

**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**

**Диссертация на соискание ученой степени доктора информатики**

Научный руководитель:

Кожокару Светлана,  
доктор хабилитат в информатике,  
профессор

Автор:

Попукайло Владимир

**КИШИНЁВ, 2017**

**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**

**Teză de doctor în informatica**

**Conducător științific:**

Cojocarui Svetlana,  
doctor habilitat în informatica,  
profesor cercetător

Autorul:

Popukaylo Vladimir

**CHIȘINĂU, 2017**

**© Popukaylo Vladimir, 2017**

## ОГЛАВЛЕНИЕ

<b>АННОТАЦИЯ</b>	6
<b>СПИСОК СОКРАЩЕНИЙ</b>	9
<b>ВВЕДЕНИЕ</b>	10
<b>1. МЕТОДОЛОГИЧЕСКИЕ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ МАЛОЙ ВЫБОРКИ</b>	17
1.1. Методы и средства принятия решений	17
1.2. Трудности применения вероятностно-статистических методов принятия решения в условиях малой выборки	23
1.3. Разграничение выборок малого и большого объема и определение их границ	28
1.4. Методы оценки плотности распределения по выборке малого объема	31
1.5. Выводы к главе 1	38
<b>2. РАЗРАБОТКА МЕТОДИК ПРИНЯТИЯ РЕШЕНИЯ ПРИ ОБРАБОТКЕ ДАННЫХ МАЛОГО ОБЪЕМА</b>	40
2.1. Нахождение грубых промахов в выборках малого объема	40
2.2. Линейная корреляция в выборках малого объема	57
2.3. Построение исходной таблицы многомерных данных малого объёма	66
2.4. Определение качества решений	77
2.5. Результаты по главе 2	80
2.6. Выводы к главе 2	81
<b>3. СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ПАССИВНЫМ ВЫБОРКАМ МАЛОГО ОБЪЕМА</b>	82
3.1. Алгоритм создания СППР по пассивным выборкам малого объема	82
3.2. Разработка программного обеспечения, реализующего вероятностно-статистические методы	84
3.3. СППР о сложности предстоящих лапароскопических операций	88
3.4. СППР о качестве пробной выпечки хлебобулочных изделий	104
3.5. Выводы к главе 3	117
<b>ОБЩИЕ ВЫВОДЫ И РЕКОМЕНДАЦИИ</b>	118
<b>СПИСОК ЛИТЕРАТУРЫ</b>	121

## **ПРИЛОЖЕНИЯ**

Приложение А1. Таблицы расчета построения исходной таблицы многомерных данных	134
Приложение А2. Описание программного обеспечения	144
Приложение А3. Расчет модели сложности операции	153
Приложение А4. Балльная оценка органолептических показателей качества хлебобулочных изделий	164
<b>ДЕКЛАРАЦИЯ ОБ ОТВЕТСТВЕННОСТИ</b>	167
<b>РЕЗЮМЕ АВТОРА</b>	168

## 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 методик 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-i 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 algoritmilor 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 al 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.

## СПИСОК СОКРАЩЕНИЙ

АЭФ – Априорно-эмпирическая функция

ЖКБ – Желчнокаменная болезнь

ЛПР – Лицо, принимающее решения

ЛХЭ – Лапароскопическая холецистэктомия

МИФ – Модифицированный индекс Фехнера

ММСБ – Модифицированный метод случайного баланса

ММТР – Многомерный метод точечных распределений

МНК – Метод наименьших квадратов

МНКО – Метод наименьших квадратов с предварительной ортогонализацией факторов

МПВ – Метод прямоугольных вкладов

МПМ – Метод последовательных медиан

МТР – Метод точечных распределений

МУН – Метод уменьшения неопределённости

ПО – Программное обеспечение

СКО – Среднеквадратическое отклонение

СППР – Система поддержки принятия решений

## ВВЕДЕНИЕ

**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 не больших 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 не решенной 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. [8-10], Шаповалова В.С. [8], Долгова Ю.А [19-26], Столяренко Ю.А. [83, 84], Кривенко М.П. [41], Большев Л.Н. [4, 5], Шор Я.Б. [92], Кобзарь А.И. [36], Орлов А.И. [56-61], Pearson E.S. [143], Efron В.М. [107], Fisher R.A. [109], de Winter J.C.F. [151, 152] 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 объема.

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 возможности 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:

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

2) 5<sup>th</sup> International Conference ICTEI-2015, Chisinau, 2015;

3) 23-a Conferință de Matematică Aplicată și Industrială CAIM 2015, ROMANIA, Suceava, 2015;

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

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

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

7) 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.

**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, приводится классификация и примеры областей их применения; также приводится схема принятия решения и классификация методов принятия решения. *Во втором параграфе* описываются трудности применения классических вероятностно-статистических методов принятия решения в условиях малой выборки, приводятся примеры областей исследований, в которых характеристики объектов не позволяют получать большие объемы экспериментальных данных; обсуждаются проблемы существующих подходов, созданных для решения проблемы ограниченности объема данных; приводятся основные проблемы анализа многомерных выборок малого объема. *В третьем параграфе* дано определение термина «малая выборка» и обоснованы границы малых и средних выборок, сделан вывод, что обосновано применение термина «малая выборка», при количестве наблюдений (экспериментов) менее 20. *В четвертом параграфе* проведен краткий обзор существующих методов обработки малых выборок, основанных на использовании каждой отдельной реализации и позволяющих

улучшить показатели при оценивании плотности распределения случайной величины; из анализа видно, что большинство таких методов основываются на использовании информации о каждой отдельной реализации и являются развитием метода прямоугольных вкладов; приведён алгоритм метода точечных распределений, в котором в качестве дополнительной информации используется знание о виде закона распределения случайной величины. *Пятый параграф* посвящен выводам и предложениям по итогам главы.

*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. Проведен анализ существующих методик обнаружения грубых промахов в одномерных выборках, на основании которого предлагается использовать для данных целей критерии Диксона (для различных законов распределения), Львовского (в случае нормального закона распределения), Ирвина (при  $n \geq 10$ ) или Титъена-Мура (как обладающего наименьшим количеством ошибок I типа). Также поднимается 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. Показывается отрицательное влияние на оценку коэффициента корреляции процедур рандомизации и увеличения объема данных методом точечных распределений. *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, доказывається однородность исходной и полученной таблиц данных. Обосновывается методика нахождения коэффициента корреляции в многомерном методе точечных распределений. *В четвертом параграфе* рассматриваются вопросы оценки качества принимаемых решений на основе вероятностно-статистического подхода; даётся определение понятию адекватности математической модели; сравнивается качество математических моделей, построенных различными способами. *Пятый параграф* резюмирует результаты главы, выводы которой приведены в *шестом параграфе*.

*В третьей главе* «Системы поддержки принятия решений по пассивным выборкам малого объема» формулируется 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. Во *втором параграфе* рассматривается вопрос разработки прикладного программного обеспечения, реализующего описанные методики; разрабатываются требования к программному обеспечению; производится выбор методов и средств для его реализации; описываются входящие в программный комплекс компоненты. В *третьем параграфе* показывается применение полученной методики для построения системы поддержки принятия решений по прогнозированию сложности лапароскопических операций, что на основании исходных данных обследования пациента расширяет возможности хирурга в прогнозировании предстоящих интраоперационных трудностей; описываются этапы создания СППР, основные компоненты программного продукта и интерфейса. В *четвертом параграфе* рассматривается применение полученной методики при создании проблемно-ориентированной компьютерной системы поддержки принятия решений о качестве пробной выпечки хлебобулочных изделий. На примере выпуска продукции «Хлеб целебный» показан процесс анализа данных, построения математических моделей и выдачи рекомендаций разработанной системой. Также в параграфе описывается логика работы и интерфейс программного продукта. *Пятый параграф* посвящен выводам и предложениям по итогам главы.

*Общие выводы и рекомендации* Анализ научной литературы по теме исследования выявил проблему статистической обработки информации в условиях многомерной малой выборки. Актуальная научная проблема, решенная в исследовании, состоит в определении вероятностно-статистических методик поддержки принятия решений на основе построения адекватных математических моделей по пассивным многомерным выборкам малого объема, полученным в различных условиях. Полученные результаты позволят улучшить качество построения систем поддержки принятия решений на основе вероятностно-статистического анализа малых объемов информации, вне зависимости от области, в которой были получены исходные данные.

# 1. МЕТОДОЛОГИЧЕСКИЕ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В УСЛОВИЯХ МАЛОЙ ВЫБОРКИ

## 1.1. Методы и средства принятия решений

Задача принятия решения – это определение оптимального способа действий для достижения наилучшего результата. Выработка плана и порядка действий по решению данной задачи – основной вопрос теории принятия решений. При этом предполагается наличие некоторой неопределённости, то есть результаты принимаемых решений заранее не известны. К настоящему времени разработаны различные способы описания неопределенностей: вероятностные модели, теория нечеткости, интервальная математика и другие. В соответствии с основной идеей системного анализа, при принятии решений сочетают как методики вербального описания проблемной ситуации, так и методы формализованного представления систем [6].

Необходимость учета для принятия решения большого числа взаимосвязанных факторов в быстро меняющейся обстановке настоятельно требуют применения вычислительной техники в процессе принятия решений [87].

Одним из видов компьютерной поддержки принятия решений являются экспертные системы, которые разрабатываются для хранения и представления знаний высококвалифицированных экспертов, для того, чтобы ими в дальнейшем могли воспользоваться другие специалисты. Обычно предполагается, что исследуемая предметная область является слабоструктурированной. Данные системы применяются для различных видов деятельности, таких как: интерпретация, прогнозирование, планирование, обучение, управление, диагностика, проектирование, отладка и др. Диапазон применения компьютерных систем очень широкий, в связи с чем активно применяется такой класс вычислительных систем, как системы поддержки принятия решения (СППР) [153].

СППР – это интерактивные, компьютерные автоматизированные системы, которые помогают лицу, принимающему решения (ЛПР), использовать информацию и модели для решения слабоструктурированных или трудно формализуемых задач [40, 131, 136, 154]. Такие системы, как правило, основываются на анализе ранее полученных данных некоторого объема, описывающих поведение изучаемого объекта, опыте работы экспертов, а также принятых в прошлом решениях и их последствиях.

При этом необходимо отметить, что такие системы создаются, чтобы помочь принять решения использующему их человеку, в таком случае называемому «лицо, принимающее решение». Данные решения представляются ему, как специалисту в данной предметной области, и являются оптимальными на основе ранее полученных данных, что необходимо в условиях высокой сложности и постоянно меняющихся условиях решаемой задачи.

Назначением СППР является внедрение в практику работы ЛПР некоторого подмножества адекватных действий, которые включают в себя [111]:

- 1) получение запросов от пользователя;
- 2) перевод на язык понятный системе;
- 3) обработку элементов знаний, имеющихся в СППР (обновление, техническое обслуживание, накопление, удаление), в соответствии с потребностями решения;
- 4) ввод/вывод сообщений в удобном для пользователя формате.

В настоящее время инструментальные средства анализа и принятия решений условно могут быть разделены на три основные группы [45]:

- 1) специализированные системы, обеспечивающие поддержку процесса принятия решений при разработке конкретных систем;
- 2) универсальные пакеты и инструменты, предназначенные для анализа разнообразных моделей и систем, содержащие также средства для многокритериального анализа;
- 3) проблемно-ориентированные системы анализа и принятия решений.

Главным недостатком систем первой группы – специализированных систем – является невозможность настройки на изменяемые условия задачи и прикладной области [45], при этом такие средства являются эффективными для поддержки принятия решений в конкретных условиях, под которые они были разработаны.

В качестве универсальных пакетов широкое распространение получили программные средства, разработанные для реализации сложных математических расчетов и моделирования в процессе решения различных задач, такие как *MatLab*, *MatCad*, *Mathematica*, *Statistica* и другие. Главный недостаток таких систем – это ориентация их интерфейса на инженеров и математиков, тогда как пользователем системы поддержки принятия решений может являться человек, не обладающий достаточным уровнем математических знаний, но являющийся специалистом в конкретной исследуемой предметной области.

Проблемно-ориентированные системы анализа и принятия решений – это информационные системы, разработанные для поддержки принятия решений по классу задач, порождаемых определённой проблемой или совокупностью схожих проблем. Эти средства разрабатываются таким образом, чтобы ориентироваться на решение некоторой группы типовых задач и обеспечивать настройку под индивидуальные особенности каждой проблемной ситуации и конкретной исследуемой предметной области. Данные программные системы характеризуются модульным подходом к реализации основных алгоритмов принятия решений и интерфейсом, ориентированным на конечного пользователя.

В настоящее время проблемно-ориентированные системы широко применяются для поддержки принятия решения в различных предметных областях, в том числе:

1) Медицине – для повышения эффективности диагностики и прогнозирования заболеваний [99, 102, 111, 132, 136, 141].

2) Промышленности – для комплексного подхода к управлению сложными системами [101, 158].

3) Экономике – для своевременного и оперативного прогнозирования экономических ситуаций и принятия решений [76].

4) Педагогике – для управления образовательными процессами и улучшения качества обучения, а так же других задач, связанных с ними [42].

5) Биологии – для обеспечения экосистемных услуг [148], контроллинга параметров сбора урожая [124] и прочее.

6) Других областях жизнедеятельности [87, 106].

В общем виде, схема процесса принятия решения состоит из следующих шагов (этапов) [95]:

1) *Описание предметной области*, в которой будет происходить дальнейшее исследование проблемы, производится ЛПР.

2) *Анализ ситуации* основан на сборе имеющейся информации об исследуемом процессе в виде формализованных данных, описывающих проблему в данной предметной области.

3) *Формулирование проблемы* позволяет выявить критерии, по которым будет производиться сравнение и выбор вариантов оптимальных решений, предоставляемых ЛПР.

4) *Анализ имеющейся информации* помогает ограничить круг возможных методов формализации проблемы, определить средства и методики исследования для принятия

решения. На этом этапе производится кодирование собранной информации в вид необходимый для обработки отобранными методами.

5) *Разработка вариантов решения* производится соответствующим математическим аппаратом, в зависимости от вида, качества и объема исходной информации.

6) *Оценка применимости вариантов решений и их последствий* осуществляется путем сравнения преимуществ и недостатков каждого из полученных решений. На данном этапе также оцениваются риски от принятия каждой из предложенных альтернатив и выдаются рекомендации для ЛПР.

7) *Выбор оптимального варианта* – предпочтение следует отдавать такому решению, которое обеспечивает достижение наилучшего результата с наиболее высокой вероятностью.

8) *Проверка и выдача решения* – на данном этапе СППР проверяет правильность полученного решения, формирует необходимую информацию и выдаёт своё заключение и рекомендации для ЛПР.

Таким образом, для создания СППР необходимо объединить опыт экспертов в исследуемой предметной области, для которой разрабатывается данная система поддержки принятия решений, и специалистов в области информационных технологий и прикладной математики.

Представленная схема принятия решения не отражает сложность данного 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 методы их обработки. Эффективность данного процесса во многом зависит от методов, применяемых при выполнении каждого из указанных этапов схемы процесса принятия решений [75].

При создании СППР для различных приложений могут быть использованы либо один из этапов генерации решений, либо их различные комбинации. На рисунке 1.1 [87] стрелками показывается возможная последовательность создания подсистем принятия решений.

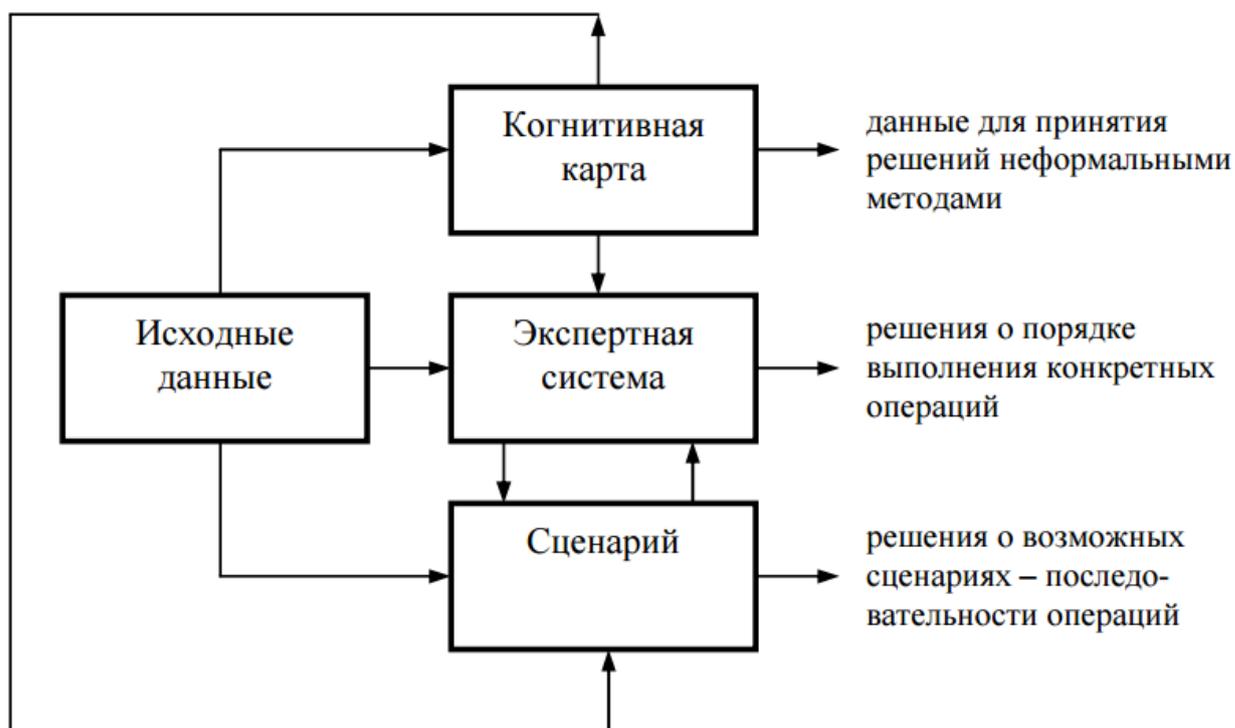


Рисунок 1.1 – Этапы генерации решений

Выделяют два основных класса методов системного анализа, использующихся в СППР: неформальные методы (эвристические или качественные) и формальные методы (математические или количественные) [11].

Трудность применения неформальных методов состоит в сложности извлечения знаний от высококвалифицированных независимых экспертов. Данные методы применяются, когда полученная информация слабо структурирована, и основываются на интуиции и знании специалистов.

В случае если имеются достоверные данные, представленные в цифрах, рекомендуется использовать формализованные, либо комплексные методы проведения исследований.

Для плохо формализуемых ситуаций применяют лингвистические, семиотические и графические представления, как удобные средства исследования структур и процессов в сложных системах. Данные подходы удобны для первого этапа постепенной формализации задач принятия решений.

Если же знания об изучаемых процессах и событиях позволяют полностью определить их поведение, то для формализации представления системы используют аналитические методы, использующие аппарат классической математики. Однако, для сложных, многокомпонентных или многокритериальных систем установить все

необходимые детерминированные взаимосвязи между исследуемыми компонентами, а также доказать правомерность применения полученных аналитических выражений практически невозможно [6].

В таких случаях, возможно применение описания исследуемой системы при помощи случайных событий либо процессов, которые могут быть описаны соответствующими вероятностными или статистическими характеристиками и закономерностями. Расширение возможностей отображения сложных систем и процессов при данном подходе, по сравнению с аналитическими методами, объясняется тем, что применение вероятностно-статистического анализа возможно, в том числе, на основе выборочного исследования.

Если статистические методы не приводят к выявлению устойчивых закономерностей, применяют логические методы, основанные на применении алгебраических методов для выражения законов формальной логики. Логические представления нашли широкое применение при исследовании и разработке автоматов различного рода, автоматических систем контроля, а также при решении широкого ряда других задач [6].

Применение статистических методов и математического программирования в системах поддержки принятия решений позволяет решить множество слабоформализованных задач благодаря возможности объединения в единой формальной модели критериев различной размерности, что очень важно при отображении исследуемых процессов. Использование аппарата математического моделирования, и в частности регрессионного анализа, требует выбора некоторой целевой функции и определения ограничений. Данные действия, по сути, являются средством постановки задачи, так как ориентация на формирование целевой функции и необходимых ограничений позволяет уточнить представление о проблемной ситуации и более точно сформулировать постановку задачи.

Применение конкретного вероятностно-статистического метода принятия решений состоит из трех этапов [58]:

- 1) переход от экономической, управленческой, технологической реальности к абстрактной математико-статистической схеме, т.е. формализация проблемы в виде построения вероятностной модели исследуемой системы управления, технологического процесса, процедуры принятия решений, в частности по результатам статистического контроля, и т.п.

2) проведение расчетов и получение выводов чисто математическими средствами в рамках вероятностной модели;

3) интерпретация математико-статистических выводов применительно к реальной ситуации и принятие соответствующего решения (например, о соответствии или несоответствии качества продукции установленным требованиям, необходимости наладки технологического процесса и т.п.), в частности, заключения (о доле дефектных единиц продукции в партии, о конкретном виде законов распределения контролируемых параметров технологического процесса и др.).

## **1.2. Трудности применения вероятностно-статистических методов принятия решений в условиях малой выборки**

Выборочное исследование позволяет принимать решения на основе экспериментальных (статистических) данных, под которыми понимают значения случайной величины, полученные в результате повторений случайного эксперимента (наблюдений над случайной величиной). Совокупность независимых случайных величин  $X_1, \dots, X_n$ , каждая из которых имеет такое же распределение, что и случайная величина  $X$ , называется случайной выборкой из генеральной совокупности  $X$  [14]. При этом число  $n$  называют объемом случайной выборки. Если объем выборки  $n$  достаточно велик ( $n \rightarrow \infty$ ) выборка считается большой, в противном случае она называется выборкой ограниченного объема [86].

Одна из основных задач вероятностно-статистических методов – это оценивание характеристик генеральной совокупности, таких как математическое ожидание, медиана, дисперсия, среднеквадратическое отклонение по результатам выборочного наблюдения. Оценки неизвестных параметров бывают двух видов: точечные и интервальные. Точечной статистической оценкой параметра генеральной совокупности называется конкретное числовое значение искомой характеристики. Точечное оценивание обычно сопровождается случайной ошибкой. При правильном формировании выборки величину ее ошибки можно рассчитать заранее. В общем случае под ошибкой выборки понимают объективно возникающее расхождение между характеристиками выборки и генеральной совокупности [81].

Для оценки, вычисленной по одной выборке, случайная ошибка конечна, и в случае, если выборка мала, то она весьма существенна. В таком случае становится важным знание точности и надежности полученной оценки. Для их уточнения проводят

интервальное оценивание, пользуясь понятиями доверительного интервала и доверительной вероятности. Интервальная оценка представляет числовые параметры, предположительно содержащие значение параметра генеральной совокупности. Интервальная оценка представляет собой интервал со случайными границами, который с заданной вероятностью накрывает неизвестное истинное значение параметра. Интервал между нижней и верхней границами такой оценки называется доверительным интервалом, а задаваемая исследователем вероятность – доверительной вероятностью. Очевидно, что чем выше доверительная вероятность, тем выше точность оценки, но в таком случае доверительный интервал становится шире. Отсюда следует, что для непрерывных случайных величин вероятность того, что точечная оценка совпадет с любым заданным значением или оцениваемым параметром, равна нулю.

Таким образом, точечная оценка имеет смысл лишь, если приведена характеристика рассеяния этой оценки (т.е. дисперсия или стандартное отклонение). В противном случае она может служить лишь в качестве исходных данных для построения исследователем интервальной оценки.

Качество статистических оценок определяется следующими их свойствами [81]:

**Несмещенность.** Оценка считается несмещенной, если при данном объеме выборки математическое ожидание ошибки равно нулю.

**Эффективность.** Оценка считается эффективной, если её ошибка, называемая ошибкой выборки, является величиной минимальной.

**Состоятельность.** Оценка считается состоятельной, если при неограниченном увеличении объема выборки её ошибка стремится к нулю.

На практике, существуют такие области, в которых характеристики исследуемых объектов позволяют получать только не большие объемы экспериментальных данных. При таком эксперименте, как правило, уровни наблюдаемых характеристик не могут быть произвольно заданы, а лишь регистрируются исследователем. В теории планирования эксперимента такой подход называется пассивным экспериментом [43]. Поскольку при пассивном эксперименте исследователь не имеет возможность задать уровень ни одного из факторов, то при проведении опытов ему остается лишь "пассивно" наблюдать за явлением и регистрировать результаты. Планирование пассивного эксперимента сводится к определению числа опытов, которые необходимо провести исследователю для решения поставленной перед ним задачи, а конечной целью пассивного эксперимента в большинстве случаев является получение математической модели исследуемого объекта в

виде уравнения регрессии, либо своевременное выявление нестабильности, определяемое, как правило, при помощи контрольных карт Шухарта [155].

Вопрос о необходимом объеме выборки в задачах различных исследований [119, 145, 146, 149] является одним из актуальных и самых часто задаваемых вопросов прикладной математической статистики [105, 149]. Это связано с тем, что при обработке данных ограниченного объема перед исследователем возникает ряд трудностей, среди которых выбор правильных статистических подходов. Сложность выбора правильной методики исследования, в условиях недостаточного количества данных, связана с тем, что классические подходы к оцениванию параметров могут не обладать должной мощностью, что увеличивает вероятность ошибок второго типа [108]. В таком случае, самым простым способом повысить статистическую значимость исследования зачастую является увеличение объема данных.

Однако, в целом ряде случаев такой подход невозможен, ввиду особенностей предметных областей изучаемых процессов. К таким областям относятся:

1) контроль качества технологического процесса, при котором требуется измерение параметров как можно меньшей доли партии при высокой вероятности получения правильного результата, что может быть связано с применением разрушающих методов контроля [83];

2) диагностика состояния сложных технологических объектов, в которых динамика протекания процессов такова, что длительность пребывания объекта в нештатных режимах кратковременна и отражается ограниченным объемом данных, на основании которого необходима идентификация состояния объекта и принятие адекватного решения по его управлению в условиях значительной неопределенности [78];

3) экологический мониторинг, при котором диагностика и прогноз изучаемого показателя производятся в реальном масштабе времени [80];

4) биологические исследования, изучающие генетически обусловленное поведение животных, что связано с невозможностью изучения большого количества особей редких видов [98, 103, 145], либо небольшие популяции растений [126, 159];

5) медицинские исследования для анализа в динамике применяемых методов лечения и лекарственных средств, при контроле за проводимыми лечебными и диагностическими операциями [55, 119, 144];

6) экономика, в задачах рыночной оценки стоимости имущества либо прогнозировании уровня развития экономики региона в условиях неполноты информации [46, 53];

7) другие области, в которых невозможно получить достаточное количество экспериментальных данных [44, 81, 116].

В таком случае ЛПП приходится принимать решение по полученным данным малого объема, что приводит к следующим распространенным проблемам при их оценивании этих данных вероятностно-статистическими методами:

- с уменьшением объема выборки увеличиваются доверительные интервалы и вероятностные ошибки [117];

- в условиях малой выборки дисперсия выборочной совокупности не может рассматриваться в качестве оценки генеральной дисперсии [16];

- в соответствии с центральной предельной теоремой [110] 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 [30, 36];

- результаты обработки данных, полученные при использовании различных критериев, например, таких как Кохрена и Граббса, являются зависимыми от объемов исследуемых выборок [38];

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

Для решения проблем, связанных с ограниченностью объёма выборок, разработаны специальные методы, такие как статистика Стьюдента [19, 143, 146], Фишера [19, 108, 109] и методы Монте-Карло [107, 117, 137]. Следует отметить, что применение статистик Стьюдента и Фишера, хоть и является широко распространенной практикой [117, 122, 124, 137], имеет достаточно серьёзные недостатки, не позволяющие с необходимой достоверностью [133] применять их в прикладных исследованиях в условиях малой выборки. Так, применение квантилей распределения Стьюдента и Фишера предполагает принятие гипотезы о нормальности закона распределения [19], при том, что погрешности большинства физических измерений отличны от нормальных [60, 61]. Применение же методов Монте-Карло, хоть и является перспективным направлением математической статистики, также имеет ряд нерешенных проблем. К ним, в частности, относится несовершенство датчиков псевдослучайных чисел [15, 32]. Кроме того, в работе [57]

указывается, что выводы, основанные на применении метода Монте-Карло, являются не строго доказанными, а лишь правдоподобными, ввиду бесконечного множества возможных отклонений случайной величины от изучаемого параметрического семейства распределений.

Принятие решений в условиях малого количества информации ещё более усложняется в многомерном случае. При проведении различных испытаний или наблюдений, зачастую, таблицы исходных данных представляют собой, так называемые сверхнасыщенные планы, часть факторов которых не влияет на целевую функцию и, в конце концов, уйдет в шум эксперимента. Отсев таких факторов может производиться на основе объективных (например, метод случайного баланса) или субъективных (экспертные методы оценок) методов, в результате чего размерность факторного пространства сокращается [19, 104].

Также, в первоначальном списке могут быть сильно коррелированные между собой факторы. Законы математической статистики утверждают, что каждая пара таких факторов должна быть разбита, то есть один из факторов должен быть отброшен, как не дающий дополнительной информации в будущей модели, а другой оставлен для дальнейшей работы. С этой целью проводится корреляционный анализ таблицы исходных данных, наиболее подходящим для этого методом является метод корреляционных плеяд и ядер [19]. Применяя данный метод, возможно произвести дополнительное сокращение факторного пространства, без существенной потери информации.

Однако, описанные выше методы хорошо изучены и работают для случаев выборок среднего и большого объема. Обработка же сверхнасыщенных планов, в которых число экспериментов относится к области малых и сверхмалых выборок, классическими методами не представляется возможной, либо приводит к большим ошибкам.

Перечислим основные проблемы анализа многомерных данных малого объема [41]:

- 1) алгоритмическая сложность процедур обработки данных;
- 2) многомерные данные – плохой объект для визуализации, поэтому весьма трудно представить себе реальную картину происходящего в процессе анализа;
- 3) с ростом размерности данных обычно существенно снижается качество применяемых методов анализа данных;
- 4) при усложнении описания объектов анализа за счет увеличения размерности возникает множество вариантов методов анализа данных, с каждым из которых связана необходимость установления значений большого числа параметров;

5) сингулярность (вырожденность) матрицы вторых моментов, если количество элементов обучающей выборки меньше или существенно меньше размерности векторного представления данных.

### **1.3. Разграничение выборок малого и большого объема и определение их границ**

Обработка любых числовых экспериментальных данных с целью получения новой или обобщающей информации, как правило, основана на методах математической статистики и теории вероятностей. Одним из фундаментальных понятий теории вероятностей является понятие случайной величины. Известно, что случайная величина проявляется только в результате большого числа экспериментов (или наблюдений). При наблюдениях в определенных условиях случайная величина обладает некоторой закономерностью, которая описывается её распределением. Распределения случайных величин зачастую либо неизвестны, либо известны только с некоторой точностью. Математическая статистика ставит и позволяет решить задачи изучения этих закономерностей по выборке данных из некоторой совокупности, называемой генеральной. Для решения этой задачи обязательным условием является репрезентативность и достаточность выборки.

Исходя из этого, для получения приемлемой ошибки при расчете параметров выборки вся классическая математическая статистика имеет дело с выборками объемом в несколько сотен измерений. В работах исследователей установлено, что нижняя граница таких выборок проходит примерно на уровне  $n=100$  измерений [92]. Таким образом, для обработки как одномерных экспериментальных данных (гистограмма, установление вида закона распределения, центральные моменты и т.п.), так и двумерных (коэффициент корреляции, корреляционное отношение, уравнение регрессии и т.п.) рекомендуются объёмы выборок  $n>100$ .

При уменьшении объемов выборок ( $n<100$ ) классические методы математической статистики становятся малопригодными, и для оценки неизвестных параметров выборки были разработаны специальные методы, такие как статистика Стьюдента, Фишера,  $\omega^2$ , методы «*bootstrapping*» и другие.

Данные методы позволяют существенно снизить рекомендуемый объем выборки и зачастую дают вполне удовлетворительные результаты, однако, как уже было отмечено, имеют ряд недостатков, не позволяющих обоснованно применять их в различных

исследованиях. Тем не менее, на практике часто имеется необходимость обработки малого числа экспериментов (наблюдений), а с дальнейшим уменьшением объема выборки ошибка оценивания значительно возрастает. Также можно предположить, что существует нижняя граница объема, при которой дальнейшее уменьшение выборки приводит к неприемлемым результатам.

Считается, что начало статистике малых выборок было положено в 1908г. после публикации работы [143] У.Госсета (Стьюдента), в которой он описал  $t$ -распределение. Он сумел определить зависимость между величиной доверительного коэффициента  $t$ , а так же численностью малой выборки  $n$  с одной стороны, и вероятностью нахождения ошибки выборки в заданных пределах с другой стороны. Эта зависимость и стала называться «распределение Стьюдента».

Гаскаров Д.В. и Шаповалов В.И. в своей работе «Малая выборка» [8] 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 [57] 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.

