данный вопрос становится еще более актуальным, если сказать о том, что массивы данных в современном мире уже измеряются тысячами и миллионами значений, что порождает большие вычисления и ставит перед исследователями вопросы эффективного использования имеющихся вычислительных мощностей. Таким образом, **тема диссертационной работы является актуальной и своевременной.**

Структура и основные результаты диссертации.

Диссертационная работа состоит из введения, двух глав, заключения, списка литературы и приложения, объем составляет 102 страницы. Список литературы состоит из 86 наименований, в том числе 9 работ автора по теме диссертации.

Первая глава диссертации имеет теоретический характер, в ней вводятся все необходимые понятия и пошагово приводится общая схема разрабатываемых в работе решающих правил. В основе построений лежит кусочно-линейная случайная ломаная – эмпирический мост, который, в свою очередь, строится по набору частичных сумм регрессионных остатков.

Суть предлагаемого алгоритма состоит в изучении эмпирического моста, построенного по регрессионной модели. Более конкретно, предлагается проводить сравнение эмпирического и теоретического моста. Для этого необходимо отыскать предельный процесс для эмпирического моста для соответствующих видов регрессионных моделей. Что и сделано автором в виде доказательства трех предельных теорем для однопараметрической и двух видов двухпараметрических регрессионных моделей. Отмечу, что сами по себе доказанные предельные теоремы, помимо того, что они обосновывают разработанный автором алгоритм, имеют высокую теоретическую ценность в области теории вероятностей и математической статистики.

Вторая глава диссертации посвящена вопросам практического применения построенных критериев. Автором дано полное методическое руководство по применению метода эмпирического моста для анализа адекватности регрессионных моделей. Кроме того, на частных примерах показаны возможности метода, а также некоторые приемы (дробление и фильтрация выборки), позволяющие провести более точный и аккуратный анализ.

Основными результатами диссертации являются:

1. Построен и теоретически обоснован алгоритм (а на его основе два статистических критерия) анализа адекватности одно- и двухпараметрических линейных регрессионных моделей на порядковые статистики.
2. Проведено сравнение и построен пример, демонстрирующий предпочтительность предложенных в диссертации критериев над известным F-критерием проверки значимости регрессионной модели.
3. Продемонстрирована применимость предложенных критериев к реальным задачам анализа данных.
4. Сформулированы методические рекомендации практическому применению предложенного алгоритма в задачах анализа данных.

Степень обоснованности, достоверность и новизна научных положений, сформулированных в диссертации.

Все полученные в диссертации результаты являются новыми, достоверными, строго и аккуратно математически обоснованными. Кроме новизны самого предлагаемого алгоритма, отметим также оригинальность и новизну изучения в качестве регрессора порядковых статистик, а также отказ от одного из основных постулатов регрессионного анализа – гомоскедастичности. Вышеуказанное подтверждается также наличием

необходимого количества научных публикаций в рецензируемых изданиях, а также апробацией на ряде российских и международных конференциях и выступлениями на научных семинарах.

Значимость для науки и практики полученных автором результатов, рекомендации по использованию.

Полученные в диссертации результаты являются значимыми и вносят весомый вклад в такие отрасли знаний как анализ данных, регрессионный анализ, теория вероятностей и математическая статистика. Результаты могут быть использованы как в дальнейших теоретических изысканиях, так и в практических исследованиях.

Отмечая актуальность, научную обоснованность, теоретическую и практическую значимость диссертационной работы, можно отметить в работе и несколько недостатков второго плана:

1. В диссертации присутствует ряд опечаток, в частности, стр.15 строка 8 снизу («пакет»), стр.22 строка 5 сверху (повторение слова «может»), стр.47 строка 2 снизу («вспомогательные»).

2. Проведенное в диссертации сравнение с F-тестом, безусловно, показывает на конкретном примере превосходство метода эмпирического моста над F-тестом. Однако с точки зрения построения теории, основанной на методе эмпирического моста, было бы правильным провести исследование и получить необходимые и достаточные условия превосходства предлагаемого метода над F-тестом. Разумеется, данное замечание следует воспринимать скорее как предложение к дополнительному перспективному исследованию.

3. На мой взгляд, текст диссертации несколько перегружен предварительными сведениями не относящимися напрямую к результатам диссертации, которые можно легко получить из соответствующих источников.

4. В доказательстве теоремы 2 при описании шагов в самом начале необходимо явно указать, что замена сумм возможна в смысле предела по вероятности.


5. Диссертационная работа выиграла бы, если бы ее результаты сравнили с результатами работы И.С. Борисова (2013 г.), в которой рассматривались схожие регрессионные модели.

Указанные недостатки, очевидно, не являются значительными и соответственно не могут повлиять на общее положительное впечатление о диссертации.

Все выносимые на защиту положения своевременно опубликованы в научных изданиях, в том числе в 4х статьях в изданиях, рекомендованных

ВАК. Автореферат полно и точно отражает основные положения и выводы диссертации.

Диссертация Е.В. Шаталина представляет собой законченное научное исследование, удовлетворяет всем требованиям ВАК, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук, Положению о присуждении ученых степеней, паспорту специальности 05.13.17, а ее автор Шаталин Евгений Викторович заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.17 — Теоретические основы информатики.

И.о. декана факультета базовой
подготовки ФГБОУ ВО «НГУЭУ»,
кандидат физико-математических наук  С. Е. Хрущев
20 октября 2017г.

Верно
Ученый секретарь ФГБОУ ВО «НГУЭУ»  В.С. Мельников

«20» октября 2017 г.



подпись
МД.

Хрущев Сергей Евгеньевич, к.ф.-м.н. по специальности 01.01.05 – теория вероятностей и математическая статистика, и.о. декана факультета базовой подготовки Федерального государственного бюджетного образовательного учреждения высшего образования «Новосибирский государственный университет экономики и управления «НИНХ» (ФГБОУ ВО «НГУЭУ»).

Адрес: 630099, г. Новосибирск, ул. Каменская, 56

Тел.: +7-(383)- 243-94-80

E-mail: s.e.hrushchev@edu.nsuem.ru