

**(ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ)**

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

УДК 519.22

**АГЕЕВА**

Елена Сергеевна

**СТАТИСТИЧЕСКИЕ ВЫВОДЫ О РЕГРЕССИОННЫХ  
МОДЕЛЯХ ПРИ НАЛИЧИИ КЛАССИФИКАЦИИ  
НАБЛЮДЕНИЙ**

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата физико-математических наук

по специальности 01.01.05 — теория вероятностей  
и математическая статистика

Минск, 2017

Работа выполнена в Белорусском государственном университете.

Научный руководитель — **Харин Юрий Семенович**,  
доктор физико-математических наук,  
профессор, член-корреспондент НАН Бе-  
ларуси, заведующий кафедрой матема-  
тического моделирования и анализа дан-  
ных Белорусского государственного уни-  
верситета.

Официальные оппоненты: **Янович Леонид Александрович**,  
доктор физико-математических наук,  
профессор, член-корреспондент НАН Бе-  
ларуси, главный научный сотрудник от-  
дела нелинейного и стохастического ана-  
лиза ГНУ “Институт математики НАН  
Беларуси”;

**Цеховая Татьяна Вячеславовна**,  
кандидат физико-математических наук,  
доцент, доцент кафедры теории вероят-  
ностей и математической статистики Бе-  
лорусского государственного универси-  
тета.

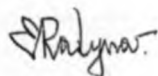
Оппонирующая организация — УО “Томельский государственный уни-  
верситет им. Ф. Скорины”.

Защита состоится 15 сентября 2017 года в 10.00 на заседании сове-  
та по защите диссертаций Д 02.01.07 при Белорусском государственном  
университете по адресу: 220030, Республика Беларусь, г. Минск, ул. Ле-  
нинградская, 8 (корпус юридического факультета), ауд. 407, тел. ученого  
секретаря: (017) 209-57-09.

С диссертацией можно ознакомиться в Фундаментальной библиотеке  
Белорусского государственного университета.

Автореферат разослан “27” июня 2017 года.

Ученый секретарь совета  
по защите диссертаций Д 02.01.07  
кандидат физ.-мат. наук доцент



Е. М. Радыно

## ВВЕДЕНИЕ

Теория статистических выводов исследует математические модели и методы построения статистических оценок параметров, статистических решений, прогнозов на основе наблюдений (статистических данных). Регрессионная модель наблюдений широко используется для описания многих процессов в технике, экономике, медицине, геонауке и других приложениях. Классическая регрессионная модель хорошо изучена в математической статистике: для нее построены статистические оценки параметров, критерии для статистической проверки гипотез согласия и гипотез о коэффициентах линейной регрессии. Однако на практике часто наблюдаются различные отклонения от этой хорошо изученной модели, которые требуется учитывать при построении статистических выводов. Для того чтобы преодолеть затруднения при статистическом анализе данных, связанные с наличием в выборке выбросов, пропусков, искажений, огрублений, требуется разработка специальных сложных моделей данных, примерами которых являются модели с гетероскедастичностью, модели с засорениями Тьюки-Хьюбера, модели с цензурированием или пропусками. Развитием этого направления в математической статистике, получившего название "робастная статистика" (англ. *robust* — "крепкий", "сильный", "устойчивый"), занимаются Р. J. Huber, F. Hampel, P. J. Rousseeuw, E. M. Ronchetti, U. Gather, P. Filzmoser, С. А. Айвазян и другие ученые. В Беларуси статистическому анализу моделей с неполными данными и другими типами искажений посвящены работы Ю. С. Харина, Е. Е. Жука, М. С. Абрамовича, В. И. Малюгина, А. Ю. Харина, И. А. Бодягина, В. А. Волошко.

Одним из возможных искажений классической модели является группирование наблюдаемых значений. В обзорной статье D. F. Heitjan<sup>1</sup> выделены три типа группирования:

- 1) округление, то есть замена истинного значения средней точкой интервала, которому оно принадлежит;
- 2) интервальное цензурирование, например, время обнаружения рецидива заболевания, о котором известно только, что оно наступило между моментами посещений врача;
- 3) группирование как таковое, то есть огрубление, вообще говоря, непрерывных переменных путем разбиения их на категории (классы) во время сбора данных или их обработки.

В диссертационной работе рассматривается регрессионная модель при наличии особого вида группирования наблюдений. Вместо точного значения зависимой переменной наблюдается только номер одного из заранее заданных непересекающихся числовых промежутков (интервалов), в который это

<sup>1</sup> Heitjan, D. F. Inference from Grouped Continuous Data: A Review / D. F. Heitjan // *Statistical Science*. — 1989. — Vol. 4, Issue 2. — P. 164-183.