Известно, что нижней границей при обработке результатов экспериментов (наблюдений) является объем  $n=3$  [92]. С верхней границей такой однозначности нет. Многие исследователи принимают величину верхней границы малой выборки равной определенному значению в диапазоне от 10 [63] до 30 [82, 145]. В некоторых источниках также фигурируют значительно большие значения [35, 92].

При этом, такие решения не являются убедительными, так как данные числа ничем не обусловлены. Таким образом, остается открытым вопрос об определении верхней границы малой выборки, которая находится примерно в диапазоне от 11 до 30 измерений.

Однако, при проведении статистических исследований, имеется настоятельная необходимость в определении границ между выборками среднего и малого объема. Это связано с необходимостью выбора оптимальных методов оценочного контроля, корреляционного и регрессионного анализа данных. Данный вопрос остро стоит при

проведении операций выборочного контроля, при котором требуется измерение параметров как можно меньшей доли партии при высокой вероятности получения правильного результата, что может быть связано с применением разрушающих процедур контроля. Кроме того, с задачами обработки малых выборок встречаются исследователи, проводящие экологический мониторинг, либо диагностику состояний сложных технологических объектов. Также с подобными проблемами сталкиваются исследователи в области медицины и экономики, где зачастую невозможно получить выборку достаточно большого объема для её обработки классическими методами.

Следовательно, задача об определении границы между выборками малого и среднего объема является актуальной. Анализ литературы по данной тематике показал, что единственным обоснованным подходом к решению данной проблемы является исследование [83], которое проводилось с помощью статистического имитационного эксперимента на базе *bootstrap*-метода, в котором рассчитывалась нижняя граница интервальной оценки наивероятнейшего коэффициента корреляции, вычисленная в пространстве преобразования Фишера с помощью критерия Пирсона [19]. Выбор данного подхода обусловило то обстоятельство, что при вычислении наивероятнейшего коэффициента корреляции используются два среднеквадратических отклонения ( $S_X$  и  $S_Y$ ), обладающие наибольшим разбросом (максимальными ошибками). Таким образом, можно утверждать, что рассмотрен наихудший случай.

В результате данного исследования было установлено, что *bootstrap*-метод удовлетворительно работает при объеме выборки  $n > 20$  независимо от вида корреляционной связи (как линейная, так и нелинейная). Данный метод так же удовлетворительно работает до объема  $n = 16$  включительно, но только в случае высокой степени тесноты линейной корреляционной связи. При объемах  $n < 16$  *bootstrap*-метод дает неудовлетворительные результаты [83].

Похожие результаты также были получены при анализе объема выборок, основанном на использовании критерия Пирсона, в работе [115], в которой верхняя граница малого объема выборки устанавливается в точке  $n = 19$ .

Таким образом, можно считать установленным, что нижней границей среднего объема выборки является  $n = 20$  измерений, а также что диапазон малой выборки лежит в пределах от  $n = 3$  до  $n = 15$  измерений. А выборки объемом  $15 < n < 20$  можно считать «неустановленной зоной», в которой выбор метода обработки определяется исследователем и зависит от параметров распределения исследуемых величин и меры тесноты линейной связи между ними.

#### 1.4. Методы оценки плотности распределения по выборке малого объема

Для устранения потерь информации при обработке малой выборки необходимо отказаться от группировки данных и перейти к методам, основанным на использовании каждой отдельной реализации (измерения) [8].

В силу того, что на практике могут быть использованы различные законы распределения, задача анализа сводится к преобразованию выборки в определенную функцию распределения.

Алгоритм преобразования, в общем случае, состоит из следующих этапов:

1. Построение эмпирической функции распределения на основе экспериментальных данных.
2. Выдвижение гипотезы о виде функции распределения.
3. Оценка параметров функции распределения по выборочным данным.
4. Проверка выдвинутой гипотезы по результатам сопоставления эмпирической и теоретической функции распределения.
5. Принятие решения о согласованности гипотезы с фактическими данными.

Впервые задачи оценивания плотности распределения случайной величины по выборке малого объема были поставлены и частично решены в работе [90].

В основе большинства методов по оценке плотности распределения лежит предварительное числовое определение эмпирической функции распределения [8]:

$$f^*(x) = \alpha \cdot f_0(x) + (1 - \alpha) \cdot f_n^*(x); \quad (1.1)$$

где  $f_0(x)$  – априорная компонента, несущая информацию о форме закона распределения;  $\alpha$  – вес (ценность) априорной информации;  $f_n^*(x)$  – эмпирическая компонента, связанная с  $i$ -ой реализацией выборки.

Для различных методов оценивания существуют разные значения коэффициента  $\alpha$  в диапазоне от 0 до 1 и разные виды функций эмпирической компоненты.

Распространённым подходом для построения оценки плотности распределения по малой выборке является применение метода прямоугольных вкладов (МПВ) [90]. Он основан на равномерном увеличении информации, полученной от каждой отдельной реализации, на конечном интервале определённой длины. В качестве дополнительной априорной информации предполагается знание интервала изменения случайной величины  $X$ . При использовании данного метода предполагается, что оцениваемая

функция распределения непрерывна и не имеет очень крутых скачков на заданном интервале.

В качестве функции вклада могут использоваться равномерное распределение, распределение Симпсона, треугольная функция, потенциальная функция, дельтаобразная функция и другие [79]. Таким образом, при применении данного метода учитывается случайный характер каждой отдельной реализации данных, а суммирование всех элементарных плотностей позволяет получить улучшенную оценку плотности распределения. Также доказано, что оптимальная оценка зависит от правильного выбора ширины вклада.

В работе [31] был предложен метод уменьшения неопределенности (МУН). В данном методе, в отличие от МПВ, предлагается использовать нормированное равномерное распределение, заданное на интервале  $[x_{i-1}, x_i+1]$ , то есть распределить скачок вероятности в точке  $x_i$  равномерно. Доказано, что МУН является частным случаем МПВ и не может дать выигрыша в эффективности, так как ширина вклада является случайной величиной и меняется лишь в зависимости от индекса  $i$ .

В работе [17] предложен метод априорно-эмпирических функций (АЭФ). В случае использования данного подхода интегральная функция распределения представляется в следующем виде:

$$F(x) = \omega F_a(x) + (1 - \omega) F_s(x) \quad (1.2)$$

где  $F_a(x)$  - априорное распределение, построенное по априорным данным;

$F_s(x)$  - эмпирическое распределение, построенное по данным выборки;

$\omega$  - коэффициент достоверности информации об априорном распределении.

Предложение отказаться от уравнивания априорной информации, полученной от единичного наблюдения, и ввести коэффициент достоверности  $\omega$  улучшает возможности оценивания данным методом.

В работе [31] был предложен метод сжатия области существования интегральных законов распределения. Основным преимуществом данного метода по сравнению с рассмотренными выше является возможность построения доверительных границ для полученной оценки, а также вычисления доверительной вероятности для каждого  $i$ -го члена последовательности. Однако, его применение позволяет получить значение оценки только в определенных точках и в дальнейшем требуется дополнительная линейная интерполяция оценки [12].

В работе [8] проведён анализ описанных выше методов, из которого следует, что

применение любого из этих подходов, либо их комбинирование позволяет получить более эффективную оценку функции плотности распределения по малой выборке по сравнению с классическим подходом. Вместе с тем в этой же работе сделан вывод, что наиболее общим и, следовательно, перспективным для исследователя является метод прямоугольных вкладов.

Г.В. Дружининым и О.В. Вороновой разработан эмпирический метод построения интегральной функции распределения, названный авторами «методом последовательных медиан» (МПМ) [27]. Данный метод также является одной из разновидностей МПВ, при этом границы интервала, на котором определена случайная величина, определяются по данным выборки, при помощи оценок медианы и моды.

Дальнейшее развитие метода прямоугольных вкладов было предложено в работе [62], в которой в качестве критерия согласия при построении оценки плотности распределения используются первые четыре момента. Определение вида аппроксимирующего распределения осуществляется путем оценки попадания расчетных моментов этого распределения в доверительные интервалы для эмпирических моментов.

Методика проверки выдвигаемых гипотез заключается в следующем. Для каждого из рассматриваемых теоретических распределений методом максимального правдоподобия определяются значения параметров, после чего вычисляются первые четыре начальных момента, соответствующие этому распределению. После этого выясняют, попадают ли эти моменты в доверительные интервалы для эмпирических моментов. При этом переходят последовательно от более высоких уровней к более низким. В качестве подходящего распределения выбирают такое, моменты которого попадают в самый узкий интервал.

В работе [13] рассматривается метод имитационного дополнения, суть которого состоит в генерации дополняющих массивов в окрестности каждого элемента исходной выборки. Такой процесс логически близок к аппроксимации ступенчатой функции распределения и позволяет свести обработку малой выборки к существующим классическим технологиям. Авторы работы утверждают, что на данный момент исследований при правильном выборе параметров вкладов, количество правильно классифицированных (отнесённых к определенному закону распределения) выборок достигает 80%.

В работе [8] также был предложен подход к определению формы и параметров вкладов каждой отдельной реализации выборки, позволяющий извлечь наибольшее количество информации о функции распределения и, таким образом, получить

оптимальные оценки значений его параметров. В таком подходе в качестве априорной информации используется заранее известный вид закона распределения.

Данная работа получила своё развитие в многочисленных трудах [8, 18, 20, 82], что привело к появлению метода точечных распределений [21, 83], в котором для различных законов распределения рассчитаны интервалы изменения случайной величины, определена ширина ядра и алгоритм его определения, а также определена вероятность появления случайной величины в интервале дискретности.

Практический алгоритм вычисления эмпирической функции распределения  $f^*(x)$  можно представить нижеследующим образом [83]:

1. Необходимо установить предполагаемый класс распределений  $f(x)$  на основе имеющейся априорной информации об исследуемой выборке, либо предварительном анализе исходных данных.

2. Определить интервал изменения  $(a, b)$  контролируемой величины  $X$  в абсолютных единицах (а.е.). Данный интервал рассчитывается на основе характеристик исследуемой выборки.

3. Вычислить оптимальное значение  $\rho$  в а.е. :

$$\rho = \rho' \cdot (b - a), \quad (1.3)$$

где вспомогательный коэффициент  $\rho'$  определяется эмпирически для каждого класса распределения  $f(x)$  и объема выборки  $n$ .

Расчету оптимального значения данной величины посвящён ряд работ различных исследователей [18, 22, 82].

4. Для каждого элемента выборки  $X_i$ , необходимо определить верхнюю  $X_{iB}$  и нижнюю  $X_{iH}$  границы интервала определения  $i$ -ой дельтаобразной функции:

$$X_{iH} = \begin{cases} a, & \text{если } X_i - \rho \leq a; \\ X_i - \rho, & \text{в остальных случаях} \end{cases} \quad (1.4)$$

$$X_{iB} = \begin{cases} b, & \text{если } X_i + \rho \geq b; \\ X_i + \rho, & \text{в остальных случаях.} \end{cases}$$

5. Для каждого элемента выборки  $X_i$  определить корректировочный коэффициент  $A_i$ , который равен:

- для нормального закона распределения  $A_i \equiv 1$ ;
- для экспоненциального, Парето, Гамма и Вейбула распределения

$$A_i = \begin{cases} 1 - \frac{X_i - \rho - a}{\rho}, & \text{если } X_{iH} = a; \\ 1, & \text{в остальных случаях.} \end{cases} \quad (1.5)$$

7. Интервал  $(a, b)$  разбить на  $k$  интервалов дискретности ( $k \in [20, 30]$ ).

При этом величина  $X'_j = a + (2j - 1) \cdot \frac{b - a}{2k}$  – центр  $j$ -го интервала дискретности.

8. Вычислить значение нормированной плотности  $f^*(X'_j)$  на интервале с центром в точке  $X'_j$ :

$$f^*(X_i) = \frac{k}{b - a} \cdot \frac{f'(X'_j)}{\sum_{j=1}^n f'(X'_j)}; \quad (1.6)$$

где  $f'(X'_j)$  – значения ненормированной оценки плотности в центре  $j$ -го интервала дискретности.

В качестве примера в таблице 1.1 приведем интервалы изменения контролируемой величины  $X$  и аппроксимационные формулы уравнения регрессии для расчета вспомогательного коэффициента  $\rho'$  (при различной форме ядра) [19]. Представлены примеры для нормального, равномерного и экспоненциального законов распределения, а также законов распределения Вейбулла и Стьюдента ( $Z_n = 1,75 + \frac{11,1}{n}$ ).

Исследования [84] доказывают, что применение метода точечных распределений позволяет увеличить количество информации, извлекаемой из выборки, что увеличивает точность расчетов до уровня, соответствующего объему выборки в 3-5 раз больше исходной.

Для примера в таблице 1.2 приведем реальные и эквивалентные объемы выборок, а также рассчитанные относительные ошибки выборочных среднеквадратических отклонений после применения МТР [21].

Таким образом, применение данного метода также позволяет сохранить информацию о виде закона распределения случайной величины, что может способствовать увеличению точности дальнейших исследований на основе параметрического оценивания. На рисунках 1.2 и 1.3 приведены примеры [83] построения эмпирических графиков плотности вероятности случайной величины, для двух различных выборок объемом  $n=8$ .

Таблица 1.1. Аппроксимационные формулы для параметра  $\rho'$

Закон распределения	Форма ядра	Уравнение регрессии	$a$	$b$
Нормальный	Прямоугольн.	$\hat{\rho}' = 0,2924 - 0,237n + 0,0012n^2$	$\bar{X} - 2,4S$	$\bar{X} + 2,4S$
	Дельтаобразн.	$\hat{\rho}' = \frac{0,3435n}{n - 0,7066}$		
Экспоненциальный	Прямоугольн.	$\hat{\rho}' = 0,1047 + \frac{0,61199}{n} - \frac{0,7405}{n^2}$	0	$3,912\bar{X}$
	Дельтаобразн.	$\hat{\rho}' = 0,3788 - 0,00065n^2 + 1,56347e^{-n}$		
Вейбулла	Прямоугольн.	$\hat{\rho}' = 0,3662 \cdot 0,9293^n \cdot 1,00198n^2$	0	$g\sqrt[3]{3,912}$
	Дельтаобразн.	$\hat{\rho}' = 0,8406 - 0,3196 \ln n + 0,0674(\ln n)^2$		
Равномерный	Прямоугольн.	$\hat{\rho}' = 0,3586 - 0,000616n - 0,000045n^2$	$X_H$	$X_B$
	Дельтаобразн.	$\hat{\rho}' = 0,6993 - 0,0012n - 0,000088n^2$		
Стьюдента	Прямоугольн.	$\hat{\rho}' = 0,1803 + \frac{0,3392}{n}$	$\bar{X} - Z_n S$	$\bar{X} + Z_n S$
	Дельтаобразн.	$\hat{\rho}' = 0,3516 + \frac{0,6510}{n}$		

Таблица 1.2. Реальные и эквивалентные объемы выборок и относительные ошибки выборочных СКО после применения МТР

Исходная выборка	$n$	3	4	5	6	7	8	9	10
	$\xi, \%$		46,6	38,9	34,1	30,7	28,2	26,2	24,5
МТР, нормальный закон	$\frac{D}{D_n}$	1,863	1,891	1,876	1,821	1,791	1,780	1,741	1,690
	$n_3$	10	14	17	20	23	25	27	29
	$\xi, \%$	23,2	19,4	17,5	16,1	15,0	14,3	13,8	13,3
МТР, закон Вейбулла	$\frac{D}{D_n}$	0,100	2,385	2,582	2,364	2,310	2,228	2,145	2,050
	$n_3$	13	23	33	34	37	40	41	42
	$\xi, \%$	20,0	15,1	12,4	12,3	11,7	11,3	11,1	11,0

На основании метода точечных распределений в работе [83] также был предложен метод построения корреляционных и регрессионных уравнений по парным выборкам малого объема. Применение метода точечных распределений позволило для этих целей использовать классический метод Чебышева [91], в котором для удобства построения уравнения регрессии предлагается сформировать и в дальнейшем обрабатывать таблицу двумерного распределения по выбранной паре случайных величин  $X$  и  $Y$ . В этой таблице в

левый столбец заносятся центры разрядов гистограммы случайной величины  $X$ , в верхнюю строку заносятся центры разрядов гистограммы случайной величины  $Y$ , в предпоследний столбец - записывают частоты разрядов  $n_j$ , а в последнюю строку заносят частоты разрядов  $n_l$  этой гистограммы. На основании полученной таблицы строится регрессионное уравнение и оценивается коридор его существования.

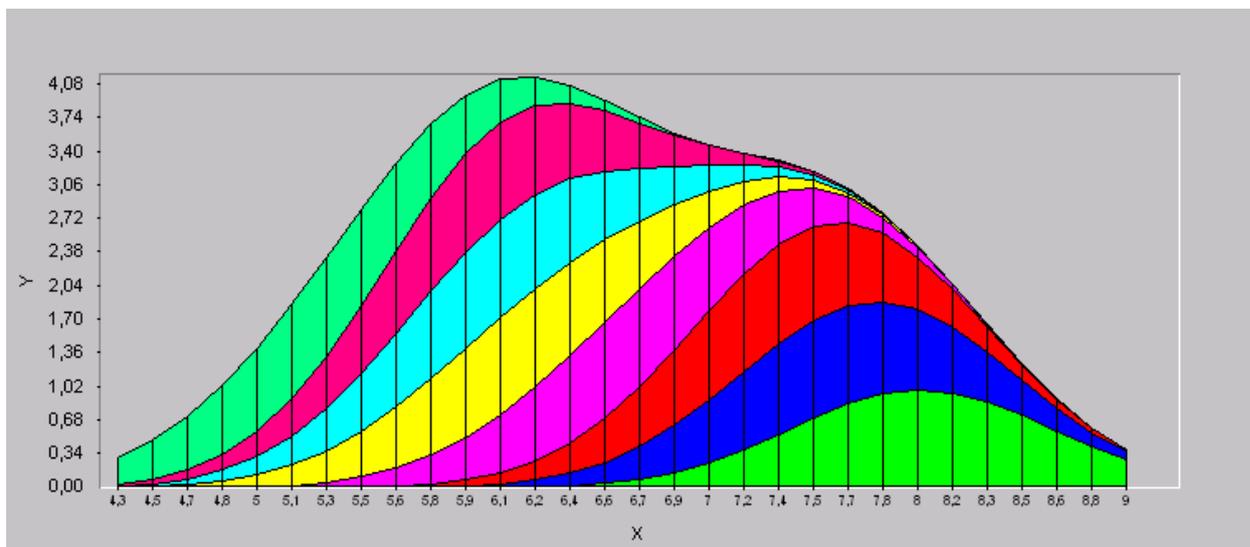


Рисунок 1.2 – Пример эмпирического графика распределения выборки  $X_1$

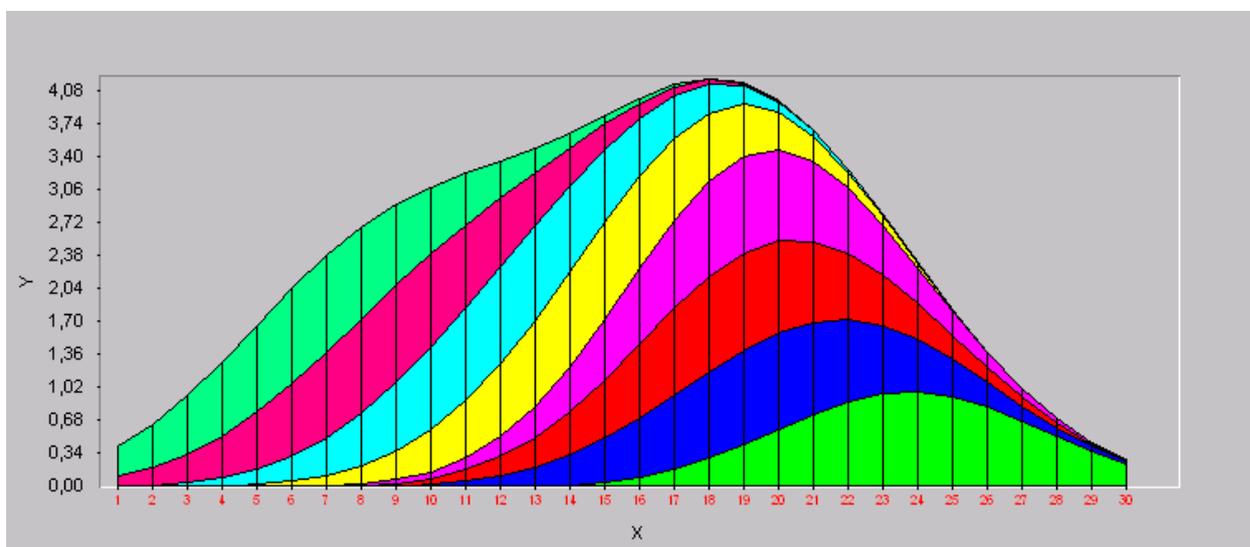


Рисунок 1.3 – Пример эмпирического графика распределения выборки  $X_2$

Из идеологии метода Чебышева понятно, что для построения гистограмм, а следовательно и таблиц двумерного распределения, требуется выборка достаточно большого объема, применение же метода точечных распределений позволяет обойти это ограничение. Кроме того, малая величина объема выборки существенным образом

расширяет коридор существования уравнения регрессии, так как каждое измерение несет очень большую информационную нагрузку. Таким образом, применение для построения регрессионного уравнения таблиц двумерного распределения по методу точечных распределений позволяет уменьшить коридор вероятностных ошибок [83]. Однако, такой подход позволяет строить регрессионные уравнения только в двухфакторном случае, а в реальном производстве на целевую функцию воздействует множество факторов, и таким образом задача математического моделирования сводится к построению наилучшего уравнения регрессии по исходной таблице многомерных данных.

Регрессионный анализ – основной метод современных вероятностно-статистических методов принятия решения, который применяют для исследований функциональной зависимости одного количественного признака от некоторого количества количественных признаков процесса (объекта). Обработка данных наблюдений методами многомерной статистики позволяет выявить тонкие особенности, характеризующие сложные связи между факторами, такие как определение степени детерминированности вариации зависимой переменной независимыми переменными, предсказание значений зависимой переменной с помощью независимой, а также определение вклада в вариацию зависимой переменной каждой из независимых. Данная информация может быть полезна ЛПР для принятия обоснованных решений при прогнозировании, принятии гипотез и выявлении скрытых взаимосвязей в исследуемых процессах.

### **1.5 Выводы к главе 1**

Анализ положения в области поддержки принятия решений выявил, что в настоящее время является актуальным создание проблемно-ориентированных СППР в различных областях деятельности человека. При этом для анализа сложных, многокомпонентных и многокритериальных систем используется вероятностно-статистический подход, применение которого для малого количества исходных данных связано с рядом трудностей.

Однако анализ такой информации позволит не ожидать накопления большего объема данных, а работать в режиме реального времени, что позволит более оперативно управлять исследуемым процессом.

На основании вышеизложенного, была сформулирована цель исследования, которая заключается в разработке методики построения систем поддержки принятия

решений по пассивным выборкам малого объема, в условиях, когда классические вероятностно-статистические методы не позволяют получить обоснованных выводов.

Для реализации данной цели было решено провести исследования возможностей синтеза существующих подходов для анализа одномерных и многомерных данных, в результате чего были поставлены следующие задачи:

1. Необходимо 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. Необходимо определить оптимальные методы принятия решения об обнаружении и устранении грубых промахов в выборках малого объема.

3. Разработать методику 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.

## **2. РАЗРАБОТКА МЕТОДИК ПРИНЯТИЯ РЕШЕНИЯ ПРИ ОБРАБОТКЕ ДАННЫХ МАЛОГО ОБЪЕМА**

### **2.1 Нахождение грубых промахов в выборках малого объема.**

При решении задач статистического анализа проблема наличия в выборке аномальных измерений имеет чрезвычайно большое значение. Присутствие единственного аномального наблюдения может приводить к оценкам, которые совершенно не согласуются с выборочными данными [47]. Грубые промахи возникают вследствие нарушения основных условий измерения или в результате ошибок экспериментатора. Они резко отличаются по величине от остальных значений, на чем и основаны критерии их исключения из рассмотрения. При обнаружении грубого промаха результат следует отбросить, а само измерение, если возможно, повторить [39]. Однако, необходимо быть уверенным, что найденная существенно отличающаяся от средних значений величина действительно является ошибкой при фиксировании случайной величины. Для проверки значимости подозрительных экспериментальных данных разработаны специальные статистические критерии. Рекомендуется проводить обязательный анализ значимости отклонения крайних значений выборки от остальных, так как если они являются выбросами, то их использование при оценке выборочных моментов и проверка различных статистических гипотез может привести к большим ошибкам [36].

Установлено, что все выборки делятся на три диапазона: выборки малого объема ( $n=3-20$  элементов) [83], выборки среднего объема ( $n=20-100$  элементов) [107] и выборки большого объема (свыше 100 элементов) [52]. Выборки каждого диапазона имеют свои особенности при обработке, которые позволяют уменьшить потери информации и тем самым повысить точность и достоверность рассчитываемых параметров.

Существует большое количество критериев, предназначенных для решения задачи отсева грубых промахов из полученной экспериментальным путем информации [36, 52, 97]. Эффективность этих методов во многом зависит от объема исследуемой выборки. Анализ современной литературы показал, что существует сравнение различных критериев только для выборок среднего и большого объема [138]. Исходя из этого, актуальна задача определения наиболее эффективных критериев для малых выборок. Так, например, в современной промышленности существует большое количество отраслей производства, в которых, в силу технологических ограничений, получение достаточно большого количества информации по различным причинам невозможно. Так при производстве

кристаллов интегральных микросхем для контроля электрических характеристик структур и качества проведения технологических операций используют специально изготавливаемые и размещаемые на подложке структуры: тестовые микросхемы и кристаллы. Основной принцип построения тестовых микросхем состоит в том, что она должна содержать все конструктивные элементы реальной микросхемы, а также должна быть изготовлена по тому же технологическому маршруту, кроме того, тестовая микросхема должна обеспечивать удобство контроля во время испытаний и оценку качества технологического процесса. Применение тестовых микросхем и кристаллов предоставляет возможность организовать эффективный технологический контроль производства интегральных микросхем и сократить трудоемкость при проведении испытаний на надежность больших интегральных схем.

Будем исследовать способность критериев распознать грубый промах при объеме выборки  $n=10$ .

Данный объем выбран, так как в промышленности существуют производства, по технологическим особенностям которых невозможно обеспечить большой объем контрольной выборки. Например, при производстве кристаллов интегральных микросхем из-за специфики топологии пластин на ней имеются от 5 до 10 тестовых ячеек, измерения в которых должны с некоторой вероятностью отражать поведение одноименных параметров 400-5000 рабочих кристаллов. Такие же ситуации встречаются и в других областях (медицина, экономика, педагогика и др.), в которых из-за особенностей предметной области невозможно набрать достаточно большой объем данных. Анализу подлежали практически все известные на сегодняшний день методы нахождения аномального измерения:

1. Метод Ирвина.
2. Критерий Стьюдента.
3. Критерий наибольшего абсолютного отклонения.
4. Критерий максимального относительного отклонения.
5. Критерий Романовского.
6. Метод вариационного размаха.
7. Критерий 3 Сигм.
8. Критерий Райта.
9. Критерий Граббса.
10.  $Q$ -критерий (Диксона).
11. Критерий Львовского.

12. Критерий Шовене.
13. Критерий Дэвида.
14. Критерий Хоглина-Иглевича.
15.  $L$ -критерий (Критерий Титъена-Мура).
16. Критерий Смоляка-Титаренко.
17. Критерий Бродского-Быцаня-Власенко.
18. Критерий Кимбера.

Все вышеперечисленные критерии предлагаются различными исследователями для определения грубых промахов в выборочных совокупностях сравнительно небольшого объема.

Изложение хода эксперимента будем вести на примере выборки, полученной при производстве кристаллов интегральных микросхем. В нашем распоряжении имеется выборка контрольных параметров объемом  $n=10$ .

$$x_i = 13,0; 13,2; 13,5; 13,7; 13,7; 14,2; 14,3; 14,5; 14,6; 16,5.$$

Проведём исследование данной выборки на наличие грубых промахов каждым из выбранных критериев.

1. Метод Ирвина.

Метод применяется для поиска грубых промахов в выборках с известной дисперсией распределения.

$$\text{Имеем: } \bar{x} = 14,12; s^2 = 0,99; s = 0,995.$$

Вычисляем (для  $x_i = 16,5$ ):

$$\tau = \frac{x_n - x_{n-1}}{s} = \frac{16,5 - 14,6}{0,995} = 1,909.$$

По таблице критических значений критерия Ирвина для  $n=10$  и уровня значимости  $\alpha = 0,95$  находим  $\tau(0,95) = 1,46$ .

Так как  $\tau = 1,909 > \tau(0,95) = 1,46$ , значение  $x_{10} = 16,5$  можно считать грубым промахом.

2. Критерий Стьюдента.

Все разновидности критерия Стьюдента являются параметрическими и основаны на дополнительном предположении о нормальности выборки данных.

Используя полученные ранее значения среднего арифметического и дисперсии, вычисляем нормированное отклонение:

$$\tau = \frac{\max |x_i - \bar{x}|}{s} = \frac{16,5 - 14,12}{0,995} = 2,391.$$

Затем вычисляем критерий отбраковки:

$$\tau_{кр}(q, \nu) = \frac{t(q, \nu) \sqrt{n-1}}{\sqrt{(n-2) + [t(q, \nu)]^2}},$$

где  $t(q, \nu)$  – критерий Стьюдента с  $q$  уровнем значимости и  $\nu = n - 2$  числом степеней свободы.

Подозреваемое число следует оставить в выборке, если  $\tau < \tau_{кр}(5\%, \nu)$ ; число можно оставить или выбросить по усмотрению исследователя, если  $\tau(5\%, q) < \tau < \tau_{кр}(0,1\%, q)$ ; а если  $\tau > \tau_{кр}(0,1\%, \nu)$ , то число нужно обязательно исключить из выборки.

В рассматриваемом примере  $\tau_{кр}(5\%, 8) = 1,87$ , а  $\tau_{кр}(0,1\%, 8) = 2,447$ , таким образом решение о наличии или отсутствии грубого промаха в данной выборке предоставляется исследователю, однако учитывая тот факт, что рассчитанное значение очень близко к табличному с уровнем значимости  $q = 0,1\%$ , а при расчете остальных критериев мы принимаем уровень значимости равный 95%, то примем решение оставить максимальный элемент в выборке.

### 3. Критерий наибольшего абсолютного отклонения.

Для нахождения грубого промаха вычисляем:

$$\tau = \frac{\max(x_i - \bar{x})}{s} = \frac{16,5 - 14,12}{0,995} = 2,391.$$

Из таблицы критических значений критерия наибольшего абсолютного отклонения выбираем значение для доверительной вероятности  $\alpha = 0,95$  и объема выборки  $n = 10$ . Так как табличное значение  $\tau(0,95) = 2,414 > \tau = 2,391$ , то гипотеза о наличии выбросов не находит подтверждения.

### 4. Критерий максимального относительного отклонения.

Для обнаружения грубого промаха вычисляем:

$$\tau = \frac{\max(x_i - \bar{x})}{s} = \frac{16,5 - 14,12}{0,995} = 2,391.$$

Из таблицы критических значений критерия максимального относительного отклонения выбираем значение для доверительной вероятности  $\alpha = 0,95$  и объема выборки  $n = 10$ . Так как табличное значение  $\tau(0,95) = 2,29 < \tau = 2,391$ , то гипотеза о признании значения  $x_{10} = 16,5$  грубым промахом подтверждается.

## 5. Критерий Романовского.

Гипотеза о наличии грубых погрешностей в подозрительных результатах подтверждается, если выполняется неравенство:

$$|x_i - \bar{x}| \geq t_p \cdot s,$$

где  $t_p$  – квантиль распределения Стьюдента при заданной доверительной вероятности с числом степеней свободы  $k = n - k_n$  ( $k_n$  – число подозрительных результатов наблюдений).

Точечные оценки распределения среднего значения и среднеквадратического отклонения вычисляются без учета подозрительных результатов наблюдений.

Отсюда получаем:  $\bar{x} = 13,856$ ;  $s = 0,328$ .

Проверим выполнение неравенства при  $t_p(0,95;9) = 2,28$ :

$$|16,5 - 13,856| \geq 2,28 \cdot 0,328, \text{ то есть } 2,644 > 1,305.$$

Так как неравенство выполнилось, гипотеза о наличии грубого промаха считается подтвержденной.

## 6. Метод вариационного размаха.

Является одним из самых простых методов исключения грубого промаха в измерениях. Для его использования определяют размах вариационного ряда упорядоченной совокупности наблюдений:  $R = 16,5 - 13,0 = 3,5$ .

Затем производят проверку подозрительного значения из ряда, используя неравенство:  $\bar{x} - z \cdot R < x_k < \bar{x} + z \cdot R$ , где  $\bar{x}$  – выборочное среднее арифметическое значение, вычисленное после исключения предполагаемого промаха;  $z$  – критериальное значение.

Нулевую гипотезу (об отсутствии грубого промаха) принимают, если указанное неравенство выполняется.

Проверим выполнение неравенства, используя полученные ранее значения среднего и размаха и табличное критериальное значение  $z = 1,3$ :

$$13,856 - 1,3 \cdot 3,5 < 16,5 < 13,856 + 1,3 \cdot 3,5,$$

$$\text{таким образом } 9,306 < 16,5 < 18,406.$$

Неравенство выполняется, следовательно, проверяемое значение не является грубым промахом.

## 7. Критерий 3 сигм.

Сущность правила трех сигм: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднеквадратического отклонения:

$$|x_k - \bar{x}| \geq 3 \cdot s, \text{ где } x_k - \text{проверяемое значение.}$$

Проверим выполнение неравенства:

$$|16,5 - 14,12| \geq 3 \cdot 0,995, \text{ то есть } 2,38 \geq 2,98.$$

Так как неравенство не выполняется, то гипотеза о наличии грубой погрешности не находит подтверждения.

#### 8. Критерий Райта.

Критерий Райта аналогичен предыдущему примеру и основан на том, что если остаточная погрешность больше четырех среднеквадратических отклонений, то результат является грубой погрешностью и должен быть исключен при дальнейшей обработке.

Очевидно, что результат проверки неравенства по критерию Райта также не подтвердит гипотезу о наличии погрешности в данной выборке.

#### 9. Критерий Граббса.

Статистики критерия Граббса предусматривают возможность проверки на наличие в выборке максимальных, либо минимальных аномальных наблюдений.

Используя найденные ранее величины вычисляем для максимального значения:

$$\tau = \frac{x_n - \bar{x}}{s} = \frac{16,5 - 14,6}{0,995} = 2,391.$$

По таблице критических значений статистик Граббса для  $n=10$  и уровня значимости  $\alpha = 0,95$  находим  $\tau(0,95) = 2,441$ .

Так как  $\tau = 2,391 < \tau(0,95) = 2,441$ , гипотеза о принятии  $x_{10} = 16,5$  грубым промахом отклоняется.

#### 10. Q-критерий (Диксона).

Широко используется [125] для быстрого выявления выпадающих наблюдений в выборках небольшого объема по отношению размаха и подразмахов.

Вычисляем:

$$Q = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{16,5 - 14,6}{16,5 - 13,0} = 0,543.$$

Из таблицы критических значений статистик Диксона находим  $Q(0,95) = 0,48$ .

Так как рассчитываемое значение больше табличного, то значение  $x_{10} = 16,5$  считается грубым промахом.

### 11. Критерий Львовского.

Предлагается [49] для нахождения аномальных наблюдений при малом числе измерений и с этой целью вводит в метод максимального относительного отклонения уточняющий коэффициент.

Вычисляем значение критерия по формуле:

$$\tau = \frac{x_n - \bar{x}}{\sqrt{\frac{n-1}{n}}s} = \frac{16,5 - 14,6}{0,949 \cdot 0,995} = 2,52.$$

Сравниваем его с табличным значением  $\tau(0,95)=2,29$  и определяем, что значение  $x_{10} = 16,5$  является грубым промахом.

### 12. Критерий Шовене.

Критерий используется для поиска грубых ошибок, в выборках распределенных как по нормальному, так и по экспоненциальному [160] закону распределения. Согласно критерию Шовене элемент выборки  $x_i$  объема  $n$  является выбросом, если вероятность его отклонения от среднего значения не более  $1/(12 \cdot n)$ .

Вычисляем значение критерия:

$$K = \frac{x_n - \bar{x}}{s} = 2,391.$$

Сравниваем его с табличным значением для объема выборки  $n=10$ .

$$K=2,391 > K_{табл}=1,96$$

По результатам сравнения приходим к выводу, что крайнее значение в выборке является выбросом.

### 13. Критерий Дэвида.

Критерий является модификацией критерия Граббса [36].

Высчитываем:

$$\tau = \frac{x_n - \bar{x}}{s} = 2,391$$

и сравниваем с табличным значением:

$\tau = 2,391 < \tau(0,95) = 2,89$ , из чего следует, что гипотеза о наличии выброса отклоняется.

### 14. Критерий Хоглина-Иглевича.

Критерий основывается на порядковых статистиках и признает наблюдение выбросом, если его значение находится вне интервала, ограниченного величинами:

$$(1+k)x_{[l]} - kx_{[n+1-l]} \text{ и } (1+k)x_{[n+1-l]} - kx_{[l]},$$

где  $x_i$  -  $i$ -ая порядковая статистика (т.е.  $i$ -ый по величине член выборки, упорядоченной по возрастанию).

Для выбора значения  $l$  используем формулу:

$$l = \frac{1}{2} \left[ \frac{n+3}{2} \right], \text{ где } [...] \text{ целая часть числа.}$$

Для  $n=10$  и  $\alpha = 0,95$ , найдем табличное значение  $k=2,4$ .

Таким образом, получаем:

$$(1+k)x_{[l]} - kx_{[n+1-l]} = x_3 - 2,4 \cdot (x_8 - x_3) = 11,1;$$

$$(1+k)x_{[n+1-l]} - kx_{[l]} = x_8 + 2,4 \cdot (x_8 - x_3) = 16,9. \text{ Все значения данной выборки}$$

попадают в рассчитанный диапазон, что позволяет с вероятностью 0,95 утверждать отсутствие выбросов в предложенной выборке.

#### 15. $L$ -критерий (Критерий Титъена-Мура)

Критерий Титъена-Мура является обобщением критерия Граббса на случай выявления нескольких выбросов в выборке [36, 147].

Для выделения  $k$  наибольших выбросов используется статистика:

$$L_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \text{ где } \bar{x}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i.$$

Рассчитав по формуле значение  $L_k$  получим:

$$L_k = \frac{13,856}{14,12} = 0,331.$$

Сравним полученный результат с табличным значением критерия:

$$L_k = 0,331 < L_k(0,95) = 0,418.$$

Так как рассчитываемое значение меньше табличного, следует признать наличие выброса значимым с достоверностью 0,95.

#### 16. Критерий Смоляка-Титаренко.

Данный критерий предложен для обнаружения грубых промахов в ограниченных выборках, извлекаемых из генеральных совокупностей, распределенных по экспоненциальному закону.

Критерий основан на отношении величины наибольшего члена к выборочному среднему.

Вычисляем (для  $x_i=16,5$ ):

$$C_1 = \frac{x_n}{\bar{x}} = \frac{16,5}{14,12} = 1,169.$$

При  $\alpha = 0,95$  и  $n=10$  находим табличное значение  $C_1(0,95)=4,449$ .

Так как  $C_1$  не превышает соответствующего критического значения, то наибольшее значение не может быть признано выбросом.

#### 17. Критерий Бродского-Быцаня-Власенко.

Критерий является аналогом критерия Диксона для случая экспоненциального распределения. Рассчитаем статистику критерия проверки на выброс для наибольшего значения выборки:

$$z = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{16,5 - 14,6}{16,5 - 13,0} = 0,543.$$

Вычисляем:

$$P = (n-1)! (1-z)^{n-2} \prod_{j=1}^{n-2} \frac{1}{1+j(1-z)} = 1,538.$$

Так как рассчитанная величина  $P=1,538$  значительно больше уровня значимости  $\alpha = 0,95$ , то можно утверждать, что наибольшее значение не является выбросом.

#### 18. Критерий Кимбера.

Кимбером была предложена последовательная процедура для выявления нескольких выбросов в выборке из экспоненциального распределения.

Для проверки гипотезы вычисляем значение критерия для последних двух элементов выборки:

$$S_1 = \frac{x_{10}}{\sum_{i=1}^{10} x_i} = \frac{16,5}{141,2} = 0,117;$$

$$S_2 = \frac{x_9}{\sum_{i=1}^9 x_i} = \frac{14,6}{124,7} = 0,117.$$

Используя табличные значения для  $\alpha = 0,95$ ,  $n=10$ , убеждаемся, что  $S_1=0,117 < S_1(0,95)=0,48$  и  $S_2=0,117 < S_1(0,95)=0,43$ .

Следовательно, на основании полученных данных, можно утверждать, что верхних выбросов в выборке не наблюдается.

Как видно из полученных вычислений, существующие критерии используют различные подходы и обладают различной мощностью при распознавании аномальных измерений в выборках малого объема.

Таким образом, рассчитав для одной и той же одномерной малой выборки, объемом  $n=10$ , 18 методов нахождения грубого промаха, применяемых различными исследователями при решении задачи поиска аномального наблюдения при небольшом числе испытаний, получили результаты, приведённые в таблице 2.1.

Таблица 2.1. - Результаты выявления грубой ошибки

Наименование критерия	Выявление ошибки
Метод Ирвина	Выявлена
Критерий Стьюдента	На усмотрение
Критерий наибольшего абсолютного отклонения	Не выявлена
Критерий максимального относительного отклонения	Выявлена
Критерий Романовского	Выявлена
Метод вариационного размаха	Не выявлена
Критерий 3 Сигм	Не выявлена
Критерий Райта	Не выявлена
Критерий Граббса	Не выявлена
Q-критерий (Диксона)	Выявлена
Критерий Львовского	Выявлена
Критерий Шовене	Выявлена
Критерий Дэвида	Не выявлена
Критерий Хоглина-Иглевича	Не выявлена
L-критерий (Критерий Титьена-Мура)	Выявлена
Критерий Смоляка-Титаренко	Не выявлена
Критерий Бродского-Быцаня-Власенко	Не выявлена
Критерий Кимбера	Не выявлена

Подобным образом было рассмотрено несколько сотен выборок, объемом  $n=10$ , что позволило сделать следующие выводы [64]:

1. Такие критерии как: наибольшего абсолютного отклонения, вариационного размаха, 3 сигм, Райта, Граббса, Дэвида, Хоглина-Иглевича, Смоляка-Титаренко, Бродского-Быцаня-Власенко и Кимбера не могут достоверно определить наличие грубого промаха в выборке малого объема.

2. Критерии Стьюдента и максимального относительного отклонения довольно точно определяют вероятность грубого промаха в выборке малого объема, но оставляют

решение об исключении из выборки на усмотрение исследователя. При применении этих методов рекомендуется проверить выборку дополнительным критерием.

3. Критерии Романовского и Шовене хорошо распознают наличие грубого промаха в выборке малого объема, однако, в некоторых случаях выдают за выброс не аномальный результат.

4. Лучше всего в определении грубых промахов в малых выборках показали себя критерии Ирвина, Львовского, Диксона и Титьена-Мура. Учитывая различные подходы для определения грубого промаха, лежащие в основе этих методов, для получения достоверного результата рекомендуется проверять выборку как минимум двумя из данных критериев.

Однако, малыми являются выборки, содержащие менее 20 элементов, в связи с чем существует необходимость проверки полученных результатов на выборках различного объема ( $n=5-20$  элементов).

Для решения поставленной задачи был проведён статистический эксперимент, в ходе которого для каждого из исследуемых значений  $n$  с помощью генератора случайных чисел были получены массивы данных, содержащие несколько тысяч выборок с нормальным законом распределения, и заранее заданными средними величинами и дисперсиями. После чего, каждая выборка упорядочивалась по возрастанию, и на место максимального элемента добавлялось значение, из диапазона  $[1\sigma; 5\sigma]$ , где  $\sigma$  – это среднеквадратическое отклонение, то есть показатель рассеивания случайной величины. Из теории математической статистики известно, что значения нормально распределенной случайной величины с вероятностью более 99% лежат в интервале  $(\bar{x}-3\sigma; \bar{x}+3\sigma)$ , где  $\bar{x}$  – истинная величина среднего арифметического. При этом в интервале  $(\bar{x}-2\sigma; \bar{x}+2\sigma)$  находится более 95% значений, а в интервале  $(\bar{x}-1\sigma; \bar{x}+1\sigma)$  более 68%. Таким образом, значения выше  $\bar{x}+3\sigma$  с большой вероятностью окажутся аномальными измерениями, в то время как значения менее  $\bar{x}+2\sigma$  таковыми являться не будут.

Для исследования были выбраны критерии, показавшие наилучшие результаты в обнаружении грубых промахов при объеме выборки в 10 элементов: Диксона, Ирвина, Титьена-Мура, Львовского, Романовского и Шовене. Также была рассмотрена мощность критерия Граббса, так как его значение при известном генеральном СКО рекомендуется использовать в качестве базового критерия в ГОСТ Р ИСО 5725-2–2002 «Точность (правильность и прецизионность) методов и результатов измерений».

Проиллюстрируем полученные ранее выводы данными, полученными в ходе статистического эксперимента (Рисунок 2.1). По оси  $X$  на графике располагаются

значения величины  $i$  из выражения  $\bar{x}+i\sigma$ , а по оси  $Y$  процент обнаружения аномальных измерений.

На графике видно, что при  $n=10$ , процент обнаружения аномальных измерений критериями Диксона, Ирвина и Львовского совпадают с точностью до 1%. Критерий Титьена-Мура обладает пониженной мощностью при обнаружении аномальных измерений, что компенсируется меньшим количеством ошибок I типа (принятие гипотезы о наличии грубой ошибки в случае её отсутствия). Оставшиеся критерии использовать не рекомендуется, из-за низкой мощности (критерий Граббса при неизвестном генеральном СКО), либо высоком уровне ошибок I типа (так, критерии Шовене и Романовского считают грубой ошибкой значение  $\bar{x}+2\sigma$  в 39% и 60% соответственно, что не может быть признано удовлетворительным). Таким образом, нами были подтверждены результаты предыдущего исследования.

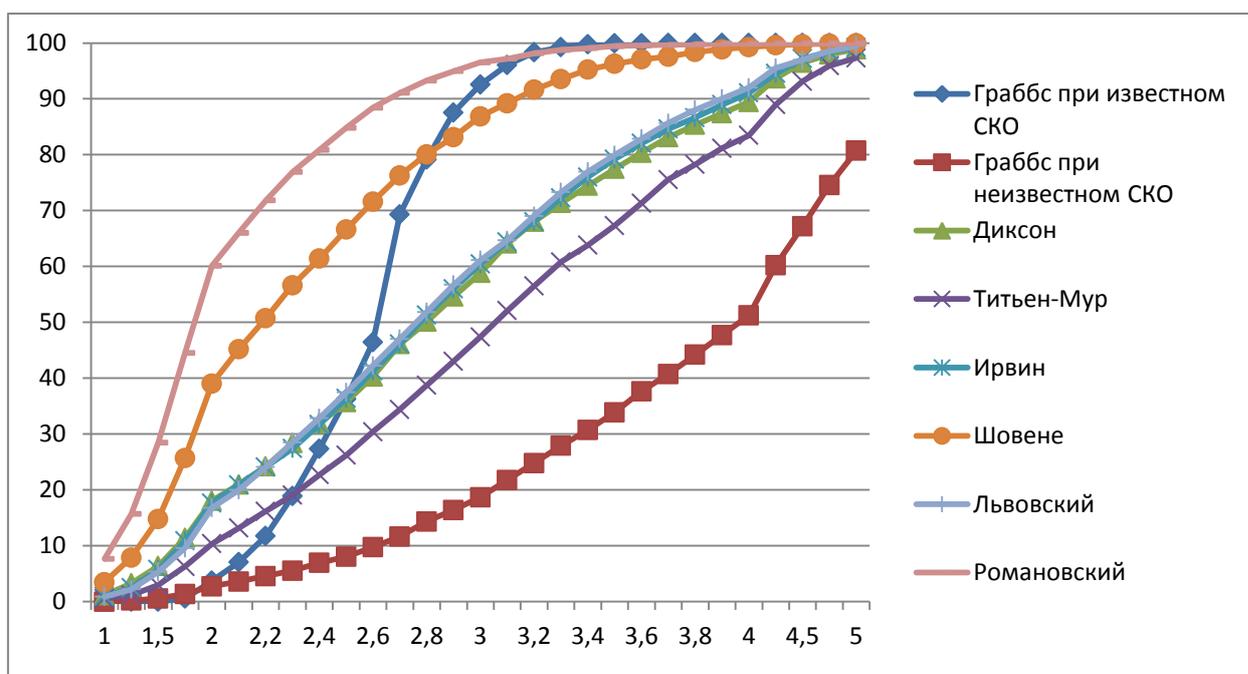


Рисунок 2.1 – Мощность критериев обнаружения аномальных измерений при  $n=10$

Проиллюстрируем, как изменяется мощность критериев обнаружения аномальных измерений при уменьшении исследуемого объема выборки, для чего на рисунках 2.2 и 2.3 изображены графики, полученные по данным статистического эксперимента для  $n=7$  и  $n=5$ , соответственно.

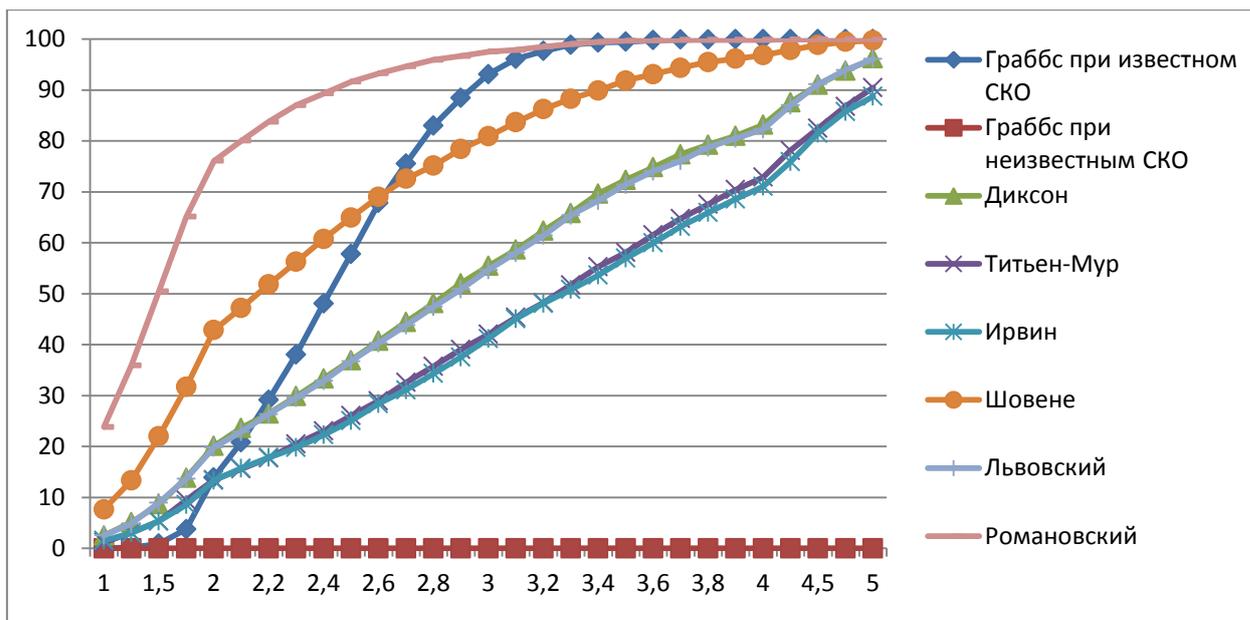


Рисунок 2.2 – Мощность критериев обнаружения аномальных измерений при  $n=7$

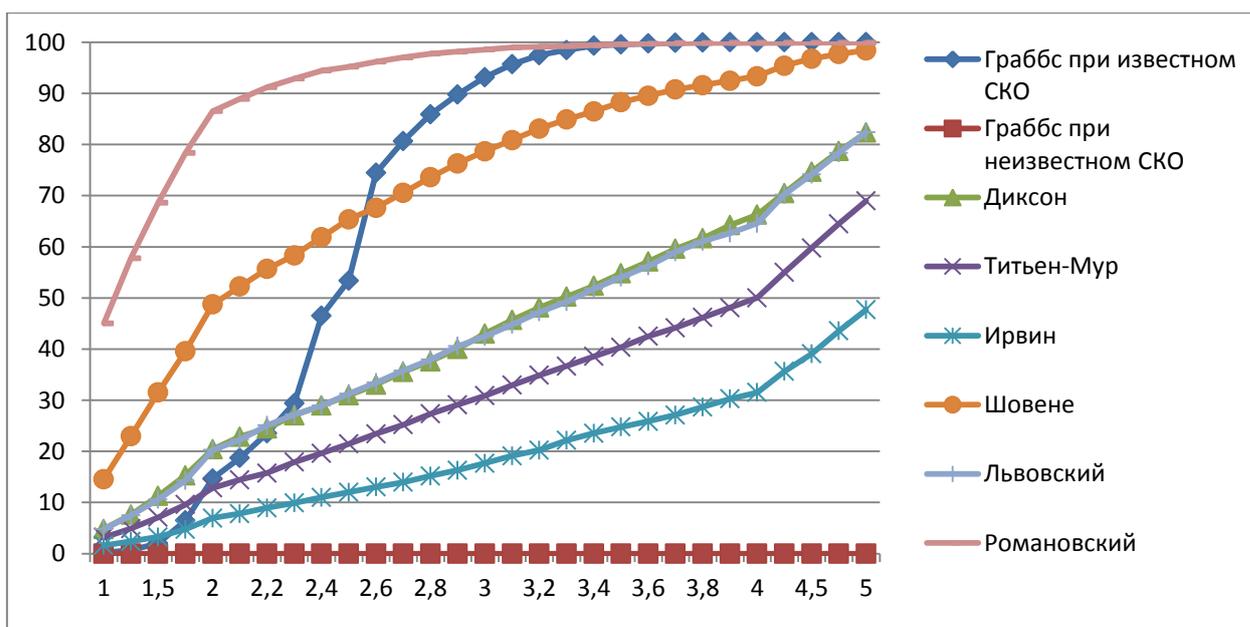


Рисунок 2.3 – Мощность критериев обнаружения аномальных измерений при  $n=5$

Как видно из полученных графиков, критерий Граббса при неизвестном генеральном СКО не способен обнаруживать аномальные измерения при объемах выборок менее 10 элементов. При использовании критериев Шовене и Романовского количество ошибок I типа увеличивается с уменьшением объёмов выборок, что указывает на зависимость данного показателя от объёма выборки. Данные критерии считают значение  $\bar{x}+2\sigma$  аномальным результатам, при объёма выборки  $n=7$  в 42,84% (Шовене) и

76,08% (Романовского), а при объеме выборки  $n=5$  в 48,76% (Шовене) и 86,52% (Романовского). Также на графиках видна прямая зависимость мощности критерия Ирвина от объема исследуемой выборки.

На основании полученных данных, можно рекомендовать для обнаружения аномальных измерений в выборках объемом менее 10 значений использовать, либо критерий Диксона, либо критерий Львовского, как дающие оптимальные и статистически неразличимые результаты.

Проиллюстрируем, как изменяется мощность критериев обнаружения аномальных результатов при увеличении исследуемого объема выборки. Как видно из графиков, полученных по данным статистического эксперимента, для  $n=13$  (Рисунок 2.4),  $n=16$  (Рисунок 2.5) и  $n=20$  (Рисунок 2.6), наилучшие результаты по обнаружению аномальных измерений в выборках при  $n>10$  показывают критерии Ирвина и Львовского.

Следует отметить, что при увеличении объемов исследуемых данных значительно возрастает точность критерия Граббса при неизвестном среднеквадратическом отклонении, а также значительно уменьшается количество ошибок I типа у критериев Шовене и Романовского.

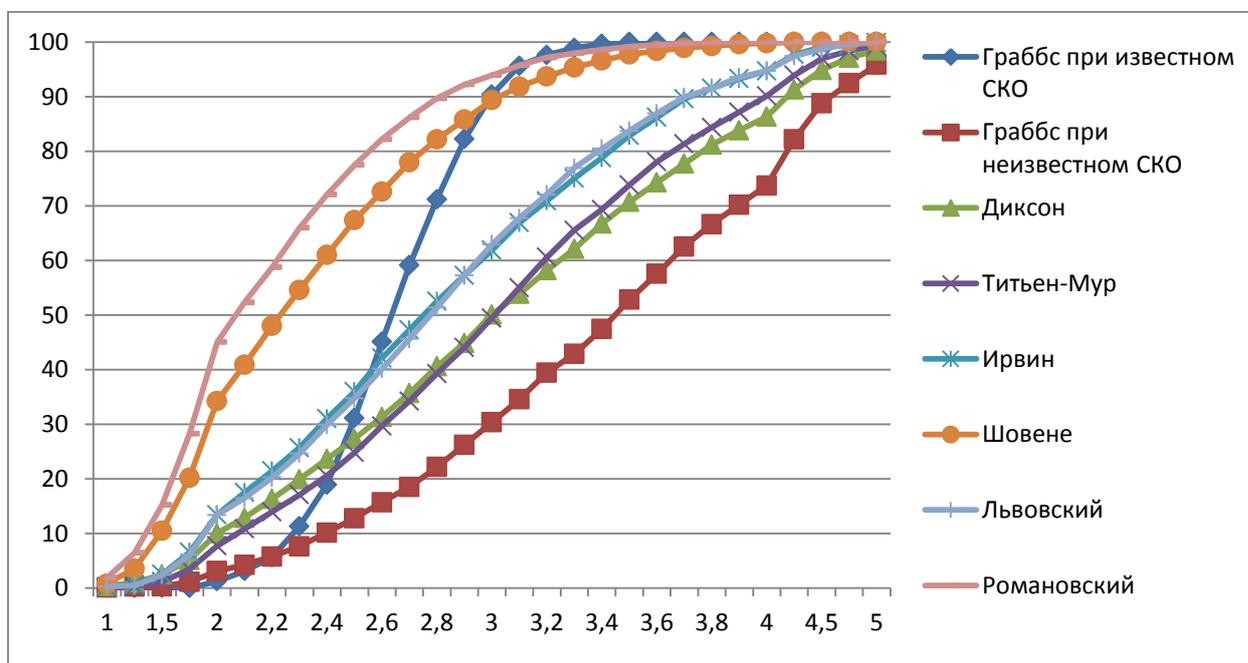


Рисунок 2.4 – Мощность критериев обнаружения аномальных измерений при  $n=13$

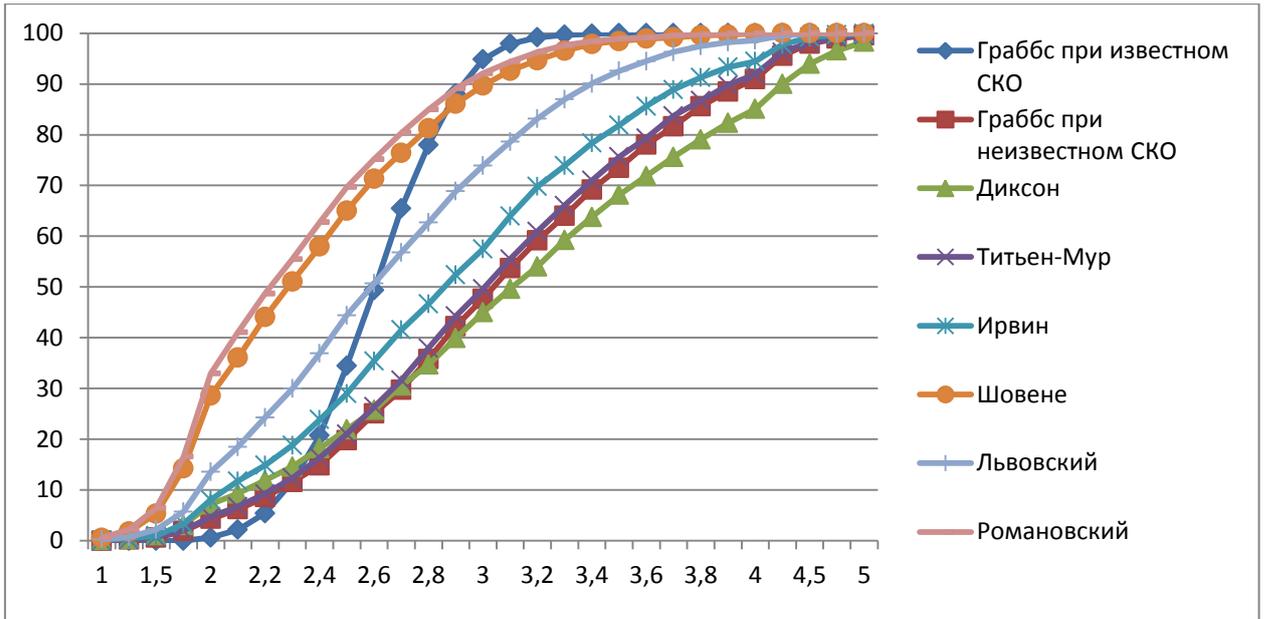


Рисунок 2.5 – Мощность критериев обнаружения аномальных измерений при  $n=16$

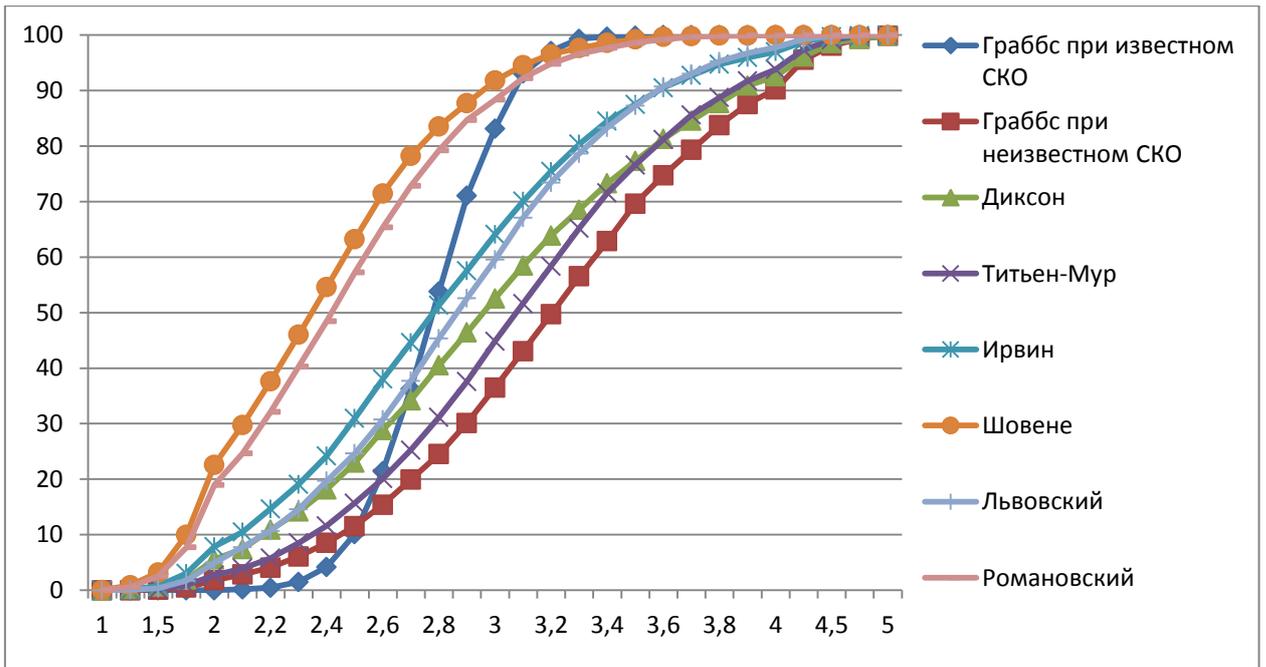


Рисунок 2.6 – Мощность критериев обнаружения аномальных измерений при  $n=20$

Для сравнения, на рисунке 2.7 приведена мощность исследуемых критериев для  $n=30$ . Полученные данные подтверждают тенденции, описанные ранее.

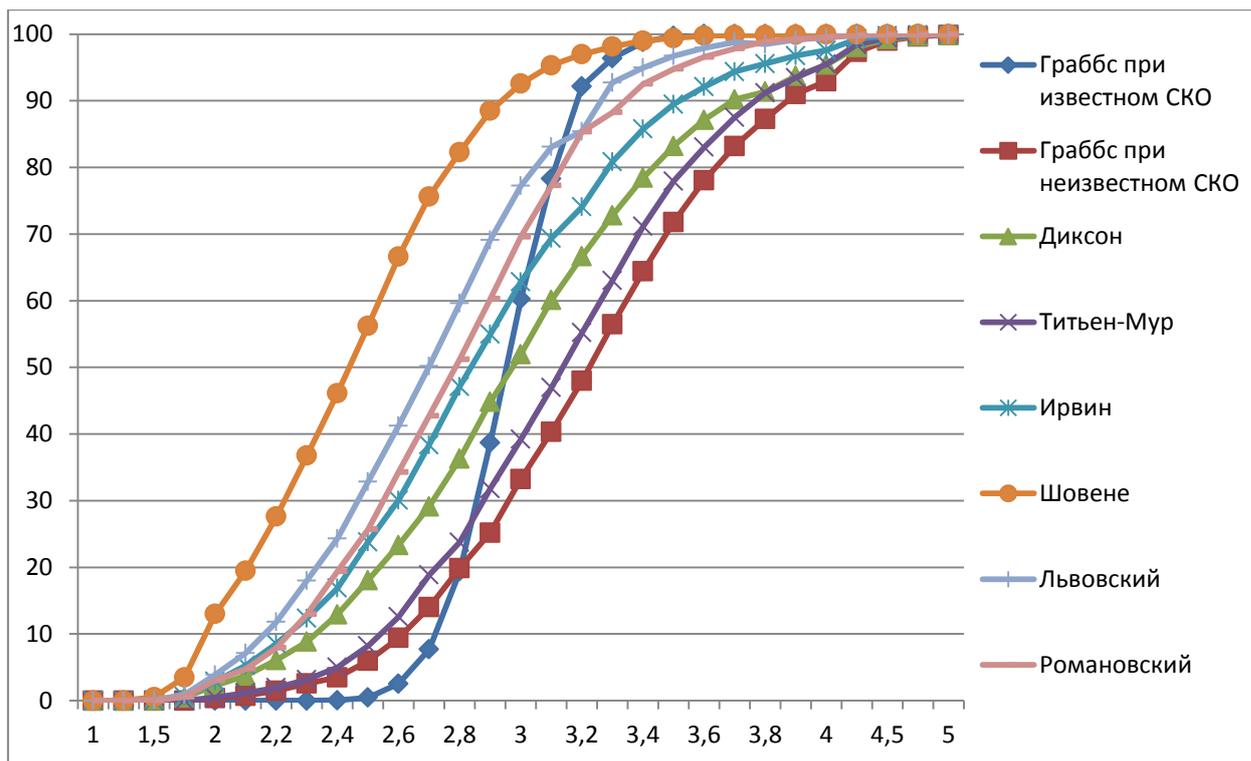


Рисунок 2.7 – Мощность критериев обнаружения аномальных измерений при  $n=30$

Таким образом, на основании проведённого исследования установлено:

1. Оптимальным критерием для обнаружения аномальных результатов наблюдений для случая малого количества исходной информации является критерий Львовского, который показывает оптимальные результаты вне зависимости от объема данных, что связано с использованием поправочного коэффициента, основанного на значении величины выборки. При этом точность критерия возрастает с увеличением количества исследуемой информации.

2. Критерий Диксона можно рекомендовать при обработке выборок с объемом  $n \leq 10$ , так как он является простым в расчетах и даёт результаты статистически неотличимые от критерия Львовского. При этом следует обратить внимание, что для объемов выборок  $n > 10$  исследовалась мощность статистик Диксона для проверки одного сомнительного наблюдения независимо от следующего по величине, как это рекомендуется в [36, с.550], при данных объемах выборок мощность критерия немного снижается, однако его использование возможно наряду с другими отобранными критериями. Направлением дальнейших исследований может быть проверка данных рекомендаций и изучение мощности различных критериев Диксона в зависимости от объема выборки.

3. Мощность критерия Ирвина имеет прямую зависимость от объема выборки, что связано с использованием величины выборочной дисперсии, которая при малом объеме выборки достоверно не отражает величину дисперсии генеральной совокупности. Однако, критерий Ирвина можно применять для случаев выборок объемом  $n \geq 10$ .

4. Критерий Титьена-Мура может быть рекомендован для обнаружения аномальных значений в малых выборках при  $n > 5$ , так как он довольно хорошо распознает ошибки в значениях более  $\bar{x} + 4\sigma$  и обладает наименьшим количеством ошибок I типа.

5. Критерий Граббса при неизвестном генеральном СКО возможно применять при объемах выборки  $n \geq 15$ .

6. Критерии Шовене и Романовского не рекомендуется применять при объеме исследуемых данных  $n \leq 20$ .

В данном случае, кроме проблемы устойчивости критериев именно к объему выборки, также добавляется проблема устойчивости критерия к отклонению от нормального закона распределения, т.к. на данный момент не существует методики, позволяющей с достаточной вероятностью определять вид закона распределения малой выборки. Следует отметить, что критерий Диксона может быть трансформирован для применения поиска грубых промахов в выборках с экспоненциальным законом распределения или распределением Вейбулла. Таким образом, если априорные данные позволяют утверждать вид закона распределения выборки, данный критерий можно считать оптимальным [68, 70].

Все вышесказанное относится к одномерным выборкам, для парных выборок методов расчетного выявления грубых промахов не существует. Между тем парные выборки составляют основу множества других статистических методов (уравнений парной и множественной регрессии, корреляционных плеяд и т.п.), расчеты по которым могут быть значительно искажены при наличии даже одного грубого промаха.

В работе [23] доказывается, что при расчете значений линейной корреляционной связи целесообразно использовать модифицированный индекс Фехнера, как меру тесноты, свободную от влияния вида закона распределения выборок и наличия в них грубых промахов. Однако, данный критерий работает только с данными объемом  $n > 40$ , поэтому не может быть использован для случая малых выборок. В связи с этим, для обнаружения парных грубых промахов при малом объеме исходной информации, рекомендуется проводить построение таблиц двумерного распределения, опираясь на методику расчета коэффициента корреляции по методу П.Л. Чебышева [19, 91].

## 2.2 Линейная корреляция в выборках малого объема

При решении задач вероятностно-статистического оценивания пассивных данных большой интерес представляет задача проверки гипотезы о значимости корреляционной связи между случайными величинами, т.е. значимости отклонения коэффициента корреляции от нуля. Это обусловлено тем, что определение коэффициента связи при отборе факторов позволяет существенно сократить факторное пространство при сохранении того же качества получаемой математической модели. Коэффициент корреляции при этом показывает, насколько связь между случайными величинами близка к строго линейной. При этом применение корреляционного анализа предъявляет к исходной информации определенные требования – число наблюдений или объектов должно быть достаточно большим, а исходные данные должны быть однородными, то есть они должны отражать наиболее характерные черты изучаемой совокупности. Кроме того, данные должны быть выражены количественно и измерены достаточно точно. Однако, при решении большого количества задач не всегда можно получить достаточно большой набор данных для их последующего анализа.

Существуют различные методы, позволяющие повысить точность расчета коэффициента корреляции при ограниченном наборе данных, при этом отсутствует точная методология, какие методы и критерии лучше применять и в каких случаях.

Для определения наличия зависимости между двумя переменными используется коэффициент корреляции. Зависимость подразумевает любые согласованные изменения, которые могут объясняться различного рода причинами. Если переменные – количественные и равноценные в своих независимых наблюдениях, то важнейшими эмпирическими мерами тесноты их линейной взаимосвязи являются коэффициент прямой корреляции знаков австрийского психолога Г.Т.Фехнера и коэффициент парной корреляции английского статистика-биометрика К.Пирсона.

Наиболее точной мерой линейной корреляционной связи является коэффициент корреляции Пирсона, рассчитанный по парной выборке достаточно большого объема, распределённой по нормальному закону и не содержащей грубых промахов. Однако, при несоблюдении хотя бы одного из этих условий величина коэффициента корреляции может значительно отличаться от реальной [114, 150]. Согласно определению, коэффициент корреляции представляет собой первый основной смешанный момент. Для вычисления коэффициента корреляции Пирсона надо найти первый начальный смешанный момент и первые два начальных момента каждого ряда распределения. Коэффициент корреляции по

своей абсолютной величине не может быть больше единицы. Знак перед коэффициентом корреляции указывает на положительную, либо отрицательную связь.

Коэффициент Фехнера – это оценка степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от средних значений факторного и результативного признаков. Для определения индекса Фехнера вычисляют средние  $X$  и  $Y$ , а затем для каждой пары определяют знаки отклонений. Данный критерий ведет себя так же, как коэффициент корреляции, то есть изменяется в пределах от -1 до 1.

Преимуществами индекса Фехнера являются: простота его вычисления, более низкая зависимость от вида закона распределения и от наличия в выборке грубых промахов. Однако, при этом его точность хуже, чем у коэффициента корреляции Пирсона и относительное их совпадение наблюдается только при значениях, больших 0,8 [19]. При меньших значениях наблюдается заметное отклонение от аналогичных величин коэффициента корреляции. Этот недостаток существенно уменьшается, если использовать модифицированный индекс Фехнера (МИФ) [18], рассчитанный по формуле

$$f^* = \pm \sqrt{\frac{|v-w|}{v+w}} \pm 0,051, \quad (2.1)$$

в которой знаки “+” берутся в обоих случаях при  $v > w$ , а знаки “-” – при  $v < w$ , где  $v$  – количество совпадающих знаков  $sign(X_i - X_{cp})$  и  $sign(Y_i - Y_{cp})$ , а  $w$  – различных.

Исследования робастности индекса Фехнера к грубым промахам (до 5% от объема выборки) показало значительно меньшее их влияние по сравнению с влиянием на коэффициент корреляции (в 2,0 – 2,5 раза). Аналогичные исследования робастности индекса Фехнера к изменениям закона распределения выборки каждого из двух параметров показали незначительные дополнительные отклонения от результатов, полученных при их нормальном распределении.

МИФ практически совпадает с коэффициентом корреляции при  $r > 0,3$  в парных выборках объемом  $n > 40$ , не зависит от вида закона распределения и достаточно устойчив против наличия в выборке грубых промахов [18].

Проблема объема выборки при исследовании коэффициента корреляции поднимается различными исследователями [107, 127, 151], однако оптимального метода нахождения значения корреляционной связи для выборок малого объема на данный момент не предложено.

Для случая малых выборок применение коэффициента Пирсона не дает приемлемых результатов в связи с невозможностью точно определить вид закона

распределения и обнаружить грубые промахи. Несмотря на это, он используется при анализе малых выборок [83]. Применение же модифицированного индекса Фехнера не представляется возможным по причине малого числа наблюдений.

Для решения этой проблемы предлагается применять процедуры виртуального увеличения выборки, такие как «bootstrap» [107], либо метод точечных распределений.

Будем исследовать способность коэффициентов корреляции верно определить уровень линейной связи для парных выборок при 10 наблюдениях. Выбранный объем был обоснован в первом пункте данной главы.

Для проведения исследования из выборок большого объема, полученных при производстве кристаллов интегральных микросхем, распределенных по нормальному закону распределения, по таблице равномерно распределенных случайных величин получим выборки объемом 10 элементов. После чего рассчитаем коэффициенты корреляции для парных выборок большого объема и эквивалентных им выборок малого объема.

Далее применим процедуру рандомизации и оценим наивероятнейший коэффициент корреляции методом «bootstrap», как наивероятнейший по экспериментальным данным и нескольким искусственным выборкам, организованным из строк исходной выборки с помощью таблицы равномерно распределенных случайных чисел, по методике изложенной в [19]. После чего определим коэффициенты корреляции, используя метод точечных распределений и классический метод Чебышева.

Рассмотрим алгоритм расчета на конкретном числовом примере.

Пусть дана парная выборка объемом  $n=10$  (таблица 2.2)

Таблица 2.2 – Исходная парная выборка

Номер измерения, $i$	1	2	3	4	5	6	7	8	9	10
$X_i$	247,03	231,19	223,24	221,70	238,64	240,42	237,47	227,54	220,44	240,07
$Y_i$	206,15	200,78	199,38	193,02	200,11	207,82	201,29	198,47	197,87	202,10

Согласно методу точечных распределений каждое измерение выборки считается центром виртуального распределения с заданным законом распределения.

Для создания корреляционной таблицы необходимо первоначально определить границы существования выборок каждого параметра ( $a_X$ ,  $b_X$  и  $a_Y$ ,  $b_Y$ ), а также интервалов перекрытия каждого ядра  $\pm p_X$  и  $\pm p_Y$  с одновременным выбором вида ядра, выбором коэффициента  $p'$  и количества интервалов дискретности каждого отрезка ( $a$ ,  $b$ ).

Результаты расчетов по алгоритму для случая нормального закона распределения следующие:  $p'=0,36962$ ,  $a_x=210,619$ ,  $b_x=254,929$ ,  $p_x=16,3778$ ,  $a_y=190,682$ ,  $b_y=210,716$ ,  $p_y=7,40454$ .

Каждый отрезок  $(a, b)$  следует разбить на 30 интервалов дискретности и найти центры для каждого такого интервала. Затем необходимо определить условие накрывания каждой  $i$ -ой дельтаобразной функции интервалом задания  $\pm p$  каждого центра  $j$ -го интервала дискретности.

Частоты средних интервалов дискретности следует объединить по три, чтобы таким образом получить 10 групп интервалов дискретности. Следующим шагом формируется таблица, в столбцах которой располагаются центры 10 групп интервалов, а в строках – экспериментальные значения соответствующей выборки. В ячейку, находящуюся на пересечении центра группы и исходного значения рассматриваемой выборки, заносится число, соответствующее условию накрывания интервалов задания для данной выборки. Таким образом, формируются таблицы для обеих выборок  $X$  и  $Y$  (таблицы 2.3 и 2.4).

Затем, следует сформировать таблицу двумерного распределения (основа – метод Чебышева), клетки которой заполняются по формуле:

$$d_{jl} = \sum_{i=1}^n n_{jl} \cdot n_{li} \quad (2.2)$$

где  $n_{jl}$ ,  $n_{li}$  – ячейки таблиц 2.3 и 2.4 соответственно, а  $n$  – объем первоначальной парной выборки.

Таблица 2.3 – Суммарные виртуальные частоты для выборки случайной величины  $X$

		247,03	231,19	223,24	221,70	238,64	240,42	237,47	227,54	220,44	240,07	$n_i$	$X_i \cdot n_i$	$X_i^2 \cdot n_i$
1	212,83	0,00	0,01	0,52	0,83	0,00	0,00	0,00	0,09	1,16	0,00	2,62	557,67	118690,72
2	217,27	0,00	0,13	1,66	2,14	0,00	0,00	0,00	0,54	2,49	0,00	6,97	1513,93	328924,97
3	221,70	0,00	0,69	2,82	2,93	0,03	0,01	0,05	1,70	2,86	0,01	11,10	2461,00	545595,63
4	226,13	0,00	1,94	2,56	2,14	0,24	0,11	0,37	2,84	1,75	0,13	12,09	2733,92	618215,24
5	230,56	0,04	2,91	1,24	0,84	1,03	0,62	1,37	2,53	0,57	0,69	11,83	2728,45	629066,72
6	234,99	0,29	2,33	0,32	0,17	2,37	1,83	2,65	1,21	0,10	1,94	13,20	3102,51	729058,27
7	239,42	1,16	0,99	0,04	0,02	2,90	2,88	2,76	0,31	0,01	2,91	13,98	3346,53	801227,18
8	243,85	2,49	0,23	0,00	0,00	1,90	2,43	1,53	0,04	0,00	2,33	10,95	2669,03	650846,85
9	248,28	2,86	0,03	0,00	0,00	0,66	1,09	0,45	0,00	0,00	1,00	6,09	1511,43	375260,56
10	252,71	1,75	0,00	0,00	0,00	0,12	0,26	0,07	0,00	0,00	0,23	2,43	614,67	155336,11
	$\Sigma$	8,59	9,26	9,17	9,07	9,25	9,23	9,26	9,26	8,94	9,24	91,26	21239,14	4952222,26

Получившаяся таблица двумерного виртуального распределения представлена в таблице 2.5.

Таблица 2.4 – Суммарные виртуальные частоты для выборки случайной величины  $Y$

		206,15	200,78	199,38	193,02	200,11	207,82	201,29	198,47	197,87	202,10	$n_l$	$Y_l \cdot n_l$	$Y_l^2 \cdot n_l$
1	191,68	0,00	0,00	0,03	2,55	0,01	0,00	0,00	0,08	0,15	0,00	2,82	540,54	103613,40
2	193,69	0,00	0,06	0,23	2,83	0,12	0,00	0,03	0,49	0,75	0,01	4,51	873,78	169241,16
3	195,69	0,00	0,39	1,01	1,68	0,64	0,00	0,25	1,60	2,02	0,12	7,70	1507,02	294910,03
4	197,69	0,01	1,39	2,34	0,53	1,86	0,00	1,07	2,79	2,92	0,64	13,56	2679,93	529805,91
5	199,70	0,11	2,67	2,91	0,09	2,89	0,02	2,40	2,60	2,26	1,87	17,81	3556,70	710263,21
6	201,70	0,62	2,74	1,92	0,01	2,40	0,16	2,89	1,30	0,93	2,89	15,86	3198,85	645209,42
7	203,70	1,83	1,50	0,68	0,00	1,07	0,78	1,86	0,34	0,20	2,39	10,66	2171,50	442342,20
8	205,71	2,88	0,44	0,13	0,00	0,25	2,07	0,64	0,05	0,02	1,06	7,54	1550,21	318888,98
9	207,71	2,42	0,07	0,01	0,00	0,03	2,93	0,12	0,00	0,00	0,25	5,83	1210,87	251510,08
10	209,71	1,08	0,01	0,00	0,00	0,00	2,21	0,01	0,00	0,00	0,03	3,35	702,24	147268,92
	$\Sigma$	8,97	9,26	9,26	7,68	9,26	8,16	9,26	9,26	9,25	9,26	89,63	17991,63	3613053,32

Таблица 2.5 – Таблица двумерного виртуального распределения

		191,68	193,69	195,69	197,69	199,70	201,70	203,70	205,71	207,71	209,71	$n_l$	$X_l - X_{cp}$	$Y_l - Y_{cp}$
1	212,83	2,32	3,39	4,42	5,33	4,48	2,24	0,64	0,10	0,01	0,00	22,93	-19,90	-9,04
2	217,27	5,90	8,56	11,22	14,00	12,40	6,60	2,03	0,36	0,04	0,00	61,12	-15,47	-7,04
3	221,70	8,10	11,94	16,55	22,33	21,40	12,46	4,29	0,89	0,14	0,03	98,13	-11,04	-5,04
4	226,13	6,02	9,49	15,26	23,80	25,99	17,62	7,35	2,03	0,59	0,27	108,44	-6,61	-3,03
5	230,56	2,48	4,65	10,07	19,96	26,90	22,72	12,04	4,82	2,49	1,47	107,59	-2,18	-1,03
6	234,99	0,61	1,74	6,05	16,21	27,42	28,08	18,19	10,07	7,08	4,47	119,92	2,25	0,97
7	239,42	0,12	0,73	3,85	12,58	24,20	27,72	21,18	15,32	12,45	7,77	125,92	6,69	2,98
8	243,85	0,03	0,33	2,03	7,11	14,54	18,32	17,26	16,22	13,96	8,17	97,97	11,12	4,98
9	248,28	0,01	0,11	0,67	2,43	5,28	7,81	10,06	12,01	10,43	5,55	54,35	15,55	6,98
10	252,71	0,00	0,02	0,12	0,47	1,16	2,29	4,22	5,90	5,06	2,48	21,73	19,98	8,99
	$n_j$	25,58	40,98	70,24	124,23	163,77	145,87	97,26	67,73	52,25	30,21	818,11	-	-

Исследование связи между двумя случайными величинами начнем с вычисления смешанного центрального момента первого порядка:

$$\mu_{1/1} = \frac{1}{N} \sum_{j=1}^k \sum_{l=1}^k (X_j - \bar{X}) \cdot (Y_l - \bar{Y}) \cdot n_{jl} = \frac{19922,9}{818,106} = 19922,9. \quad (2.3)$$

Смешанный основной момент первого порядка является коэффициентом корреляции и находится по формуле:

$$r_{1/1} = \frac{\mu_{1/1}}{S_X \cdot S_Y} = \frac{24,3525}{42,8364} = 0,5685 \quad (2.4)$$

Следующим шагом найдём модифицированный индекс Фехнера по данным таблицы двумерного виртуального распределения (таблица 2.5).

Для расчета индекса Фехнера необходимо вычислить таблицу знаков  $sign(X_i - X_{cp})$  и  $sign(Y_i - Y_{cp})$ .

Определим число совпадающих и несовпадающих знаков в двумерной таблице виртуального распределения. Для этого суммируем частоты  $n_{jl}$ , которые стоят в ячейках

на пересечении совпадающих и несовпадающих знаков величин  $X$  и  $Y$  соответственно. Решив эту задачу, получим  $v=589,041$ , а  $w=229,065$ , где  $v$  – количество совпадающих знаков в таблице, а  $w$  – несовпадающих. Тогда согласно формуле 2.1 величина МИФ будет равна:

$$f^* = \pm \sqrt{\frac{|5891,041 - 229,065|}{589,041 + 229,065}} \pm 0,051 = 0,71433.$$

Таким образом, были обработаны более ста парных выборок [66] с коэффициентом корреляции от 0,1 до 1.

Результаты проделанной работы для наглядности представлены на графиках (рисунки 2.8-2.11), на которых изображены линии тренда, аппроксимируемые по двум точкам. По оси  $X$  данных графиков располагаются значения коэффициентов корреляции Пирсона по первоначальной выборке большого объема, а по оси  $Y$  на различных графиках расположены значения следующих величин: коэффициента корреляции Пирсона по выборке малого объема (рисунок 2.8), наивероятнейшего коэффициента корреляции после применения «*bootstrapping*» (рисунок 2.9), коэффициента корреляции Пирсона по выборке малого объема после применения метода точечных распределений (рисунок 2.9), модифицированного индекса Фехнера после применения метода точечных распределений (рисунок 2.10). Пунктирными линиями на графиках обозначены доверительные интервалы для коэффициента корреляции, полученного по первоначальной выборке [5 с.250].

Для оценки силы связи в теории корреляции применяется шкала английского статистика Чеддока: слабая — от 0,1 до 0,3; умеренная — от 0,3 до 0,5; заметная — от 0,5 до 0,7; высокая — от 0,7 до 0,9; весьма высокая (сильная) — от 0,9 до 1,0 [3].

Как видно из рисунка 2.7, практически все случаи, когда коэффициент корреляции Пирсона, рассчитанный по малой выборке, показывает слабую или умеренную корреляционную связь, на самом деле оказываются выборками с данными видами связей. Также данный критерий почти безошибочно указывает на высокую и весьма высокую (сильную) связь. Однако, в случаях связей величины линейной корреляционной связи в диапазоне  $0,5 < r < 0,75$  наблюдается заметная тенденция к завышению значения коэффициента корреляции. Кроме того, разброс значений коэффициентов корреляции до  $r=0,6$  в первоначальной выборке большого объема приводит к довольно большому разбросу при определении связи по парной выборке объемом  $n=10$ . Таким образом, применение данного подхода позволяет признавать линейную зависимость между факторами сильносвязанной при пороговом значении не ниже  $r=0,75$ .

Процедура нахождения наивероятнейшего коэффициента корреляции посредством «*bootstrapping*» не даёт улучшения результата, а только увеличивает разброс получаемых значений коэффициента корреляции, при общем сохранении тренда (рисунок 2.9).

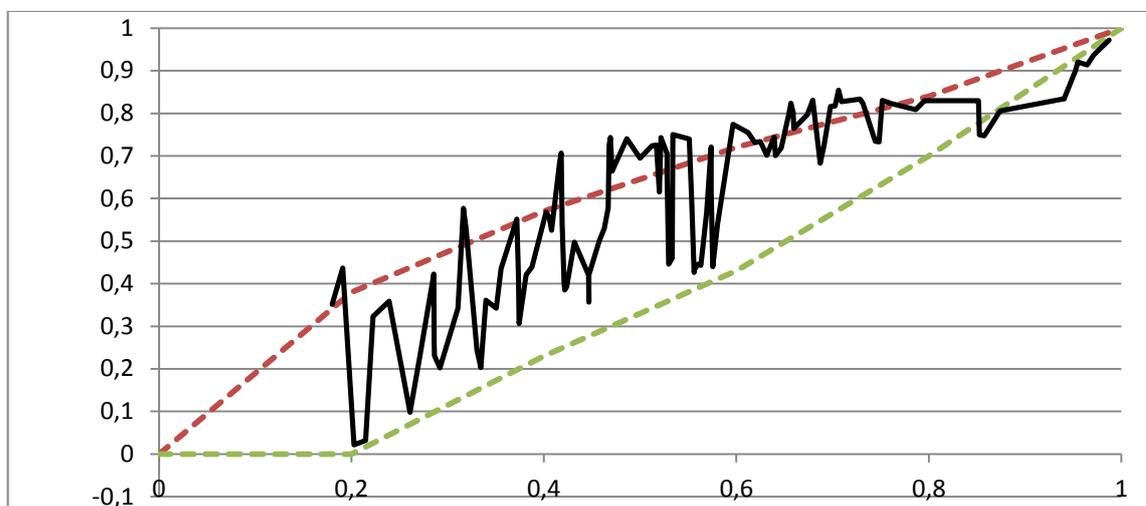


Рисунок 2.8 – Коэффициент корреляции Пирсона по малой выборке

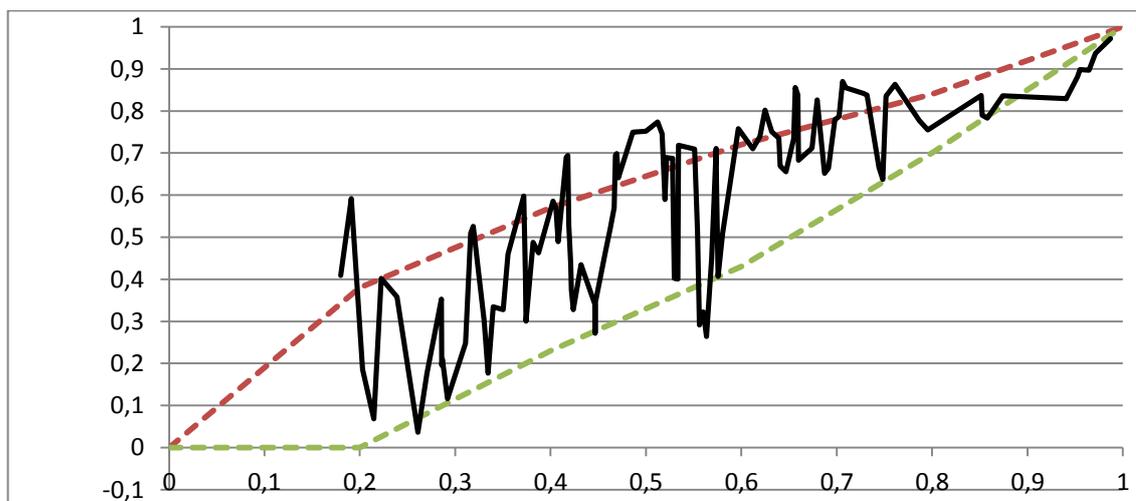


Рисунок 2.9 – Наивероятнейший коэффициент корреляции после применения «*bootstrapping*»

Из рисунка 2.10 видно, что после применения метода точечных распределений, общая тенденция определения значений коэффициента корреляции сохраняется, разброс рассчитанных значений значительно снижается, однако в случаях заметной и более тесных связей явно преобладает тенденция к значительному снижению значения коэффициента корреляции, что не позволяет успешно использовать данный метод на практике. Таким образом, необходимо разработать дополнительный поправочный коэффициент, который учтёт обнаруженный недостаток.

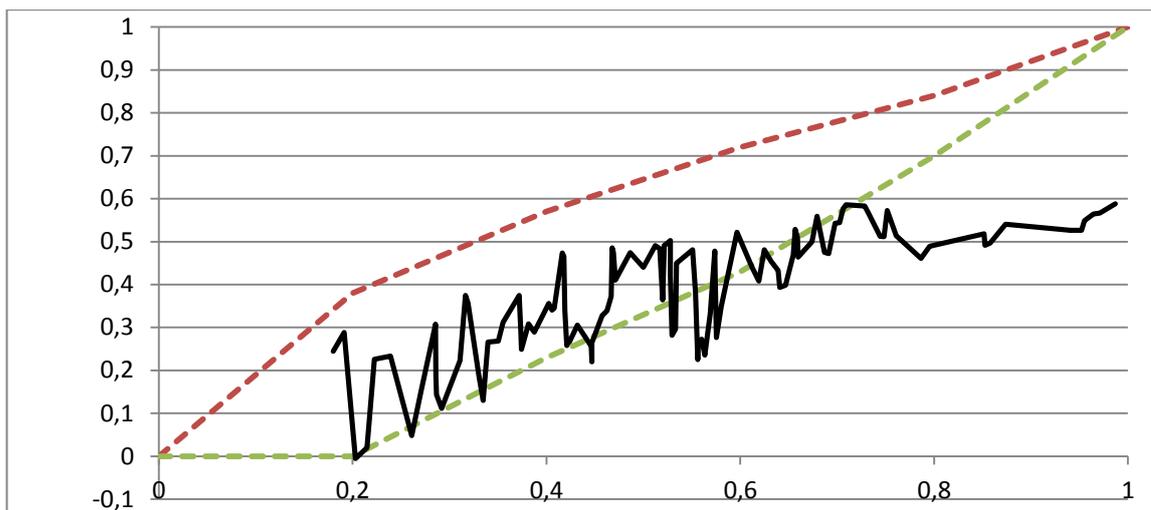


Рисунок 2.10 – Коэффициент корреляции Пирсона после применения МТР

Применение модифицированного индекса Фехнера после метода точечных распределений (рисунок 2.11) при значениях первоначального коэффициента корреляции до 0,3 также даёт на выходе большой разброс, который даже увеличивается по сравнению с нахождением коэффициента корреляции Пирсона по малой выборке. Однако, при более высоких значениях (от 0,3 до 0,75), разброс значительно уменьшается, и это приводит к тому, что почти все значения попадают в доверительный интервал. Но, как видно из графика, данный подход также не позволяет обнаружить сильную линейную зависимость между величинами. Сделаем вывод, что неспособность определить сильную связь по выборкам малого объема является недостатком всего подхода, основанного на синтезе метода точечных распределений и метода Чебышева.

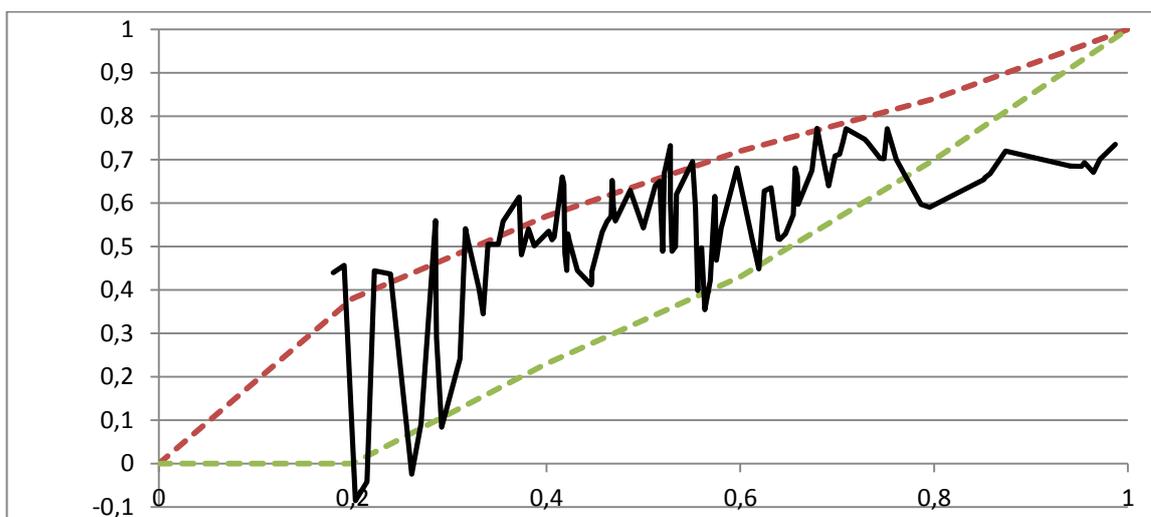


Рисунок 2.11 – Модифицированный индекс Фехнера после применения МТР

Проведенное исследование показало, что при решении задач корреляционного анализа применительно к выборкам малого объема ни один из рассмотренных коэффициентов не показал приемлемых результатов, позволяющих с высокой степенью надежности делать какое-либо утверждение о виде линейной корреляционной связи между величинами. Применение процедур виртуального увеличения количества исследуемой информации, таких как «*bootstrapping*» и метод точечных распределений не увеличивает точности при нахождении наивероятнейшего коэффициента корреляции Пирсона. Расчет модифицированного индекса Фехнера после применения метода точечных распределений даёт удовлетворительные результаты для разграничения слабой и высокой связей, однако не позволяет определить сильную корреляционную связь. Также, в связи с относительно высокой трудоёмкостью данного метода, применение его в задачах контроля качества микроэлектронной продукции, вместо нахождения коэффициента Пирсона по классическому методу, не рекомендуется.

Для случаев обработки таблиц многомерных данных следующим шагом корреляционного анализа является построение корреляционной матрицы.

Элементы матрицы – коэффициенты корреляции между  $x_i$  и  $x_j$  симметричны относительно главной диагонали, на которой стоят единицы.

Непосредственный анализ корреляционной матрицы представляет значительную трудность, так как корреляционные связи между факторами образуют деревья, цепи, циклы и другие фигуры графов. Для выделения главных зависимостей следует прибегнуть к одному из методов анализа таких матриц, простейшим из которых является метод корреляционных плеяд [19].

Приведём общий алгоритм метода корреляционных плеяд:

1. Необходимо найти в корреляционной матрице максимальный не диагональный элемент (по модулю).
2. Вычеркнуть из корреляционной матрицы столбцы, номера которых равны индексам найденного элемента.
3. Внести связь (между факторами, с номерами индекса максимального элемента) в граф корреляционных плеяд.
4. Продолжить поиск максимального не диагонального элемента по строкам, номера которых равны индексам предыдущего максимального элемента.
5. Продолжить поиск, пока не будут вычеркнуты все столбцы.
6. Выбрать пороговое значение  $r$  и проанализировать полученный граф. Если  $r > r_{\text{порог}}$ , то считаем связь сильной и объединяем группы факторов в плеяды.

7. Выбрать для каждой из плеяд 1 фактор – представитель плеяды. Остальные члены плеяды из таблицы исходных данных удалить.

8. Повторить работу алгоритма. Если в каждой плеяде содержится только по одному фактору, то цель работы достигнута.

Исходя из полученных данных, сделаем вывод, что при корреляционном анализе многомерных выборок малого объема, для определения сильно связанных факторов необходимо использовать предварительный корреляционный анализ по классическому методу, используя коэффициент корреляции Пирсона и метод корреляционных плеяд. Данная методика может быть использована для определения связей  $r > 0,75$ . При нахождении таких связей, факторы, находящиеся в одной плеяде, можно убрать из последующего рассмотрения, как не дающие дополнительной информации, а также для сокращения факторного пространства, оставив лишь один фактор, характеризующий данную плеяду.

### **2.3 Построение исходной таблицы многомерных данных малого объема**

В многомерном анализе крайне редко удастся опереться на физическую природу данных и общность некоторых распределений, встречающихся в теории вероятностей и математической статистике. Поэтому обычно проблему выбора модели приходится решать на основе результатов пассивного эксперимента.

Известно, что для обеспечения заданной точности оценки в случае одно- либо двухмерных данных достаточен объем выборки  $n=4$  или  $n=19$ , соответственно [29]. Однако, при увеличении размерности требуемый объем выборки существенно возрастает и уже в четырехмерном, пятимерном случаях необходимый объем выборки измеряется сотнями элементов [140].

Существуют два подхода в формировании моделей на основе данных: параметрический и непараметрический. Параметрические модели задаются с помощью фиксированного набора параметров, определяющего все многообразие данных из заданного семейства распределений; эти модели преимущественно приводят к высокоэффективным (мощным) выводам, но оказываются неустойчивыми при возникающих по тем или иным причинам отклонениях от выбранных предположений. Рост размерности признакового пространства чаще всего приводит к существенному росту размерности параметрического пространства, что также неблагоприятно влияет на результативность статистических процедур. На практике, кроме этого, не всегда

априорные предположения можно представить в виде параметрической модели. Для решения подобных задач необходимо обращаться к непараметрическим методам оценивания распределения данных и, в частности, оценивания плотности распределения. Непараметрический анализ оказывается достаточно гибким, не притязательным в смысле априорных предположений [41].

Один из самых старых из разработанных методов моделирования по пассивным данным – метод наименьших квадратов (МНК) [156], который базируется на подборе такого уравнения регрессии, в котором сумма квадратов разности между уравнением и экспериментальными данными была бы наименьшей из всех возможных. Однако, при обработке результатов пассивного эксперимента вследствие неортогональности пассивного планирования, коэффициенты регрессии, полученные МНК, становятся комплексными характеристиками всей совокупности аргументов. В этом случае оценка значимости коэффициентов регрессии по критерию Стьюдента становятся некорректной, а признание незначимым и отбрасывание какого-либо из найденных коэффициентов регрессии требует пересчета оставшихся коэффициентов заново.

Проблема существенного упрощения процедуры определения коэффициентов регрессии может быть решена путем подбора для каждой регрессионной задачи своей специальной системы линейно-независимых функций  $\psi(X)$  таких, чтобы матрица системы нормальных уравнений  $X^T X$  была единичной, результатом реализации этой идеи стал МНКО [139]. В случае применения данного метода нет необходимости искать обратную матрицу, а также можно отбрасывать незначимые коэффициенты регрессии без учета остальных. Выбор системы функций осуществляется с использованием ортогональных полиномов Чебышева таким образом, чтобы кривая  $Y(X)$  разлагалась по выбранной системе функций в ряд, быстро сходящийся в каждой точке  $X_{kj}$ . При этом система функций должна быть определена на том интервале значений переменной  $X_{kj}$ , на котором расположены экспериментальные точки.

Анализ особенностей метода наименьших квадратов с предварительной ортогонализацией факторов (МНКО), как в теоретическом плане, так и в плане практического применения, позволяет обратить внимание на следующее.

1. Эффективность метода зависит от порядка следования факторов (эффектов) друг за другом при расчете коэффициентов модели. В случае расположения их в порядке убывания по степени значимости эффективность метода возрастает. Поэтому следует считать целесообразным, перед применением МНКО, предварительно расположить

исследуемые факторы (эффекты) в порядке убывания значимости (степени влияния) по отношению к целевой функции.

2. Существенной особенностью и преимуществом МНКО является то обстоятельство, что в силу перехода данных в заведомо ортогональную систему координат можно получать эффективные оценки коэффициентов и для коррелированных факторов и для квадратных членов.

3. Другой существенной особенностью и преимуществом МНКО является то, что для получения модели не требуется длинной таблицы исходных данных, лишь бы координаты точек факторного пространства были бы достаточно далеки друг от друга [1].

Также одним из самых известных и удобных методов моделирования пассивных экспериментов является метод случайного баланса (МСБ), предложенный Ф.Е. Саттерзвайтом [135]. Метод был разработан специально для сверх насыщенных планов и, как указано в ГОСТ24026-80 «Исследовательские испытания. Планирование эксперимента. Термины и определения», основывается на статистическом оценивании и некоторых эвристических соображениях. МСБ широко применяется для отсеивания небольшого числа значимых факторов на шумовом поле. Сущность метода случайного баланса заключается в построении матрицы планирования со случайным распределением уровней факторов, в проведении эксперимента по матрице и в специфичной обработке данных эксперимента. В дальнейшем данный метод был развит до Модифицированного метода случайного баланса (ММСБ) [19, 24], в котором сложная и громоздкая графоаналитическая процедура оценок коэффициентов модели заменена более легкой аналитической.

Любые факторные планы, в том числе ММСБ, одним из предварительных этапов в планировании производят предварительную ортогонализацию факторов, то есть переход от координат с абсолютными единицами измерения факторов  $X_k$  к координатам с относительными единицами  $x_k$ , где единичной мерой служит достаточно произвольно выбранный шаг  $\Delta X_k$ , свой для каждого конкретного фактора. Поэтому в идеологии метода считается целесообразным всю область  $X_{kmax} - X_{kmin}$  разбить на три части (приблизительно, но необязательно, одинаковые по вероятности попадания в них), как показано на рисунке 2.12 [24]. Причем за центр всей группировки следует принять среднюю арифметическую  $\bar{X}_k$ , высчитанную по большому количеству данных.

Тогда все значения  $X_k \leq \bar{X}_k - z \cdot S_k$  отнесем к области  $x_k = -1$ , все значения  $X_k \geq \bar{X}_k + z \cdot S_k$  – к области  $x_k = +1$ , остальные значения – к области  $x_k = 0$ . Схема преобразования

координат факторов из именованной области в относительную приведена на рисунке 2.6. Величина  $z$  может меняться в пределах 0,25–0,5 по желанию исследователя.

Необходимая точность выделения значимых факторов для построения модели может быть обеспечена за счет увеличения числа опытов в каждой такой локализованной области факторного пространства. Именно этим обусловлена невозможность применения ММСБ для выборок малого объема.

Продлав работу по преобразованию координат факторов, получаем из таблицы экспериментальных данных таблицу планирования эксперимента, каждая строка которой представляет собой координату точки факторного пространства в виде набора относительных величин  $-1, 0, +1$ , и результат эксперимента в этой точке  $Y$ .

Общее требование для любых факторных планов – гомоскедастичность – в плане ММСБ нарушается, поэтому для расчетов оценок коэффициентов регрессии  $b_k$  и их дисперсии  $D_k$  следует использовать специальные выражения, учитывающие поправки на это нарушение.

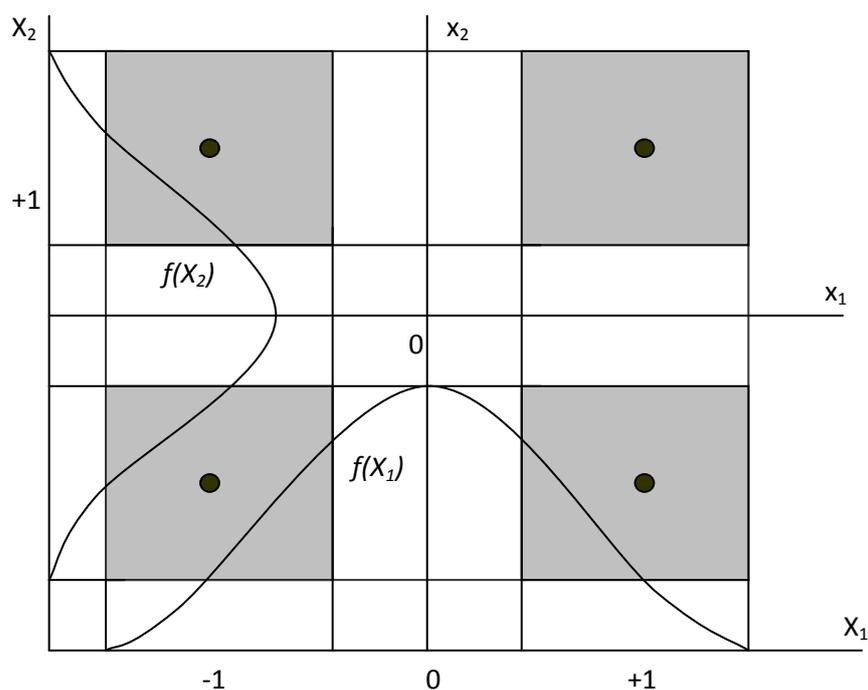


Рисунок 2.12 – Схема преобразования координат факторов

Такая оценка коэффициентов регрессии (математической модели) имеет минимальную систематическую и случайную ошибки, при этом никаких ограничений на равенство (неравенство) дисперсий частных выборок не накладывается.

Если из первоначальной матрицы планирования изъять столбцы, соответствующие незначимым факторам, то новая матрица перестаёт быть сверхнасыщенным планом, у неё появляются совпадающие строки, которые необходимо совместить. При этом строки имеют разные объемы выборок результатов  $m_j$ .

Это позволяет проверить числовые значения на однородность (отсутствие аномальных результатов) любым из известных способов [19].

Из плана эксперимента можно извлечь дополнительную информацию о влиянии парных взаимодействий, которые иногда могут быть больше влияния каждого фактора в отдельности. С этой целью в план эксперимента включаются столбцы парных взаимодействий, каждая координата которых получается простым перемножением кодов координат исходных факторов. Проверка парных взаимодействий на значимость ничем не отличается от соответствующей процедуры для каждого из основных факторов. Рассматривать взаимодействия более высокого порядка (тройные, четверные и т.д.) не имеет смысла, так как их влияние на выходную величину заведомо меньше влияния шумового фона.

Исследования показали, что процедура ММСБ дает хорошие результаты даже при значительном отклонении распределения выходной величины от нормального закона, лишь бы само распределение оставалось унимодальным.

Однако, так как модифицированный метод случайного баланса является отсеивающим методом, то его применение к малым выборкам невозможно (особенностью ММСБ является неременное требование разделения экспериментальных данных на три области, объем каждой из которых превышает верхнюю границу объема малой выборки).

Для решения этой проблемы ниже представлен метод, который позволяет использовать преимущества метода точечных распределений для анализа многофакторных таблиц пассивного эксперимента [128, 130].

На начальном этапе необходимо провести непараметрический анализ таблицы исходных данных, для чего с помощью метода точечных распределений построим для каждого  $X_i$  и  $Y$  таблицы расчета ненормированных плотностей вероятностей в виртуальной области по алгоритму, который был изложен в п.1.4.

Далее для каждой строки  $l$  исходных экспериментальных данных необходимо построить виртуальные таблицы данных, в которые следует вносить одновременно величины двух столбцов  $X_{ij}$  из соответствующей таблицы ненормированных плотностей вероятностей, построенных по методу точечных распределений, и столбца  $X_{il}$ .

Выравнивание (стыковка) пар столбцов  $X_{ij}$  и  $X_{il}$  (а также )  $Y_j$  и  $Y_l$  должно происходить по уровню максимальной плотности вероятности.

Таким образом, получаются  $l$  виртуальных таблиц, состоящих из данных и вероятностей их попадания в заданный интервал, которые содержат большое количество пробелов в значениях, в связи с чем не могут быть обработаны.

Исходя из этого, из всех таблиц, найденных в предыдущем пункте настоящего алгоритма, удаляются не полностью заполненные строки, т.е. строки, в которых вероятность попадания в заданный интервал одного или нескольких из факторов не соответствует критериям. После чего, производится удаление столбцов, обозначающих ненормированные плотности вероятности в виртуальной области.

Состыковка отредактированных таблиц происходит в порядке нумерации строк таблицы исходных данных. Полученная виртуально увеличенная таблица в 15-20 раз длиннее таблицы исходных данных, что позволяет добиться большей точности и достоверности при её обработке.

Рассмотрим результат работы описанного выше алгоритма на примере. Пусть в результате производства  $n=8$  единиц (партий) изделий получены указанные в таблице 2.6 числовые значения контрольных параметров ( $X_i$  – параметры, контролируемые в ходе технологического процесса;  $Y$  – выходной показатель качества изделия). Все наименования размерностей для простоты опущены.

Таблица 2.6 – Таблица исходных данных

№ партии	Факторы $X_i$				Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	
1	0,695	89,65	66,71	-27,29	57,18
2	0,644	99,40	68,58	-32,09	75,48
3	0,674	108,50	64,97	-36,08	79,12
4	0,695	92,50	67,71	-28,32	72,03
5	0,711	95,80	66,11	-28,90	76,34
6	0,685	100,90	68,13	-27,45	72,22
7	0,692	102,60	65,78	-30,21	81,90
8	0,697	90,60	66,85	-31,83	55,94

В таблице 2.7 показан пример таблицы ненормированной плотности вероятности в виртуальной области, построенный по методу точечных распределений для фактора 1. Таблицы для факторов 2-4 и для выходной величины даны в приложении А1.

Следующим шагом необходимо построить таблицы состыковки полученных данных для каждой строки первоначальной таблицы исходных данных. На этом же этапе происходит выравнивание по максимальному значению плотности вероятности.

Таким образом, получено 5 таблиц состыкованных данных (количество факторов + выходная величина). Результаты работы представлены в приложении А1. В таблице 2.8 изображен пример состыковки виртуальных значений для первой строки таблицы исходных данных.

Таблица 2.7 – Таблица ненормированных плотностей вероятности для фактора 1

$X_{1j}$	$X_{1f}$							
	0,644	0,674	0,685	0,692	0,695	0,695	0,697	0,711
0,64143	0,9779	0,0277						
0,64466	0,9985	0,0545						
0,64788	0,9504	0,0997						
0,65111	0,843	0,1702	0,0206					
0,65434	0,6968	0,2709	0,0417					
0,65757	0,5367	0,4016	0,0787	0,0182				
0,6608	0,3853	0,555	0,1382	0,0373	0,0192	0,0192	0,0119	
0,66403	0,2577	0,7147	0,2263	0,0711	0,0391	0,0391	0,0254	
0,66725	0,1609	0,8573	0,3448	0,1262	0,0741	0,0741	0,0502	
0,67048	0,0935	0,959	0,4904	0,2091	0,1311	0,1311	0,0929	
0,67371	0,0506	0,9997	0,65	0,3229	0,2162	0,2162	0,1599	
0,67694	0,0256	0,9712	0,8029	0,4647	0,3321	0,3321	0,2567	0,0198
0,68017	0,012	0,8793	0,9242	0,6232	0,4756	0,4756	0,384	0,0403
0,6834		0,7419	0,9914	0,7789	0,6346	0,6346	0,5352	0,0762
0,68662		0,5838	0,9912	0,9068	0,7887	0,7887	0,6948	0,1342
0,68985		0,4279	0,9236	0,9845	0,9143	0,9143	0,8413	0,2205
0,69308		0,2922	0,802	0,9961	0,9876	0,9876	0,9494	0,3378
0,69631		0,186	0,649	0,9392	0,9942	0,9942	0,9984	0,4823
0,69954		0,1103	0,4895	0,8252	0,9327	0,9327	0,9784	0,6416
0,70277		0,061	0,344	0,6757	0,8154	0,8154	0,8936	0,7954
0,706		0,0314	0,2253	0,5156	0,6644	0,6644	0,7605	0,919
0,70922		0,0151	0,1377	0,3671	0,5049	0,5049	0,6037	0,9894
0,71245			0,0784	0,2434	0,3574	0,3574	0,4464	0,9929
0,71568			0,0416	0,1503	0,2357	0,2357	0,3075	0,9287
0,71891			0,0205	0,0865	0,1449	0,1449	0,1975	0,8094
0,72214				0,0464	0,083	0,083	0,1182	0,6575
0,72537				0,0232	0,0443	0,0443	0,0659	0,4977
0,72859					0,0221	0,0221	0,0343	0,3515
0,73182							0,0166	0,2311
0,73505								0,1416

После отсечения не полностью заполненных строк и удаления столбцов, заполненных значениями ненормированной плотности вероятности, совмещаем таблицы и получаем виртуальную таблицу данных (Приложение А1), состоящую из 122 строк, что в 15,25 раз больше, чем было в первоначальной таблице.

Получившуюся таблицу можно обрабатывать методами математического моделирования по пассивным данным.

Проанализируем полученную и исходную выборки на однородность. Для этого рассчитаем  $t$ -распределение Стьюдента [19] для каждого из факторов и выходной величины (таблица 2.9).

Табличное значение критерия Стьюдента с числом степеней свободы  $df=128$  при доверительной вероятности 95%  $t_{(табл)}=1,657$ . Сравнив рассчитанные значения с табличным, видно, что гипотеза о равенстве центров распределения принимается. Таким образом, выборки можно признать однородными.

Таблица 2.8 – Состыковка виртуальных значений для 1 строки

$X_1$	$P_{X1}$	$X_1$	$P_{X2}$	$X_1$	$P_{X3}$	$X_1$	$P_{X4}$	$Y$	$P_Y$
0,6608	0,0192			64,4904	0,0115	-32,65247	0,0115		
0,66403	0,0391			64,68745	0,0245	-32,17623	0,0245		
0,66725	0,0741			64,8845	0,0487	-31,69998	0,0488		
0,67048	0,1311			65,08155	0,0902	-31,22374	0,0904		
0,67371	0,2162	82,84657	0,2255	65,2786	0,1559	-30,74749	0,1562		
0,67694	0,3321	83,8928	0,3442	65,47565	0,2511	-30,27125	0,2515		
0,68017	0,4756	84,93903	0,4896	65,6727	0,3768	-29,79501	0,3774	49,64024	0,4324
0,6834	0,6346	85,98525	0,6491	65,86975	0,5271	-29,31876	0,5277	51,18567	0,5887
0,68662	0,7887	87,03148	0,802	66,0668	0,6871	-28,84252	0,6878	52,73109	0,7469
0,68985	0,9143	88,07771	0,9235	66,26385	0,8348	-28,36627	0,8354	54,27652	0,8831
0,69308	0,9876	89,12393	0,9911	66,4609	0,9453	-27,89003	0,9456	55,82195	0,9732
0,69631	0,9942	90,17016	0,9913	66,65795	0,9975	-27,41379	0,9976	57,36738	0,9995
0,69954	0,9327	91,21639	0,9241	66,855	0,9811	-26,93754	0,9809	58,91281	0,9567
0,70277	0,8154	92,26262	0,8028	67,05205	0,8993	-26,4613	0,8988	60,45824	0,8534
0,706	0,6644	93,30884	0,65	67,2491	0,7683	-25,98505	0,7676	62,00367	0,7095
0,70922	0,5049	94,35507	0,4905	67,44615	0,6117	-25,50881	0,611	63,5491	0,5498
0,71245	0,3574	95,4013	0,345	67,6432	0,4539	-25,03257	0,4532	65,09453	0,397
0,71568	0,2357	96,44752	0,2261	67,84025	0,3139	-24,55632	0,3133	66,63996	0,2672
0,71891	0,1449	97,49375	0,1381	68,0373	0,2023	-24,08008	0,2019	68,18539	0,1676
0,72214	0,083	98,53998	0,0786	68,23435	0,1215	-23,60384	0,1212	69,73082	0,098
0,72537	0,0443	99,5862	0,0417	68,4314	0,068	-23,12759	0,0678	71,27625	0,0534
0,72859	0,0221	100,63243	0,0206	68,62845	0,0355			72,82168	0,0271
				68,8255	0,0173			74,36711	0,0128

Определим, равны ли выборочные дисперсии исходной и виртуально увеличенной выборки. Для этого по формуле 1.22 и рассчитанным значениям эмпирических дисперсий, рассчитаем дисперсионное отношение, которое представляет собой  $F$ -распределение Фишера. При этом в числитель помещаем большую из рассчитанных дисперсий. Сравним полученные значения (Таблица 2.9) с критическим значением  $F_{(табл)}(q=0,95, df_1=8, df_2=122)=2,97$ . Так как все расчетные значения меньше табличного, то гипотеза о равенстве выборочных дисперсий принимается.

Рассчитанные значения средних, дисперсий, средневзвешенных дисперсий, а также расчетные значения критериев Стьюдента и Фишера приведены в таблице 2.9.

Таким образом, проведенные расчеты показывают, что и средние арифметические и дисперсии исходных и виртуально увеличенных выборок статистически неразличимы, а значит могут быть признаны однородными.

Рассмотрим, как ведёт себя коэффициент корреляции Пирсона после виртуального увеличения выборки многомерным методом точечных распределений. Исследование проводилось на том же самом наборе данных (п. 2.2), полученном при производстве кристаллов интегральных микросхем.

Таблица 2.9 – Сравнение исходной и виртуально увеличенной выборки

	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	$Y$
Среднее по исходным данным	0,685429	97,49375	66,855	-30,2713	71,27625
Дисперсия по исходным данным	0,000462	42,75746	1,516743	8,859698	93,29503
Среднее после ММТР	0,693434	97,21075	67,01978	-29,268	71,79565
Дисперсия после ММТР	0,000477	52,79439	2,194549	9,99335	112,3361
Средневзвешенная дисперсия	0,000476	52,24549	2,157482	9,931354	111,2948
Расчетное значение критерия Стьюдента	1,004987	0,107281	0,307384	0,87227	0,134901
Расчетное значение критерия Фишера	1,033946	1,234741	1,446883	1,127956	1,204095

Для этого сравним значения коэффициентов корреляции между факторами исходной большой выборки и коэффициентов корреляции виртуально увеличенной выборки.

На рисунке 2.13 изображена линия тренда, аппроксимируемая по двум точкам, на которой по оси  $X$  располагаются значения коэффициентов корреляции Пирсона по

первоначальной выборке большого объема, а по оси  $Y$  – значения коэффициента корреляции Пирсона после виртуального увеличения многомерной малой выборки.

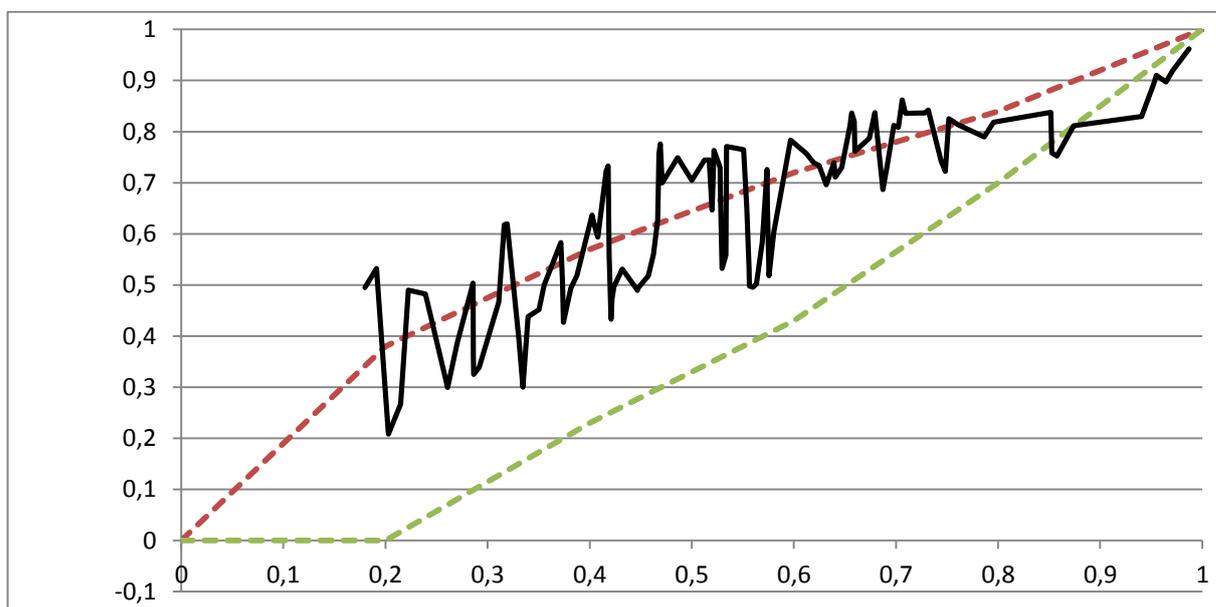


Рисунок 2.13 – Коэффициент корреляции после применения ММТР

Как видно из рисунка, применение многомерного метода точечных распределений, существенно снижает разброс между рассчитанными коэффициентами корреляции, особенно в случае слабых связей, при общем сохранении вида линии тренда.

Также необходимо обратить внимание на тот факт, что расчет коэффициента корреляции Пирсона по малой выборке в ряде случаев показывал слабую отрицательную связь в случаях наличия слабой положительной между факторами. Этот недостаток усугубляется после применения виртуального увеличения выборки «*bootstrapping*», но он исчезает, если производить увеличение выборки многомерным методом точечных распределений.

Проведенные исследование показали, что разработанная методика дает более точные показатели при оценивании линейного коэффициента корреляции, вне зависимости от исследуемого объема выборки.

В качестве примера приведем графики отношения оцениваемого коэффициента корреляции по выборкам малого объема и оцениваемого коэффициента корреляции после виртуального увеличения выборки многомерным методом точечных распределений (значения по оси  $Y$ ) к коэффициенту корреляции генеральной совокупности (значения по оси  $X$ ) для объема выборки  $n=6$  (рисунок 2.14) и  $n=16$  (рисунок 2.15).

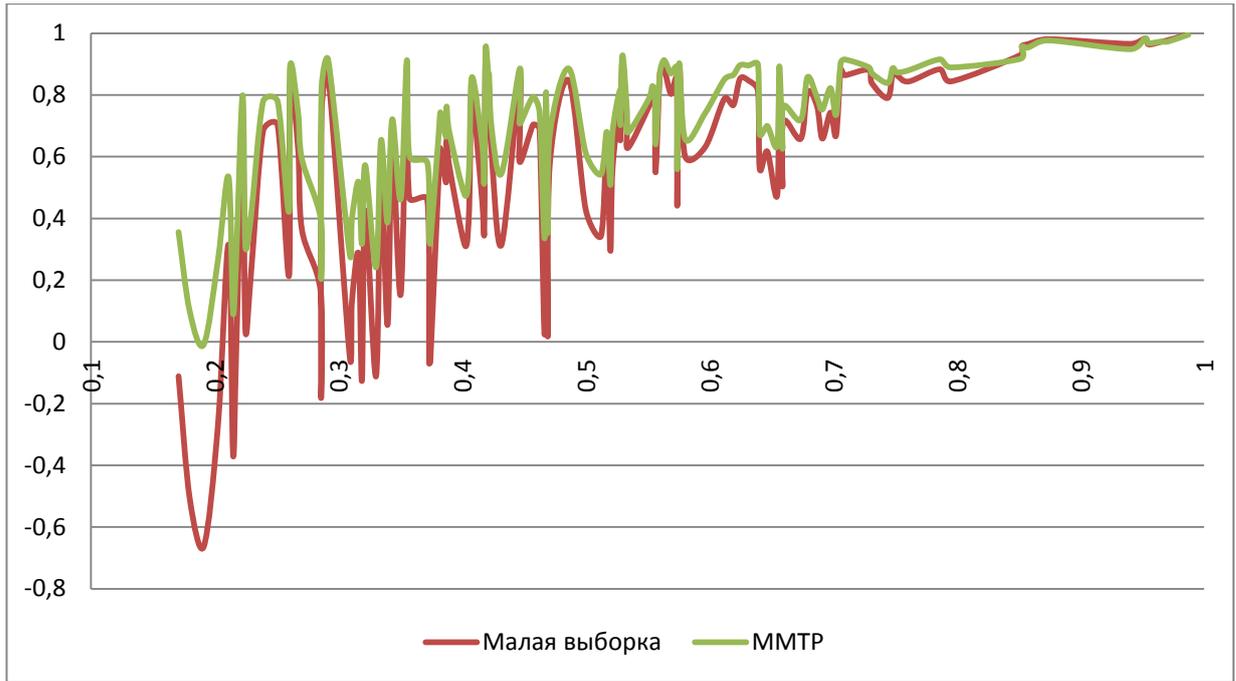


Рисунок 2.14 – Оценка коэффициентов корреляции при  $n=6$

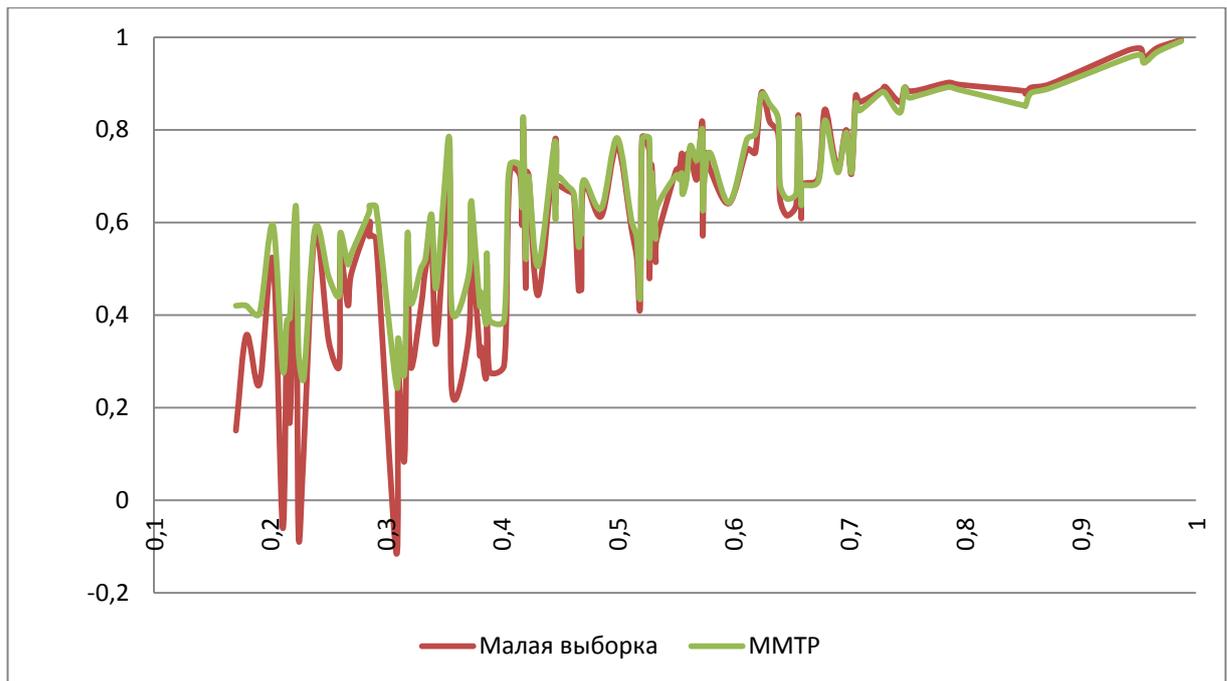


Рисунок 2.15 – Оценка коэффициентов корреляции при  $n=16$

Приведенные графики подтверждают, что при уменьшении объема исходных данных оценивание выборочного коэффициента корреляции приводит к большим ошибкам и его значение не может с достаточной достоверностью подтверждать наличие значимой наблюдаемой линейной связи между исследуемыми случайными величинами.

Применение разработанной методики позволяет сузить доверительный интервал при оценивании коэффициента корреляции Пирсона по данным малого объема, а также избежать ошибочного определения отрицательной корреляционной связи.

Таким образом, можно рекомендовать рассчитывать коэффициент корреляции после виртуального увеличения объема выборки предложенной методикой [37]. 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  $7 \leq n \leq 10$  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  $r = 0,9$ .

Таким образом, 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.

## 2.4 Определение качества решений

Известно, что вероятностно-статистические методы, как и другие методы математики, применяются не непосредственно к реальным объектам, а к их формализованному виду, то есть математическим моделям. Для одного и того же реального объекта (процесса) возможно построение нескольких моделей, поэтому получив решение математической задачи, необходимо разобраться в его реальном смысле, т.е. проанализировать полученные результаты.

Во-первых, полученная модель должна отражать наиболее существенные свойства реального изучаемого объекта с концептуальной точки зрения. Если математическая модель пригодна по этому признаку, то говорят, что она обладает функциональностью или целенаправленностью.

Во-вторых, модель должна быть простой в реализации и требовать для своего создания минимального количества ресурсов. Это связано с тем, что построенную математическую модель необходимо использовать на практике, а сложные модели зачастую практически невозможно реализовать, кроме того, они редко дают хорошую сходимость результатов.

В третьих, самым важным требованием к математической модели является требование ее правильного соответствия изучаемому реальному объекту. Совпадение свойств модели и соответствующих свойств моделируемого объекта называется адекватностью математической модели.

Под адекватностью понимается [54]:

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 соответствии с тем, ставится второе условие или нет, говорят соответственно о количественных или качественных моделях. Вместо количественной адекватности также говорят о точности модели.

Нахождение математической модели любым из известных методов регрессионного анализа и подтверждение ее адекватности результатам эксперимента неизбежно ставит вопросы об области, то есть коридоре существования построенной модели и об её информационной ценности. Действительно, как всякая регрессия, уравнение модели в любой форме, в том числе и полиномиальной, представляет собой уравнение геометрического места точек математических ожиданий многомерных распределений, полученных в соответствующих сечениях факторного пространства. Поскольку одной из основных предпосылок обсуждённых методов является предварительная проверка целевой функции на соответствие её распределения нормальному закону, то вопрос о коридоре существования модели может свестись к вопросу о крутости функции плотности вероятности условного распределения  $f(Y / X_1, X_2, \dots, X_n)$ , которое заведомо должно быть нормальным.

Это обстоятельство позволяет вместо неудобной характеристики крутости условного распределения свести вопрос к его дисперсии, среднеквадратическому отклонению или, через известные статистические соотношения, к ширине коридора существования  $\pm \Delta Y$  при определенной доверительной вероятности [19].

Оценка адекватности модели реальному объекту производится по близости результатов расчетов экспериментальным данным. Одним из самых распространенных методов такой проверки является оценивание по дисперсиям отклонений откликов модели от среднего значения реальных откликов систем. Сравнение дисперсий, в случае ограниченного объема выборки, проводят с помощью проверки гипотезы о согласованности  $F$ -критерием Фишера.

Как известно, на одном и том же массиве исходных данных можно построить несколько адекватных математических моделей, причем необязательно с одним и тем же перечнем значимых факторов. В силу адекватности все эти модели имеют право на существование, однако, точность описания ими выходной величины различна. В принципе ее можно оценить по величине коридора существования модели и/или по количеству членов регрессии, однако в большом количестве случаев по вышеупомянутым причинам сама величина коридора является оценкой, иногда довольно грубой.

Существует большое количество различных критериев, позволяющих сравнивать качество построенных регрессионных моделей [142]. Поэтому предлагается о качестве модели судить по количеству информации, которое она может дать, то есть по информационной ёмкости [19].

В соответствии с вышеизложенным, для оценки применимости полученных решений, оценим качество математических моделей, получаемых после регрессионного анализа увеличенных таблиц данных предложенной методикой. Для этого сравним и проанализируем значения расчетного критерия Фишера и информационной ёмкости для моделей, построенных по увеличенной по предложенной методике таблице данных.

Для примера рассмотрим адекватные математические модели, которые можно построить по данным, рассмотренным в пункте 2.4. При этом заметим, что используя и МНКО, и ММСБ невозможно построить адекватную математическую модель по исходной выборке данных.

Виртуальное увеличение выборки многомерным методом точечных распределений позволяет объединить в плеяды часть исследуемых входных величин, выделить наиболее значимые из них и на основании такого решения сократить факторное пространство и построить математические модели, уравнения регрессии которых продемонстрированы в таблице 2.10.

Модели, построенные по методу наименьших квадратов с предварительной ортогонализацией факторов, как правило, обладают большей информационной ёмкостью, так как применение формулы расчета информационной ёмкости в случае варьирования

факторов на трёх уровнях (что следует из методики расчета ММСБ) даёт более грубый результат, чем в случае непрерывного распределения факторов. Однако, расчетный уровень критерия адекватности Фишера почти всегда значительно ниже для моделей построенных модифицированным методом случайного баланса и в ряде случаев, при пассивных экспериментах с малым числом наблюдений, построить адекватную математическую модель возможно только применив ММСБ.

Таблица 2.10 – Сравнительная таблица качества математической модели

Наименование метода построения модели	Модель	Критерий Фишера	Информационная ёмкость
МНКО после ММТР	$Y = -86,68 + 55,148X_1 + 1,2368X_2$	0,9513	61,15%
	$Y = -37,97 + 1,234X_2 + 0,34848X_4$	0,9601	65,35%
ММСБ после ММТР	$Y = 71,796 + 3,401X_1 + 11X_2$	0,1504	23,9%
	$Y = 71,796 + 11,143X_2 + 3,598X_3 - 5,841X_2X_3$	0,7906	37,6%
	$Y = 71,796 + 11X_2 + 4,49X_4$	0,5378	24,3%

Расчеты показывают, что виртуальное увеличение объема выборки многомерным методом точечных распределений во многих случаях позволяет получать адекватные математические модели для сверхнасыщенных планов [67], а также более качественные модели даже в тех случаях, когда невозможно построить адекватную математическую модель по исходным данным малого объема [73, 74].

Таким образом, произведя оценку применимости полученных решений, можно сделать 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.

## 2.5 Результаты по главе 2

1. Исследована процедура поиска грубых ошибок в выборках малого объёма и найдены критерии лучше всего справляющиеся с данной проблемой, даны рекомендации по применению критериев в зависимости от объемов выборки.

2. Исследованы возможности существующих подходов к нахождению линейной корреляционной связи в выборках малого объёма, по результатам работы предложена процедура линейного корреляционного анализа для задач обнаружения сильно связанных

факторов и дальнейшего сокращения факторного пространства, без существенной потери информации.

3. 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 превосходящие по длине исходные, 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*».

4. Проанализировано качество моделей, построенных с использованием разработанного метода, сделан вывод, что применение процедуры виртуального увеличения объёма 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 объёма.

## **2.6 Выводы к главе 2**

1. Необходимо сформулировать методику построения СППР по пассивным выборкам малого объёма для случаев, когда классические методы не позволяют получить модель необходимого качества по исходным данным.

2. В связи с трудоёмкостью процесса обработки выборок малого объёма (применением метода точечных распределений) и со сложностью составления увеличенной таблицы данных для дальнейшего моделирования необходимо разработать комплекс специализированных компьютерных программ, реализующий предложенные методы.

3. Необходимо произвести апробацию полученных методик и разработать СППР для различных областей деятельности исследователей.

### **3. СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ПАССИВНЫМ ВЫБОРКАМ МАЛОГО ОБЪЕМА**

#### **3.1 Алгоритм создания СППР по пассивным выборкам малого объема**

На основе полученных результатов была поставлена задача разработать методику построения систем поддержки принятия решений по пассивным выборкам малого объема для случаев, когда классические методы не позволяют получить адекватную модель или модель необходимого качества.

Данная методика включает в себя следующие операции:

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, сбоям 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 искать одним из следующих критериев:

- a) Метод Ирвина.
- b) Критерий Львовского.
- c) Критерий Гитьена-Мура.
- d) Критерий Диксона (для нормального распределения).

Кроме того, как показано в теоретической части исследования, применение критерия Гитьена-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 [85], известная из характеристик исследуемого процесса, приблизительно оцененная по гистограмме, либо предварительно рассчитанная одним из методов получения оценок по малым выборкам (рассмотренных в параграфе 1.4).

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 виртуально увеличенной таблице 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, определяются возможные варианты математических моделей, которые содержит в себе многомерная виртуальная выборка.

7. Следующим шагом является построение математических моделей по получившейся таблице данных любыми из известных и подходящих для данного эксперимента методов регрессионного анализа многомерных пассивных данных.

8. Проверка адекватности и качества построенных моделей, определение оптимальных математических моделей, и их последующая интерпретация, совместно с профильными специалистами в конкретной области исследования.

9. Обобщение полученной информации в СППР, основанную на собранных данных, математической модели исследуемого процесса и формальных знаниях экспертов в изучаемой области.

### **3.2 Разработка программного обеспечения, реализующего вероятностно-статистические методы**

Рассмотренные в предыдущих главах методы обработки пассивной экспериментальной информации очень трудоёмки при использовании статистических пакетов общего назначения. Так процесс построения таблиц расчета ненормированных плотностей вероятности в виртуальной области и дальнейшая их состыковка требует от исследователя большой внимательности и занимает значительное время. Существующие же специализированные программные продукты [7, 82], реализующие метод точечных распределений, не позволяют экспортировать данные в удобной для дальнейшей обработки форме. Кроме того, данные программные продукты не удобны для дальнейшего масштабирования, в том числе добавления новых методов, для различных законов распределения.

Следовательно, целесообразно разработать комплекс программных продуктов, который будет реализовывать методы математического моделирования по пассивным выборкам малого объема. Причем при разработке программного комплекса следует учесть следующие требования:

1. Программный комплекс должен состоять из модулей, реализующих различные методы моделирования по пассивным выборкам малого объема.

2. Программный комплекс должен использовать единообразные входные и выходные интерфейсы, так чтобы результаты работы одного метода могли быть исходными данными для других методов.

3. Необходимо обеспечить легкость импорта и экспорта данных в привычные для исследователей статистические пакеты общего назначения.

4. Программные модули должны использовать преимущества объектно-ориентированного программирования, такие как наследование, для обеспечения возможности дальнейшего масштабирования программного продукта на основе уже написанных математических модулей.

СППР, как автоматизированные информационные системы, должны позволять человеку без профессиональной математической подготовки адаптироваться к изменению применяемых вычислительных моделей и при этом общаться с ЛПР на специфическом для управляемой предметной области языке, а также выдавать результаты, которые наиболее полно будут отображать информацию об исследуемом объекте.

#### **Выбор методов и средств для реализации программного продукта**

Программный продукт был написан на языке *C#* в среде *Microsoft Visual Studio 2010 Express*, вследствие чего предполагается работа программы на компьютерах под управлением ОС *Windows* с предустановленной платформой *.NET Framework*. Выбор языка программирования связан с тем, что на момент начала написания программного продукта *C#* был самым динамично развивающимся языком и, несмотря на спад популярности в последнее время, данный язык и сейчас сохраняет своё место в числе пяти самых используемых языков программирования по оценкам индекса *TIOBE* (<http://www.tiobe.com/tiobe-index/>).

Очевидным недостатком данного подхода является необходимость запуска приложений под ОС *Windows*. Однако, при необходимости, код программ может быть легко портирован под *Unix*-системы, посредством перекомпиляции в среде *Mono* или *DotGNU*.

Форматом обрабатываемых и полученных данных является *.csv* (от англ. *Comma Separated Values* – значения, разделённые запятыми), что позволит исследователю легко импортировать полученные результаты в сторонние приложения для проведения дальнейшего анализа и построения многомерного уравнения регрессии. Импорт из данного формата поддерживается большинством табличных редакторов и инженерных

пакетов. Так, дальнейшая обработка полученных данных может вестись в таких популярных статистических приложениях, как *MS Excel* (с надстройкам для анализа данных), *Statistica*, *Stata*, а также считываться стандартными функциями языков программирования, активно применяющихся при анализе данных (*R*, *Python*).

### Разработка архитектуры и логики программного продукта

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. Общая структура разработанного комплекса прикладных программ представлена на рисунке 3.1.

Каждый из артефактов, входящих в общий узел *MathModel*, является отдельным файлом с исходным кодом программы. Все приложения являются консольными, что упрощает их внедрение в различное программное обеспечение.

Рассмотрим структуру и логику работы каждой программы, входящей в программный комплекс.

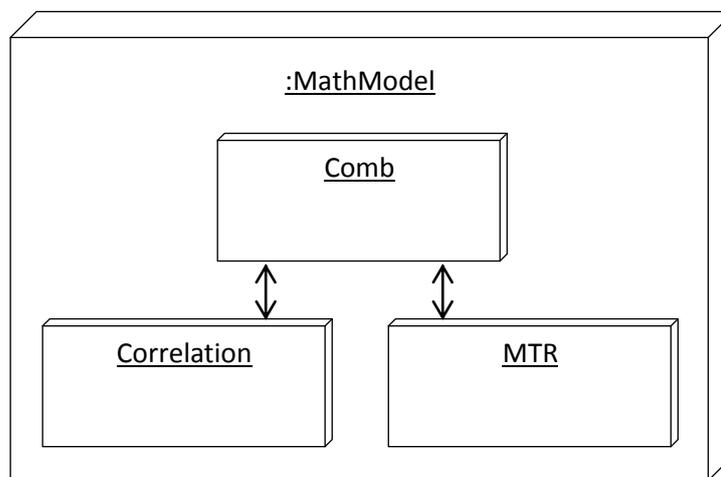


Рисунок 3.1 – Структура программного комплекса

#### 1. *Correlation*

Программа *Correlation* позволяет проанализировать величину линейной корреляционной связи, рассчитанной по критерию Пирсона. Для этого строится корреляционная матрица. Элементы матрицы – коэффициенты корреляции между  $x_i$  и  $x_j$  симметричны относительно главной диагонали, на которой стоят единицы. Чтобы графически представить связи между факторами на основе корреляционной матрицы строится граф корреляционных плед. Таким образом, логика выполнения программы

сводится к расчету коэффициентов корреляции между парами столбцов и интерпретации полученных результатов в виде корреляционной матрицы. После чего программа разделяет полученные значения на плеяды, в соответствии с заданным пороговым коэффициентом корреляции и сохраняет результат в виде матрицы смежности и списка рёбер графа корреляционных плеяд. Программа работает с данными, полученными из внешнего файла, полное имя которого пользователь вводит с клавиатуры (имя по умолчанию: «*data.csv*»). Корреляционная матрица сохраняется в файле с именем, которое пользователь вводит с клавиатуры (имя по умолчанию: «*result.csv*»). Матрица смежности графа корреляционных плеяд сохраняется в файле с именем «*Pleyads*»+имя файла корреляционной матрицы (по умолчанию «*Pleyads50result.csv*»). Также перед именем каждого файла пишется номер шага упрощения графа. Файл с максимальным номером – результат работы программы в виде матрицы смежности. Список рёбер графа корреляционных плеяд сохраняется в файле с именем «*Total*»+имя файла корреляционной матрицы (по умолчанию «*Total50result.csv*»). Цифры в имени файлов графа показывают, с каким пороговым коэффициентом корреляции рассчитывался данный граф. Пользователь может повторить построение графа корреляционных плеяд, введя новый пороговый коэффициент корреляции. После выполнения работы программа выдаёт на экран имена сохранённых файлов и информацию о хранимых в них данных. Логика работы программы наглядно представлена в приложении А2.

## 2. *Mtr*

Программа *Mtr* [71] реализует алгоритм метода точечных распределений, подробно описанный в разделе 1.4.

На вход программы подаётся таблица исходных данных, полученная из внешнего файла, полное имя которого пользователь вводит с клавиатуры (имя по умолчанию: «*in.csv*»). При чтении данных программа разделяет факторы (столбцы исходного файла) и запрашивает у пользователя вид закона распределения для каждого из них. В зависимости от вида закона распределения выбирается алгоритм расчета таблицы ненормированной плотности вероятности в виртуальной области, после чего производятся вычисления и построенные таблицы сохраняются под именами «*i.csv*», где *i* – номер фактора в таблице исходных данных.

Общий алгоритм работы программы представлен в приложении А2.

## 3. *Comb*

Программа *Comb* [72] реализует алгоритм виртуального увеличения многомерных таблиц данных, описанный в разделе 2.3

На вход программы подаётся таблица исходных данных, полученная из внешнего файла, полное имя которого пользователь вводит с клавиатуры (имя по умолчанию: «*in.csv*») и таблицы, построенные после применения метода точечных распределений (имена файлов «*i.csv*», где *i* – номер фактора в таблице исходных данных).

Описание реализованных программ и входящих в них классов, функций и алгоритмов приведено в приложении А2.

Все программы работают с одинаковым типом файлов и единообразной системой имён. Так на вход программы *Mtr* подаётся файл с именем *in.csv*, содержащий таблицу исходных данных. На выходе образуются *n* файлов, где *n* – количество факторов в эксперименте. Все эти файлы при необходимости, можно подать на вход программы *Comb*, которая произведёт их состыковку и сохранит в файле *out.csv* для их последующей обработки.

Весь программный комплекс состоит из консольных приложений, не требующих от пользователя специальных знаний и умений для обработки данных.

Работа данного программного продукта была протестирована на различных массивах данных. Результаты вычислений совпали с рассчитанными контрольными примерами.

Описанное выше программное обеспечение может служить основой для создания различных проблемно-ориентированных СППР. Примеры некоторых из них опишем в следующих параграфах.

### **3.3 СППР о сложности предстоящих лапароскопических операций**

В данном параграфе рассмотрим задачу построения СППР разработанной для применения в лечебном процессе отделения Эндоскопической и малоинвазивной хирургии. Данная система позволяет расширить возможности хирургов в прогнозировании предстоящих интраоперационных трудностей, используя принципы доказательной медицины.

Основными этапами схемы процесса принятия решения, описанными в параграфе 1.1, являются: описание предметной области, анализ ситуации, формулирование проблемы, анализ имеющейся информации, разработка вариантов решения, оценка применимости вариантов решений и их последствий, выбор оптимального варианта, проверка и выдача решения.

Первым этапом схемы процесса принятия решения является описание предметной области, в которой будет проходить дальнейшее исследование проблемы.

Желчнокаменная болезнь (ЖКБ) относится к наиболее распространенным заболеваниям в мире и занимает третье место после сердечнососудистых заболеваний и сахарного диабета [33].

Актуальность проблемы ЖКБ обусловлена увеличением числа пациентов с развитием более сложных форм заболевания, а также в связи с высокой послеоперационной летальностью.

Одним из самых частых осложнений при ЖКБ является холецистит. При этом если диетический подход и ультразвуковые методики не дают положительного результата для лечения назначается операция по удалению желчного пузыря (холецистэктомия).

В настоящее время широкое распространение получили операции, при которых вместо одного широкого разреза выполняются отдельные проколы, через которые возможно введение различных рабочих инструментов хирурга. Данные операции носят название лапароскопических и могут быть выполнены при всех формах желчнокаменной болезни и её осложнениях.

В экономически развитых странах удельный вес лапароскопических холецистэктомий (ЛХЭ) в общей структуре холецистэктомий составляет 85– 95%. Несмотря на это, результаты ЛХЭ не всегда удовлетворяют хирургов, когда идет речь об осложненных формах, технических интраоперационных трудностях или сопутствующих заболеваниях у пациентов. ЛХЭ у таких больных всегда чревата теми или иными осложнениями в ходе операции или в послеоперационном периоде. До настоящего времени часто имеет место лишь слепое использование ЛХЭ без попыток предсказать эти трудности [33].

ЛХЭ, выполняемая избирательно, имеет значительные преимущества и может быть более безопасной и эффективной благодаря правильной предоперационной работе, знанию возможных осложнений и надлежащей подготовке [134].

Профилактическое направление всегда было присуще эффективной медицине, поэтому прогнозирование интраоперационных трудностей операции не только повысит уровень научности имеющихся знаний, накопленных при обследовании больного, но и приведет к уменьшению числа осложнений, инвалидности, летальности, а также материальных затрат лечебного учреждения [89].

В связи с этим актуальна проблема разработки объективной системы прогнозирования предстоящих трудностей при ЛХЭ, которая поможет избежать непредвиденных нежелательных осложнений.

Дополнительные трудности вызывает тот факт, что ЖКБ у мужчин встречается примерно в два раза реже, чем у женщин [50], при этом риск отказа от продолжения операции эндохирургически и её завершения открытым способом в 2,5 раза выше [113, 157]. Более того, известно, что принадлежность к мужскому полу является единственным статистически значимым фактором такого риска [96, 100, 112]. В связи с этим для прогнозирования уровня сложности предстоящих лапароскопических операций был выбран самый трудный случай, а именно предстоящее оперирование мужчин с острым обструктивным калькулёзным холециститом. Данный выбор обусловлен тем фактом, что при хроническом течении болезни проведение ЛХЭ в настоящее время является штатной процедурой и интраоперационные осложнения сведены к минимуму. Однако, при наличии конкрементов, нарушается проходимость органов и воспалительные процессы переходят в острую форму, что влечёт за собой увеличение сложности операции и повышается риск её перехода в открытую форму.

Вторым этапом схемы процесса принятия решения является анализ ситуации об исследуемом процессе. Такой анализ показал невозможность быстрого набора большого количества статистических данных для проведения исследования классическими методами. Однако в связи с необходимостью предположить уровень сложности предстоящей лапароскопической операции, было предложено проводить анализ методами математического моделирования по пассивным выборкам малого объема, используя данные, полученные по результатам операций, проведенных мужчинам в течении года.

На третьем этапе происходит формулирование проблемы и определение критериев, по которым будет производиться сравнение и выбор вариантов оптимальных решений, предоставляемых ЛПП.

Уровень сложности лапароскопической операции было решено кодировать значениями от 1 до 4 по характеристикам, изложенным в работе [34]. Данное кодирование основано на многолетнем опыте экспертов, практикующих врачей-хирургов, проводивших данное исследование.

I степень – ЛХЭ длительностью не более 60 минут, которая проводится, как правило, при хроническом неосложнённом калькулёзном (или бескаменном) холецистите без каких-либо особенностей и осложнений у больного, без лишнего веса или с ожирением I степени.

II степень – ЛХЭ длительностью от 60 до 90 минут, проведенная на фоне умеренных патологоанатомических изменений и интраоперационных осложнений:

- умеренный или выраженный перихолецистит при хроническом холецистите, плоскостные спайки в области желчного пузыря после предыдущих операций или перенесённого гепатита;

- напряжённый желчный пузырь больших размеров, требующий пункции (водянка, острый холецистит, застойный желчный пузырь);

- умеренные трудности, связанные с аномалиями анатомии;

- микроперфорации желчного пузыря электрокрючком или зажимом, не требующие использования контейнера;

- не длительное артериальное, венозное или капиллярное кровотечение, требующее неоднократного применения электрокоагуляции и последующего визуального контроля;

- небольшие надрывы печени при тракции желчного пузыря, микроповреждения её инструментами.

II степень сложности ЛХЭ также присваивается операции, выполненной без вышеперечисленных критериев у больного с ожирением II степени в связи с трудностями при введении троакаров, уменьшением свободного пространства для пневмоперитонеума, жировыми отложениями в области желчного пузыря и гепатодуоденальной связки.

III степень – ЛХЭ длительностью более 90 минут, выполненная при наличии следующих обстоятельств, выраженных патологоанатомических изменений и интраоперационных осложнений:

- выраженные трудности, связанные с аномалиями анатомии;

- рыхлый или плотный подпечёчный инфильтрат при остром деструктивном холецистите;

- перфорация желчного пузыря электрокрючком или зажимом, некроз стенки пузыря с её фрагментацией, разрывами, гангреной, требующие использования контейнера для эвакуации желчного пузыря и конкрементов;

- длительное артериальное, венозное или капиллярное кровотечение, требующее неоднократного применения электрокоагуляции, аспирации крови и сгустков из области операции и последующего визуального контроля;

- выраженные инфильтративные изменения в области шейки желчного пузыря, крайне затрудняющие дифференциацию элементов треугольника Кало;

- интимное предлежание задней стенки желчного пузыря к его ложу в печени при склерозе пузыря, подостром холецистите, интраоперационно сопровождающееся множественными повреждениями печени в области ложа, оставлением задней стенки желчного пузыря;

- разрыв желчного пузыря при эвакуации его через рану брюшной стенки с выпадением конкрементов, излитием содержимого пузыря в брюшную полость.

III степень сложности ЛХЭ присваивается и операции, выполненной без вышеперечисленных критериев, у больного с ожирением III степени в связи с трудностями при введении троакаров, уменьшением свободного пространства для пневмоперитонеума, жировыми отложениями в области желчного пузыря и гепато-дуоденальной связки.

IV степень – ЛХЭ вне зависимости от длительности операции, сопровождающаяся выраженными патологоанатомическими изменениями в зоне операции и интраоперационными осложнениями, требующими обязательной конверсии на лапаротомию.

IV степень сложности также присваивается ЛХЭ, которая в послеоперационном периоде потребовала повторных лечебных видеолапароскопий или лапаротомии для ликвидации осложнений операции.

Как видно из разработанной классификации, степень сложности определяется хирургом постфактум, на основании определённых характеристик проведённой операции. Однако, возможность заранее спрогнозировать предстоящие интраоперационные трудности и выдача рекомендаций ведущему операцию хирургу позволит более тщательно провести подготовительные работы, что в итоге приведёт к снижению числа осложнений. При этом большинство факторов не только не поддаются строгой формализации, но и не могут быть оценены до начала операции.

В связи с этим, было необходимо отобрать входные факторы, от которых может зависеть сложность операции и влияние которых затруднительно оценить хирургу в ходе подготовки к операции. Данная процедура проводилась совместно с практикующими хирургами отделения эндоскопической и малоинвазивной хирургии. Экспертная группа состояла из специалистов в области математического моделирования (д.т.н. профессор Ю.А. Долгов и к.т.н. доцент Столяренко Ю.А) и медицины (к.м.н. хирург-эндоскопист Звягинцев В.В.; к.м.н. хирург-эндоскопист Фомов Г.В.; врач высшей категории, заведующий отделения ЭиМХ Горпинюк В.П., к.м.н. профессор Ставинский Р.А.). Таким образом, из 42 исследуемых перед операцией признаков были отобраны 9, которые вероятно влияют на степень сложности операции, могут быть формализованы и измерены количественно:

1. Возраст.
2. Количество лет камненосительства.

3. Длина желчного пузыря, мм.
4. Ширина желчного пузыря, мм.
5. Толщина стенки желчного пузыря, мм.
6. Максимальный размер конкрементов, мм.
7. Конкремент в шейке желчного пузыря, мм.
8. Время от начала приступа острого холецистита, сутки.
9. Лейкоциты крови, ед/л.

Кроме перечисленных факторов, предикторами, которые необходимо учитывать при прогнозировании также являются наличие у пациента ожирения, либо других выявленных анатомических изменений.

Таким образом, была поставлена задача разработать проблемно-ориентированную компьютерную СППР, опирающуюся на многолетний опыт и знания экспертов в данной области и использующую вероятностно-статистические методы для обоснованной помощи ЛПР в принятии решения о сложности предстоящей лапароскопической операции.

Общая структура СППР представлена на рисунке 3.2.

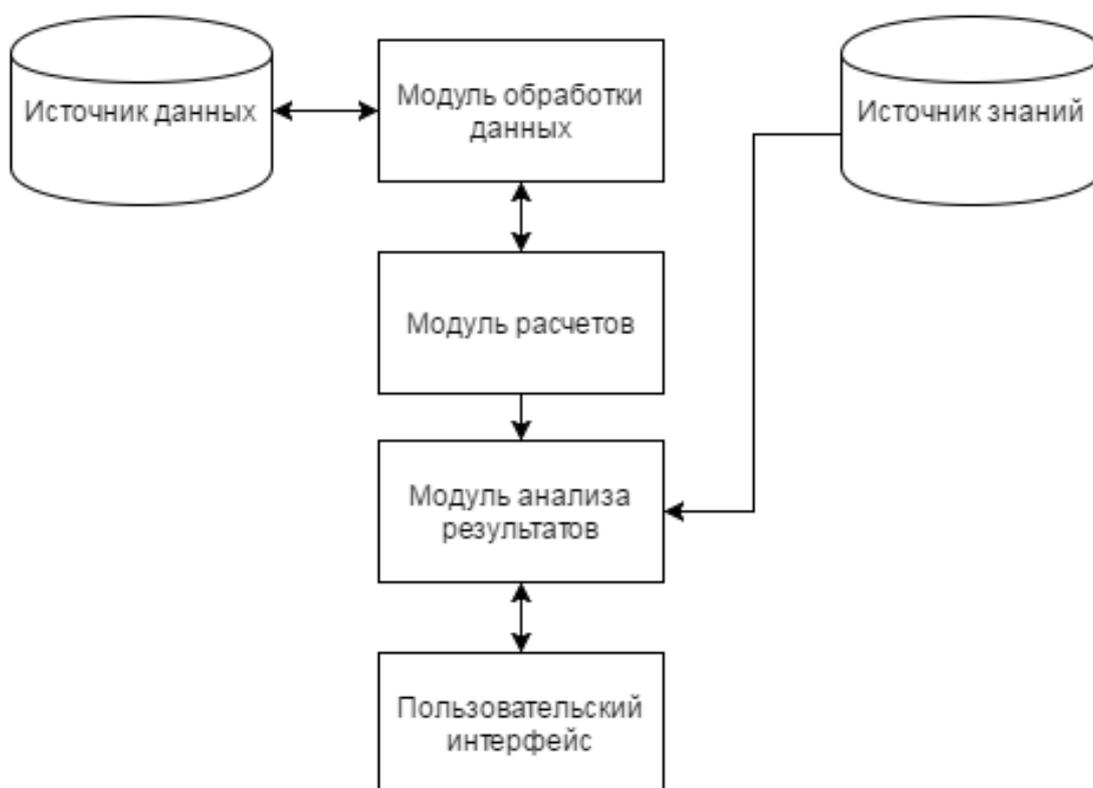


Рисунок 3.2 – Общая структура СППР

На этапе анализа имеющейся информации данные, собранные экспертами, были кодированы и записаны в таблицу, необходимую для дальнейшей обработки математическими методами.

В таблице 3.1 представлены исходные данные для моделирования, полученные при оперировании острых обструктивных калькулёзных холециститов у мужчин за один календарный год, где в строках располагаются сводные данные о восьми проведённых операциях, а в столбцах – предикторы, индексы которых равняются соответствующим значениям описанных ранее входных факторов. В таблице исходных данных отсутствует информация об операциях, проведённых на первом и четвёртом уровнях сложности. Это объясняется тем, что первый уровень сложности при острых холециститах не присваивается в соответствии с методикой [34], а четвертый уровень связан с переходом операции в открытую форму, что является самым нежелательным развитием событий, которого в исследуемый период времени удалось избежать. Однако математическое моделирование будет производиться средствами регрессионного анализа, который позволяет аналитически описать связь между предикторами и зависимой величиной, а значит и предсказывать значения, основываясь на зависимости результативного признака от величины факторов.

В качестве независимой переменной  $Y$  для прогнозирования выбран уровень сложности операции, который определён хирургом эмпирически после окончания проведения операции.

Таблица 3.1 – Исходные данные

№	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$Y$
1	33	1	98	38	3	7	7	7	9,8	2
2	44	1	90	38	4	12	12	5	8,9	3
3	43	1	110	40	5	20	10	5	9,5	3
4	30	7	110	37	2	15	0	6	8,9	2
5	65	15	120	45	5	20	15	8	9	3
6	66	10	110	55	5	10	10	5	12,6	3
7	70	7	110	43	5	15	15	14	9,7	3
8	52	1	90	40	4	22	15	3	12,5	3

Следующим этапом является разработка вариантов решения проблемы.

Применение к исходным данным классических подходов не позволяет построить математическую модель, включающую в себя информацию о большинстве исследуемых параметров, которую возможно признать статистически значимой, так как исходный набор данных является сверхнасыщенным планом.

Проверим исходные данные на наличие аномальных измерений. Для этого воспользуемся критерием Диксона, который в исследовании [68] показал самую высокую мощность при объеме исследуемых данных в 8 элементов.

В таблице 3.2 представлены рассчитанные значения критерия Диксона для проверки одного сомнительного наблюдения ( $r_{10}$ ), а также для проверки одного сомнительного наблюдения независимо от одного противоположного крайнего наблюдения ( $r_{11}$ ). При этом проверялись, как максимальное, так и минимальное значение каждого из исследуемых параметров и выходной величины.

Таблица 3.2 – Расчетные значения критерия Диксона

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$Y$
$MAX r_{10}$	0,1	0,357	0,333	0,556	0	0,133	0	0,545	0,027	0
$MAX r_{11}$	0,108	0,357	0,333	0,588	0	0,167	0	0,667	0,027	0
$MIN r_{10}$	0,075	0	0	0,056	0,333	0,2	0,467	0,182	0	0
$MIN r_{11}$	0,083	0	0	0,125	0,333	0,231	0,467	0,4	0	0

При сравнении рассчитанных значений с критическими значениями для данного метода ( $r_{10}=0,59$ ,  $r_{11}=0,683$ ) при доверительной вероятности, рекомендуемой для медицинских исследований, равной 0,01 следует обратить внимание, что все минимальные значения значительно ниже критических, что может с точностью свидетельствовать об отсутствии ошибок измерений, которые могут повлиять на качество статистических оценок.

При анализе максимальных элементов можно выделить два параметра, расчетные значения которых ниже критических, но находятся близко к границе чувствительности применяемого метода при использовании проверки одного сомнительного наблюдения независимо от одного противоположного крайнего наблюдения. Этими параметрами являются  $X_4$  (Ширина желчного пузыря, мм) и  $X_8$  (Время от начала приступа острого холецистита, сутки).

На основе анализа полученных результатов было принято решение провести повторную проверку на наличие аномального измерения при помощи  $L$ -критерия Титьена-Мура, применение которого, как показано в параграфе 2.1, приводит к меньшему количеству ошибок I типа, то есть вероятности определить значение аномальным, в то время, как оно таковым не является. Расчетные значения для данного критерия приведены в таблице 3.3

Таблица 3.3 – Расчетный уровень  $L$ -критерия Титъена-Мура

Расположение выброса	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$Y$
Наибольший элемент	0,735	0,459	0,677	0,208	0,901	0,726	0,876	0,202	0,577	0,952
Наименьший элемент	0,714	0,888	0,698	0,883	0,419	0,617	0,323	0,807	0,899	0,571

В отличие от критерия Диксона, для принятия гипотезы о наличии аномального измерения в соответствии с  $L$ -критерием Титъена-Мура, рассчитанное значение должно быть меньше критического для заданного уровня значимости ( $L=0,202$ ). Как видно из полученных значений, максимальные значения ширины желчного пузыря и времени от начала приступа острого холецистита также находятся в пределах погрешности метода. В связи с чем данные показатели были повторно проверены вместе с экспертами-хирургами, непосредственно проводящими операции, которые подтвердили правильную запись таблицы исходных данных.

Для проверки гипотезы о нормальности в данных малого объема рекомендуется применять непараметрические критерии согласия, такие как тест Шапиро-Уилка. Применяв данный критерий к таблице входных данных, с доверительной вероятностью 99%, мы можем отклонить гипотезу о нормальности генеральной совокупности из которой получена исследуемая величина только для фактора  $X_9$ , который характеризует количество лейкоцитов в крови. При уровне значимости в 95% также может быть отклонена гипотеза о нормальности распределения данных о ширине желчного пузыря и толщине его стенки.

Применение критерия Шапиро-Уилка к зависимой переменной нецелесообразно, так как уровень сложности операции является дискретной величиной. Построим графики плотности вероятности для трёх признаков, распределение которых может значимо отличаться от нормального, чтобы визуально оценить степень отклонения и влияющие на неё величины (рисунок 3.3).

Как видно из графика, количество лейкоцитов в крови, по имеющимся данным, имеет ярко выраженное двумодальное распределение, из-за чего и были получены такие результаты при применении теста Шапиро-Уилка. Однако, с уверенностью отклонить гипотезу о нормальности данного параметра невозможно из-за малого количества исследуемых данных. Так, если бы в нашей таблице было всего на один элемент больше, то подобный график мог бы сгладиться до унимодального, с небольшой асимметрией. Основываясь на полученной информации, следует сделать вывод, что с необходимой

достоверностью невозможно ни принять, ни отвергнуть гипотезу о нормальности распределений генеральных совокупностей исследуемых величин. Учитывая тот факт, что для построения регрессионных моделей нормальность предикторов не является строго необходимым требованием, перейдем к анализу зависимостей в данных.

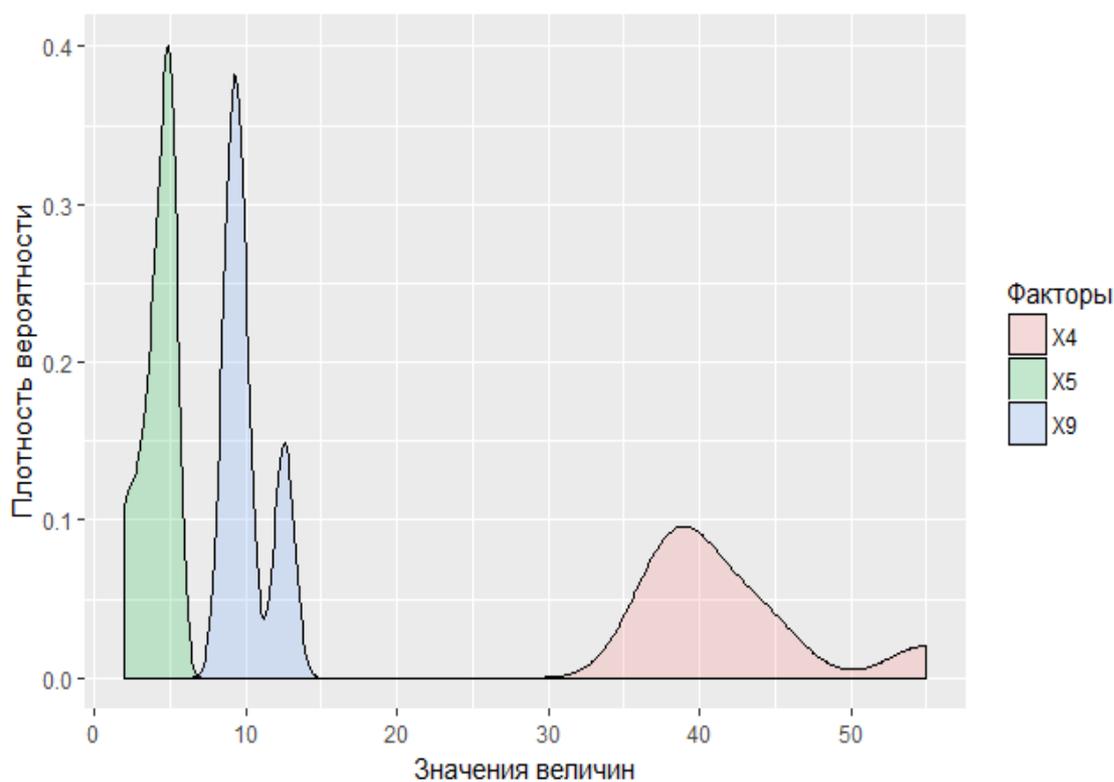


Рисунок 3.3 – Плотности вероятности исследуемых величин

Проведенный корреляционный анализ полученных данных выявил сильную линейную корреляционную связь между факторами  $X_1$ ,  $X_4$ ,  $X_5$  и  $X_7$ , а также между факторами  $X_2$  и  $X_3$  (таблица 3.4). Экспертами было отмечено, что каждый из данных факторов важен для рассмотрения, поэтому нельзя «отбрасывать» из рассмотрения ни один из них и необходимо строить модели для каждого случая. Кроме того, проведенные исследования показали, что при столь малом объеме выборки коэффициент корреляции Пирсона не всегда показывает достоверный уровень линейной связи между анализируемыми факторами. В связи с вышеизложенным, было принято решение применить многомерный метод точечных распределений для уточнения характеристик выборки и дальнейшего построения регрессионных уравнений на основе виртуально увеличенной таблицы данных.

Таблица 3.4 – Корреляционная матрица

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
$X_1$	1	0,5982	0,3849	0,7468	0,8296	0,2234	0,7608	0,4512	0,3651
$X_2$	0,5982	1	0,8061	0,6175	0,3028	0,1050	0,1179	0,3654	-0,0513
$X_3$	0,3849	0,8061	1	0,4417	0,3188	0,1770	-0,0843	0,4512	-0,2632
$X_4$	0,7468	0,6175	0,4417	1	0,6232	-0,1369	0,3004	0,0653	0,5826
$X_5$	0,8296	0,3028	0,3188	0,6232	1	0,3080	0,7999	0,2425	0,2535
$X_6$	0,2234	0,1050	0,1770	-0,1369	0,3080	1	0,4050	-0,1343	0,0312
$X_7$	0,7608	0,1179	-0,0843	0,3004	0,7999	0,4050	1	0,2285	0,2682
$X_8$	0,4512	0,3654	0,4512	0,0653	0,2425	-0,1343	0,2285	1	-0,3788
$X_9$	0,3651	-0,0513	-0,2632	0,5826	0,2535	0,0312	0,2682	-0,3788	1

Проведём процедуру виртуального увеличения таблицы исходных данных на основе алгоритма, изложенного в параграфе 2.3. В связи с невозможностью достоверно оценить вид закона распределения по исходным данным малого объема и отсутствием априорной информации по этому вопросу, было принято решение использовать метод точечных распределений с параметрами, характерными для нормального закона распределения случайных величин. При этом алгоритм метода предполагает изначальное определение границ изменения величин, что не позволяет полученным данным принимать значения, не согласующиеся с реальными характеристиками исследуемого объекта. В частности, при построении таблиц ненормированной плотности вероятности в виртуальной области нижняя граница факторов «Количество лет камненосительства» и «Конкремент в шейке желчного пузыря» была сдвинута на уровень нуля, в связи с невозможностью отрицательных значений данных. В результате преобразований и состыковки полученных таблиц ненормированных плотностей вероятностей получим итоговую виртуально увеличенную таблицу данных, содержащую 88 строк, представленную в приложении А3. Итоговая таблица была просмотрена экспертами на наличие многомерных грубых ошибок, которые выбивались бы из общей совокупности проявлений заболевания. Было подтверждено, что одновременное возрастание всех показателей при росте сложности операции в отдельных частях таблицы не противоречит реальной клинической картине.

Повторный корреляционный анализ по виртуально увеличенной выборке не подтвердил гипотезу о наличии сильной линейной зависимости между факторами, первоначально объединенными в плеяду. По вновь рассчитанным значениям, в соответствии с алгоритмом, возможно признать сильной только связь между факторами  $X_1$  и  $X_5$ , которая больше порогового значения 0,8, определённого для данного метода. Однако, эксперты высказали мнение, что и эта связь, возможно, определена ошибочно, в

связи с чем было принято решение оставить для дальнейшего рассмотрения все полученные данные.

По полученной виртуальной таблице данных была построена модель методом наименьших квадратов. В соответствии с алгоритмом метода, в ходе расчетов, фактор  $X_9$  был отброшен, как незначимый, в результате чего была получена модель:

$$Y=0,3406+0,0147X_1-0,0355X_2+0,0175X_3-0,0405X_4+0,43X_5-0,0607X_6+0,1276X_7-0,1034X_8,$$

которая является статистически значимой при проверке на таблице виртуально увеличенных данных.

Проверить значимость данной модели для исходных данных не предоставляется возможной, так как количество входящих в неё факторов равно количеству наблюдений, что делает невозможным расчет статистики Фишера из-за отрицательного значения степени свободы.

Таким образом, проведенная оценка применимости вариантов решений показала возможность применения разработанной модели для выдачи рекомендаций ЛПР по предварительному прогнозированию сложности ЛХЭ.

На основании данных, полученных вероятностно-статистическими методами, была реализована проблемно-ориентированная СППР, учитывающая предыдущий опыт экспертов в данной области.

Для выдачи рекомендаций ЛПР, используется подход, сочетающий применение полученной математической модели и нескольких продукционных правил. Это связано с тем, что в модель не были включены некоторые номинативные факторы, например, такие как степень ожирения.

При наличии у пациента анатомических аномалий или ожирения соответствующей степени предсказанная сложность операции автоматически увеличивается до указанной в вышеизложенных рекомендациях.

На рисунке 3.4 показан интерфейс окна программы для ввода информации о предстоящей операции. Была реализована проверка правильности при вводе данных. Во всех полях, кроме «ФИО», должны быть введены числовые данные из диапазона, указанного экспертами. Для устранения ошибок при обработке дробных чисел, при вводе пользователем в качестве знака разделителя точки, она автоматически заменяется запятой. Если введены не корректные данные, выводится сообщение об ошибке, после которого следует исправить неточности и повторить ввод.

Рисунок 3.4 – Интерфейс окна ввода информации о пациенте

После ввода информации о предстоящей операции и нажатии кнопки «Подтвердить данные» открывается окно с консультациями для ЛПП (рисунок 3.5), в котором приводится вероятность прогнозируемой степени сложности операции, её описание и рекомендации по проведению ожидаемой операции.

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.

Если данные в главном окне программы были введены корректно, то они сразу записываются в файл, который используется для дальнейшего ретроспективного анализа работы СППР. Для просмотра и редактирования данного файла по кнопке «Архив данных», расположенной в левом верхнем углу главного окна программы, можно перейти

на дочернюю форму (рисунок 3.6), на которой в виде таблицы отображается информация о проведенных операциях.

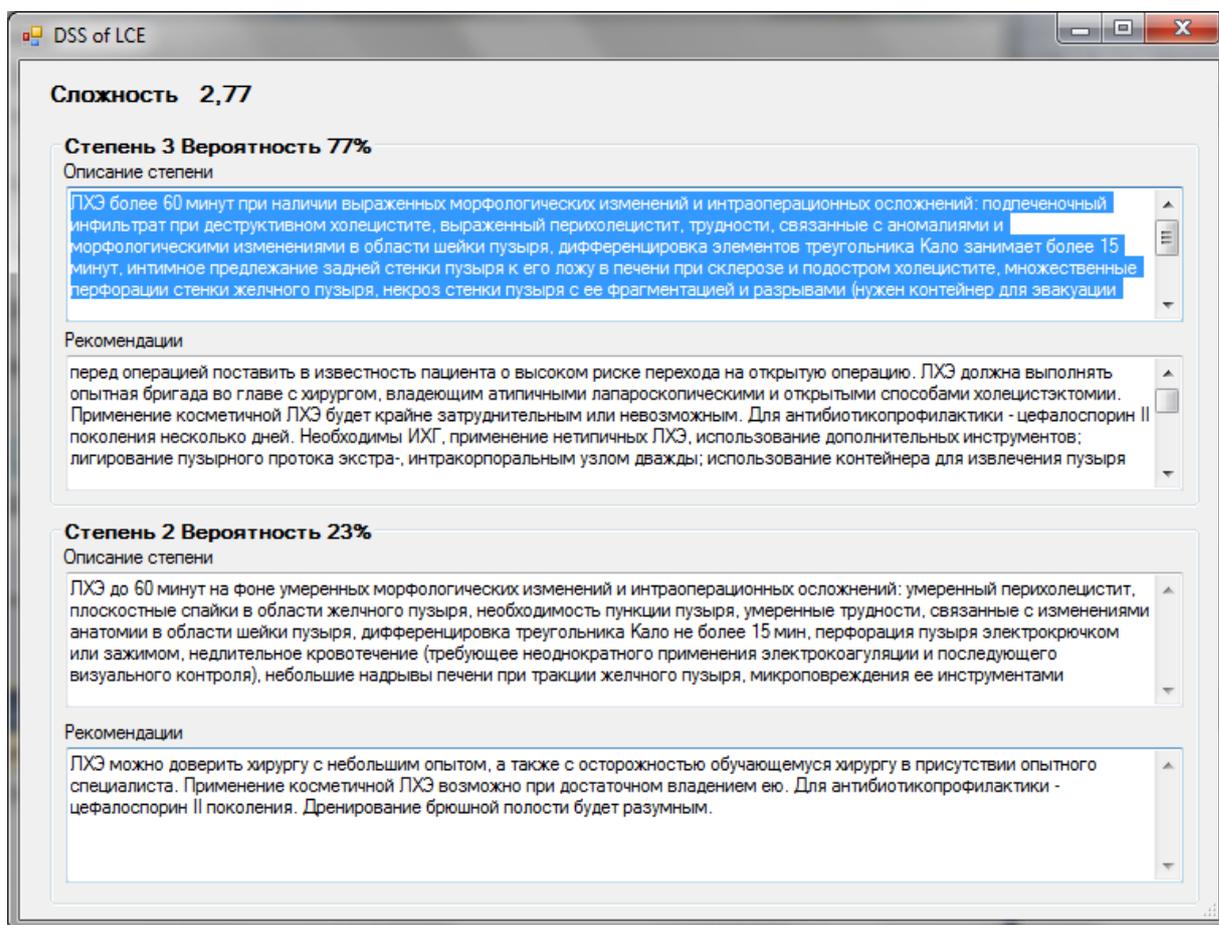


Рисунок 3.5 – Рекомендации для ЛПП

Данную таблицу можно редактировать, что может быть применимо в случае ошибочно набранных данных, либо для внесения фактического уровня сложности уже проведенной операции, что может быть использовано для дальнейшего ретроспективного анализа. Сохранение изменений производится по кнопке, расположенной в правом верхнем углу окна.

По прошествии времени, у мужчин с острым холециститом было проведено ещё 11 операций, сложность которых оценивалась построенной моделью.

Результаты оценивались следующим образом:  $Y < 2,5$  – II уровень сложности,  $2,5 \leq Y \leq 3,4$  – III уровень сложности,  $Y > 3,4$  – IV уровень сложности. Первый уровень сложности операции, при остром холецистите, не может быть назначен, в соответствии с методикой [34].

Сохранить изменения

	№ карты	ФИО	Возраст	Кол-во лет камненосит	Длина желчного пузыря	Ширина желчного пузыря	Толщина стенки желчного пузыря	Максимальный размер конкрементов	Конкремент в шейке желчного пузыря	Время от начала приступа	Расчетная сложность операции	Фактическая сложность операции
▶	1	Ива...	33	1	98	38	3	7	7	7	2,51	2
	2	Пет...	44	1	90	38	4	12	12	5	3,19	3
	3	Ива...	43	1	110	40	5	20	10	5	3,09	3
	4	Пет...	30	7	110	37	2	15	0	6	2	2
	5	Вас...	65	15	120	45	5	20	15	8	3,61	3
	6	Гри...	66	10	110	55	5	10	10	5	3,27	3
	7	Анд...	70	7	110	43	5	15	15	14	4,07	3
	8	Але...	52	1	90	40	4	22	15	3	3,08	3

Рисунок 3.6 – Информация о прошедших операциях

Таким образом, разработанная СППР правильно распознала сложность операции и выдала точные рекомендации в 73% случаев при последующих операциях, что может быть признано удовлетворительным при раннем прогнозировании и на этапе набора информации для более точного статистического анализа. Следует отметить, что среди проанализированных операций было две, в ходе которых было принято решение об их переводе в открытую фазу. Разработанная система успешно справилась с предсказанием заданной степени сложности, указав, соответственно, риск конверсии в 73% и 100% для каждой из операций. Наиболее же частой ошибкой математической модели является предсказание первого уровня сложности, вместо второго, что может быть исправлено добавлением ограничивающего правила, запрещающего выдавать рекомендации по первому уровню сложности при острых холециститах.

Проверка построенной на раннем этапе модели, на полученном итоговом наборе данных (таблица 3.5) показала, что модель является адекватной, статистически значимой и входящие в неё факторы описывают выходную величину примерно на 70%, при величине среднеквадратичной ошибки оценки равной 0,851.

Для сравнения приведём (таблица 3.6) значения коэффициента смешанной корреляции( $r^2$ ), среднеквадратичной ошибки оценки( $S_{xy}$ ), табличного( $F_t$ ) и расчетного( $F$ ) значений критерия Фишера, рассчитанные на том же наборе данных, для полученной модели( $Y_0$ ) и моделей( $Y_1$ - $Y_2$ ), которые могли бы быть построены без применения предложенного метода.

В приведенной таблице  $Y_1$  – модель, построенная методом наименьших квадратов по исходному набору данных без учета корреляции,  $Y_2$  – вариант модели, построенной методом наименьших квадратов, по исходному набору данных, с выделением факторов, характеризующих каждую из плеяд ( $X_1$  и  $X_2$  соответственно).

Таблица 3.5 – Итоговый набор данных

№	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$Y$
1	33	1	98	38	3	7	7	7	9,8	2
2	44	1	90	38	4	12	12	5	8,9	3
3	43	1	110	40	5	20	10	5	9,5	3
4	30	7	110	37	2	15	0	6	8,9	2
5	65	15	120	45	5	20	15	8	9	3
6	66	10	110	55	5	10	10	5	12,6	3
7	70	7	110	43	5	15	15	14	9,7	3
8	52	1	90	40	4	22	15	3	12,5	3
9	69	10	160	60	7	9	9	12	9,8	4
10	35	3	100	40	4	20	0	7	11	2
11	54	8	100	40	4	20	0	12	9	2
12	63	8	110	40	8	10	0	7	10	3
13	45	4	120	40	4	15	0	5	10	3
14	60	5	110	45	4	15	0	3	9,8	2
15	70	10	120	60	7	15	10	9	9,8	3
16	80	4	130	50	5	4	0	7	11	3
17	40	2	110	40	5	7	7	10	9	3
18	68	3	120	40	5	36	36	10	10,1	4
19	47	1	95	40	4	25	25	21	9,8	3

Таблица 3.6 – Сравнительная характеристика качества моделей

	$Y_0$	$Y_1$	$Y_2$
$r^2$	0,702	0,444	0,469
$S_{xy}$	0,851	0,781	0,690
$F$	6,147	1,2565	2,29
$F_t$	3,07	3,012	3,025

Анализ величины корреляционной связи по итоговой таблице данных показал, что предположение об отсутствии высокого уровня линейной зависимости между величинами находит своё подтверждение. Видно, что в данном примере, при увеличении объема данных численные значения коэффициента корреляции Пирсона уменьшаются и, как показано в таблице 3.7 не выходят из области заметных связей, по классификации, приведенной в параграфе 2.4.

Таблица 3.7 – Корреляционная матрица

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$
$X_1$	1	0,5298	0,5552	0,6776	0,6583	-0,0304	0,2230	0,1443	0,2575
$X_2$	0,5298	1	0,5107	0,5843	0,4322	-0,1262	-0,1385	0,0558	-0,0893
$X_3$	0,5552	0,5107	1	0,6853	0,5284	-0,2330	-0,0523	0,0974	-0,0764
$X_4$	0,6776	0,5843	0,6853	1	0,5948	-0,3090	-0,0215	0,0658	0,2802
$X_5$	0,6583	0,4322	0,5284	0,5948	1	-0,1664	0,0694	0,1459	0,0714
$X_6$	-0,0304	-0,1262	-0,2330	-0,3090	-0,1664	1	0,6671	0,2415	0,0005
$X_7$	0,2230	-0,1385	-0,0523	-0,0215	0,0694	0,6671	1	0,4412	0,0554
$X_8$	0,1443	0,0558	0,0974	0,0658	0,1459	0,2415	0,4412	1	-0,3004
$X_9$	0,2575	-0,0893	-0,0764	0,2802	0,0714	0,0005	0,0554	-0,3004	1

Таким образом, применение методики вероятностно-статистической обработки пассивных выборок малого объема позволило построить компьютерную проблемно-ориентированную СППР для прогнозирования сложности лапароскопических операций на тот момент, когда было невозможно набрать данные, достаточные для моделирования классическими вероятностно-статистическими методами.

### 3.4 СППР о качестве пробной выпечки хлебобулочных изделий

В настоящее время рынок хлебобулочных изделий нуждается в поиске путей повышения эффективности. Несмотря на то, что общее потребление хлебобулочных изделий в целом является стабильным [94], производители хлеба сталкиваются с новыми рисками, связанными, с одной стороны, с тенденциями к увеличению спроса на здоровую пищу и возврату к домашней выпечке, а с другой стороны возросшей конкуренции со стороны предприятий фаст-фуда [120].

При этом сохраняется тенденция низкой маржи и традиционно высокая конкуренция между самими предприятиями хлебопекарного рынка. По данным статистики, потребление традиционных видов хлеба на душу населения с каждым годом уменьшается, популярность же нетрадиционных видов хлеба, наоборот, устойчиво растёт [123]. Кроме того, качество и свежесть продукции всё больше влияет на покупательский спрос, так более 40% опрошенных [120] отмечают приоритет качества над ценовым фактором, а более 70 процентов называют одним из главных критериев хлебобулочного изделия – свежесть.

В связи с вышеизложенным, для того, чтобы предприятие могло оставаться успешным и конкурентоспособным на данном рынке, необходимо постоянное увеличение

ассортимента изделий. Данный шаг позволяет охватить большее разнообразие потребительских вкусов и увеличить объем производства.

В настоящее время часть показателей качества хлебобулочных изделий, согласно существующим ГОСТам и техническим условиям, определяется экспертными методами измерений. В пищевой промышленности широко распространенными являются органолептические методы, которые в ряде случаев не обеспечивают единства измерений. Для решения этой проблемы разрабатываются различные математические подходы [77, 88], призванные помочь автоматизировать процесс принятия решения, на основе количественного подхода. В настоящее время разработаны различные системы поддержки принятия решений [93, 106], позволяющие ЛПП принимать более достоверные решения на основе полученных на различных этапах производства данных.

Если на предприятии производится выпечка широкого ассортимента булочных изделий, то применяется порционная методика приготовления теста, которая предполагает отдельный замес теста для каждой партии продукции. На этапе запуска новой продукции хлебопекарные производства, как правило, используют метод пробной лабораторной выпечки [ГОСТ 27669-88], на основе которого для конкретного наименования устанавливаются физико-химические показатели [ГОСТ Р 52462-2005]. Как правило, при проведении пробной выпечки используется рецептура на 100 кг муки, что позволяет получить воспроизводимые результаты и, если заключительные анализы оказываются положительными, наладить процесс выпуска новой продукции с аналогичными характеристиками. Однако, если выходные анализы оказываются не соответствующими нормам (неудовлетворительными), то производится повторный пробный замес, с откорректированной рецептурой, либо параметрами технологического процесса. Обычно, такая работа ведётся на основе знаний и опыта технологов предприятия. Качество готовой хлебобулочной продукции одновременно определяется по различным показателям, среди которых [ГОСТ 31805-2012]:

1. Объем выпеченного изделия.
2. Влажность изделия.
3. Кислотность изделия.
4. Пористость изделия.
5. Внешний вид изделия.
6. Состояние мякиша.
7. Вкусовые качества.

Оценка качества готовой продукции и выдача рекомендаций по оперативному регулированию рецептуры и технологического процесса по пробной партии является сложной аналитической процедурой, в связи с чем, было принято решение разработать СППР [69], которая позволит своевременно принимать более обоснованные решения, основываясь на информации малого объема.

При этом, так как производится 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. Использование большого количества образцов из каждой партии также является невозможным, так как при правильной организации производственного процесса основные характеристики партии являются однородными и практически не имеют вариаций контролируемых параметров, что и отражается в лабораторных актах.

Более того, использование данных эксперимента, которые, фактически, не являются статистически независимыми, приведёт к так называемой проблеме «мнимых повторностей» или псевдорепликации [121]. Это означает, что при построении регрессионной модели на основе такого набора данных оценка среднего квадрата отклонений наблюдаемых значений от предсказанных моделью величин будет иметь тенденцию превысить истинную дисперсию ошибок наблюдений. Такая переоценка снизит мощность статистических тестов и сделает консервативной проверку гипотезы о значимости коэффициентов модели, что может привести к неверной интерпретации результатов [118].

В связи с вышеизложенным, для оперативного регулирования рецептуры и технологического процесса выпечки было решено использовать методы, предложенные для многомерного моделирования по выборкам малого объема.

К основным факторам, влияющим на формирование качества хлебобулочной продукции, относятся:

1. Качество основного сырья:
  - a. Вид муки.
  - b. Влажность муки.
  - c. Зольность муки.
  - d. Количество клейковины.
  - e. Качество клейковины
2. Технология изготовления:

- a. Количество муки
- b. Количество воды
- c. Количество добавок (дрожжи, соль, сахар и др.).
- d. Время выпечки.
- e. Температура выпечки.

На этапе анализа ситуации в области исследуемого объекта, было выявлено, что основными количественными характеристиками качества являются объем выпеченной продукции и её влажность. Внешний вид изделия включает в себя ряд показателей, среди которых основными являются: поверхность корки (гладкая, неровная, с трещинами, с подрывами, рваная), а также цвет корки (бледная, светло-желтая, светло-коричневая, коричневая, темно-коричневая). Состояние мякиша также определяется органолептическими методами по следующим показателям: цвет, равномерность окраски, эластичность и пористость. Вкусовые качества, согласно ГОСТу определяются путем разжевывания и отражают вкус, хруст, комкуемость и крошковатость. Показатели, влияющие на формирование качества, являются количественными переменными, что позволяет использовать все полученные данные для построения математических моделей, описывающих изготовление изделий.

При изучении различных подходов к оценке органолептических показателей качества хлебобулочных изделий было выявлено, что наиболее приемлемой для определения потребительских характеристик хлеба является методика, разработанная в Московском технологическом институте пищевой промышленности. Данная методика подразумевает использование пятибалльной шкалы для каждого из показателей качества. Расшифровка соответствия баллов словесным определениям по каждому из видов хлебобулочных изделий приведена в приложении А4. Для учета значимости показателей в формировании качества хлеба установлены их весовые коэффициенты, таким образом, что общая максимально возможная сумма равна 100 баллам.

Таким образом, анализируемые показатели можно разбить на две группы:

1. Количественные непрерывные величины:
  - 1.1. Объем выхода.
  - 1.2. Влажность мякиша.
  - 1.3. Кислотность.
  - 1.4. Пористость.
2. Балльные величины:
  - 2.1. Форма.

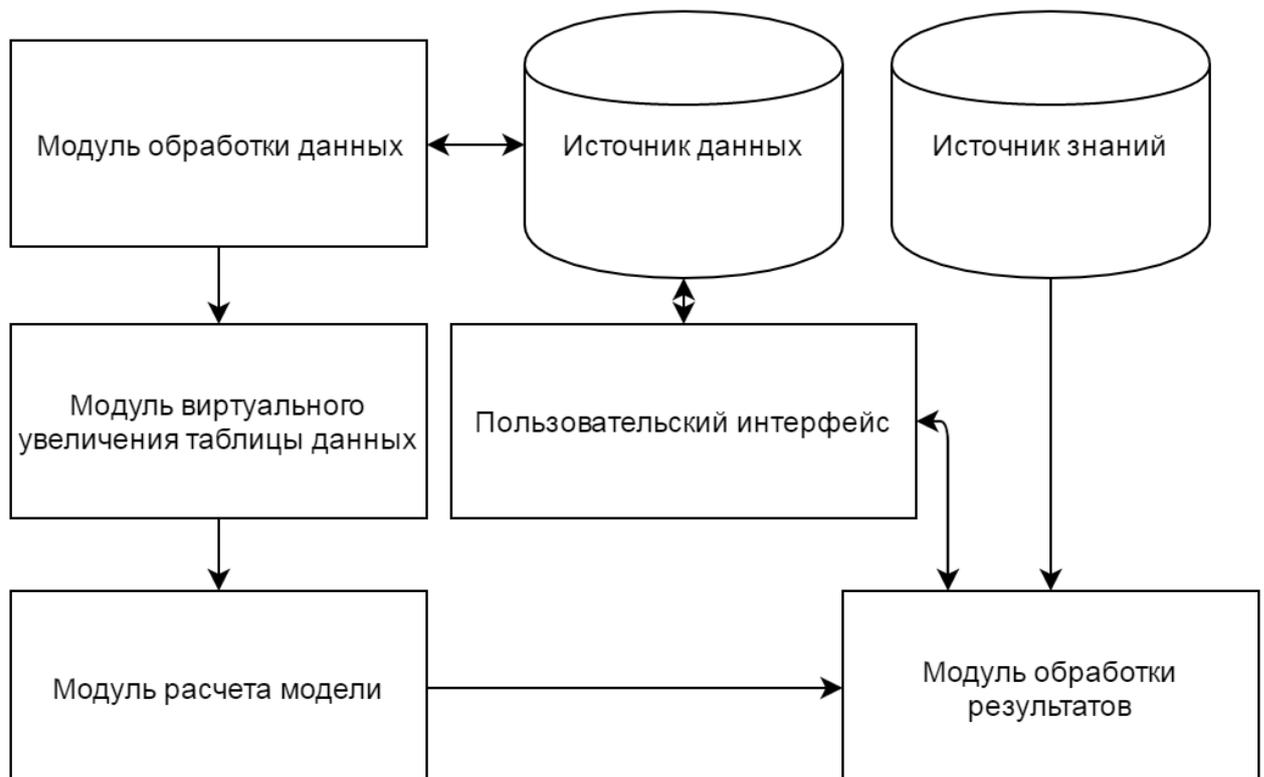
- 2.2.Цвет.
- 2.3.Поверхность.
- 2.4.Состояние мякиша.
- 2.5.Запах.
- 2.6.Вкус.

На этапе анализа имеющейся информации было принято решение строить отдельные математические модели для каждого из выходных показателей, которые будут образовывать систему уравнений, наиболее полно описывающую изучаемый производственный процесс.

Разрабатываемая система должна точно и достоверно описывать влияние изменений, как в технологическом процессе, так и в рецептуре на качество получаемого хлебобулочного изделия. Это обусловлено тем фактом, что как климатические условия, так и качество муки может варьироваться от партии к партии, что приводит к задаче принятия решения в условиях неопределённости.

На этапе ознакомления с техническим процессом было установлено, что время, в течение которого производится изделие, и температура – это величины, которая зависит как от веса изделия, так и от климатических условий в день выпечки.

Общая структура СППР представлена на рисунке 3.7:



### Рисунок 3.7 – Общая структура СППР

Рассмотрим процесс принятия решения о регулировании технологического процесса и рецептуры выпечки на конкретном примере по данным, полученным при изготовлении продукции «Хлеб целебный».

В таблицах 3.8-3.9 приведём исходные данные и результаты, полученные по четырём пробным замесам, каждый из которых был произведён через сутки после предыдущего.

Таблица 3.8 – Выходные параметры

№ замеса	Объем	Влажность	Кислотность	Пористость
1	145	47,8	6,4	52
2	146	48	6	51
3	145	47,9	6,1	51
4	150	46	6,2	53

Таблица 3.9 – Входные параметры

№ замеса	Время (X <sub>1</sub> )	Температура (X <sub>2</sub> )	Мука (X <sub>3</sub> )	Вода (X <sub>4</sub> )	Влажность (X <sub>5</sub> )	Кол-во клейковины (X <sub>6</sub> )	ИДК (X <sub>7</sub> )	Зольность (X <sub>8</sub> )
1	40	220	100	45	14	24	50	0,71
2	40	220	100	47	13,5	23	45	0,71
3	40	220	100	47	13,5	23	45	0,71
4	40	220	100	42	12	26	50	0,71

Таким образом, исходные данные являются сверхнасыщенным планом, в связи с чем к ним невозможно применить классические методы статистического анализа. Так, попытка построить линейное уравнение регрессии методом наименьших квадратов для выходного объема выпеченного хлеба неизбежно приводит к отбрасыванию важных факторов, таких как влажность муки или её клейковина. Так, при объеме данных всего в четыре строки, есть возможность построить модель с максимальным числом предикторов не более двух. Более того, подтвердить адекватность и качество получаемых моделей не представляется возможным. Для примера рассмотрим уравнения, полученные при рассмотрении влияния на выходной объем вариаций следующих пар величин:

- 1) «Количество воды» и «Влажность муки».
- 2) «Количество воды» и «Количество клейковины в муке».
- 3) «Влажность муки» и «Количество клейковины в муке».

Для первого случая полученное при помощи МНК [51, 156] уравнение регрессии будет иметь вид:  $Y=186,5455-0,2727X_4-2,0909X_5$ . Скорректированный множественный

коэффициент корреляции такого уравнения равен 0,9118, что является очень хорошим результатом. Однако вероятность незначимости данного уравнения по критерию Фишера равняется 0,1715, что превышает допустимое значение, в связи с чем полученное уравнение не может быть признано адекватным. Более того, если рассчитать значимость влияния каждого отдельного фактора, входящего в уравнение по критерию Стьюдента, то полученные результаты укажут, что даже при вероятности в 90% возможно утверждать о значительном влиянии на выходную величину только свободной переменной полученного уравнения.

Для второго случая, полученное уравнение равняется:  $Y = -401 + 6X_4 + 11,5X_6$ . При таком подходе, скорректированный коэффициент корреляции и расчетное значение критерия Фишера оказываются идентичными предыдущему примеру. Вероятность незначимости же каждого отдельного слагаемого колеблется в интервале от 0,218 до 0,306, что не может быть признано удовлетворительным.

При рассмотрении третьего варианта, в котором для рассмотрения выбираются факторы «Влажность муки» и «Количество клейковины в муке», построенная модель имеет вид:  $Y = 161 - 2X_5 + 0,5X_6$ . Значимость данной модели статистически неотличима от двух предшествующих.

Таким образом, очевидно, что классический подход к построению математических моделей не позволяет спрогнозировать качество изготавливаемых изделий, в связи с чрезвычайно малым количеством исходных данных.

При применении к исходным данным процедуры виртуального увеличения таблицы при помощи метода точечных распределений необходимо выбрать только те величины, которые имели хоть какие-то вариации во время проведенных замесов. Таким образом, в таблице останутся факторы:  $X_4$  – количество воды,  $X_5$  – влажность муки,  $X_6$  – количество клейковины в муке и  $X_7$  – ИДК. Проведем проверку на нормальность каждой из входных величин, а также известных значений объема выпеченного хлеба при помощи теста Шапиро-Уилка. Данная процедура показывает, что для уровня значимости в 99% мы не можем отбросить гипотезу о принадлежности генеральных совокупностей исследуемых величин к семейству нормальных распределений. При этом при 95% вероятности можно утверждать, о том, что величина ИДК распределена по закону, отличающемуся от нормального. Продemonстрируем результаты данной проверки на графике плотностей вероятностей для каждой из входных величин (рисунок 3.8).

Следующим шагом проведем процедуру виртуального увеличения объема исследуемых данных по алгоритму, изложенному в параграфе 2.3. Результатом данной

процедуры будет таблица данных, содержащая 72 строки. Повторим процедуру поиска оптимальной линейной регрессионной модели по рассчитанным данным методом наименьших квадратов.

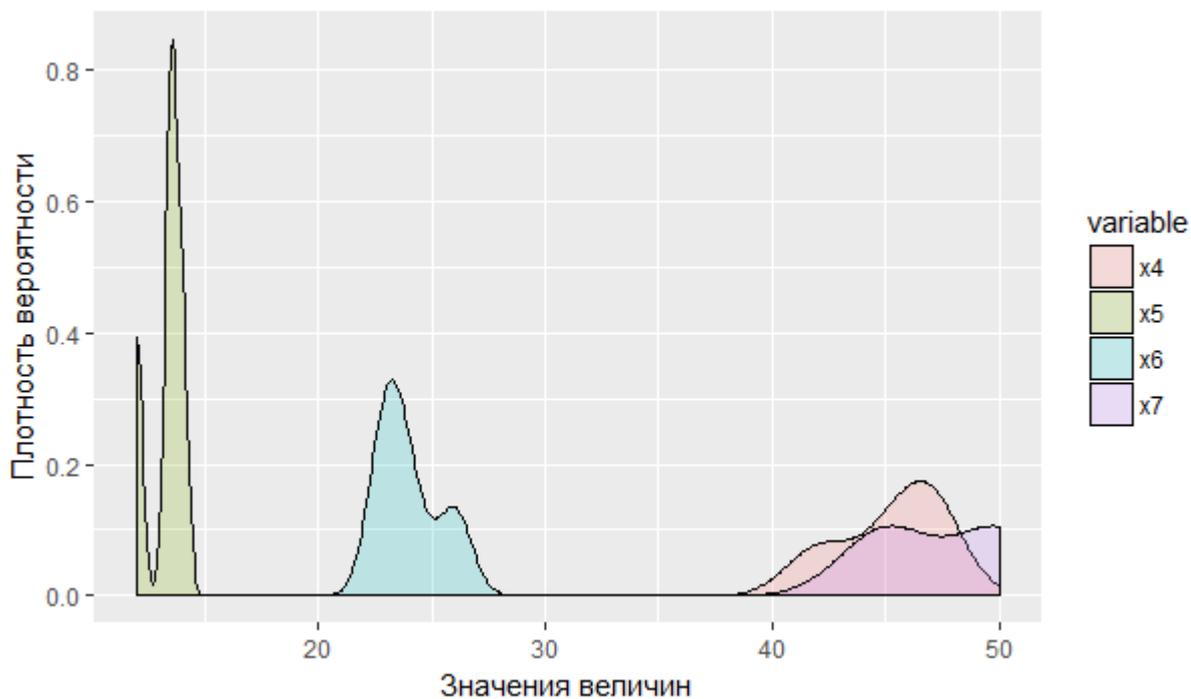


Рисунок 3.8 – Плотности вероятности входных параметров

Полученная модель будет иметь вид:  $Y=99,95176+0,67319X_4-1,87531X_5+1,70685X_6$ . Полученная модель является как адекватной в целом по критерию Фишера, так и каждый из входящих в неё слагаемых является значимым, с вероятностью ошибки близкой к нулю. Скорректированный коэффициент корреляции  $R^2=0.9889$ , что говорит о очень высоком объяснении вариаций выходной величины исследуемыми величинами. Проиллюстрируем полученные результаты на графике, изображенном на рисунке 3.9.

Интерпретируя результаты полученной модели можно отметить, что добавление воды в изделие будет увеличивать объем готовой продукции, что согласуется с реальным технологическим процессом. При этом плановый выход готовой продукции увеличивается при использовании более сухой муки, что отражено в отрицательном знаке перед вторым слагаемым, при этом уравнение предсказывает изменение объема на 1,875 единиц при вариации показателя влажности муки на 1%. Последним слагаемым полученной модели является количество клейковины в муке, которое также прямо пропорционально итоговому выходу изделия. На основании полученных результатов можно сделать вывод,

что полученное регрессионное уравнение адекватно описывает исследуемую предметную область.

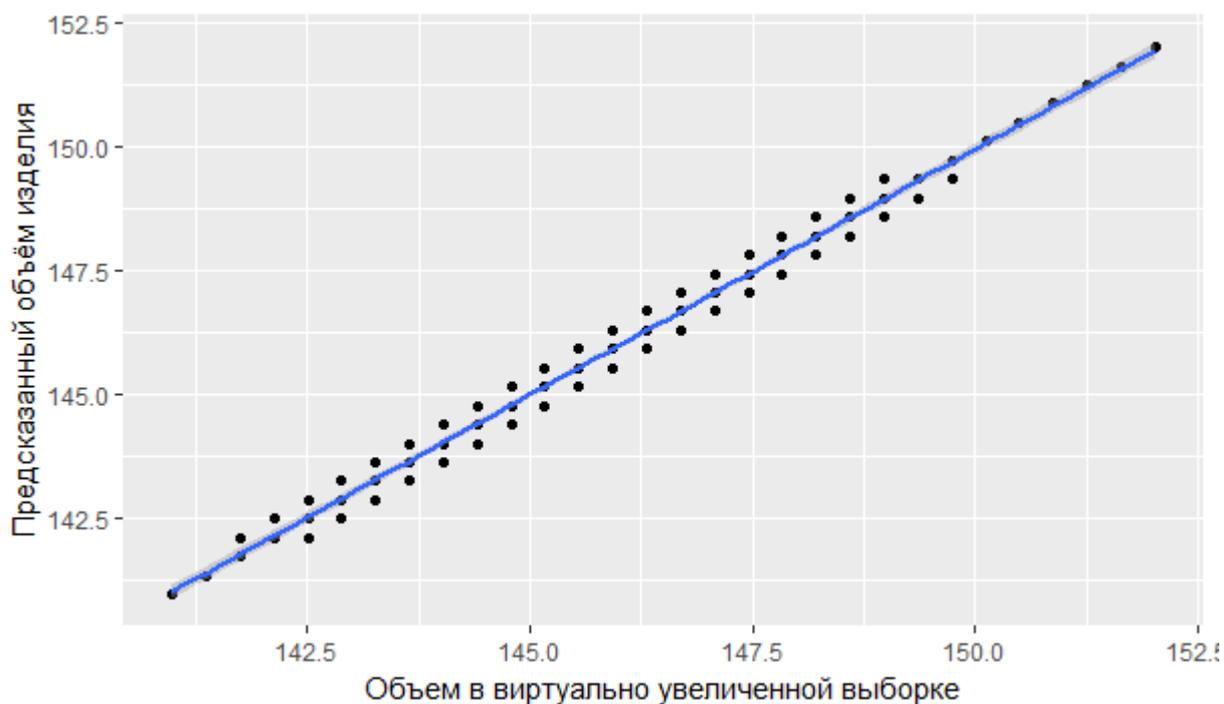


Рисунок 3.9 – Линейная аппроксимация результатов

Таким образом, показано, что применение виртуального увеличения объема данных, основанного на методе точечных распределений, может улучшить качество статистического анализа таблиц, полученных при изготовлении хлебобулочных изделий.

Аналогичным образом возможно построение моделей для каждого выходного количественного параметра.

В таблице 3.10 приведены модели, описывающие вариации влажности, кислотности и пористости для хлебобулочного изделия «Хлеб целебный», построенные по результатам четырёх пробных замесов.

Таблица 3.10 – Результаты моделирования

Параметр	Полученная модель
Влажность	$Y=30,75+0,2133X_4+0,5842X_5-0,0003096X_5X_6$
Кислотность	$Y=4,043233-0,067208X_4+0,263764X_5+0,070327X_6$
Пористость	$Y=38,02+0,2016X_4-0,6691X_5+0,005474X_5X_6$

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 продукционными. 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.

В таблице 3.11 рассмотрены некоторые продукционные правила, составленные для разрабатываемой СППР.

Таблица 3.11 – Примеры правил, включенных в базу знаний СППР

<b>Предпосылки</b>	<b>Заключение</b>
Зольность > 1,6 AND Цвет изделия ≤ 3	Для осветления мякиша рекомендуется добавить аскорбиновую кислоту
Кислотность > 4,5 <sup>0</sup> AND (Цвет изделия ≤ 3 OR Поверхность ≤ 3)	Рекомендуется уменьшить температуру брожения
Пористость > 40%	Рекомендуется увеличить температуру брожения

Таким образом, в разрабатываемой СППР предсказание всех количественных параметров опирается на математическую модель, построенную в соответствии с разработанными методами и алгоритмами. Однако, часть исследуемых параметров анализируются только органолептически и являются балльными, что не позволяет применить к ним вышеописанный математико-статистический аппарат. Для управления вариациями таких данных применяются продукционные правила, анализирующие как количественные, так и балльные показатели [129].

На рисунках 3.10-3.12 показан интерфейс СППР о качестве пробной выпечки хлебобулочных изделий. При открытии программы инженеру-технологу предлагается выбор между двумя действиями: выбрать уже существующее изделие, с известными характеристиками, либо создать новое.

При переходе к созданию нового изделия становится активной часть окна, в которой предлагается заполнить информацию о проведенной пробной выпечке изделия. В верхней части данной формы размещается две таблицы: информация о входных параметрах и информация о результатах анализа. После их заполнения необходимо нажать кнопку «Добавить в план эксперимента», после чего все введенные данные скопируются в строку результирующей таблицы (рисунок 3.10). Если в рецептуре используется пшеничная мука различных типов, то в итоговую таблицу помещаются средние арифметические характеристики муки, с учётом процентного соотношения объёмов каждого вида в смеси.

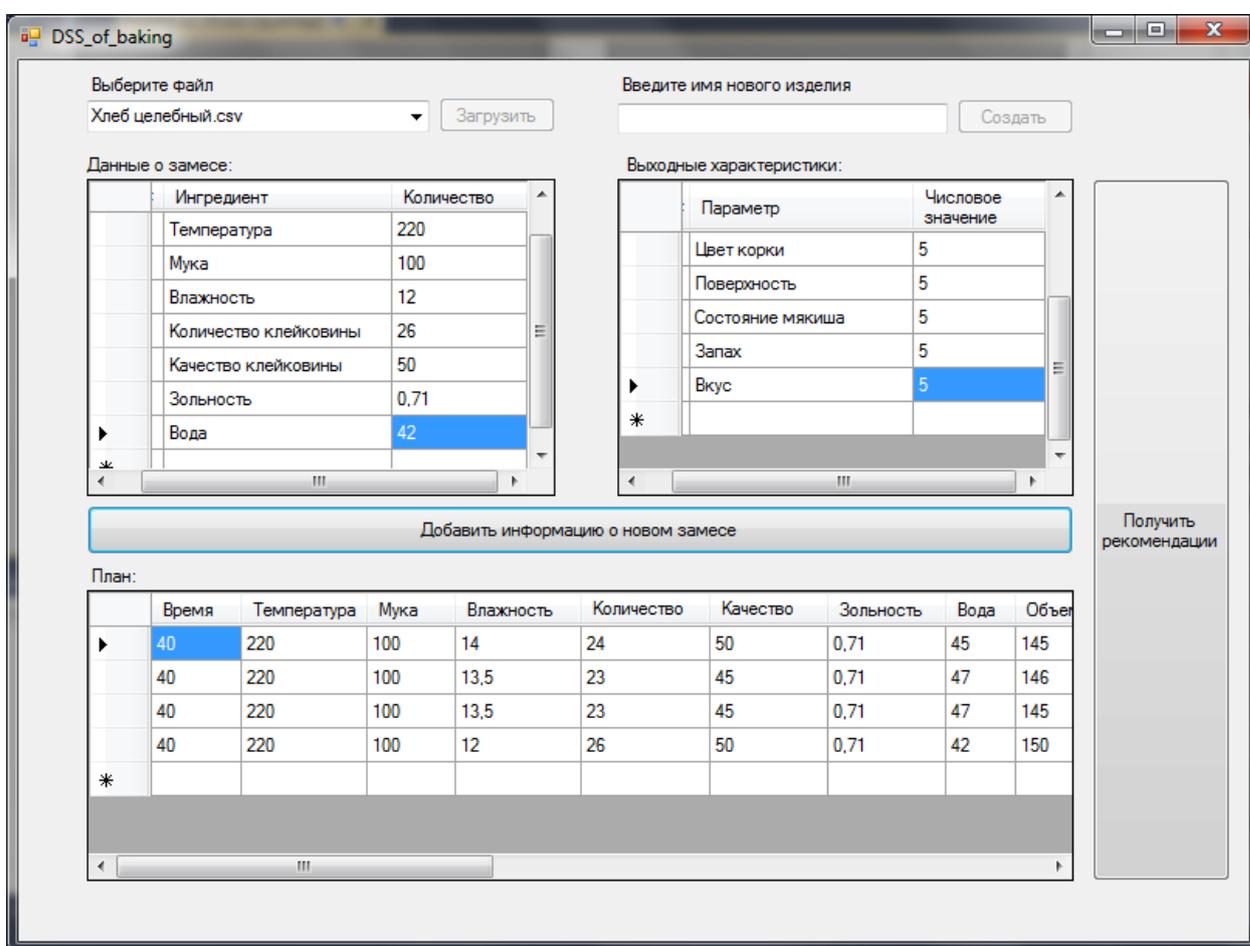


Рисунок 3.10 – Главное окно СППР о качестве пробной выпечки

Таким образом, технолог может добавлять информацию об очередном замесе каждого из продуктов, которая будет сохраняться в .csv файлах. При выборе уже существующего изделия данные из этих файлов выгружаются в таблицу на форму, после чего их можно редактировать, либо добавлять информацию о новых замесах.

Если число строк итоговой получившейся таблицы более двух, то к данным автоматически применяется критерий Диксона для проверки на наличие грубых промахов. При получении результата, сигнализирующего о возможной ошибке, под таблицей появляется предупреждающая надпись (рисунок 3.11), указывающая на переменную с подозрительным значением с рекомендацией для ЛПР обратить внимание и перепроверить данные.

При нажатии на кнопку «Получить рекомендации» вызываются скрипты, написанные на языках *C#* и *R*, которые анализируют полученную информацию при помощи математико-статистических методов. Результаты статистического анализа показываются в окне с рекомендациями (рисунок 3.12). В этом же окне располагаются инструменты интерактивного прогнозирования качества выбранного хлебобулочного изделия. Так, изменяя входные параметры, инженер-технолог может в режиме реального времени посмотреть какие будут последствия для основных прогнозируемых показателей.

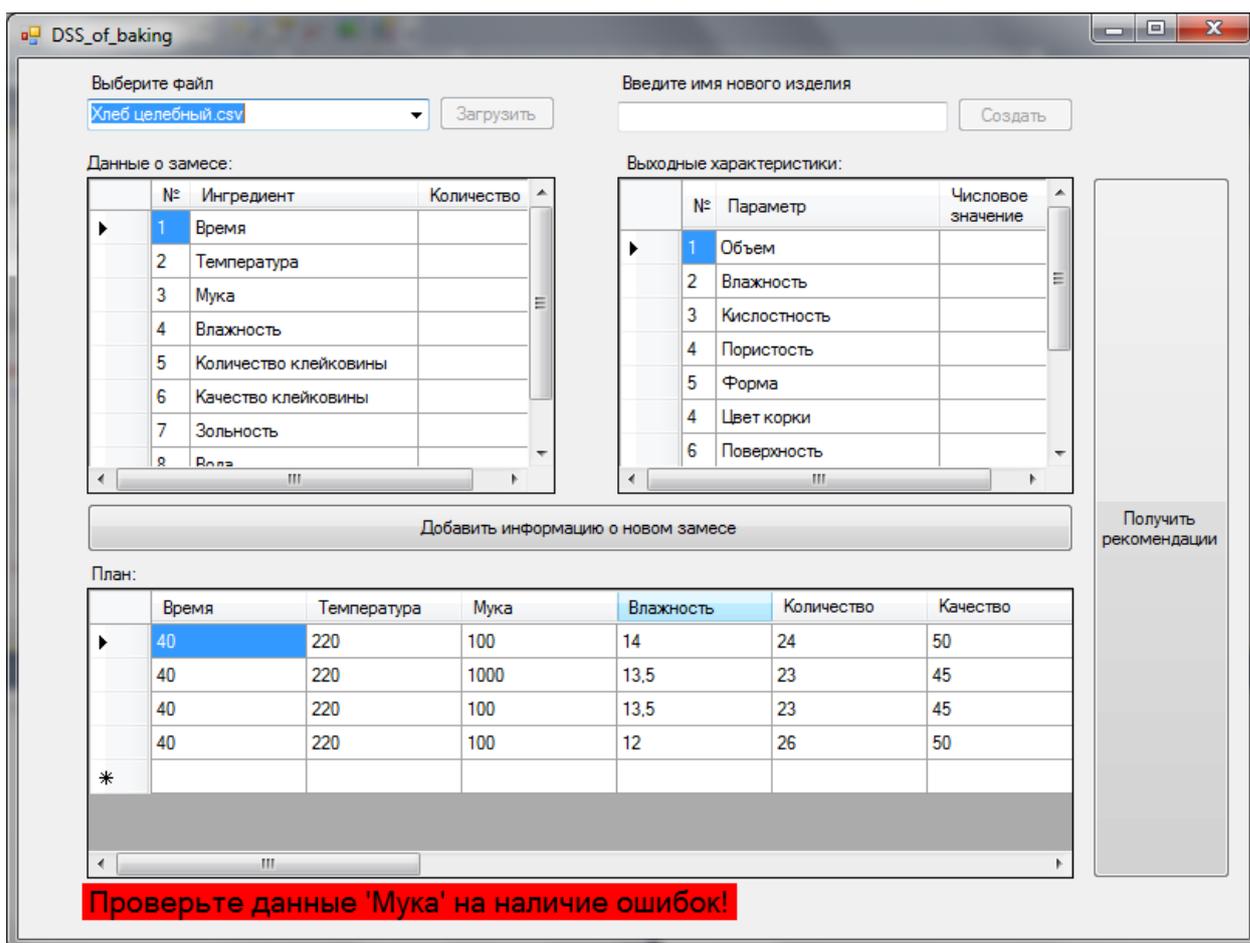


Рисунок 3.11 – Предупреждение для ЛПР о возможной ошибке в данных

Для построения математических моделей используется линейный регрессионный анализ, в котором учитывается не только влияние каждого отдельного изменяемого фактора, но и все их взаимодействия. Если какой-либо из входных параметров не изменялся в течение всего производственного цикла, то он не включается в рассмотрение, так как его вариации не могут быть проанализированы и оценены. Анализ на наличие грубых промахов проводится по всем входным и выходным параметрам, если существует подозрение на обнаружение такого значения, программа выводит предупреждение для ЛПП, с рекомендацией проверить введенные данные, при этом данное сообщение не влияет на остальную процедуру анализа, которая проводится в соответствии с алгоритмом.

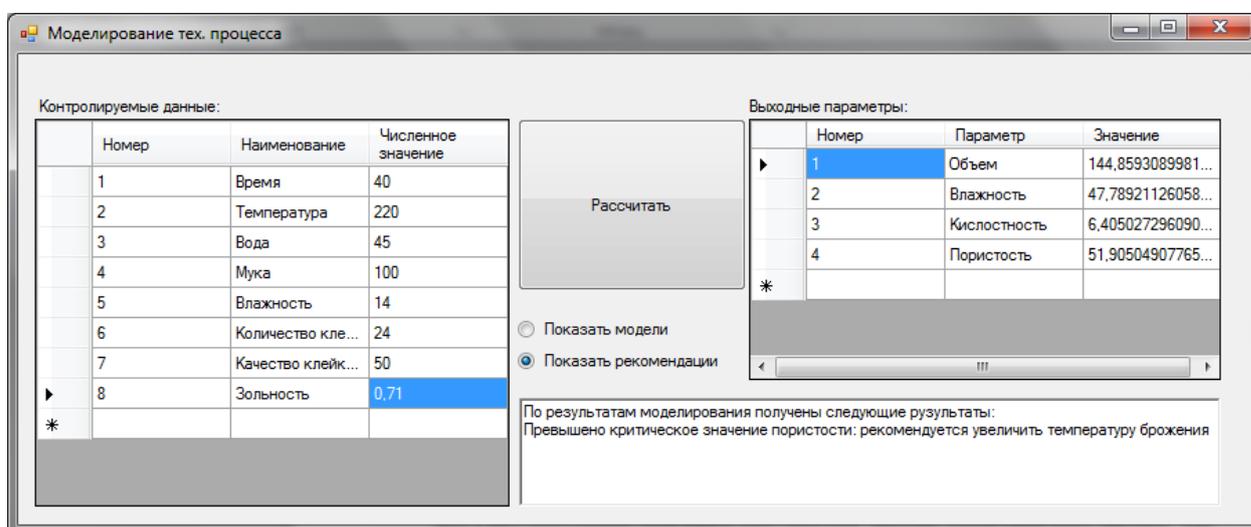


Рисунок 3.12 – Результаты моделирования и рекомендации для ЛПП

Резюмируя вышеописанное, следует отметить, что применение разработанной методики вероятностно-статистической обработки пассивных выборок малого объема позволило построить компьютерную проблемно-ориентированную СППР для помощи инженеру-технологу в оперативном регулировании рецептуры при пробной выпечке хлебобулочных изделий.

### 3.5 Выводы к главе 3

1. Применение разработанной методики к выборкам малого объема, полученным в различных условиях, во многих случаях позволяет сравнительно улучшить качество

получаемых математических моделей [73, 74], сохранив первоначальный закон распределения случайных величин.

2. 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.

3. Применение виртуального увеличения количества 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.

4. Для получения статистически значимых математических моделей необходимо применять различные методики построения уравнений регрессии.

5. Разработанное программное обеспечение помогает ускорить обработку пассивных данных малого объема по разработанной методике, выполняя самые трудоемкие участки вычислений и преобразований данных.

6. 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 [64], выявлены методы обладающие наибольшей мощностью [68], даны рекомендации для выявления грубых промахов в выборках малого объема [70].

2. Разработана методика виртуального увеличения таблицы многомерных пассивных данных на основе метода точечных распределений [128], используя информацию о виде закона распределения каждой отдельной одномерной случайной величины [85] 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 [130].

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  $r=0,75$  [66].

4. Показано, что применение виртуального увеличения таблицы пассивных данных позволяет более точно определять величину линейной корреляционной связи [65].

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 [37, 67].

6. Разработано программное обеспечение, которое позволяет значительно упростить и ускорить процесс обработки многомерных выборок малого объема на основе метода точечных распределений [71, 72].

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 невозможно 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 [69, 73, 74, 129].

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 выработать 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.

Актуальная научная проблема, решенная в исследовании, состоит в определении 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, Титьена-Мура или Диксона [68, 70]. 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 [128], 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 [65]. Следует отметить, что не было обнаружено методов, позволяющих проводить корреляционный анализ данных малого

объема с достаточной точностью и достоверностью [66], 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 не возможен или экономически не эффективен.

## СПИСОК ЛИТЕРАТУРЫ

1. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. 2-е изд., перераб. и доп. Москва: Наука, 1976, 279 с.
2. Алексеев С. И. Исследование систем управления: Учебно-методический комплекс. Москва: Изд. центр ЕАОИ, 2008, 195 с.
3. Бараз В.Р. Использование MS Excel для анализа статистических данных: учебное пособие / В.Р. Бараз, В.Ф. Пегашкин. – 2-е изд., перераб. и доп. – Нижний Тагил : НТИ (филиал) УрФУ, 2014. –181с.
4. Большев Л.Н. К вопросу о различии по малым выборкам нормального и равномерного типов распределений. В: Теория вероятностей и её применения, 1965, т.10 №4, с.764-765.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, Главная редакция физико-математической литературы, 1983, 416с.
6. Валуев С.А., Волкова С.Н. и др. Системный анализ в экономике и организации производства. Л: Политехника, 1991, 398с.
7. Васюткина И.А. Программно-математические средства статистического моделирования на основе данных пассивного эксперимента. Диссертация на соискание ученой степени к.т.н. М.: МГАПИ, 2004, 251 с.
8. Гаскаров Д. В., Шаповалов В. И. Малая выборка. Москва: Статистика, 1978, 248 с.
9. Гаскаров Д.В. Интеллектуальные информационные системы. Москва: Высшая школа, 2003, 431 с.
10. Гаскаров Д.В., Фомин В.В., Вихров Н.М. Математические основы теории информационных систем: учебное пособие. Санкт-Петербург: СПбГУВК, 1996, 81 с.
11. Гайдукова Н.В. исследование и применение методов системного анализа при формировании портфеля проектов. В: Вісник ЛДУ БЖД, 2013, № 8, с. 60-65.
12. Гвоздев Статистический анализ надежности: Методические указания к лабораторным работам по курсам «Надежность АСОИиУ», «Надежность и эффективность в технике», «Теория надежности систем». Уфа: Уфимск. гос. авиац. техн. ун-т, 2001, 45 с.
13. Горбунова Е.Б. Метод непараметрической оценки закона распределения случайного параметра по малому числу наблюдений В: Инженерный вестник Дона, 2014, №3. <http://ivdon.ru/ru/magazine/archive/n3y2014/2516>

14. Горяинов В.Б. Математическая статистика: Учеб. Для вузов / В.Б.Горяинов, И.В. Павлов, Г.М. Цветкова и др. Под ред. В.С.Зарубина, А.П. Крищенко. – 3-е изд., исправл, М.: Изд-во МГТУ им. Н.Э.Баумана, 2008. 424 с.
15. 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.
16. Громыко Г.Л. Теория статистики: Учебник, 2-е изд., перераб. и доп. Москва: ИНФРА-М, 2011, 476 с.
17. Демаков И.П. Потепун В.Е. Статистические методы определения вида закона распределения при анализе точности и надежности промышленных изделий по результатам эксперимента. Л: ЛДНТП, 1970, 39 с.
18. Долгов А. Ю. Повышение эффективности статистических методов контроля и управления технологическими процессами изготовления микросхем. Диссертация на соискание ученой степени к.т.н, Москва, 2000, 218 с.
19. Долгов Ю.А. Статистическое моделирование: Учебник для вузов. – 2-е изд., доп, Тирасполь: Полиграфист, 2011, 352 с.
20. Долгов Ю.А. Методы выборочного контроля и математического моделирования для управления групповыми технологическими процессами. Диссертация на соискание ученой степени док. техн. наук, Кишинев, 1990, 331 с.
21. Долгов Ю.А., Долгов А.Ю., Столяренко Ю.А. Метод повышения точности вычисления параметров выборки малого объема. В: Вестник ПГУ, 2010, с. 232-242.
22. Долгов Ю. А., Долгов А. Ю., Ваняшкин М. М., Деткова А. В. Определение границ и ширины ядра эквивалентных выборок в методе точечных распределений. В: Радіоелектронні і комп'ютерні системи, 2010, № 6, с. 161–165.
23. Долгов Ю.А. Влияние парных грубых промахов на коэффициент корреляции 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. 152-153.
24. Долгов Ю.А. Модифицированный метод случайного баланса В: Электронное моделирование, 1987, № 4, с. 79-84.
25. Долгов Ю.А., Нарбутовичюс В.И. Нахождение законов распределения антропометрических параметров В: Док. итоговой АК ППС ИТФ за 2013 г. – Тирасполь, 15-17 янв. 2014 г./ Приднестр. Гос. Ун-т, Тирасполь: «ТесЛайн», 2014, с. 133-134.

26. Долгов Ю.А., Окушко Р.В., Нарбутавичюс В.И. Математическое моделирование гендерных компонент соматотипа современного человека. Этап 1: период юношества, женщины В: Scientific Research in the 21<sup>st</sup> Century. Moscow, Russia, 2015, с. 103-112.
27. Дружинин Г.В., Воронова О.В. Сравнение методов построения эмпирической функции распределения по малому числу наблюдений. В: Надежность и контроль качества, 1983, № 1.
28. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник./ Под ред. И.И. Елисеевой. – 5-е изд., перераб. и доп, М.: Финансы и статистика, 2004, 656с.
29. Епанечников В.А. Непараметрическая оценка многомерной плотности вероятности В: Теория вероятностей и ее применения, 1969, Т.14, № 1, с. 156-161.
30. Еременко В.С., Куц Ю.В., Мокийчук В.М. Оценка однородности выборок малого объема, В: Системи обробки інформації, 2006, №7(56), с.26-29.
31. Еременко И.В., Свердлик А.Н. Об одном методе построения законов распределения величин при малом числе испытаний. В: Некоторые вопросы специального применения вычислительной техники, Л: ЛВИКА им. А.Ф. Можайского, 1963, с.18-29.
32. Ермаков С.М. О датчиках случайных чисел. В: Заводская лаборатория. Диагностика материалов, 1993, т.59, №7, с.48-50.
33. Звягинцев В.В. Прогнозирование сложности лапароскопической холецистэктомии. Диссертация на соискание ученой степени к.м.н. Нижний Новгород: Нижегородская государственная медицинская академия министерства здравоохранения российской федерации, 2014, 150с.
34. Звягинцев В. В., Горпинюк В. П., Ставинский Р. А., Фомов Г. В., Чалык Ю. В. К вопросу о характеристиках лапароскопической холецистэктомии В: Медицинский альманах, 2012, №1, с.86-89.
35. Кендал М., Стюарт А. Теория распределений. Москва: Наука, 1966, 588 с.
36. Кобзарь А.И.. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2012, 816 с.
37. Кожокару С.К., Попускайло В.С. Метод принятия решений в условиях малой выборки. В: Научный обозреватель, 2016, №3(63), с.79-80.
38. Козыр Е.В. Этапы обработки данных при малых выборках. В: Механіка гіроскопічних систем, 2010, №21, с.103-111.
39. Кокунин, В.А. Статистическая обработка данных при малом числе опытов. В: Укр. биохим. Журнал, 1975, №47(6), с. 776-790.

40. Кравченко Т. К., Исаев Д. В. Экономические информационные системы. В: Информатика, Т. 1., 2012, № 3, с. 278
41. Кривенко М.П. