

**АКАДЕМИЯ НАУК РЕСПУБЛИКИ МОЛДОВА
ИНСТИТУТ МАТЕМАТИКИ И ИНФОРМАТИКИ**

На правах рукописи
У.Д.К. 004.415.2

ПОПУКАЙЛО ВЛАДИМИР СЕРГЕЕВИЧ

**ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ
ПО ПАССИВНЫМ ВЫБОРКАМ
МАЛОГО ОБЪЁМА**

**122.03 –МОДЕЛИРОВАНИЕ, МАТЕМАТИЧЕСКИЕ МЕТОДЫ И
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ**

**Автореферат диссертации на соискание ученой степени
доктора информатики**

КИШИНЁВ, 2017

Работа выполнена в лаборатории системного программирования Института Математики и Информатики Академии Наук Молдовы

Научный руководитель:

Кожокару Светлана, доктор хабилитат в информатике, профессор, Институт Математики и Информатики Академии Наук Молдовы

Официальные оппоненты:

1. Косташи Илие, доктор хабилитат в информатике, профессор, Молдавская Экономическая Академия.
2. Кэпэцине Георге, доктор технических наук, профессор, Государственный Университет Молдовы.

Состав Специализированного Ученого Совета:

1. Гайндрик Константин, председатель, доктор хабилитат в информатике, член-корреспондент АН Молдовы, профессор, Институт Математики и Информатики АН Молдовы.
2. Цицкиев Инга, научный секретарь, доктор в информатике, Институт Математики и Информатики АН Молдовы.
3. Мишкой Георге, доктор хабилитат физико-математических наук, академик АН Молдовы, профессор, Институт Математики и Информатики АН Молдовы и Международный Независимый Университет Молдовы.
4. Катаранчук Сергей, доктор хабилитат физико-математических наук, профессор, Государственный Университет Молдовы.
5. Палагин Владимир, доктор технических наук, профессор, Черкасский Государственный Технологический Университет, Украина.
6. Корлат Андрей, доктор физико-математических наук, доцент, Институт Математики и Информатики АН Молдовы.

Защита диссертации состоится 22 сентября 2017, в 15:00 часов, на заседании Специализированного Ученого Совета D 01.122.03-04 при Институте Математики и информатики АН Молдовы по адресу: ауд. 340, ул. Академическая, 5, г. Кишинёв, Республика Молдова, MD-2028.

С диссертацией и авторефератом можно ознакомиться в Центральной научной библиотеке АНМ «Андрей Лупан», а также на сайте С.Н.А.А. (www.cnaa.md).

Автореферат разослан 12.07.2017 г.

Ученый секретарь Специализированного Ученого Совета

Цицкиев Инга, доктор в информатике, доцент _____

Научный руководитель

Кожокару Светлана, доктор хабилитат в информатике, профессор _____

Автор

Попукайло Владимир _____

© Попукайло Владимир, 2017

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. В современном мире большое развитие получили системы поддержки принятия решений (СППР) – интерактивные компьютерные автоматизированные системы, помогающие лицу, принимающему решения (ЛПР) при анализе различных трудно формализуемых задач.

При этом, универсальные решения, как правило, ориентированы на математиков и инженеров, что затрудняет их использование специалистами в различных областях. В связи с этим, широкое распространение получили проблемно-ориентированные СППР, которые разрабатываются для конкретной прикладной задачи и настраиваются под индивидуальные особенности каждой проблемной ситуации.

При анализе сложных систем применяют вероятностно-статистический подход на основе выборочных методов, который расширяет возможности аналитических методов.

Тем не менее, на практике существуют такие области, в которых характеристики исследуемых объектов требуют анализа небольших объемов экспериментальных данных, обработка которых классическими вероятностно-статистическими методами зачастую не даёт приемлемых результатов.

Хотя задача статистической обработки малых выборок имеет одинаковую природу, в различных областях она решается различными способами, с учетом специфики рассматриваемых систем. Кроме того, остаётся нерешенной задача многомерного анализа таблиц пассивного эксперимента. Таким образом, актуальна задача разработки подходов к поддержке принятия решений по пассивным выборкам малого объема, которые будут широко применимы для различных классов задач.

Цель и объекты исследования.

Объект исследования – пассивные выборки малого объема, полученные в различных условиях.

Цель исследования. Разработать и обосновать специфические методы и приёмы, используемые в системах поддержки принятия решений при статистической обработке пассивных выборок малого объема, позволяющие получить более точную и объективную информацию об исследуемом процессе.

Для реализации поставленной цели были сформулированы следующие **задачи исследования:**

1) провести анализ различных подходов и методов нахождения грубых промахов в одномерных выборках малого объёма;

2) исследовать возможности использования метода точечных распределений для обработки выборок малого объема;

3) разработать методику применения возможностей метода точечных распределений для случая многомерного моделирования по пассивным данным малого объема;

4) используя разработанную методику,

– построить адекватные математические модели;

– численно оценить качество построенных моделей.

5) проанализировать способность коэффициентов корреляции верно определить уровень линейной корреляционной связи в случае малого числа наблюдений;

6) проверить разработанную методику на данных, полученных в различных областях исследований;

7) разработать алгоритмы и программное обеспечение, реализующее трудоёмкие методики;

8) разработать системы поддержки принятия решений для ряда конкретных прикладных областей, в которых получение данных большого объема не представляется возможным.

Теоретико-методологической основой исследования являются труды авторов по системному анализу, системам поддержки принятия решения, математической статистике и математическому моделированию. Поставленные в исследовании задачи разрабатывались, опираясь на труды Гаскарова Д.В., Шаповалова В.С., Долгова Ю.А., Столяренко Ю.А., Кривенко М.П., Большева Л.Н., Шора Я.Б., Кобзаря А.И., Орлова А.И., Pearson E.S., Efron B.M., Fisher R.A., de Winter J.C.F. и др.

Методы исследования: анализ научно-методической литературы и практического опыта, систематизация ранее полученных результатов по проблеме исследования, сравнение существующих подходов к решению задачи, методы теории принятия решений, системного анализа, теории вероятностей и математической статистики, моделирование, синтез существующих вероятностно-статистических подходов и методов.

Научная новизна исследования:

1) Выявлены надёжные критерии определения грубых промахов и их возможности при определении аномального измерения в выборках малого объёма, предложена методика определения грубых промахов на основе синтеза существующих подходов.

2) Даны рекомендации по проведению линейного корреляционного анализа многомерных выборок малого объёма.

3) Разработана методика построения математических моделей, основанная на алгоритме виртуального увеличения многомерной выборки, используя двумерные таблицы плотности распределения, получаемые методом точечных распределений.

4) Предложены рекомендации по созданию систем поддержки принятия решений по данным малого объема, на их основе разработаны такие системы для ряда областей.

Важная научная проблема, решенная в работе: определение вероятностно-статистических методик поддержки принятия решений на основании выборок малого объема, полученных в ходе пассивного эксперимента.

Теоретическая значимость исследования заключается в методике улучшения качества принимаемых решений на основе предложенных подходов и алгоритмов, основанных на вероятностно-статистических методах обработки информации.

Практическая значимость результатов работы. Предложенные в диссертационной работе методики поддержки принятия решения могут быть широко применены при статистическом анализе в различных областях исследований.

Значимость результатов заключается в следующих элементах:

1) отобраны критерии, наиболее точно определяющие возможное аномальное измерение в одномерной выборке малого объема;

2) разработаны рекомендации по проведению линейного корреляционного анализа таблиц пассивного эксперимента в случае малого количества исходных данных;

3) создана и опробована методика построения математических моделей, основанная на виртуальном увеличении объема многомерной выборки малого объема, позволяющая строить математические модели, в том числе для сверхнасыщенных планов;

4) разработаны алгоритмы и программное обеспечение, реализующее данные методики;

5) разработаны СППР для конкретных прикладных областей, в которых невозможно получить данные большого объема.

Достоверность и обоснованность результатов работы обеспечиваются использованием методов исследования, адекватных целям и задачам работы, репрезентативностью экспериментальных данных, корректностью их обработки с применением методов математической статистики.

Внедрение результатов работы. Результаты представленной работы внедрены в учебный процесс кафедр «Биологии и физиологии человека» и «Терапии №2» медицинского факультета Университета им. Т.Г. Шевченко (г. Тирасполь), в работу научно-исследовательской лаборатории «Математическое моделирование» Университета

им. Т.Г. Шевченко (г. Тирасполь), в учебную программу магистратуры по направлению «Информатика и вычислительная техника» того же университета.

Апробация результатов работы. Результаты исследования были представлены на заседании кафедры «Информационных технологий и автоматизированного управления производственными процессами» Университета им. Т.Г. Шевченко (г. Тирасполь), семинаре лабораторий «Системы программирования» и «Информационные системы» Института математики и информатики Академии Наук Республики Молдова, а также на следующих конференциях:

– VIII Международная конференция «Математическое моделирование в образовании, науке и производстве», Тирасполь, 2013;

– 5th International Conference ICTEI-2015, Chisinau, 2015;

– 23-a Conferință de Matematică Aplicată și Industrială CAIM 2015– CAIM, ROMANIA, Suceava, 2015;

– IX Международная конференция «Математическое моделирование в образовании, науке и производстве», г. Тирасполь, 2015;

– Всероссийская научно-практическая конференция «Информационно-телекоммуникационные системы и технологии», РФ, Кемерово, 2015;

– Международная научная конференция докторантов «Тенденции развития современной науки: видения молодых исследователей», Кишинёв, 2016.

– The Fourth Conference of Mathematical Society of the Republic of Moldova dedicated to the centenary of Vladimir Andrunachievici (1917-1997), Кишинёв, 2017.

Основные положения, выносимые на защиту:

1) Методика нахождения аномальных значений в выборках малого объема, позволяющая наиболее точно проводить процедуру анализа пассивных данных на наличие грубых промахов.

2) Методика, основанная на виртуальном увеличении объема выборки методом точечных распределений, позволяющая строить адекватные математические модели по пассивным выборкам малого объема.

3) Методика обнаружения величины линейной корреляционной связи по выборкам малого объема, позволяющая объективно определять сильную корреляционную связь не ниже значения коэффициента корреляции $r=0,8$; процедуры виртуального статистического увеличения информации («*bootstrapping*») ухудшают точность проводимого корреляционного анализа, увеличивая разброс получаемых оценок.

По теме диссертационного исследования опубликовано 16 работ.

Структура и содержание диссертации. Работа состоит из введения, трех глав, заключения, выводов, списка литературы из 160 наименований, 4 приложений, 23 таблиц, 30 рисунков.

Ключевые слова: теория принятия решений, системы поддержки принятия решений, малая выборка, пассивный эксперимент, корреляционный анализ, регрессионный анализ, грубые промахи, метод точечных распределений.

Содержание диссертационной работы. Классические вероятностно-статистические методы поддержки принятия решений не всегда дают приемлемые результаты при обработке данных малого объема. Поэтому целью исследования является разработка и обоснование специфических методов и приёмов статистической обработки пассивных выборок малого объема, которые позволят получать более объективную информацию для принятия решений.

Во введении обоснована актуальность темы, определён объект исследования, сформулированы цель, и задачи исследования, определена теоретико-методологическая основа диссертационной работы, определены методы исследования, раскрыта научная новизна, теоретическая и практическая значимость исследования, сформулированы основные положения, выносимые на защиту, приведены сведения об апробации и внедрении результатов.

В первой главе «Методологические и теоретические основы поддержки принятия решений в условиях малой выборки» рассматривается проблематика анализа сложных систем и построения СППР.

В первом параграфе рассматриваются методы и средства принятия решений, даётся определение СППР, описываются задачи, решаемые такими системами, приводится их классификация. Показывается, что проблемно-ориентированные СППР широко распространены в различных областях исследований, таких как: медицина, промышленность, экономика, педагогика, биология и других. Также приводится схема принятия решения [25], которая состоит из следующих этапов: описание предметной области, анализ ситуации, формулирование проблемы, анализ имеющейся информации, разработка вариантов решения, оценка применимости вариантов решения и их последствий, выбор оптимального варианта решения, а также проверка и выдача решения. Представленная схема принятия решения не отражает сложности всего процесса. На практике данный процесс может быть итерационным, например оценка применимости вариантов решений и их последствий может быть проведена несколько раз, до момента, когда риски от принятия решений

будут минимальными. Также схема допускает параллельность выполнения отдельных процедур и операций, например, таких как выявление вида данных, описывающих исследуемый процесс и методов их обработки. Эффективность данного процесса во многом зависит от методов, применяемых при выполнении каждого из указанных этапов схемы процесса принятия решений. В параграфе также приводится классификация методов принятия решения. Рассматриваются отличия формальных и неформальных методов, описываются этапы принятия решений с использованием вероятностно-статистического подхода.

Во втором параграфе описываются трудности применения классических вероятностно-статистических методов принятия решения в условиях малой выборки, доказывається актуальность вопроса о необходимом объеме выборки для различных исследований; приводятся примеры областей исследований, в которых характеристики объектов не позволяют получать большие объемы экспериментальных данных. В таком случае ЛПР приходится принимать решение по полученным данным малого объема, что приводит к следующим распространенным проблемам при их оценивании вероятностно-статистическими методами:

- с уменьшением объема выборки увеличиваются доверительные интервалы и вероятностные ошибки [28];

- в условиях малой выборки дисперсия выборочной совокупности не может рассматриваться в качестве оценки генеральной дисперсии [3];

- в соответствии с центральной предельной теоремой [26] предположение о нормальном распределении всех возможных значений справедливо только при значительном объеме выборки;

- классические подходы к оцениванию законов распределения выборок, как параметрические, такие как критерий Пирсона и RS -критерий, так и непараметрические, к примеру, критерии типа Колмогорова-Смирнова и W -критерий Шапиро-Уилка не позволяют с достаточной точностью судить о принадлежности генеральной совокупности к конкретному закону распределения [5, 6];

- результаты обработки данных, полученные при использовании различных критериев, например, таких как Кохрена и Граббса, являются зависимыми от объемов исследуемых выборок [8];

- оценка неизвестных параметров, произведенная путем максимизации функции правдоподобия, может быть сильно смещена для малых выборок;

- применение квантилей распределения Стьюдента и Фишера предполагает принятие гипотезы о нормальности закона распределения [4], при том, что погрешности большинства физических измерений отличны от нормальных [10, 11];

- применение методов Монте-Карло сталкивается с несовершенностью датчиков псевдослучайных чисел, а также бесконечным множеством возможных отклонений случайной величины от изучаемого параметрического семейства распределений [2, 9].

Кроме этого приводятся основные проблемы анализа многомерных выборок малого объема, приводящих к тому, что с ростом размерности данных обычно существенно снижается качество применяемых методов анализа данных.

В третьем параграфе рассмотрен вопрос определения термина «малая выборка» и определения границ малых и средних выборок. Гаскаров и Шаповалов в своей работе «Малая выборка» [1] пишут, что выборку следует считать малой, если при её обработке методами, основанными на группировке наблюдений, нельзя достичь заданной точности и достоверности. Следовательно, при обработке малых выборок следует отказаться от данных методов и использовать информацию о каждой отдельной реализации выборки. Профессор А.И. Орлов в работе [9] утверждает, что: под "малой выборкой" понимают такую выборку, для которой нельзя применять выводы, основанные на предельных теоремах. В каждой конкретной задаче, таким образом, возникает необходимость разделить конечные объемы выборки на два класса – те, для которых можно применять предельные теоремы, и те, для которых делать это нельзя из-за риска получения неверных выводов. Указано, что различные исследователи принимают величину верхней границы малой выборки равной определенному значению в диапазоне от 10 до 30, при этом не опираются ни на какие числовые данные. Следовательно, задача об определении границы между выборками малого и среднего объема является актуальной. Анализ литературы по данной тематике показал, что одним из немногих обоснованных подходов к решению данной проблемы является исследование [23], которое проводилось с помощью статистического имитационного эксперимента на базе *bootstrap*-метода, в котором рассчитывалась нижняя граница интервальной оценки наивероятнейшего коэффициента корреляции. Во втором исследовании по этой тематике [27] проводился анализ объема выборок, основанный на использовании критерия Пирсона. Таким образом, опираясь на результаты двух независимых исследований можно считать установленным, что верхней границей малой выборки является значение $n=20$.

В четвертом параграфе проведен краткий обзор существующих специализированных методов обработки малых выборок, основанных на использовании каждой отдельной

реализации и позволяющих улучшить показатели при оценивании плотности распределения случайной величины, таких как:

1. Метод прямоугольных вкладов;
2. Метод уменьшения неопределенности;
3. Метод сжатия области существования интегральных законов распределения;
4. Метод последовательных медиан;
5. Метод имитационного дополнения;
6. Метод точечных распределений.

Из анализа видно, что большинство таких методов основываются на использовании информации о каждой отдельной реализации и являются развитием метода прямоугольных вкладов. Также приведён полный алгоритм метода точечных распределений [23], в котором в качестве дополнительной информации используется знание о виде закона распределения случайной величины. Приведены интервалы измерения контролируемой величины X и аппроксимационные формулы для расчета вспомогательных коэффициентов по методу точечных распределений для нормального и экспоненциального законов, а также закона распределения Вейбулла.

Пятый параграф посвящен выводам и предложениям по итогам главы. На основании изложенного в главе материала делается вывод, что в настоящее время является актуальным вопрос создания проблемно-ориентированных СППР, предназначенных для обработки многомерных данных малого объёма. Для достижения этой цели необходимо:

1. Разработать специализированную методику поддержки принятия решений для случаев обработки многомерных данных малого объема, основанную на вероятностно-статистическом подходе, а также анализе, синтезе и развитии существующих методов.
2. Определить оптимальные методы принятия решения об обнаружении и устранении грубых промахов в выборках малого объема.
3. Разработать методологию построения виртуальной многомерной таблицы данных по выборкам малого объема на основе метода точечных распределений.
4. Проанализировать и выявить оптимальную методику определения величины линейной корреляционной связи по выборкам малого объема.
5. Проанализировать особенности методов математического моделирования по пассивным данным при работе с выборками малого объема.

Во второй главе «Разработка методик принятия решений при обработке данных малого объема» поднимается вопрос о применимости различных вероятностно-статистических методов для обработки данных малого объема.

В первом параграфе рассматривается проблема нахождения грубых промахов в выборках малого объема. Данная процедура является обязательной для первичной обработки таблиц пассивного эксперимента, так как наличие единственного аномального измерения может привести к большим ошибкам при оценке выборочных моментов и проверке различных статистических гипотез. Проведен анализ существующих методик обнаружения грубых промахов в одномерных выборках, таких как: метод Ирвина, критерий Стьюдента, критерий наибольшего абсолютного отклонения, критерий максимального относительного отклонения, критерий Романовского, метод вариационного размаха, критерий 3 Сигм, критерий Райта, критерий Граббса, Q -критерий (Диксона), критерий Львовского, критерий Шовене, критерий Дэвида, критерий Хоглина-Иглевича, L -критерий (Критерий Титьена-Мура), критерий Смоляка-Титаренко, критерий Бродского-Бацаня-Власенко, критерий Кимбера. На основании проведенного анализа [12, 16] предлагается использовать для данных целей критерии Львовского (в случае нормального закона распределения), Диксона (для различных законов распределения), Ирвина (при $n \geq 10$) или Титьена-Мура (в качестве обладающего наименьшим количеством ошибок I типа), как позволяющих наиболее точно определить наличие аномальных измерений в выборках малого объема. Также рассматривается вопрос о влиянии парных грубых промахов и их нахождении в таблице пассивного эксперимента.

Во втором параграфе анализируются возможности различных подходов для оценки величины линейной корреляционной связи между двумя величинами в случае малого количества наблюдений. Рассматриваются классический подход определения величины линейной корреляционной связи критерием Пирсона, применение *bootstrapping* как метода рандомизации и виртуального увеличения обрабатываемой информации, а также применения метода Чебышева при увеличении количества информации методом точечных распределений. Показывается, что ни один из этих методов не обладает достаточной точностью для применения классического корреляционного анализа в случае выборок малого объема [14]. В частности, применение *bootstrap*-метода увеличивает разброс значений рассчитанного коэффициента корреляции Пирсона по сравнению с классическим методом. А применение методов точечных распределений и Чебышева приводит к невозможности определить высокий уровень линейной зависимости между величинами, как критерием Пирсона, так и модифицированным индексом Фехнера. Рекомендуется при определении сильносвязанных факторов не использовать процедуры рандомизации и применять пороговый коэффициент $r=0,75$, так как значения меньше данного коэффициента не могут указать на сильную линейную корреляционную связь с необходимой точностью. Также

приводится алгоритм выделения главных связей в корреляционной матрице методом корреляционных плед.

В третьем параграфе рассматривается методика построения виртуально увеличенной таблицы пассивного эксперимента на основании величин плотности вероятностей, полученных из таблиц двумерного виртуального распределения методом точечных распределений. При проведении преобразований строятся виртуальные таблицы данных, для каждой строки l таблицы исходных данных, в которые вносятся столбцы X_{ij} и X_{il} из соответствующих таблиц ненормированных плотностей вероятностей. Корреляционный и регрессионный анализ таких таблиц не представляется возможным из-за большого количества пробелов в значениях, в связи с чем не полностью заполненные строки удаляются из рассмотрения, а оставшиеся таблицы объединяются в порядке нумерации строк исходных данных. Также приводится числовой пример расчета описанного алгоритма [29] и доказывається однородность исходной и полученной таблиц данных, на основании неразличимости средних арифметических и дисперсий выборок для каждого из факторов и выходной величины по критериям Стьюдента и Фишера. Обосновывается методика нахождения коэффициента корреляции после виртуального увеличения информации многомерным методом точечных распределений. Показывается, что применение такого подхода позволяет сузить доверительный интервал при оценивании коэффициента корреляции Пирсона по данным малого объема, а также избежать ошибочного определения отрицательной корреляционной связи [13]. В случае обнаружения в выборках $n > 10$ линейных корреляционных связей $r > 0,75$ также следует признать данную наблюдаемую связь сильной и значимой. В данных объемом $10 \geq n \geq 7$ наблюдаемую связь следует признать значимой при $r \geq 0,8$. При объеме исследуемой выборки $n \leq 6$, значение порогового коэффициента корреляции следует принять на уровне $r = 0,9$. Делается вывод, что применение виртуального увеличения таблицы исходных данных предложенным методом, в отличие от методов *bootstrapping*, позволяет сохранить знание о виде закона распределения случайной величины и о величине линейной корреляционной связи между исследуемыми факторами, что позволит принимать более обоснованные решения по малому количеству исходных данных, а также улучшить качество статистического оценивания параметрическими методами.

В четвертом параграфе рассматриваются вопросы оценки качества принимаемых решений на основе вероятностно-статистического подхода; даётся определение понятию адекватности математической модели, под которым понимается: 1) правильное качественное описание рассматриваемых свойств объекта (процесса): например, возможности на

основании исследования модели сделать правильные выводы о направлении изменений каких-либо количественных характеристик этих свойств, об их взаимосвязи, о характере колебаний объекта или процесса, устойчивости его состояния или эволюционировании и т. п. и, в итоге, на основании этих данных принять верное решение; 2) правильное количественное описание рассматриваемых свойств с некоторой разумной точностью.

В параграфе сравнивается качество математических моделей, построенных различными способами. Делается вывод, что применение разработанной методики в ряде случаев помогает получить более качественные результаты математического моделирования, что может позволить исследователю принять более эффективные решения при анализе многомерных данных малого объема.

Пятый параграф посвящен результатам проделанной работы. Отмечается, что: исследована процедура поиска грубых ошибок в выборках малого объёма и найдены критерии лучше всего справляющиеся с данной проблемой, даны рекомендации по проведению процедуры отсеивания грубых промахов; исследованы возможности существующих подходов к нахождению линейной корреляционной связи в выборках малого объема, по результатам работы предложена процедура линейного корреляционного анализа для задач обнаружения сильно связанных факторов и дальнейшего сокращения факторного пространства без существенной потери информации; разработан метод построения виртуальной многомерной выборки, используя метод точечных распределений и стыковку данных по максимальным значениям плотностей вероятностей; показано, что применение данного метода позволяет получить таблицы в более чем 10 раз длиннее исходной, со статистически неразличимыми параметрами распределений исследуемых факторов; проиллюстрировано, что применение данного метода позволяет более точно оценить величину линейной корреляционной связи, в сравнении с процедурами *bootstrapping*-анализа.; проанализировано качество моделей, построенных с использованием разработанного метода, сделан вывод, что применение процедуры виртуального увеличения объема выборки, основанной на методе точечных распределений, в ряде случаев помогает получить более качественные результаты математического моделирования, что может позволить исследователю принять более эффективные решения при анализе многомерных данных малого объема.

Шестой параграф посвящен выводам по итогам главы, в которых говорится, что на основании полученных данных необходимо сформулировать методику построения СППР по пассивным выборкам малого объёма для случаев, когда классические методы не позволяют получить модель необходимого качества по исходным данным. В связи с

трудоёмкостью процесса обработки выборок малого объема (применением метода точечных распределений) и со сложностью составления виртуальной таблицы данных для дальнейшего моделирования необходимо разработать комплекс специализированных компьютерных программ, реализующий предложенные методы. Рекомендуется проверить разработанные методики и разработать СППР для различных областей деятельности исследователей.

В *третьей главе* «Системы поддержки принятия решений по пассивным выборкам малого объема» формируется, реализуется и проверяется методика построения СППР по пассивным выборкам малого объема.

В *первом параграфе*, на основе проанализированной информации и полученных данных, формулируется методика, позволяющая обеспечить математический аппарат поддержки принятия решения при обработке таблиц пассивного эксперимента, в случае малого количества исходных данных.

Приведём данный алгоритм полностью:

1. Отсев грубых промахов. Как показано в параграфе 2.1, в качестве предварительной обработки таблиц пассивного эксперимента, на этапе анализа полученных данных, при наличии априорной информации о виде закона распределения случайных величин, входящих в факторное пространство, при объеме выборки $n \leq 10$, рекомендуется применять критерий Диксона для нахождения аномальных измерений, связанных с усталостью персонала, сбоем оборудования, неполадок в техническом процессе и т.д. В случае отсутствия необходимой априорной информации и невозможностью с необходимой точностью определить вид закона распределения случайной величины, также возможно искать выбросы критерием Львовского. При объеме выборки $n \geq 10$ поиск аномальных результатов также возможно провести одним из следующих критериев:

1. Метод Ирвина.
2. Критерий Львовского.
3. Критерий Титьена-Мура.
4. Критерий Диксона (для нормального распределения).

Кроме того, как показано в теоретической части исследования, применение критерия Титьена-Мура обладает меньшим количеством ошибок I типа, поэтому его использование рекомендуется для подтверждения гипотезы о наличии грубой ошибки в данных малого объема.

В случае обнаружения грубой ошибки в одной из исследуемых величин, из исходной таблицы пассивных многомерных данных удаляется вся строка, содержащая аномальное измерение. Если удалить строку не предоставляется возможным, ввиду малого количества исходной информации, значение можно заменить на близкое по статистическим характеристикам к основной группе данных.

2. Последующая обработка таблицы ставит целью обнаружение двумерных парных выбросов. С этой целью необходимо построить таблицы двумерного распределения для каждой из пар входных величин X_i . Проанализировав таблицы на наличие резко выделяющихся значений, в случае их обнаружения, также необходимо удалить содержащие их строки из исходной выборки. Следует отметить, что наличие парных грубых промахов следует искать в случае, если исходные данные получены в ходе какого-либо технологического процесса, где условия получения данных делают возможным ошибки такого типа.

3. Следующий шаг алгоритма связан с разработкой вариантов решения исследуемой задачи. Подготовку к дальнейшему регрессионному анализу следует начинать с нахождения величины линейной корреляционной связи между факторами, входящими в таблицу пассивного эксперимента. Для этих целей необходимо провести корреляционный анализ методом корреляционных плеяд, при пороговом коэффициенте корреляции Пирсона $r=0,75$. Для определения факторов, характеризующих плеяды, можно воспользоваться экспертными методами оценки. При невозможности определить характеризующие факторы, либо равнозначности факторов для исследования, рекомендуется определить все возможные варианты комбинаций слабосвязанных факторов для дальнейшего регрессионного анализа и сформировать из них различные таблицы для их последующей обработки.

4. После выделения из таблицы исходных данных факторов, связанных друг с другом слабой линейной зависимостью, проводится процедура увеличения объема многомерных данных, основанная на применении метода точечных распределений для каждого из входных факторов и выходной величины, а также оценке значений величин ненормированных плотностей вероятностей в виртуальной области. Для этих целей используется априорная информация о виде распределения случайной величины, известная из характеристик исследуемого процесса, приблизительно оцененная по гистограмме, либо предварительно рассчитанная одним из методов получения оценок по малым выборкам.

Итогом этой операции служит формирование виртуально увеличенной таблицы данных, отражающей информацию из исходной многомерной выборки, которая длиннее первоначальной в 10-20 раз, в зависимости от качества исходных данных и закона распределения случайных величин факторов.

5. По таблице виртуально увеличенных данных повторно определяются коэффициенты корреляции для всех факторов, по принципу «каждый с каждым», после чего проводится анализ полученных результатов методом корреляционных плеяд при пороговом значении коэффициента Пирсона $r=0,8$. В случае дополнительного обнаружения факторов, связанных сильной линейной зависимостью, также рекомендуется убрать их из дальнейшего рассмотрения, оставив только по одному фактору, наиболее точно характеризующему каждую из плеяд.

6. После выделения из виртуально увеличенной выборки слабосвязанных факторов, определяем возможные варианты математических моделей, которые содержит в себе многомерная виртуальная выборка.

7. Строим математические модели по получившей виртуальной таблице данных любыми из известных и подходящих для данного эксперимента методов регрессионного анализа многомерных пассивных данных.

8. Проверяем адекватность и качество построенных моделей, определяем оптимальные математические модели, приступаем к их интерпретации, совместно с профильными специалистами в конкретной области исследования.

9. Обобщаем полученную информацию в СППР, основанную на собранных данных, на математической модели исследуемого процесса и формальных знаниях экспертов в изучаемой области.

Во *втором параграфе* рассматривается вопрос о реализации полученной методики в виде программного обеспечения; разрабатываются требования к программному обеспечению; производится выбор методов и средств для его реализации; описываются входящие в программный комплекс компоненты, реализующие метод точечных распределений, метод корреляционных плеяд и метод состыковки таблиц ненормированной плотности вероятности в виртуальной области. Программное обеспечение для математического моделирования и анализа данных по пассивным выборкам малого объема состоит из трёх независимых друг от друга программ, объединенных общими интерфейсами работы с входными и выходными данными [19, 20]. Каждая подпрограмма является отдельным консольным приложением, работающим с csv-файлами, что упрощает их внедрение в различное прикладное программное обеспечение.

В *третьем* параграфе показывается применение полученной методики для прогнозирования сложности лапароскопических операций у мужчин с острыми холециститами, что на основании исходных данных обследования пациента расширяет возможности хирурга в прогнозировании предстоящих интраоперационных трудностей. В параграфе приводятся классификация сложности операций, предложенная экспертами. В ходе подготовки первоначального анализа данных были проведены проверки на наличие грубых промахов в исходных данных, а также на нормальность исследуемых величин. Рассмотрена проблема определения величины линейной корреляционной связи в выборках малого объёма, показаны преимущества определения величины коэффициента корреляции Пирсона после виртуального увеличения выборки многомерным методом точечных распределений. На основе полученной модели разработана компьютерная система поддержки принятия решений. Система была апробирована на операциях, проведённых в течение следующего года: модель правильно распознала сложность предстоящей операции в 73% случаях, что может быть признано удовлетворительным при раннем прогнозировании. Проверка полученной модели на итоговом наборе данных показала, что модель является адекватной, статистически значимой и входящие в неё факторы описывают выходную величину примерно на 71%, при величине среднеквадратичной ошибки оценки равной 0,712. Проведено сравнение значений коэффициента смешанной корреляции (r^2), среднеквадратичной ошибки оценки (S_{xy}) и расчетного (F) значения критерия Фишера, рассчитанных на итоговом наборе данных для полученной модели и моделей, которые могли бы быть построены без применения предложенного метода. Показано, что полученная модель обладает наилучшими характеристиками, а значит является наиболее качественной. Так как система использует вероятностно-статистические методы, следует отметить, что предсказываемый уровень сложности операции всегда имеет некоторую погрешность, в связи с чем, кроме наиболее вероятного результата для ЛПП выводятся рекомендации и для менее вероятного исхода событий. Таким образом, в окне с рекомендациями сверху указывается рассчитанная по введенным данным степень сложности операции, после чего значение округляется и в специальных текстовых блоках выводится информация об описании предсказанной степени сложности операции и консультации по её проведению. В нижней части окна приводится описание и консультация по проведению операции в случае менее вероятного исхода событий, для того, чтобы ЛПП на основе своих знаний и опыта смог принять более подходящее для него решение.

В *четвертом параграфе* рассматривается применение полученной методики при создании проблемно-ориентированной компьютерной системы поддержки принятия решений о качестве пробной выпечки хлебобулочных изделий. Оценка качества готовой продукции и выдача рекомендаций по оперативному регулированию рецептуры и технологического процесса по пробной партии является сложной аналитической процедурой. При этом, так как осуществляются не более 5 производственных циклов для каждого вида изделий, а параметры всех изделий в одной пробной выпечке идентичны, что характерно для любого группового процесса, то данные, по которым следует принять решение, являются как малой выборкой, так и сверхнасыщенным планом. В параграфе описываются основные входные и выходные параметры, влияющие на качество хлебобулочных изделий, относящиеся как к рецептуре, так и к технологическому процессу. На примере выпуска продукции «Хлеб целебный» показан процесс анализа данных, построения математических моделей и выдачи рекомендаций разработанной системой. Доказывается, что применение виртуального увеличения данных, основанного на методе точечных распределений, может улучшить качество статистического анализа таблиц, полученных при изготовлении хлебобулочных изделий. Кроме знаний о количественных параметрах, которые получаются при помощи моделирования, в словарь терминов разрабатываемой СППР также необходимо добавить знания, полученные из нормативно-технической документации, а также от экспертов в данной предметной области. На основе данных знаний был сформулирован свод правил, который связывает отдельные характеристики технологического процесса и рецептуры, позволяющий объяснить возникновение тех или иных явлений, а также спрогнозировать поведение изучаемых параметров. Большая часть правил, включенных в базу знаний, являются производными. Таким образом, левая часть правила представляет собой некоторый набор предпосылок, а в правой части содержится заключение, которое можно сделать на их основе. Такое правило срабатывает, если выполняются предпосылки, связанные между собой набором булевых операций. При этом каждая предпосылка – это определенное, выраженное количественно, значение одного из исследуемых параметров. Таким образом, в разрабатываемой СППР предсказание всех количественных параметров опирается на математическую модель, построенную в соответствии с разработанными методами и алгоритмами. Однако, часть параметров анализируются только органолептически и являются балльными, что не позволяет применить к ним вышеописанный математико-статистический аппарат. Для управления вариациями таких данных применяются производные правила,

анализирующие как количественные, так и балльные показатели [17, 30]. Также в параграфе описывается логика работы и интерфейс программного продукта.