значение попало. Такое искажение данных будем называть классификацией. Классификация естественным образом возникает при сборе данных, например, при проведении социологических опросов: респондентам не всегда нужно указывать точный ответ, достаточно только выбрать соответствующий интервал. Другая причина появления классифицированных наблюдений связана с удобством хранения информации, когда точное значение не представляет интерес.

Стоит отметить так называемую интервальную регрессию<sup>2</sup>, которую также называют “регрессией при наличии группирования”. Эта модель предполагает, что все значения зависимой переменной наблюдаются с точностью до некоторого интервала. Исследуемая в диссертационной работе регрессионная модель при наличии классификации наблюдений подпадает под это определение, однако, интервальная регрессия является более общим случаем, в котором на интервалы не накладываются никакие ограничения, в частности, их неслучайность и зафиксированность. В связи с этим статистический анализ для интервальной регрессии не получил теоретического развития и ограничивается численными методами построения оценок максимального правдоподобия и некоторых их аппроксимаций (см. работы J. Burridge<sup>3</sup>, S. B. Caudill and J. D. Jackson<sup>4</sup>, M. B. Stewart<sup>5</sup>), при этом вероятностные свойства полученных оценок не исследованы.

Таким образом, актуальной задачей является развитие вероятностно-статистического анализа регрессионной модели при наличии классификации наблюдений с учетом ее особенностей.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Связь работы с крупными научными программами (проектами) и темами

Тема диссертации соответствует направлению фундаментальных и прикладных научных исследований “12.1. Физические и математические методы и их применение для решения актуальных проблем естествознания, техники, новых технологий, экономики и социальных наук”, определенному Перечнем приоритетных направлений фундаментальных и прикладных научных исследований Республики Беларусь в области естественных, технических, гумани-

<sup>2</sup> Long, J. S. Regression Models for Categorical and Limited Dependent Variables Using Stata / J. S. Long, J. Freese. — 2-nd ed. — Stata Press, 2014. — 589 p.

<sup>3</sup> Burridge, J. A Note on Maximum Likelihood Estimation for Regression Models using Grouped Data / J. Burridge // J.R. Statist.Soc. Series B (Methodological). — 1981. — Vol. 43, Issue 1. — P. 41–45.

<sup>4</sup> Caudill, S. B. Heteroscedasticity and grouped data regression / S. B. Caudill, J. D. Jackson // Southern Economic Journal. — 1993. — Vol. 60, Issue 1. — P. 128–135.

<sup>5</sup> Stewart, M. B. On least squares estimation when the dependent variable is grouped / M. B. Stewart // Review of Economic Studies. — 1983. — Vol. 50. — P. 737–753.

тарных и социальных наук на 2011–2015 годы. Результаты диссертационного исследования использованы при выполнении в Белорусском государственном университете и в Учреждении БГУ “Научно-исследовательский институт прикладных проблем математики и информатики” следующих научно-исследовательских работ:

1. НИР “Разработка методов робастного статистического анализа случайных последовательностей и полей при наличии неоднородностей и цензурирования” ГПНИ “Конвергенция”, подпрограмма “Математические методы” (2011–2015 гг.), номер госрегистрации 20111047;

2. НИР “Разработка вероятностных моделей, методов и алгоритмов статистического анализа и прогнозирования дискретных временных рядов” ГПНИ “Конвергенция”, подпрограмма “Методы математического моделирования сложных систем” (2016–2020 гг.), номер госрегистрации 20162616;

3. НИР “Робастные статистические выводы и их применение в компьютерных системах моделирования и анализа данных” (2011–2015 гг.), номер госрегистрации 20120355;

4. Грант Министерства образования Республики Беларусь 2014, НИР “Регрессионное прогнозирование при наличии классификации зависимой переменной” (2014 г.), номер госрегистрации 20140775;

5. Грант Министерства образования Республики Беларусь 2015, НИР “Робастное оценивание параметров множественной регрессии при наличии классификации зависимой переменной” (2015 г.), номер госрегистрации 20150722;

6. Грант БГУ 2011, НИР “Робастный статистический анализ марковских случайных последовательностей” (2011 г.);

7. Грант БГУ 2012, НИР “Статистический анализ данных сложной структуры” (2012 г.);

8. Грант БГУ 2013, НИР “Малопараметрические вероятностно-статистические модели и их применения в задачах анализа данных” (2013 г.);

9. Грант БГУ 2014, НИР “Вероятностно-статистический анализ марковских зависимостей в дискретных данных” (2014 г.);

10. Грант БГУ 2015, НИР “Статистический анализ дискретных временных рядов и случайных последовательностей” (2015 г.).

### **Цель и задачи исследования**

Цель диссертационной работы: построение статистических выводов (статистических оценок, решающих правил и прогнозирующих статистик) о регрессионных моделях при наличии классификации наблюдений. Для достижения поставленной цели требуется решить следующие основные задачи.

2. Для регрессионной модели при наличии классификации наблюдений доказаны новые достаточные условия сильной состоятельности и асимптотической нормальности оценки максимального правдоподобия параметров модели, что дает возможность делать выводы о точности оценивания коэффициентов регрессии и дисперсии случайных ошибок наблюдений [2, 3, 9, 13, 15, 16, 18, 19, 21, 22, 23, 24].

3. Для подстановочной прогнозирующей статистики в рамках регрессионной модели при наличии классификации наблюдений построены новые оценки асимптотического смещения и среднеквадратического риска, что позволяет не только построить прогноз зависимой переменной, но и оценить точность этого прогноза [2, 3, 7, 13, 16, 23].

4. Впервые построены статистические критерии проверки простой нулевой гипотезы против простой и сложной альтернатив об истинном значении параметров регрессионной модели при наличии классификации наблюдений; доказано, что с увеличением объема выборки вероятности ошибок I-ого рода этих критериев стремятся к заданному уровню значимости, а мощности критериев стремятся к единице, что позволяет контролировать точность принимаемых решений о значениях параметров модели [4, 10, 11, 12, 20].

5. Впервые построены статистические критерии проверки гипотез согласия функции регрессии с некоторым параметрическим семейством для регрессионной модели при наличии классификации наблюдений, основанные на модификации  $\chi^2$ -статистики, в случае простой и сложной нулевых гипотез; доказано, что с увеличением объема выборок вероятности ошибок I-ого рода этих критериев стремятся к заданному уровню значимости, что дает возможность контролировать точность принимаемых решений о принадлежности функции регрессии заданному параметрическому семейству функций [2, 3, 6, 14, 17, 21, 22, 23, 25].

## **Рекомендации по практическому использованию результатов**

Практическая значимость диссертационной работы состоит в том, что теоретические результаты могут применяться при статистическом анализе и прогнозировании регрессионных данных при наличии классификации наблюдений в таких областях, как экономика, финансы, техника, медицина. Полученные результаты также целесообразно использовать в учебном процессе при чтении специальных курсов по математической и прикладной статистике на математических факультетах университетов.

## СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ УЧЕНОЙ СТЕПЕНИ

### Статьи в научных изданиях в соответствии с пунктом 18

#### Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь

1. *Агеева, Е. С.* Статистическое оценивание параметров множественной линейной регрессии при наличии случайного цензурирования / Е. С. Агеева, Ю. С. Харин // Вестн. Беларус. гос. ун-та. Сер. 1, Физ., матем., информат. — 2012. — № 1. — С. 72–78.
2. *Агеева, Е. С.* Состоятельность оценки максимального правдоподобия параметров множественной регрессии по классифицированным наблюдениям / Е. С. Агеева, Ю. С. Харин // Доклады Национальной академии наук Беларуси. — 2012. — Т. 56, № 5. — С. 11–19.
3. *Ageeva, H.* ML estimation of multiple regression parameters under classification of the dependent variable / H. Ageeva, Yu. Kharin // Lithuanian Mathematical Journal. — 2015. — Vol. 55, № 1. — P. 48–60.
4. *Агеева, Е. С.* Проверка простых гипотез для регрессионной модели при наличии классификации зависимой переменной / Е. С. Агеева // Труды института математики. — 2015. — Т. 23, № 1. — С. 3–11.

#### Статьи в сборниках научных трудов

5. *Агеева, Е. С.* О статистическом оценивании параметров регрессии при наличии случайного цензурирования / Е. С. Агеева, Ю. С. Харин // Современные информационные компьютерные технологии: сборник научных статей / ГрГУ им. Я. Купалы. — Гродно, 2011. — С. 199–204.
6. *Агеева Е. С.* Критерий согласия для регрессионной модели при наличии классификации зависимых наблюдений / Е. С. Агеева // Теория вероятностей, случайные процессы, математическая статистика и их приложения: сборник научных статей / РИВШ; под редакцией Н.Н. Труша, Г.А. Медведева, Ю.С. Харина. — Минск, 2014. — С. 11–16.
7. *Агеева, Е. С.* Прогнозирование регрессионных временных рядов при наличии классификации наблюдений / Е. С. Агеева, Ю. С. Харин // Экономика. Моделирование. Прогнозирование: сб. науч. тр. / НИЭИ Мин-ва экономики Респ. Беларусь. — Минск, 2015. — Вып. 9. — С. 178–183.

## Статьи в сборниках материалов научных конференций

8. *Агеева, Е. С.* Статистический анализ множественной линейной регрессии при наличии случайного цензурирования / Е. С. Агеева, Ю. С. Харин // Информационные системы и технологии: материалы VI Международной конференции, Минск, 24–25 ноября, 2010 г. / Белорус. гос. ун-т; редкол.: А. Н. Курбацкий [и др.]. — Минск, 2010. — С. 120–124.

9. *Агеева, Е. С.* Статистическое оценивание параметров множественной регрессии при наличии классификации наблюдений / Е.С. Агеева, Ю.С. Харин // Международный конгресс по информатике: информационные системы и технологии: материалы международного научного конгресса, Минск, 31 октября–3 ноября 2011 г.: в 2 ч. / БГУ, ОИПИ НАН Беларуси, НТА «Инфо-парк»; редкол.: С. В. Абламейко (отв. ред.) [и др.]. — Минск, 2011. — Ч. 1. — С. 22–26.

10. *Агеева, Е. С.* Проверка сложных гипотез для множественной регрессии при наличии классификации наблюдений / Е. С. Агеева, Ю. С. Харин // 70-ая научная конференция студентов и аспирантов Белорус. гос. ун-та: сб. работ, Минск, 15–18 сентября, 2013 г.: в 3 ч. — Минск, 2013. — Ч. 1. — С. 160–164.

11. *Агеева, Е. С.* Проверка простой нулевой гипотезы и сложной альтернативной гипотезы для множественной регрессии при наличии классификации наблюдений / Е. С. Агеева // Международный конгресс по информатике: информационные системы и технологии: материалы международного научного конгресса, Минск, 4-7 ноября, 2013 г. — Минск: БГУ, 2013. — С. 12–16.

12. *Ahejeva, H.* On hypothesis testing for regression model under classification of dependent variable / H. Ahejeva // Computer data analysis and modeling: proceeding of the 10th International conference, Minsk, September 11–14, 2013: in 2 vol. / Ministry of Education of the Republic of Belarus; ed. board: Prof. Dr. S. Aivazian, Prof. Dr. P. Filzmoser, Prof. Dr. Yu. Kharin. — Minsk, 2013. — Vol. 1. — P. 52–55.

13. *Агеева, Е. С.* Прогнозирование линейных регрессионных временных рядов при классификации зависимой переменной / Е. С. Агеева, Ю. С. Харин // Проблемы прогнозирования и государственного регулирования социально-экономического развития: материалы XV Междунар. науч. конференции, Минск, 23–24 октября, 2014 г. — Минск, 2014. — С. 188–189.

14. *Агеева, Е. С.* Проверка сложных гипотез согласия для регрессионной модели при наличии классификации зависимой переменной / Е. С. Агеева // Статистические методы анализа экономики и общества: труды 6-ой Междунар. научно-практической конференции студентов и аспирантов, Москва, 12–15 мая, 2015 г. — Москва, Россия, 2015. — С. 37–38.



15. Харин, Ю. С. Об одном подходе к построению робастной оценки линейной регрессионной модели на основе классификации зависимой переменной / Ю. С. Харин, Е. С. Агеева // Проблемы прогнозирования и государственного регулирования социально-экономического развития: материалы XVI Международ. науч. конф., Минск, 23 октября, 2015 г.: в 3 т. / НИЭИ М-ва экономики Респ. Беларусь; редкол.: А.В. Червяков [и др.]. — Минск, 2015. — Т. 1. — С. 150–155.

16. Ageeva, H. Forecasting of regression model under classification of the dependent variable / H. Ageeva // Computer data analysis and modeling: proceeding of the 11th International conference, Minsk, September 6–10, 2016. / Publishing center of BSU; ed. board: Prof. Dr. S. Aivazian, Prof. Dr. P. Filzmoser, Prof. Dr. Yu. Kharin. — Minsk, 2016. — P. 117–120.

17. Агеева, Е. С. Хи-квадрат критерий для регрессионной модели при наличии классификации наблюдений / Е. С. Агеева, Ю. С. Харин // Международный конгресс по информатике: информационные системы и технологии — International Congress on Computer Science: Information Systems and Technologies [Электронный ресурс] : материалы междунар. науч. конгресса, Республика Беларусь, Минск, 24–27 окт. 2016 г. / редкол.: С. В. Абламейко (гл. ред.), В. В. Казаченок (зам. гл. ред. [и др.]). — Минск : ВГУ, 2016. — С. 416–420. — 1 электрон. опт. диск (DVD-RW).

### Тезисы докладов

18. Aheyeva, H. S. On asymptotic properties of ML estimators for the regression parameters under classification of observations / H. S. Aheyeva, Yu. S. Kharin // Modern Stochastics: Theory and Applications III: conference materials of the Intern. conf., Kiev, Sept. 10–14, 2012. — Kiev, Ukraine, 2012. — P. 77.

19. Агеева, Е. С. Об асимптотических свойствах ОМП регрессионных коэффициентов при наличии классификации наблюдений / Е. С. Агеева, Ю. С. Харин // XI Белорусская математическая конференция: сб. тезисов Междун. науч. конф., Минск, 4–9 ноября, 2012 г. — Минск, 2012. — С. 43.

20. Aheyeva, H. S. Hypothesis testing in the multiple regression model under classification of the dependent variable / H. S. Aheyeva, Yu. S. Kharin // International Conference on Robust Statistics: book of abstr. of the Intern. conf., St. Petersburg, July 8–12, 2013. — St. Petersburg, Russia, 2013. — P. 11.

21. Ageeva, H. Statistical analysis for regression model under grouping distortion / H. Ageeva, Yu. Kharin // 11th International Vilnius Conference on Probability Theory and Mathematical Statistics: book of abstr. of the Intern. conf., Vilnius, 30 June – 3 July, 2014. — Vilnius, Lithuania, 2014. — P. 25.

22. Ageeva, H. Statistical inferences for multiple regression under grouping / H. Ageeva, Yu. Kharin // International Conference on Robust Statistics: book of abstr. of the Intern. conf., Martin-Luther-University Halle-Wittenberg, August 10–15, 2014. — Halle (Saale), Germany, 2014. — P. 18.

23. Ageeva, H. Forecasting of regression time series under classification of the dependent variable / H. Ageeva, Yu. Kharin // Probability, Reliability and Stochastic Optimization: book of abstr. of the Intern. conf., Kyiv, April 7–10, 2015. — Kiev, Ukraine, 2015. — P. 37.

24. Kharin, Yu. Statistical Analysis of Time Series Based on Incomplete Discrete Data / Yu. Kharin, H. Ageeva, H. Rudakouskaya // Data Analysis Methods for Software Systems: book of abstr. of 7th Intern. Workshop, Druskininkai, Lithuania, December 3–5, 2015. — Druskininkai, Lithuania, 2015. — P. 30–31.

25. Ageeva, E. C. Проверка гипотез согласия для регрессионной модели при наличии классификации зависимой переменной / Е. С. Агеева // XII Белорусская математическая конференция: материалы Международной научной конференции. Минск, 5–10 сентября, 2016 г.: в 5 ч. — Минск, 2016. — Ч. 4. — С. 3.

## РЭЗЮМЭ

Агеева Алена Сяргееўна

Статыстычныя вывады аб рэгрэсійных мадэлях пры наяўнасці класіфікацыі назіранняў

*Ключавыя словы:* рэгрэсія, групаванне, класіфікацыя, ацэнка максімальнай праўдападобнасці, параметрычныя гіпотэзы, гіпотэзы адпаведнасці,  $\chi^2$ -статыстыка.

Мэтай дысертацыйнай работы з'яўляецца пабудова статыстычных вывадаў (статыстычных ацэнак, рашаючых правілаў і прагназуючых статыстык) аб рэгрэсійных мадэлях пры наяўнасці класіфікацыі назіранняў. Пры даследаванні выкарыстоўваліся метады тэорыі імавернасцей, матэматычнай статыстыкі, рэгрэсійнага аналізу, камп'ютэрнага мадэлявання.

У дысертацыйнай рабоце атрыманы наступныя новыя навуковыя вынікі. Для рэгрэсійнай мадэлі пры наяўнасці класіфікацыі назіранняў устаноўлены неабходныя і дастатковыя ўмовы ідэнтыфікацыі мадэлі, знойдзены дастатковыя ўмовы моцнай кансістэнтнасці і асімптатычнай нармальнасці ацэнкі максімальнай праўдападобнасці. Для падстаноўчай прагназуючай статыстыкі прапанаваны ацэнкі яе асімптатычнага зрушэння і сярэднеквадратычнага рыску. Пабудаваны статыстычныя крытэрыі для праверкі прастай нулявой гіпотэзы супраць прастай і складанай альтэрнатыў аб сапраўдных значэннях параметраў рэгрэсійнай мадэлі пры наяўнасці класіфікацыі назіранняў; даказана, што іх імавернасці памылкі I-ага роду імкнуцца да зададзенага ўзроўню значнасці, а магутнасці імкнуцца да 1. Пабудаваны статыстычныя крытэрыі для праверкі гіпотэз адпаведнасці для рэгрэсійнай мадэлі пры наяўнасці класіфікацыі назіранняў, заснаваныя на мадыфікацыі  $\chi^2$ -статыстыкі; даказана, што іх імавернасці памылкі I-ага роду імкнуцца да зададзенага ўзроўню значнасці.

Атрыманыя вынікі могуць быць выкарыстаны пры статыстычным аналізе і прагназаванні рэгрэсійных дадзеных пры наяўнасці класіфікацыі назіранняў у такіх галінах, як эканоміка, фінансы, тэхніка, медыцына, а таксама ў навучальным працэсе.

## РЕЗЮМЕ

Агеева Елена Сергеевна

Статистические выводы о регрессионных моделях при наличии классификации наблюдений

*Ключевые слова:* регрессия, группирование, классификация, оценка максимального правдоподобия, параметрические гипотезы, гипотезы согласия,  $\chi^2$ -статистика.

Цель диссертационной работы заключается в построении статистических выводов (статистических оценок, решающих правил и прогнозирующих статистик) о регрессионных моделях при наличии классификации наблюдений. При исследовании использовались методы теории вероятностей, математической статистики, регрессионного анализа, компьютерного моделирования.

В диссертационной работе получены следующие новые научные результаты. Для регрессионной модели при наличии классификации наблюдений установлены необходимые и достаточные условия идентифицируемости, найдены достаточные условия сильной состоятельности и асимптотической нормальности оценки максимального правдоподобия. Для подстановочной прогнозирующей статистики построены оценки ее асимптотического смещения и среднеквадратического риска. Построены статистические критерии для проверки простой нулевой гипотезы против простой и сложной альтернатив об истинных значениях параметров регрессионной модели при наличии классификации наблюдений; доказано, что их вероятности ошибок I-ого рода стремятся к заданному уровню значимости, а мощности стремятся к 1. Построены статистические критерии проверки гипотез согласия для регрессионной модели при наличии классификации наблюдений, основанные на модификации  $\chi^2$ -статистики; доказано, что их вероятности ошибок I-ого рода стремятся к заданному уровню значимости.

Полученные результаты могут быть использованы при статистическом анализе и прогнозировании регрессионных данных при наличии классификации наблюдений в таких областях, как экономика, финансы, техника, медицина, а также в учебном процессе.

## SUMMARY

Ageeva Helena

Statistical inferences on regression models under classification of observations

*Key words:* regression, grouping, classification, maximum likelihood estimator, parametric hypotheses, goodness-of-fit testing,  $\chi^2$  statistic

The goal of this dissertation is to construct statistical inferences (statistical estimators, statistical tests and forecasting statistics) on regression models under classification of observations. The techniques used include methods of probability theory, mathematical statistics, regression analysis, computer simulation.

In this dissertation the following new scientific results are obtained. Necessary and sufficient conditions for identifiability of the regression model under the classification of observations are established. Sufficient conditions for consistency and asymptotic normality of the MLE are found. Asymptotic bias and squared error risk are obtained for the plug-in forecasting statistic for the considered model. Statistical tests for simple null hypothesis versus simple and composite alternative hypotheses for true value of the regression model parameter under the classification of observations are constructed; their probabilities of type I error are proven to converge to the given significance level and their powers are proven to converge to 1. Statistical tests for goodness-of-fit hypotheses for the regression model under the classification of observations are constructed based on the modification of  $\chi^2$  statistic; their probabilities of type I error are proven to converge to the given significance level.

The obtained results can be used in statistical analysis and forecasting of real data under classification of observations in areas such as economics, finance, technology, medicine, as well as in the educational process.

