На правах рукописи

Семёнова Мария Александровна

РАЗРАБОТКА АЛГОРИТМИЧЕСКОГО ОБЕСПЕЧЕНИЯ И ИССЛЕДОВАНИЕ ОБОБЩЕННЫХ МОДЕЛЕЙ ПРОПОРЦИОНАЛЬНЫХ ИНТЕНСИВНОСТЕЙ

Специальность 05.13.17 – Теоретические основы информатики

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

Новосибирск – 2015

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Новосибирский государственный технический университет»

Научный руководитель:	кандидат технических наук, доцент Чимитова Екатерина Владимировна
Официальные оппоненты:	Кошкин Геннадий Михайлович, доктор физико-математических наук, профессор, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет», кафедра теоретической кибернетики, старший научный сотрудник;
	Огородников Василий Александрович, доктор физико-математических наук, старший научный сотрудник, Федеральное государственное бюджетное учреждение науки Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, лаборатория стохастических задач, главный научный сотрудник
Ведущая организация:	Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Томский государственный университет систем управления и радиоэлектроники»

Защита состоится «15» октября 2015 г. в 14^{00} часов на заседании диссертационного совета Д 212.173.06 при Федеральном государственном бюджетном образовательном учреждении высшего образования «Новосибирский государственный технический университет» по адресу: 630073, Новосибирск, пр. К. Маркса, 20.

С диссертацией можно ознакомиться в библиотеке Новосибирского государственного технического университета и на сайте http://www.nstu.ru.

Автореферат разослан « » сентября 2015 г.

Ученый секретарь
диссертационного совета Фаддеенков Андрей Владимирович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Современное состояние и актуальность темы исследования. Методам анализа надежности и выживаемости на сегодняшний день посвящено множество работ, направленных как на получение новых знаний и содержащих результаты инновационных исследований в области медицины, биологии, страхования, социологии, экономики, в задачах анализа надежности и долговечности, так и на обучение специалистов, работающих в данных областях.

Исследуемой случайной величиной в таких задачах является время до наступления некоторого события, которое, как правило, называется отказом. При этом в качестве отказа может выступать как выход из строя технического устройства, так и смерть тяжелобольного пациента, или же наступление ремиссии заболевания при условии получения некоторого вида лечения.

Специфика задач анализа надежности и выживаемости заключается в наличии цензурированных данных, возникновение которых обусловлено условиями проведения эксперимента. Среди отечественных публикаций, посвященных анализу цензурированных данных, стоит отметить работы Скрипника В.М., Благовещенского Ю.Н., Приходько Ю.Г., Назина А.Е., Ушакова И.А., Острейковского В.А., Антонова А.В., Никулина М.С., Аронова И.З. и др.

При построении моделей надежности и выживаемости учитывается зависимость вероятности наступления отказа от значений объясняющих переменных, называемых также ковариатами. В качестве ковариат могут выступать как внутренние свойства объектов исследования (например, возраст, пол или наличие хронических заболеваний), так и условия проведения эксперимента (например, вид терапии или наличие вспомогательных видов лечения). Одной из первых моделей зависимости вероятности отказа от ковариат является модель пропорциональных интенсивностей Кокса. В основе данной модели заложено предположение пропорциональности рисков, которое заключается в том, что отношение функций интенсивности для объектов с разными значениями ковариат постоянно во времени. Достоверность построенной модели зависит от того, выполняется ли данное предположение, следовательно, его проверка является обязательным этапом построения модели пропорциональных интенсивностей.

Для проверки предположения о пропорциональности рисков существует множество графических методов, которые не рассматриваются в настоящей работе вследствие отсутствия количественной меры их сравнения. Другой способ проверки предположения пропорциональности рисков заключается во включении в модель зависимых от времени ковариат с последующей проверкой гипотезы о незначимости добавленных переменных. Однако такой подход может быть применен лишь при известной форме зависимости ковариат от времени. Множество методов проверки предположения о пропорциональности рисков основаны на сравнении полученных результатов для разных значений объясняющей переменной. Данные методы хорошо зарекомендовали себя в

случае бинарных объясняющих переменных, например, при сравнении рисков в экспериментальной и контрольной группах. Если же мы имеем дело с количественными переменными, то возникает проблема, связанная с неоднозначностью способа категоризации этих переменных, вследствие чего результаты применения таких методов часто противоречат друг другу.

В этой связи особый интерес представляет статистический критерий проверки гипотезы о пропорциональности рисков, называемый в настоящей работе критерием Никулина по имени одного из авторов. Известно, что предельным распределением статистики критерия Никулина при справедливости нулевой гипотезы является хи-квадрат распределение. Однако вопрос о корректности использования предельного распределения для вычисления достигнутого уровня значимости при небольших объемах выборок, с которыми, вообще говоря, приходится иметь дело в задачах анализа выживаемости, до сих пор не исследовался.

Несмотря на преимущества и популярность модели пропорциональных интенсивностей, предположение, заложенное в ее основу, на практике зачастую не выполняется, и возникает необходимость построения более сложных моделей. Основная трудность построения обобщенных моделей Ксая или SCE (Simple Cross-Effect model) заключается в необходимости одновременного оценивания регрессионных параметров, в том числе обобщающих, и неизвестной базовой функции риска. Более того, при построении вероятностных моделей выживаемости часто возникает задача определения ковариат, оказывающих статистически значимое влияние на функцию выживаемости, для чего проверяется гипотеза о незначимости регрессионных параметров с использованием критерия отношения правдоподобия или критерия Вальда.

Для проверки гипотезы о виде параметрической модели пропорциональных интенсивностей Кокса и её обобщений предлагается использование критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова или Андерсона-Дарлинга по выборкам остатков Кокса-Снелла. Необходимо отметить, что при проверке гипотезы о виде обобщенных параметрических моделей пропорциональных интенсивностей по цензурированным данным, выборки остатков оказываются многократно цензурированными, поэтому используются модификации непараметрических критериев, опирающиеся на оценки Каплана-Мейера.

Другой подход к проверке гипотезы о виде параметрических моделей по цензурированным данным основан на преобразовании исходной цензурированной выборки в псевдополную и использовании классических критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова или Андерсона-Дарлинга. Использование классических критериев при проверке данной гипотезы по цензурированным выборкам значительно упрощает процедуру проверки по сравнению с использованием модифицированных критериев согласия. Однако вопрос о корректности применения такого подхода в случае проверки сложных гипотез остается открытым. Кроме того, существует необходимость исследования мощности классических критериев,

использующих при проверке гипотез о виде моделей пропорциональных интенсивностей переход к псевдополным выборкам. Наконец, для проверки гипотезы о виде параметрической модели пропорциональных интенсивностей по цензурированным данным существует критерий типа хи-квадрат.

В связи с наличием различных подходов к проверке гипотезы о виде параметрических моделей пропорциональных интенсивностей актуальной задачей является сравнительный анализ различных критериев согласия.

В условиях цензурированности наблюдений и ограниченности объёмов выборок свойства оценок и критериев проверки статистических гипотез зачастую существенно отличаются от асимптотических.

Цель и задачи исследований. Целью данной диссертационной работы является исследование свойств критериев проверки статистических гипотез относительно обобщенных моделей пропорциональных интенсивностей, а также разработка алгоритмов, обеспечивающих корректность применения критериев в случае цензурированных данных.

В соответствии с поставленной целью предусмотрено решение следующих задач.

1. Исследование распределений статистик критериев отношения правдоподобия и Вальда, используемых для проверки гипотез о незначимости параметров моделей Кокса, Ксая и SCE в условиях неизвестной базовой функции выживаемости.

2. Исследование распределений статистик и мощности критериев, используемых для проверки предположения о пропорциональности рисков на основе полупараметрических моделей.

3. Исследование распределений статистик и мощности критериев проверки гипотезы о виде параметрической модели пропорциональных интенсивностей Кокса.

4. Разработка алгоритмов моделирования распределений статистик модифицированных критериев согласия для цензурированных выборок, в том числе для случайно цензурированных выборок при неизвестном распределении моментов цензурирования.

5. Разработка программного обеспечения построения и проверки гипотез о виде обобщенных моделей пропорциональных интенсивностей по цензурированным данным.

Методы исследований. Для решения поставленных задач использовались методы математической статистики, теории вероятностей, математического программирования и статистического моделирования.

Научная новизна диссертационной работы заключается в:

– результатах исследования распределений статистик критериев отношения правдоподобия и Вальда, используемых при проверке гипотезы о равенстве нулю параметров моделей пропорциональных интенсивностей, Ксая и SCE в условиях неизвестной базовой функции выживаемости;

– результатах сравнительного анализа критериев, используемых для проверки предположения о пропорциональности рисков в условиях неизвестной базовой функции выживаемости;

– разработанных алгоритмах моделирования распределений статистик модифицированных критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для цензурированных выборок при проверке гипотез о виде параметрических моделей Кокса, Ксая и SCE по выборкам остатков;

– предлагаемой методике проверки гипотезы о виде параметрических моделей выживаемости, в основе которой лежит преобразование исходной цензурированной выборки в псевдополную;

– результатах сравнительного анализа критериев проверки гипотезы о виде параметрической модели пропорциональных интенсивностей и ее обобщений по цензурированным выборкам.

Основные положения, выносимые на защиту:

1. Результаты исследования распределений статистик и мощности критериев отношения правдоподобия, Вальда и Никулина, используемых при проверке предположения о пропорциональности рисков в условиях неизвестной базовой функции выживаемости.
2. Методика проверки гипотез о виде параметрических моделей Кокса, Ксая и SCE, в основе которой лежит преобразование исходной цензурированной выборки в псевдополную.
3. Алгоритмы моделирования распределений статистик модифицированных критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для цензурированных выборок при проверке гипотез о виде параметрических моделей Кокса, Ксая и SCE по выборкам остатков.
4. Результаты сравнительного анализа мощности критериев проверки гипотез о виде параметрической модели пропорциональных интенсивностей.
5. Программное обеспечение построения параметрических и полупараметрических моделей пропорциональных интенсивностей Кокса и ее обобщений.

Личный творческий вклад автора заключается в проведении исследований, обосновывающих основные положения, выносимые на защиту, и в реализации полученных результатов в разрабатываемой программной системе статистического анализа данных типа времени жизни «LiTiS».

Практическая ценность и реализация результатов заключается в формировании рекомендаций по проверке статистических гипотез о параметрах и виде вероятностных моделей пропорциональных интенсивностей и их обобщений по цензурированным выборкам. Разработанные алгоритмы моделирования распределений статистик критериев согласия для цензурированных выборок реализованы в программной системе статистического анализа данных типа времени жизни «LiTiS» (Свидетельства о государственной регистрации программы для ЭВМ № 2012618138 (2012 г.), № 2012618143 (2012 г.), № 2014661905 (2015 г.). – М.: Федеральная служба по интеллектуальной собственности (Роспатент)).

Результаты проведенных исследований и разработанное программное обеспечение были внедрены в практику деятельности гематологического

центра ФГКУ «Главного военного клинического госпиталя имени академика Н.Н. Бурденко», использованы в научно-исследовательской деятельности кафедры проектирования технологических машин ФГБОУ ВО «Новосибирский государственный технический университет» и в учебном процессе на механико-технологическом факультете, а также нашли практическое применение в учебном процессе на факультете прикладной математики и информатики ФГБОУ ВО «Новосибирский государственный технический университет», что подтверждается соответствующими справками о внедрении.

Исследования и разработка программного обеспечения проводились при поддержке Министерства образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» (по государственным контрактам № П1190 от 27 августа 2009 г., № П2611 от 26 ноября 2009 г. и № П950 от 20 августа 2009 г, № 02.740.11.5187 от 12 марта 2010 г., соглашения № 14.B37.21.0860 от 6 сентября 2012 г.) и в рамках проектной части государственного задания (проект № 2.541.2014/К).

Соответствие диссертации паспорту научной специальности. Содержание диссертации соответствует п.5 области исследований «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» паспорта специальности научных работников 05.13.17 – «Теоретические основы информатики» по техническим наукам.

Апробация результатов диссертации. Результаты работы докладывались на международной конференции “Accelerated life testing, reliability-based analysis and design”, Клермон-Ферран, Франция, 2010г.; международной конференции “Computer data analysis and modeling: complex stochastic data and systems”, Минск, Белоруссия, 2010г.; международном семинаре “Applied methods of statistical analysis”, Новосибирск, 2011г. и 2013г.; международном семинаре по моделированию “International Workshop on Simulation”, Римини, Италия, 2013г.; международной конференции “Advanced Mathematical and Computational Tools”, Санкт-Петербург, 2014г.; международной научно-технической конференции “Актуальные проблемы электронного приборостроения”, Новосибирск, 2012г. и 2014г.; международной научно-практической интернет-конференции “Перспективные инновации в науке, образовании, производстве и транспорте”, Одесса, Украина, 2011г.; научной сессии НИЯУ МИФИ, секции “Современные проблемы надежности. Анализ надежности оборудования АЭС”, Обнинск, 2015г.; Российской научно-технической конференции “Обработка информационных сигналов и математическое моделирование”, Новосибирск, 2012г., 2013г. и 2014г.; Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”, Новосибирск, 2010г. и 2011г.; всероссийской научной конференции молодых ученых “Наука. Технология. Инновации”, Новосибирск, 2010г.; конференции молодых исследователей “Progress through innovative technologies”, Новосибирск, 2012г.

Публикации. По результатам диссертационных исследований опубликованы 24 печатных работы, в том числе 5 статей в научных журналах и изданиях, рекомендуемых ВАК РФ, 2 статьи в сборниках научных трудов, 2 статьи в рецензируемых Международных изданиях, 12 публикаций в материалах Международных и Российских конференций, 3 свидетельства о государственной регистрации программы для ЭВМ.

Структура и объем диссертации. Общий объем диссертационной работы составляет 154 страницы, основная часть изложена на 150 страницах и состоит из введения, 4-х разделов основного содержания, включая 26 таблиц и 43 рисунка, заключения, списка использованных источников из 115 наименований и приложения.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследования, определены научная новизна и практическая ценность работы, дано краткое содержание работы по разделам.

В первом разделе приводятся соотношения для получения оценок максимального правдоподобия (ОМП) параметров рассматриваемых полупараметрических (с неизвестным базовым распределением отказов) и параметрических моделей, а также описываются критерии проверки статистических гипотез относительно обобщенных моделей пропорциональных интенсивностей.

Пусть время наступления отказа T_x – непрерывная положительно определенная случайная величина. Время наступления отказа зависит от объясняющих переменных, заданных вектором ковариат $x=\left[x_1,x_2,...,x_m\right]^T$, каждый элемент которого определен на отрезке $\left[x_i^{\min},x_i^{\max}\right]$, $i=1,2,...,m$, границы задаются условиями эксперимента.

Закон распределения случайной величины T_x определяется одной из следующих функций: функцией плотности распределения $f_x(t)$, функцией распределения $F_x(t)$, функцией выживаемости $S_x(t)=P(T_x\geq t)=1-F_x(t)$, функцией интенсивности $\lambda_x(t)=\lim\limits_{h\to 0}\left(P\left\{t\leq T_x<t+h\,|\,T_x>t\right\}/h\right)$ и кумулятивной функцией риска $\Lambda_x(t)=\int\limits_0^t\lambda_x(u)du=-\ln\left(S_x(t)\right)$.

Кумулятивная функция риска для модели пропорциональных интенсивностей имеет следующий вид

$$\Lambda_x(t;\beta)=\exp\left(\beta^T\cdot x\right)\Lambda_0(t), \tag{1}$$

где $\beta=\left[\beta_1,\beta_2,...,\beta_m\right]^T$ – вектор неизвестных параметров модели, $\Lambda_0(t)$ – базовая кумулятивная функция риска. Данную модель можно использовать

только при выполнении предположения пропорциональности рисков, в противном случае полученные выводы могут оказаться некорректными.

Ксай предложил обобщенную модель пропорциональных интенсивностей в виде

$$\Lambda_x(t;\beta,\gamma)=\exp\left(\beta^T\cdot x\right)\left\{\Lambda_0(t)\right\}^{\exp\left(\gamma^T\cdot x\right)}. \qquad (2)$$

Параметры β и $\gamma=\left[\gamma_1,\gamma_2,...,\gamma_m\right]^T$ являются m-мерными, а сама модель представляет собой обобщение модели пропорциональных интенсивностей с заменой степени 1 базовой функции риска $\Lambda_0(t)$ на $\exp\left(\gamma^T\cdot x\right)$. Функции интенсивности, соответствующие данной модели, при разных значениях ковариат пересекаются для любых значений параметров β и $\gamma\neq 0$.

SCE-модель или модель с пересечением функций выживаемости, позволяющая получить не только пересекающиеся, но и расходящиеся функции интенсивности, может быть записана следующим образом

$$\Lambda_x(t;\beta,\gamma)=\left(1+\exp\left(\left(\beta+\gamma\right)^T\cdot x\right)\cdot\Lambda_0(t)\right)^{\exp\left(-\gamma^T\cdot x\right)}-1. \qquad (3)$$

Если не задана параметризация базовой функции риска, вероятностные модели называются полупараметрическими. Если же вводится предположение о виде закона распределения отказов, то есть параметризация базовой функции риска, то модель является параметрической. В качестве базового распределения при построении вероятностных моделей надежности и выживаемости чаще всего используются экспоненциальное распределение, распределение Вейбулла, гамма-распределение, логнормальное распределение и другие.

Цензурированная выборка с ковариатами может быть записана в виде

$$\mathrm{X}_n=\left\{\left(t_1,\delta_1,x^1\right),\left(t_2,\delta_2,x^2\right),...,\left(t_n,\delta_n,x^n\right)\right\}, \qquad (4)$$

где n – объем выборки, $t_i=\min\left(T_i,C_i\right)$ – время T_i до наступления отказа i-го объекта или момент цензурирования C_i, δ_i – индикатор цензурирования, $x^i=\left[x_1^i,x_2^i,...,x_m^i\right]^T$ – вектор значений ковариат, $i=1,2,...,n$.

Цензурированные выборки, встречающиеся на практике, можно разделить на три основных типа и их комбинации. Если время эксперимента ограничено, то есть наблюдение за объектами ведется до заранее определенного момента времени c, тогда $\forall\delta_i=0$: $C_i=c$, и незавершенные до окончания эксперимента наблюдения называются цензурированными первого типа (I типа). Если эксперимент продолжается до наступления определенного количества (r) отказов, и наблюдение за остальными объектами прекращается в момент r-го отказа, то незавершенные наблюдения называются цензурированными второго типа (II типа), и $\forall\delta_i=0$: $C_i=T_{(r)}$, где $T_{(r)}$ – время последнего отказа. Возможны ситуации, в которых цензурирование происходит в один момент времени – однократное цензурирование, или в различные моменты времени –

многократное цензурирование. В случае многократного цензурирования моменты выбытий C_i могут быть зафиксированы, например, если при тестировании некоторых изделий в определенные моменты времени из исследования выводилось по несколько объектов. В противном случае, если C_i – случайная величина из некоторой функции определения $F_x^C(t)$, то наблюдения со значением C_i, $\delta_i = 0$ называются цензурированными третьего типа или случайно цензурированными. Данный тип цензурирования, в свою очередь, делится на независимое (неинформативное) цензурирование, когда закон распределения $F_x^C(t)$ не зависит от $F_x(t)$ – функции распределения значений T_i. При зависимом (информативном) цензурировании третьего типа случайная величина C_i может зависеть от закона распределения $F_x(t)$. В настоящей работе рассматриваются выборки с независимым цензурированием.

Для проверки гипотезы о параметрах модели $H_0 : \mu = \mu^0$ можно использовать критерий отношения правдоподобия со статистикой

$$LR = 2\left(\ln L(\hat{\mu}) - \ln L\left(\mu^0\right)\right) \tag{5}$$

и критерий Вальда, статистика которого может быть записана следующим образом

$$W = \left(\hat{\mu} - \mu^0\right)^T I(\hat{\mu})\left(\hat{\mu} - \mu^0\right), \tag{6}$$

где $L(\mu)$ – функция правдоподобия, $I(\mu) = \left[-\dfrac{\partial^2 \ln L(\mu)}{\partial \mu_i \partial \mu_j}\right]_{s \times s}$, $i, j = 1, 2, ..., s$, s – размерность вектора параметров модели $\mu = \left(\beta^T, \gamma^T\right)^T$, $\hat{\mu}$ – оценка максимального правдоподобия (ОМП). При справедливости нулевой гипотезы статистики (5) и (6) асимптотически распределены по закону χ^2 с числом степеней свободы равным количеству оцениваемых параметров модели.

Статистическую гипотезу о виде модели в общем случае можно сформулировать следующим образом

$$H_0 : \Lambda_x(t) = q\left(x, \Lambda_0(t); \beta, \gamma\right), \tag{7}$$

где функция $q\left(x, \Lambda_0(t); \beta, \gamma\right) = \exp\left(\beta^T \cdot x\right)\Lambda_0(t)$ для модели пропорциональных интенсивностей, $q\left(x, \Lambda_0(t); \beta, \gamma\right) = \exp\left(\beta^T \cdot x\right)\left\{\Lambda_0(t)\right\}^{\exp\left(\gamma^T \cdot x\right)}$ для модели Ксая и $q\left(x, \Lambda_0(t); \beta, \gamma\right) = \left(1 + \exp\left(\left(\beta + \gamma\right)^T \cdot x\right)\Lambda_0(t)\right)^{\exp\left(-\gamma^T \cdot x\right)} - 1$ для SCE-модели.

Универсальный способ проверки гипотезы о виде параметрической вероятностной модели выживаемости вида $\Lambda_x(t) = q\left(x, \Lambda_0(t; \theta); \beta, \gamma\right)$ основан на построении, так называемых, остатков Кокса-Снелла, которые вычисляются следующим образом

$$R_i = \Lambda_{x^i}\left(t_i;\hat{\theta},\hat{\beta},\hat{\gamma}\right),\ i=1,2,...,n, \tag{8}$$

где $\hat{\theta},\hat{\beta},\hat{\gamma}$ – ОМП параметров предполагаемой модели. Если модель верна, то полученная выборка остатков $\mathrm{R}_n = \left\{\left(R_1,\delta_1\right),...,\left(R_n,\delta_n\right)\right\}$ принадлежит стандартному экспоненциальному распределению.

Для проверки гипотезы H_0 о принадлежности выборки остатков экспоненциальному распределению по цензурированным выборкам при вычислении статистик модифицированных критериев типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга на наблюдаемом интервале $\left[0,\tau\right]$, где τ – время последнего полного наблюдения (отказа), вместо эмпирической функции распределения $F_n\left(t\right)$ предлагается использовать оценку функции распределения Каплана-Мейера $\hat{F}_n\left(t\right)$.

Обозначим через $0 < a_1 < a_2 < ... < a_k = \tau < 1$, $k \le n$, неповторяющиеся значения полных наблюдений $\left(R_i,\delta_i = 1\right)$ в выборке остатков. Тогда оценку Каплана-Мейера можно вычислить по формуле

$$\hat{F}_n\left(t\right) = 1 - \prod_{a_i \le t}\left(1 - \frac{d_i}{r_i}\right),$$

где $d_i = \sum\limits_{R_j = a_i} \delta_j$, r_i – количество наблюдений, для которых $R_j \ge a_i$, $j = 1,...,n$.

Статистика модифицированного критерия типа Колмогорова с поправкой Большева имеет вид

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{9}$$

где $D_n = \sup\limits_{0 \le t \le \tau}\left|\hat{F}_n\left(t\right) - F\left(t\right)\right|$.

Модификация статистики Крамера-Мизеса-Смирнова определяется следующим образом

$$S_\omega = n\int_0^\tau \left(F\left(t\right) - \hat{F}_n\left(t\right)\right)^2 dF\left(t\right). \tag{10}$$

В модифицированном критерии Андерсона-Дарлинга статистика задается выражением

$$S_\Omega = n\int_0^\tau \frac{\left(F\left(t\right) - \hat{F}_n\left(t\right)\right)^2}{F\left(t\right)\cdot\left(1 - F\left(t\right)\right)} dF\left(t\right). \tag{11}$$

В выражениях (9), (10) и (11) $F\left(t\right)$ – функция распределения стандартного экспоненциального закона, $\hat{F}_n\left(t\right)$ – оценка Каплана-Мейера функции распределения остатков. Следует заметить, что при проверке сложной гипотезы описанные критерии согласия даже для полных данных теряют свойство

свободы от распределения, и распределения статистик $G(S \mid H_0)$ критериев согласия зависят от вида проверяемой вероятностной модели, от количества параметров, оцениваемых по выборке, от метода оценивания параметров и других факторов.

Проверяемая гипотеза отвергается при больших значениях статистик. Аналитические выражения для распределений статистик рассматриваемых критериев в случае проверки сложной гипотезы неизвестны, поэтому вычисление критических значений статистик, необходимых для принятия решения с использованием данных критериев, возможно с опорой на численные оценки распределений статистик, получаемые с использованием статистического моделирования.

Учитывая сложность применения модифицированных критериев – отсутствие аналитического вида распределений статистик при справедливости нулевой гипотезы, в настоящей работе предложена методика применения классических критериев согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга к цензурированным данным, предусматривающая предварительное преобразование исходной цензурированной выборки (4) в псевдополную выборку $\left(\hat{t}_1, x^1\right), \left(\hat{t}_2, x^2\right), ..., \left(\hat{t}_n, x^n\right)$.

Для проверки гипотезы вида (7) в работах Никулина и Багдонавичуса предложен критерий типа хи-квадрат, который предполагает разбиение наблюдаемого интервала $[0, \tau]$ на k непересекающихся интервалов $I_1, I_2, \ldots, I_k$ граничными точками $0 < a_1 < a_2 < ... < a_k = \tau$. Статистика критерия хи-квадрат для параметрических моделей может быть записана в форме

$$Y_x^2 = \sum_{j=1}^{k} \frac{\left(U_j - e_j\right)^2}{U_j} + Q, \tag{12}$$

$U_j = \sum_{i: X_i \in I_j} \delta_i$ – количество наблюдаемых, и $e_j = \sum_{i=1}^{n} \int_{I_j} \lambda_{x^i}\left(u; \hat{\theta}, \hat{\beta}, \hat{\gamma}\right) Y_i(u) du$ – количество ожидаемых отказов в интервале I_j, $j = 1, ..., k$, $Y_i(u) = 1$, если $t_i \geq u$.

Предельным законом распределения статистики Y_x^2 является χ_r^2-распределение. Проверяемая гипотеза H_0 отвергается при заданном уровне значимости α, если $Y_x^2 > \chi_{r,1-\alpha}^2$, где $r = k - 1$ в случае базового экспоненциального распределения и распределения Вейбулла, и $r = k$ для базового логнормального распределения, гамма-распределения и обобщенных распределений Вейбулла и гамма.

Для проверки предположения о пропорциональности рисков предлагается использовать критерий Никулина, названный по имени одного из авторов, статистика которого может быть записана как

$$T = n^{-1} U^T \hat{D}^{-1} U, \tag{13}$$

где $U=\left[U_1,U_2,...,U_m\right]^T$, $U_k=\sum_{i=1}^{n}\delta_i\left[-x_k^i\ln\left(1+\exp\left(\hat{\beta}^T x^i\right)\right)-\frac{\hat{S}_1(t_i)}{S_0(t_i)}\right]$, $k=1,...,m$,

$\hat{S}_1(t_i)=\sum_{j=1,t_j\geq t_i}^{n}-x_k^j\exp\left(\hat{\beta}^T x^j\right)\ln\left(1+\exp\left(\hat{\beta}^T x^j\right)\hat{\Lambda}_0(t_i)\right)$, $S_0(t_i)=\sum_{j=1,t_j\geq t_i}^{n}\exp\left(\hat{\beta}^T x^j\right)$,

m – размерность вектора параметров модели пропорциональных интенсивностей, $\hat{D}$ – ковариационная матрица вектора U. При справедливости проверяемой гипотезы статистика критерия Никулина подчиняется предельному χ_m^2-распределению.

Во втором разделе методами компьютерного моделирования исследуются статистические свойства ОМП параметров моделей пропорциональных интенсивностей, а также распределения статистик и мощность критериев отношения правдоподобия, Вальда и Никулина при построении полупараметрических моделей.

В результате исследования распределений ОМП параметров модели пропорциональных интенсивностей и её обобщений в условиях неизвестной базовой функции выживаемости показано, что на свойства ОМП параметров β моделей пропорциональных интенсивностей (1) существенное влияние оказывает наличие цензурированных наблюдений в выборке. При ограниченных объемах выборок с ростом степени цензурирования увеличивается смещение оценок, а распределение ОМП становится все более асимметричным. Распределения ОМП параметров β и γ обобщенных моделей пропорциональных интенсивностей Ксая (2) и SCE (3) при ограниченных объемах выборок оказываются далекими от нормального закона даже в случае полных выборок (без цензурирования). При этом в наибольшей степени отклонение от нормального закона наблюдается для ОМП параметров SCE-модели. Отклонение непараметрической оценки неизвестной базовой функции выживаемости от теоретической функции выживаемости уменьшается с ростом объема выборок и не зависит от вида модели.

Для обеспечения проверки гипотез вида $H_0:\mu=\mu^0$ с использованием критерия Вальда получены выражения для вторых частных производных логарифма функции правдоподобия обобщенных моделей пропорциональных интенсивностей по параметрам:

$$\frac{\partial^2\ln L(\mu)}{\partial\mu_l\partial\mu_k}=\sum_{i=1}^{n}\delta_i\left[\frac{1}{g(t_i)}\frac{\partial^2 g(t_i)}{\partial\mu_l\partial\mu_k}-\frac{1}{h(t_i)}\frac{\partial^2 h(t_i)}{\partial\mu_l\partial\mu_k}-\frac{1}{g^2(t_i)}\frac{\partial g(t_i)}{\partial\mu_k}\frac{\partial h(t_i)}{\partial\mu_l}+\right.$$

$$\left.+\frac{1}{h^2(t_i)}\frac{\partial h(t_i)}{\partial\mu_k}\frac{\partial g(t_i)}{\partial\mu_l}\right],\ k,l=1,...,s,$$

где $g(t_i)=g\left(x^i,\Lambda_0(t_i);\mu\right)$, $h(t_i)=\sum_{j:t_j\geq t_i}^{n}g\left(x^j,\Lambda_0(t_i);\mu\right)$, $\mu=\left(\beta^T,\gamma^T\right)^T$ – вектор

параметров модели, s – размерность вектора параметров модели, для модели Кокса $g\left(x,\Lambda_0(t);\mu\right)=\exp\left(\beta^T x\right)$, для модели Ксая $g\left(x,\Lambda_0(t);\mu\right)=\exp\left(\left(\beta+\gamma\right)^T x\right)\left(\Lambda_0(t)\right)^{\exp\left(\gamma^T x\right)-1}$ и для SCE-модели $g\left(x,\Lambda_0(t);\mu\right)=\exp\left(\beta^T x\right)\left(1+\exp\left(\left(\beta+\gamma\right)^T x\right)\Lambda_0(t)\right)^{\exp\left(-\gamma^T x\right)-1}$.

Показано, что отклонение реальных распределений статистики критерия отношения правдоподобия от предельного χ^2-распределения уменьшается с ростом объема выборок. Увеличение числа цензурированных элементов в выборке приводит к увеличению отклонения эмпирического распределения статистики от соответствующего предельного распределения, имеющего место при отсутствии цензурирования. Такая зависимость распределений статистики от степени цензурирования и объема выборки (от относительного количества цензурированных наблюдений в выборке) объясняется отличием в этих условиях реальных свойств ОМП параметров от асимптотических.

Аналогичные результаты, характеризующие отклонение распределений статистики критерия Вальда от соответствующего предельного распределения, получены для модели пропорциональных интенсивностей (1) и модели Ксая (2) с различными объясняющими переменными. При увеличении объема выборок эмпирические распределения статистики критерия Вальда в случае проверки гипотезы о незначимости параметров SCE-модели (3) не сходятся к предельному χ^2-распределению. Более того, действительные распределения статистики, полученные в результате моделирования, существенно отличаются по форме. Данный факт можно объяснить тем, что распределения ОМП параметров SCE-модели плохо приближаются нормальным законом даже при объеме выборок $n=200$.

Методами статистического моделирования исследованы распределения статистик и мощность критериев Вальда, отношения правдоподобия и критерия Никулина со статистикой (13), применяемых для проверки предположения о пропорциональности рисков. Показано, что при справедливости нулевой гипотезы отклонения распределений статистик критерия отношения правдоподобия и критерия Никулина от соответствующих предельных χ^2-распределений уменьшаются с ростом объема выборок. В то же время с ростом степени цензурирования соответствующие отклонения увеличиваются, причём в наибольшей степени в случае критерия отношения правдоподобия.

На основании результатов исследования распределений статистик и мощности критериев отношения правдоподобия, Вальда и Никулина, применяемых для проверки предположения о пропорциональности рисков, при построении полупараметрической модели пропорциональных интенсивностей Кокса целесообразно рекомендовать использование критериев отношения правдоподобия и Никулина.

В третьем разделе исследуются распределения статистик критериев Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга и критерия типа хи-квадрат, применяемых для проверки гипотез о виде параметрических обобщенных моделей пропорциональных интенсивностей. Проводится сравнительный анализ мощности этих критериев.

Методами статистического моделирования показано, что с ростом объема выборок уменьшается отклонение распределений статистики критерия типа хи-квадрат (12) от соответствующего предельного χ^2-распределения при проверке гипотезы о виде параметрической модели пропорциональных интенсивностей в случае полных и цензурированных данных.

Показано, что в случае полных данных (без цензурирования) распределения $G(S \mid H_0)$ статистик непараметрических критериев согласия, применяемых при проверке гипотезы о виде параметрической модели пропорциональных интенсивностей Кокса к выборкам остатков, не зависят от размерности вектора ковариат и плана эксперимента. Отсюда следует, что для вычисления достигнутого уровня значимости в таких ситуациях можно использовать имеющиеся аппроксимации предельных распределений статистик непараметрических критериев согласия, предназначенные для использования в случае проверки различных сложных гипотез о виде распределения (для выборок без ковариат).

Показано, что распределения статистик модифицированных критериев Колмогорова (9), Крамера-Мизеса-Смирнова (10) и Андерсона-Дарлинга (11) зависят не только от объема выборок и степени цензурирования, но и от схемы цензурирования, а также от закона распределения моментов цензурирования (для случайно цензурированных выборок). Влияние вышеописанных факторов на распределения статистик модифицированных критериев согласия можно объяснить, в том числе тем, что выборка остатков, по которой проверяется гипотеза о виде модели как для цензурированных данных I-го или II-го типа, так и для случайно цензурированных данных, оказывается многократно цензурированной. Добавим, что значения цензурированных наблюдений и их расположение в вариационном ряду выборки остатков определяется типом, схемой и моделью цензурирования. Применение модифицированных критериев согласия для проверки гипотезы о виде параметрической модели выживаемости требует знания распределений их статистик при конкретной совокупности факторов, соответствующих анализируемой выборке и влияющих на распределения статистик. Найти требуемые распределения в соответствующих условиях можно только численно, опираясь на статистическое моделирование.

Для моделирования требуемых распределений статистик модифицированных критериев Колмогорова (9), Крамера-Мизеса-Смирнова (10) и Андерсона-Дарлинга (11), необходимых для применения этих критериев к анализу рассматриваемых моделей выживаемости, был разработан алгоритм, описываемый следующей последовательностью действий:

1) задать параметры моделирования:

− объем генерируемых выборок n,

− количество генерируемых выборок N,

– параметрическую модель, соответствующую проверяемой гипотезе H_0 (в качестве значений параметров θ, β и γ берутся ОМП, полученные по той выборке, относительно которой проверяется гипотеза),
– если необходимо – параметрическую модель, соответствующую конкурирующей гипотезе H_1;

2) смоделировать полную выборку с ковариатами в соответствии с моделью, соответствующей H_0 (или H_1 при вычислении мощности критериев);

3) цензурировать выборку, полученную на предыдущем шаге;

4) найти оценки максимального правдоподобия $\hat{\theta}, \hat{\beta}, \hat{\gamma}$ параметров модели, соответствующей гипотезе H_0;

5) вычислить остатки (8) и значения статистик (9), (10), (11);

6) пункты 2-5 повторить N раз.

Для использования описанного алгоритма моделирования распределений статистик для случайно цензурированных выборок необходимо определить, каким образом будет производиться моделирование случайно цензурированных выборок. Предположим, что распределение моментов цензурирования не зависит от параметров распределения моментов отказов. Тем не менее, распределение моментов цензурирования может зависеть от значений ковариат. Чтобы учесть это, для моделирования моментов цензурирования $C_1, C_2, ..., C_n$ необходимо подбирать параметрическую модель, описывающую распределение моментов цензурирования $F_x^C(t)$ в зависимости от значений ковариат. Однако подбор некоторой параметрической модели для моментов цензурирования не представляется разумным, поскольку эта задача вновь сопряжена с проблемой проверки гипотезы о виде модели по случайно цензурированной выборке. Поэтому был разработан непараметрический способ моделирования случайно цензурированных данных на основе оценки базовой функции риска по цензурированным наблюдениям.

Пусть вид зависимости моментов цензурирования от ковариат соответствует модели пропорциональных интенсивностей. То есть при проверке гипотезы о виде параметрической модели выживаемости для моделирования случайно цензурированных данных мы строим полупараметрическую модель Кокса (1) по исходным данным, в которых цензурированные наблюдения рассматриваются как полные и наоборот.

Таким образом, непараметрическая оценка функции риска для моментов цензурирования имеет вид

$$\hat{\Lambda}_0^C(t) = \sum_{i=1, t \ge t_i}^{n} \left[(1-\delta_i) \Big/ \sum_{j=1, t_j \ge t_i}^{n} \exp\left(\left(\hat{\beta}^C \right)^T x \right) \right],$$

где $\hat{\beta}^C$ – ОМП параметров модели, полученные по исходным данным, в которых цензурированные наблюдения рассматриваются как полные и наоборот.

Предложенный алгоритм моделирования случайно цензурированной выборки, соответствующей модели пропорциональных интенсивностей Кокса (1), имеет следующий вид:

1) моделируются значения $T_i = F_x^{-1}\left(\xi_i;\hat{\theta},\hat{\beta},\hat{\gamma}\right)$, где $\xi_i \succ Rav[0,1]$, $i=1,...,n$;

2) вычисляется значение $\Lambda_0^i = -\ln\left(\vartheta_i\right) \Big/ \exp\left(\left(\hat{\beta}^C\right)^T x^i\right)$, где $\hat{\beta}^C$ – ОМП, полученные по исходным данным, в которых цензурированные наблюдения рассматриваются как полные и наоборот, а $\vartheta_i \succ Rav[0,1]$, $i=1,...,n$;

3) моделируются значения C_i, $i=1,...,n$:

– если $\hat{\Lambda}_0^C\left(c_k\right) \le \Lambda_0^i \le \hat{\Lambda}_0^C\left(c_{k+1}\right)$, то $C_i = c_k + \dfrac{\left(c_{k+1}-c_k\right)\cdot\left(\Lambda_0^i - \hat{\Lambda}_0^C\left(c_k\right)\right)}{\left(\hat{\Lambda}_0^C\left(c_{k+1}\right)-\hat{\Lambda}_0^C\left(c_k\right)\right)}$,

где c_k – упорядоченные по возрастанию значения моментов цензурирования в исходных данных, $k=1,...,r-1$, а r – количество цензурированных наблюдений;

– если $\Lambda_0^i \le \hat{\Lambda}_0^C\left(c_1\right)$ или $\Lambda_0^i \ge \hat{\Lambda}_0^C\left(c_r\right)$, то $C_i = \dfrac{c_1 \cdot \Lambda_0^i}{\hat{\Lambda}_0^C\left(c_1\right)}$ или $C_i = c_r + c_r\left(\Lambda_0^i - \hat{\Lambda}_0^C\left(c_r\right)\right)$, соответственно;

4) вычисляются значения $t_i = \min\left(T_i, C_i\right)$, $\delta_i = 1\left\{T_i \le C_i\right\}$, $i=1,...,n$.

Применение предложенного алгоритма моделирования случайно цензурированных выборок делает возможным получение корректного результата проверки гипотез о виде параметрических моделей выживаемости путем вычисления достигнутого уровня значимости по смоделированному распределению статистики модифицированного критерия. При проверке гипотезы о виде модели по случайно цензурированной выборке замена пунктов 2 и 3 алгоритма моделирования распределений статистик на алгоритм моделирования случайно цензурированных выборок приведет к получению корректного результата проверки гипотезы.

В этом же разделе сформулирована методика применения классических критериев типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга к псевдополным выборкам.

Для каждого цензурированного наблюдения $(t_i = C_i, \delta_i = 0)$ методом обратной функции моделируется значение $\hat{T}_i = F_{x^i}^{-1}\left(\xi_i;\hat{\theta},\hat{\beta},\hat{\gamma}\right)$, где $\xi_i \succ Rav\left[F_{x^i}\left(C_i;\hat{\theta},\hat{\beta},\hat{\gamma}\right),1\right]$. Соответствующие цензурированные наблюдения в выборке заменяются на смоделированные. Таким образом, мы получаем псевдополную выборку $\left(\hat{t}_1,x^1\right),\left(\hat{t}_2,x^2\right),...,\left(\hat{t}_n,x^n\right)$, в которой $\hat{t}_i = T_i$, если $\delta_i = 1$, и $\hat{t}_i = \hat{T}_i$, если $\delta_i = 0$, $i=1,...,n$. Смоделированные значения $\hat{T}_i$ принадлежат

параметрической модели, соответствующей проверяемой гипотезе H_0, со значениями параметров $\hat{\theta}$, $\hat{\beta}$, $\hat{\gamma}$, которые являются ОМП, полученными по исходной цензурированной выборке.

После замены значений цензурированных наблюдений смоделированными значениями $\hat{t}_i$, необходимо вновь оценить параметры по полученной выборке без цензурированных наблюдений $\left(\hat{t}_1, x^1\right), \left(\hat{t}_2, x^2\right), ..., \left(\hat{t}_n, x^n\right)$. Затем вычислить значение статистики классического критерия на основе псевдополной выборки и параметрической модели с переоцененными параметрами.

Показано, что в случае малой степени цензурирования (менее 20% цензурированных наблюдений) преобразование цензурированной выборки в псевдополную позволяет использовать классические критерии согласия и вычислять достигнутые уровни значимости, используя соответствующие аппроксимации предельных распределений, полученные для полных данных без ковариат.

Методами статистического моделирования проведен сравнительный анализ мощности рассматриваемых критериев проверки гипотезы о виде параметрической модели пропорциональных интенсивностей относительно близких конкурирующих гипотез о базовом распределении и обобщенных моделях. В ходе анализа показано, что среди рассмотренных критериев наиболее предпочтительными по мощности являются модифицированные критерии типа Крамера-Мизеса-Смирнова и Андерсона-Дарлинга. Для проверки гипотезы о виде параметрической модели пропорциональных интенсивностей по выборкам с малой степенью цензурирования (менее 20% цензурированных наблюдений) целесообразно рекомендовать использование классических критериев Крамера-Мизеса-Смирнова и Андерсона-Дарлинга, опирающихся на преобразование цензурированной выборки в псевдополную.

Следует отметить, что для построения моделей надежности и выживаемости и осуществления корректных выводов с использованием рассматриваемых критериев необходимо программное обеспечение, позволяющее в интерактивном режиме моделировать распределения статистик применяемых критериев.

Четвертый раздел содержит краткое описание разработанного программного обеспечения и описание основных этапов статистического анализа выживаемости пациентов гематологического центра ФГКУ «Главного военного клинического госпиталя имени академика Н.Н. Бурденко», основанного на построении вероятностных моделей выживаемости.

Приводится описание разработанного программного обеспечения, опирающегося, в том числе, на результаты исследований, представленных в предыдущих разделах, и позволяющего осуществлять построение моделей пропорциональных интенсивностей и проверку их адекватности. Представлено описание существующих наиболее популярных программных продуктов «STATISTICA», «SPSS Statistics» и программной системы на основе языка программирования R. Показано, что послужило основанием для разработки

программной системы статистического анализа данных типа времени жизни «LiTiS», а также описаны основные этапы ее разработки, зарегистрированные версии и личный вклад автора.

Подробное описание графического интерфейса пользователя (основных частей рабочего окна, панели инструментов, реализованных инструментов моделирования и мастера анализа, реализованного на основе результатов исследований, изложенных в разделах 2 и 3) позволяет сформировать представление о функциональных возможностях «LiTiS». Модули программной системы обеспечивают решение задач построения обобщенных моделей пропорциональных интенсивностей по цензурированным выборкам и исследования статистических закономерностей при проверке гипотез относительно рассматриваемых моделей. Каждая версия программной системы имеет пользовательский интерфейс и зарегистрирована как программа для ЭВМ (номера государственной регистрации 2012618138, 2012618143 и 2014661905).

В данной работе решена задача построения вероятностной модели при исследовании выживаемости пациентов с множественной миеломой в гематологическом центре ФГКУ «Главного военного клинического госпиталя имени академика Н.Н. Бурденко». Целью исследования являлось сравнение времени ответа на лечение в двух группах пациентов. Лечение первой группы заключалось в химиотерапии с велкейдом (бортезомибом), тогда как вторая группа подвергалась обычной химиотерапии.

На основе непараметрических оценок Каплана-Мейера были сформулированы предварительные выводы о распределении времени достижения ответа на лечение. Например, показано, что вероятность получения общего ответа на лечение, который заключает в себе и ремиссию, и частичный ответ (улучшение некоторых показателей), и стабилизацию, выше вероятности прогрессирования заболевания в любой момент времени.

Для оценки эффективности применения препарата велкейд в зависимости от схемы химиотерапии на основе информационного критерия Акаике предложено использование параметрической SCE-модели с базовым логнормальным распределением. Достигнутые уровни значимости для модифицированных критериев типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга позволяют утверждать, что гипотеза о виде построенной SCE-модели не отвергается при заданном уровне значимости $\alpha = 0.1$.

В ходе исследования, проведенного с использованием разработанного программного обеспечения построения обобщенных моделей пропорциональных интенсивностей и алгоритма моделирования распределений статистик модифицированных критериев проверки гипотезы о виде модели, показано, что время получения ответа на лечение в группе пациентов, получавших химиотерапию с велкейдом, меньше, чем аналогичное время в контрольной группе пациентов. В то же время показано, что использование данного препарата (независимо от дозы) связано с распространённостью и выраженностью полинейропатии в виде чувствительных расстройств.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В соответствии с поставленными задачами исследования получены следующие основные результаты:

1. Для задачи проверки гипотезы о равенстве нулю регрессионных параметров модели пропорциональных интенсивностей и ее обобщений в условиях неизвестной базовой функции выживаемости по цензурированным выборкам показано, что с ростом объема выборок уменьшается отклонение распределений $G(S \mid H_0)$ статистики отношения правдоподобия и статистики Никулина от соответствующих предельных χ^2-распределений.

2. Показано, что в случае проверки гипотезы о равенстве нулю регрессионных параметров модели пропорциональных интенсивностей и модели Ксая распределения статистики Вальда при справедливости нулевой гипотезы сходятся к соответствующему предельному χ^2-распределению, однако в случае SCE-модели распределения статистики критерия Вальда существенно отличаются от предельного χ^2-распределения при всех рассмотренных объемах выборок.

3. Показано, что для проверки гипотезы о выполнении предположения пропорциональности интенсивностей при построении полупараметрической модели пропорциональных интенсивностей Кокса более предпочтительно использование критерия отношения правдоподобия и критерия Никулина.

4. Впервые показано, что в случае полных данных (без цензурирования) распределения $G(S \mid H_0)$ статистик критериев согласия, применяемых к выборкам остатков при проверке гипотезы о виде параметрической модели пропорциональных интенсивностей Кокса, не зависят от размерности вектора ковариат и плана эксперимента. Следовательно, для проверки гипотезы о виде параметрической модели Кокса по выборкам остатков можно использовать классические критерии согласия для выборок без ковариат.

5. Разработан алгоритм моделирования распределений статистик модифицированных критериев согласия типа Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для цензурированных выборок, в том числе случайно цензурированных выборок с неизвестным распределением моментов цензурирования, использование которого позволяет корректно оценивать достигнутый уровень значимости при проверке гипотезы о виде параметрической модели пропорциональных интенсивностей и её обобщений по выборкам остатков.

6. Предложен метод проверки гипотез о виде параметрических моделей выживаемости, опирающийся на преобразование исходной цензурированной выборки в псевдополную. Показано, что в случае малой степени цензурирования (не более 20%) применение имеющихся результатов для полных выборок не приведет к существенным ошибкам в оценивании достигнутого уровня значимости при проверке гипотез по псевдополным выборкам.

7. В результате сравнительного анализа мощности критериев, используемых для проверки гипотезы о виде параметрической модели пропорциональных интенсивностей, показано, что в большинстве рассмотренных случаев модифицированные критерии типа Крамера-Мизеса-Смирнова и Андерсона-Дарлинга являются наиболее предпочтительными.

8. На базе программной системы статистического анализа данных типа времени жизни «LiTiS» разработано программное обеспечение построения вероятностных моделей выживаемости по цензурированным выборкам. Различные версии данной системы были зарегистрированы в виде объектов интеллектуальной собственности как программы для ЭВМ.

Результаты проведенных исследований и разработанное программное обеспечение были внедрены в практику деятельности гематологического центра ФГКУ «Главного военного клинического госпиталя имени академика Н.Н. Бурденко», использованы в научно-исследовательской деятельности кафедры проектирования технологических машин ФГБОУ ВО «Новосибирский государственный технический университет» и в учебном процессе на механико-технологическом факультете, а также нашли практическое применение в учебном процессе на факультете прикладной математики и информатики ФГБОУ ВО «Новосибирский государственный технический университет», что подтверждается соответствующими справками о внедрении.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Издания из Перечня ВАК ведущих рецензируемых научных изданий для опубликования основных научных результатов диссертаций:

1. Лемешко Б.Ю., Чимитова Е.В., Ведерникова М.А. Модифицированные критерии согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлинга для случайно цензурированных выборок. Ч.2 // Научный вестник НГТУ. – 2013. – № 1(50). – С. 3-16.

2. Balakrishnan N., Chimitova E., Galanova N., Vedernikova M.: Testing goodness-of-fit of parametric AFT and PH models with residuals // Communication in Statistics – Simulation and Computation. – 2013. – Vol. 42. No. 6. – P. 1352-1367.

3. Чимитова Е.В., Ведерникова М.А., Галанова Н.С. Непараметрические критерии согласия в задачах проверки адекватности моделей надежности по цензурированным данным // Вестник ТГУ. Управление, вычислительная техника и информатика. – 2013. – № 4(25). – С. 115-124.

4. Balakrishnan N., Chimitova E., Vedernikova M.: An empirical analysis of some nonparametric goodness-of-fit tests for censored data // Communication in Statistics – Simulation and Computation. – 2014. – Vol. 44. No. 4. – P. 1101-1115.

5. Чимитова Е.В., Семёнова М.А. Проверка адекватности параметрических моделей надежности по усеченным слева и цензурированным справа данным // Доклады академии наук высшей школы Российской Федерации. – 2015. – № 1(26). – С. 104-120.

Международные рецензируемые издания:

6. Semenova M., Chimitova E., Rukavitsyn O., Bitukov A. Models with cross-effect of survival functions in the analysis of patients with multiple myeloma // Topics in Statistical Simulation. Springer Proceedings in Mathematics & Statistics. – 2014. – Vol. 114. – P. 457-463.

7. Semenova M., Chimitova E. Testing statistical hypotheses for generalized semiparametric proportional hazards models with cross-effect of survival functions // Advanced Mathematical and Computational Tools in Metrology and Testing X. Series on Advances in Mathematics for Applied Sciences. – 2015. – Vol. 86. – P. 350-358.

Сборники научных трудов:

8. Чимитова Е.В., Ведерникова М.А. Проверка адекватности модели пропорциональных интенсивностей Кокса по случайно цензурированным выборкам // Сборник научных трудов НГТУ. – Новосибирск: Изд-во НГТУ, 2010. – №4(62). – С. 103-108.

9. Ведерникова М.А. Критерии согласия в задачах проверки адекватности параметрических моделей надежности и выживаемости // Международный конкурс научных работ по приоритетным направлениям развития науки, технологий и техники в Российской Федерации. Сборник трудов. М.: НИИ Радиоэлектроники и лазерной техники, 2012. – С. 7-14.

Другие издания:

10. Ремнева М.М., Ведерникова М.А. Исследование вопросов построения моделей пропорциональных интенсивностей Кокса и проверки их адекватности: материалы конференции // Российская научно-техническая конференция «Информатика и проблемы телекоммуникаций». – Новосибирск, 2010. – Т. 1. – С. 87-90.

11. Chimitova E., Chuyanova E., Galanova N., Vedernikova M. Computer approach to the choice of parametric ALT-models // Proceedings of the Third International Conference «Accelerated Life Testing, Reliability-based Analysis and Design». – Clermont-Ferrand, 2010. – P. 111-116.

12. Chimitova E., Remneva M., Vedernikova M. Computer approach in constructing survival regression models and testing adequacy // Proceedings of the Ninth International Conference «Computer Data Analysis and Modeling: Complex Stochastic Data and Systems». – Minsk: Publ.center of BSU, 2010. – P. 100-103.

13. Ведерникова М.А. Исследование критериев проверки гипотезы о согласии с моделью пропорциональных интенсивностей Кокса // Материалы всероссийской научной конференции молодых ученых «Наука. Технология. Инновации». – Новосибирск: Изд-во НГТУ, 2010. – Ч. 1. – С. 9-11.

14. Ведерникова М.А., Чимитова Е.В. О проверке предположения пропорциональности рисков в полупараметрической модели Кокса // Российская научно-технической конференция «Информатика и проблемы телекоммуникаций». – Новосибирск, 2011. – Т. 1. – С. 51-53.

15. Chimitova E., Liero H., Vedernikova M. Application of classical Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests for censored samples // Proceedings of the International Workshop «Applied Methods of Statistical Analysis. Simulations and Statistical Inference». – Novosibirsk: Publ. house of NSTU, 2011. – P. 176-185.

16. Чимитова Е.В., Ведерникова М.А. Компьютерный подход к проверке статистической гипотезы о согласии по цензурированным данным // Сборник научных трудов по материалам международной научно-практической конференции «Перспективные инновации в науке, образовании, производстве и транспорте ‘2011». – Одесса: Черноморье, 2011. – Т. 6. – С. 18-27.

17. Ведерникова М.А., Чимитова Е.В. Вопросы проверки адекватности модели пропорциональных интенсивностей Кокса // Материалы Российской научно-технической конференции «Обработка информационных сигналов и математическое моделирование». – Новосибирск, 2012. – С. 16-18.

18. Чимитова Е.В., Ведерникова М.А. Критерии согласия в задачах проверки адекватности параметрических моделей надежности и выживаемости // Материалы XI международной конференции «Актуальные проблемы электронного приборостроения». – Новосибирск, 2012. – Т. 6, С. 58-65.

19. Ведерникова М.А., Дёмин В.А., Чимитова Е.В. Критерии согласия в задачах проверки адекватности параметрических моделей надежности и выживаемости // Материалы Российской научно-технической конференции «Обработка информационных сигналов и математическое моделирование». – Новосибирск, 2013. – С. 38-40.

20. Semenova M., Bitukov A. Parametric models in the analysis of patients with multiple myeloma // Proceedings of the International Workshop «Applied Methods of Statistical Analysis. Applications in Survival Analysis, Reliability and Quality Control». – Novosibirsk: Publ. house of NSTU, 2013. – P.250-256.

21. Семёнова М.А., Чимитова Е.В. Исследование распределений статистики критерия хи-квадрат для параметрических обобщенных моделей пропорциональных интенсивностей // Материалы XII международной конференции «Актуальные проблемы электронного приборостроения». – Новосибирск, 2014. – Т. 6, С. 52-58.

Свидетельства о государственной регистрации программы для ЭВМ:

22. Чимитова Е.В., Румянцев А.В., Ведерникова М.А., Галанова Н.С. Система статистического анализа данных типа времени жизни «LiTiS 1.0» // Свидетельство о государственной регистрации программы для ЭВМ № 2012618138. – М.: Роспатент. – 2012.

23. Чимитова Е.В., Румянцев А.В., Ведерникова М.А., Галанова Н.С. Система статистического анализа данных типа времени жизни «LiTiS 1.1» // Свидетельство о государственной регистрации программы для ЭВМ № 2012618143. – М.: Роспатент. – 2012.

24. Чимитова Е.В., Румянцев А.В., Семёнова М.А., Галанова Н.С. Демин В.А. Система статистического анализа данных типа времени жизни «LiTiS 1.2» // Свидетельство о государственной регистрации программы для ЭВМ № 2014661905. – М.: Роспатент. – 2015.

Подписано в печать 7.07.2015 г. Формат 60 х 84 х 1/16
Бумага офсетная. Тираж 100 экз. Печ. л. 1.5.
Заказ № 1123

Отпечатано в типографии
Новосибирского государственного технического университета
630073, г. Новосибирск, пр-т К. Маркса, 20