Пятый параграф посвящен выводам и предложениям по итогам главы, в которых отмечается, что: применение разработанной методики к выборкам малого объема, полученным в различных условиях, во многих случаях позволяет сравнительно улучшить качество получаемых математических моделей, сохранив первоначальный закон распределения случайных величин; применение разработанной методики, не имея априорной информации о виде закона распределения, на основании анализа гистограмм, либо вынесения предположения о виде закона распределения, также может помочь улучшить качество получаемых математических моделей; применение виртуального увеличения информации с учетом вида закона распределения, на основании метода точечных распределений, в ряде случаев позволяет более точно определить величину линейной корреляционной связи коэффициентом корреляции Пирсона; для получения статистически значимых математических моделей необходимо применять различные методики построения уравнений регрессии; разработанное программное обеспечение помогает ускорить обработку пассивных данных малого объема по разработанной методике, выполняя самые трудоемкие участки вычислений и преобразований данных; разработанное программное обеспечение и его применение для поддержки принятия решений в конкретных прикладных областях исследований подтверждает полученные в диссертационной работе результаты.

ОБЩИЕ ВЫВОДЫ И РЕКОМЕНДАЦИИ

Результаты проделанной работы:

1. Проанализированы существующие критерии для обнаружения грубых промахов в выборках малого объема [12], выявлены методы обладающие наибольшей мощностью [16], даны рекомендации для выявления грубых промахов в выборках малого объема [18].

2. Разработана методика виртуального увеличения таблицы многомерных пассивных данных на основе метода точечных распределений [29], используя информацию о виде закона распределения каждой отдельной одномерной случайной величины [24] и величину ненормированной плотности вероятности в виртуальной области. Разработанная методика, позволяет получить таблицу данных, статистически однородную исходной, сохраняющую вид закона распределения каждой отдельной одномерной случайной величины, при этом по количеству строк превышающую исходную в 10-20 раз [31].

3. Проанализированы существующие методики для определения величины линейной корреляционной связи в выборках малого объема, показано, что ни один из них не обладает достаточной точностью для проведения анализа классическими методами, в связи с чем рекомендуется использовать пороговое значение для определения сильной линейной зависимости коэффициентом Пирсона, равное $r=0,75$ [14].

4. Показано, что применение виртуального увеличения таблицы пассивных данных позволяет более точно определять величину линейной корреляционной связи [13].

5. Разработан алгоритм построения систем поддержки принятия решений, основанный на разработанных вероятностно-статистических методах, позволяющий улучшить качество регрессионных моделей по пассивным выборкам малого объема, когда классические методы не дают требуемой точности [7, 15].

6. Разработано программное обеспечение, которое позволяет значительно упростить и ускорить процесс обработки многомерных выборок малого объема на основе метода точечных распределений [19, 20].

7. Методика опробована в различных областях жизнедеятельности, и на выборках, полученных в различных условиях. Были построены две проблемно-ориентированные компьютерные СППР для областей, в которых практически невозможно или экономически невыгодно получение большого количества экспериментальных данных. Полученные результаты показали, что применение разработанных методик может позволить исследователям построить системы, которые не только отвечают требованиям,

предъявляемым к качеству полученных решений, но и в ряде случаев получить более достоверные результаты по сравнению с применением классических методов [17, 21, 22].

8. Системы поддержки принятия решений, разработанные автором, позволили:

8.1. Выдавать рекомендации о предстоящих лапароскопических операциях, уровне их сложности, порядке проведения и вероятности осложнений на момент, когда не было возможности набрать достаточное количество данных.

8.2. Оценивать качество пробной выпечки хлебобулочных изделий на основе интерактивной генерации моделей и вырабатывать варианты решения, адаптированные к конкретной ситуации.

Разработанные СППР могут быть успешно интегрированы в более крупные информационные системы, путём использования и модернизации созданных баз знаний и математических моделей.

Таким образом, в работе решена проблема определения вероятностно-статистических методик поддержки принятия решений на основе построения адекватных математических моделей по пассивным многомерным выборкам малого объёма, полученным в различных условиях.

Проведенное исследование поддержки принятия решений по пассивным выборкам малого объема позволило сформулировать основные рекомендации, которые могут быть полезны как с научной, так и с практической точки зрения:

1. Принятие решения о наличии грубого промаха в данных малого объема следует производить при помощи одного или нескольких из следующих критериев: Ирвина, Львовского, Титъена-Мура или Диксона [16, 18]. При этом следует дополнительно изучить влияние отклонения от нормального закона распределения на мощность данных критериев, а также возможность обнаружения групповых ошибок и вариантов их устранения.

2. Для улучшения качества применяемых решений по данным малого объема рекомендуется использовать предложенную методику виртуального увеличения объема выборки [29], позволяющую сохранить знание о виде закона распределения каждой из исследуемых случайных величин.

3. После виртуального увеличения объема выборки, основанного на применении метода точечных распределений, рекомендуется проводить корреляционный анализ с пороговым коэффициентом Пирсона равным 0,8 [13]. Следует отметить, что не было обнаружено методов, позволяющих проводить корреляционный анализ данных малого объема с достаточной точностью и достоверностью [14], в связи с чем направлением

дальнейших исследований может стать разработка методики для оценки величины связи между малыми выборками.

4. Рекомендуется применять полученные результаты для улучшения качества построения проблемно-ориентированных систем поддержки принятия решений на основе вероятностно-статистического анализа малых объемов информации, вне зависимости от области исследований, если набор достаточного количества исходных данных практически невозможен или экономически неэффективен.

БИБЛИОГРАФИЯ

1. Гаскаров Д. В., Шаповалов В. И. Малая выборка. Москва: Статистика, 1978, 248с.
2. Григорьев Ю.Д. Метод Монте-Карло: вопросы точности асимптотических решений и качества генераторов псевдослучайных чисел. В: Заводская лаборатория. Диагностика материалов, 2016, т.82, №7, с. 72-84.
3. Громыко Г.Л. Теория статистики: Учебник, 2-е изд., перераб. и доп. Москва: ИНФРА-М, 2011, 476 с.
4. Долгов Ю.А. Статистическое моделирование: Учебник для вузов. – 2-е изд., доп, Тирасполь: Полиграфист, 2011, 352 с.
5. Еременко В.С., Куц Ю.В., Мокийчук В.М. Оценка однородности выборок малого объема, В: Системи обробки інформації, 2006, №7(56), с.26-29.
6. Кобзарь А.И.. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2012, 816 с.
7. Кожокару С.К., Попукайло В.С. Метод принятия решений в условиях малой выборки. В: Научный обозреватель, 2016, №3(63), с.79-80.
8. Козыр Е.В. Этапы обработки данных при малых выборках. В: Механіка гіроскопічних систем, 2010, №21, с.103-111.
9. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло. В: Научный журнал КубГАУ, 2015, №114(10), с.27-41.
10. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. Ростов-на-Дону: Феникс, 2009, 572с.
11. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? В: Заводская лаборатория. Диагностика материалов, 1991, т.57, №7, с.64-66.
12. Попукайло В.С. Исследование критериев грубых ошибок применительно к выборкам малого объема В: Радіоелектронні і комп'ютерні системи, 2015, №3(73), с. 39-44.
13. Попукайло В.С. Исследование линейной корреляционной связи в многомерном методе точечных распределений В: Информационно-управляющие системы, 2016, №6, с. 96-98. doi:10.15217/issn1684-8853.2016.6.96
14. Попукайло В.С. Исследование линейной корреляционной связи в парных выборках малого объема. В: Технология и конструирование в электронной аппаратуре, 2016, №1, с. 27-32. doi: 10.15222/ТКЕА2016.1.27

15. Попукайло В.С. Методы математического моделирования по выборкам малого объема В: Доклады итоговой научн. конф. проф.-препод. состава инженерно-технического институт за 2015 год: г. Тирасполь 22-26 янв. 2016г./ Приднестр. гос. ун-т им. Т.Г. Шевченко; под ред.: Долгова Ю.А. Тирасполь: «TesLine», 2016, с. 150-151.

16. Попукайло В.С. Обнаружение аномальных измерений при обработке данных малого объема. В: Технология и конструирование в электронной аппаратуре, 2016, №1, с. 27-32. doi: 10.15222/ТКЕА2016.4-5.42

17. Попукайло В.С. Поддержка принятия решения о качестве пробной лабораторной выпечки. В: Проблемы современной науки и образования, 2017, №7(89) с. 43-45 doi:10.20861/2304-2338-2017-89-003

18. Попукайло В.С. Принятие решений о наличии выбросов в данных малого объема. In: Materialele Conferinței Științifice a Doctoranzilor (cu participare internațională) “Tendințe contemporane ale dezvoltării științei: viziuni ale tinerilor cercetători”, 25 mai 2016, Chișinău, Universitatea Academiei de Științe a Moldovei, 2016, p.306-310.

19. Попукайло В.С. Построение таблицы исходных данных для многомерного моделирования по выборкам малого объема. В: Тезисы IX Международной конференции «Математическое моделирование в образовании, науке и производстве». Тирасполь, 8-10 октября 2015г. Тирасполь: Изд-во Приднестровского университета, 2015, с.176-177.

20. Попукайло В.С. Программная реализация метода точечных распределений. В: Сборник материалов Всерос., научно-практической конференции «Информационно-телекоммуникационные системы и технологии», 16-17 окт. 2015 г., Кемерово [Электронный ресурс] / ФГБОУ ВПО «Кузбас. гос. техн. ун-т им. Т. Ф. Горбачева»; редкол.: Трофимов И.Е. (отв. ред.) [и др.], Кемерово, 2015. <http://sibscience.ru/page/ITSIT-2015/ITSIT/6-Matematicheskoe-modelirovanie/6042.pdf>

21. Попукайло В.С. Построение математической модели эффективности работы банка в условиях малой выборки. В: Актуальные проблемы гуманитарных и естественных наук, 2016, №3(86), с. 74-78.

22. Попукайло В.С. Прогнозирование уровня успеваемости абитуриентов в условиях малой выборки. В: Наука вчера, сегодня, завтра. Сб. статей по материалам XXXII международно-практической конференции, №3(25), 2016, с. 95-101.

23. Столяренко, Ю.А. Контроль кристаллов интегральных схем на основе статистического моделирования методом точечных распределений. Диссертация на соискание ученой степени к.т.н. М., ГУП НПЦ «СПУРТ», 2006. 191 с.

24. Столяренко Ю.А. Гординский Д.Н. Попукайло В.С. Применение метода точечных распределений для различных законов распределения В: Тезисы VII Международной конференции. 3-5 октября 2013г. – Тирасполь: Изд-во Приднестровского университета, 2013, с. 201-202.
25. Gaindric C. Luarea decizilor. Metode si tehnologii. Chisinau: Stiinta, 1998, 162p.
26. Fischer H. A History of the Central Limit Theorem: From Classical to Modern Probability Theory, Sources and Studies in the History of Mathematics and Physical Sciences, New York: Springer, 2011, p.402.
27. Gurov S. 0-event: Point and Interval Estimates. In: Mathematics and Statistics, 2013, №1(2): p. 53-58
28. Hays W. Statistics. Boston: Cengage Learning, 1994, 1136 p.
29. Popukaylo V. Analysis of statistical modeling methods for small-volume samples. In: Proceedings Proceeding of the 5th International Conference «Telecommunications, Electronics and Informatics» ,may 20-23, 2015, Chişinău, Tehnica-UTM, Moldova, 2015, p.207-211.
30. Popukaylo V. Development of the decision support system for test baking. In: Proceedings CMSM 4, June 28 – July 2, 2017 Chişinău / ed.: Mitrofan Choban [et al.]. Chişinău: Institute of Mathematics and Computer Science, 2017 (CEP USM), 572 p.
31. Popukaylo V. Small size sample mathematical modeling B: Meridian Ingineresc 2015, №4(595), с. 25-30.

АННОТАЦИЯ

Попукайло Владимир

ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ ПО ПАССИВНЫМ ВЫБОРКАМ МАЛОГО ОБЪЕМА

диссертация доктора информатики, Кишинев, 2017.

Структура диссертации: введение, три главы, общие выводы и рекомендации, библиографический список из 160 наименований, 4 приложения, 120 страниц базового текста, 30 рисунков, 23 таблицы. По материалам диссертационного исследования опубликовано 16 печатных работ.

Ключевые слова: теория принятия решений, системы поддержки принятия решений, малая выборка, пассивный эксперимент, корреляционный анализ, регрессионный анализ, грубые промахи, метод точечных распределений.

Область исследования: пассивные малые выборки, полученные в различных условиях.

Цель исследования: разработка методики построения систем поддержки принятия решений по пассивным выборкам малого объема, в условиях, когда классические вероятностно-статистические методы не позволяют получить обоснованных выводов.

Задачи исследования: 1) проанализировать различные подходы и методы обработки выборок малого объема; 2) исследовать возможности использования существующих критериев для нахождения грубых промахов в выборках малого объема; 3) исследовать возможности критериев обнаружения линейной корреляционной связи в выборках малого объема; 4) разработать методику построения систем поддержки принятия решений в условиях пассивного эксперимента и невозможности получить большой массив исходных данных; 5) апробировать разработанную методику на данных, полученных в различных условиях.

Научная новизна и оригинальность исследования – разработана оригинальная методика, позволяющая строить системы поддержки принятия решений по пассивным выборкам малого объема, дающие надёжные рекомендации для лица, принимающего решения.

Актуальность исследования состоит в том, что существует большое количество областей, в которых невозможно получить данные большого объема, в связи с чем, принятие решения с использованием классических методик крайне затруднено или невозможно.

Важная научная проблема, решенная в работе: определение вероятностно-статистических методик поддержки принятия решений, на основании выборок малого объема, полученных в ходе пассивного эксперимента.

Теоретическая значимость исследования заключается в методике улучшения качества принимаемых решений на основе предложенных подходов и алгоритмов, опирающихся на вероятностно-статистические методы обработки информации.

Практическая значимость результатов работы. Предложенные в диссертационной работе методики поддержки принятия решения могут быть широко применены при статистическом анализе в различных областях исследований, в которых нет возможности получить большое количество данных.

Внедрение результатов исследования: Результаты работы внедрены в лечебный процесс отделения Эндоскопической и малоинвазивной хирургии Республиканской клинической больницы (г. Тирасполь), учебный процесс кафедр «Биологии и физиологии человека» и «Терапии №2» медицинского факультета Университета им. Т.Г. Шевченко (г. Тирасполь), в работу научно-исследовательской лаборатории «Математическое моделирование» и в учебную программу магистратуры по направлению «Информатика и вычислительная техника» того же университета.

ADNOTARE
Popukaylo Vladimir

SUPPORTUL DECIZIILOR ÎN BAZA EȘANTIOANELOR PASIVE DE DIMENSIUNI MICI

Teză de doctor în informatică. Chișinău, 2017

Structura tezei: introducere, trei capitole, concluzii, bibliografie – 160 titluri, 4 anexe, 120 pagini de text de bază, 30 figuri, 23 tabele. Rezultatele cercetării sunt publicate în 16 lucrări științifice. **Cuvinte-cheie:** teoria luării deciziilor, sisteme pentru suport decizii, eșantioane mici, experiment pasiv, analiză corelațională, analiză regresivă, date aberante, metoda distribuțiilor punctate.

Domeniul de cercetare: eșantioane mici pasive, obținute în diferite condiții.

Scopul lucrării: elaborarea metodicii de construire a sistemelor pentru suport decizii în baza eșantioanelor de dimensiuni mici, în condițiile , când metodele clasice ale teoriei probabilităților și statisticii matematice nu permit obținerea unor concluzii fundamentate.

Obiectivele cercetării: 1) analiza diferitelor abordări și metode de procesare a eșantioanelor de dimensiuni mici; 2) cercetarea posibilității de utilizare a criteriilor existente pentru identificarea datelor aberante în eșantioane de volum mic; 3) cercetarea posibilităților de determinare a relației de corelare liniară în eșantioane de volum mic; 4) elaborarea metodicii de construire a sistemelor pentru suport decizii în condiții de experiment pasiv și imposibilitatea obținerii unui masiv mare de date inițiale; 5) aprobarea metodicii elaborate pentru date, obținute în diferite condiții; 6) fundamentarea și demonstrarea aplicabilității metodologiei propuse.

Noutatea științifică a lucrării o constituie metodologia originală de construire a sistemelor pentru suport decizii în baza eșantioanelor pasive de dimensiuni mici, care îi oferă decidentului cele mai exacte recomandări.

Actualitatea cercetării este motivată de existența unui spectru larg de domenii pentru care este imposibil de obținut un volum mare de date, fapt ce împiedică sau chiar face imposibilă luarea deciziilor prin aplicarea metodelor clasice.

Problema științifică principală, soluționată de către autor, constă în elaborarea metodicii probabilistice și statistice de suport al deciziilor în baza eșantioanelor de dimensiuni mici, obținute în baza experimentului pasiv.

Semnificația teoretică a cercetării o constituie metoda de îmbunătățire a calității deciziilor prin abordările propuse și algoritmi elaborați, bazați pe metode probabilistice și statistice de procesare a informației.

Valoarea aplicativă a cercetării: metodele propuse în teză pot fi utilizate pe larg în analiza statistică în diverse domenii, pentru care nu există posibilitatea de a obține un volum mare de date.

Implementarea rezultatelor cercetării: rezultatele cercetării sunt implementate în procesul de tratament al Secției de chirurgie endoscopică și minim invazivă a spitalului clinic republican din or. Tiraspol, procesul didactic al catedrelor “Biologia și fiziologia omului” și “Terapia №2” a facultății de medicină a Universității “T.G.Șevcenco” din or. Tiraspol, precum și în cercetările din cadrul laboratorului de modelare matematică și cursurile de masterat la specialitatea „Informatică și tehnică de calcul” ale aceleiași universități.

ANNOTATION

Popukaylo Vladimir

DECISION SUPPORT ON SMALL SIZE PASSIVE SAMPLES

Thesis of Doctor in Computer Science. Kishinev, 2017.

Structure of the thesis: Introduction, three Chapters, General conclusions and recommendations, bibliography of 160 titles, 4 appendices, 120 pages of body text, 30 figures, 23 tables. According to the dissertation research materials 16 printing works are published.

Keywords: mathematical modeling, small sample, passive experiment, correlation analysis, regression analysis, outliers, the method of point distributions.

Field of study: small size passive samples obtained in different conditions.

Research objective: to develop a construction technique of adequate mathematical models for small size passive samples, in conditions when classical probabilistic-statistical methods do not allow obtaining valid conclusions.

Research tasks: 1) to analyze various approaches and methods of small size samples processing; 2) to investigate the possibility of using existing criteria for outliers finding in small size samples; 3) to investigate the possibility of the linear correlation detection in the small size samples; 4) to develop a methodology for constructing decision support systems in cases of passive experiment and impossibility of obtaining a large amount of raw data; 5) to test the developed technique on the obtained data in various conditions; 6) to substantiate and to prove the applicability of the developed technique.

The scientific novelty and originality of the research: the original methodology for construction the decision support systems which give the most accurate recommendations for decision-makers basing on passive small samples.

The relevance of research: there are many areas where it is impossible to obtain large amounts of data, and therefore, the decision-making using classical techniques is extremely difficult or impossible.

Important scientific problem solved in the work: the determination of probabilistic and statistical methods of decision support based on small size samples obtained during the passive experiment.

The theoretical significance of the study lies in the methodology of improving the quality of decisions taken on the basis of the proposed approaches and algorithms based on probabilistic and statistical methods of data processing

The practical significance of the research results. The decision support methods, proposed in the thesis, can be widely used in the statistical analysis in various fields of research, where it is impossible to obtain a large amount of data.

The research results implementation: The results of the research is implemented in the direct care of Department of Endoscopic and Minimally Invasive Surgery of the Republican Clinical Hospital (Tiraspol), into educational process of the department "Biology and Human Physiology" and "Therapy №2» Medical Faculty of the T.G. Shevchenko University, into the work of the Research Laboratory "Mathematical Modeling" and academic Master's program in "Computer Science and Engineering" of the same university.

ПОПУКАЙЛО ВЛАДИМИР СЕРГЕЕВИЧ

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
ПО ПАССИВНЫМ ВЫБОРКАМ
МАЛОГО ОБЪЁМА

122.03 –МОДЕЛИРОВАНИЕ, МАТЕМАТИЧЕСКИЕ МЕТОДЫ И
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Автореферат диссертации на соискание ученой степени
доктора информатики

Рекомендовано к печати: 05.06.2017

Формат бумаги: 60x84 1/16

Бумага и печать офсетные

Тираж 50 экз.

Печатных листов: 1,75

Заказ № ____

Типография «ВлаСаНа» SRL,
MD-3200, г. Бендеры, ул.Ленина, 29.

**ACADEMIA DE ȘTIINTE A MOLDOVEI
INSTITUTUL DE MATEMATICĂ ȘI INFORMATICĂ**

Cu titlu de manuscris

C.Z.U: 004.415.2

POPUKAYLO VLADIMIR

**SUPPORTUL DECIZIILOR ÎN BAZA EȘANTIOANELOR PASIVE
DE DIMENSIUNI MICI**

122.03 – MODELARE, METODE MATEMATICE, PRODUSE PROGRAM

Autoreferatul teză de doctor în informatica

CHIȘINĂU, 2017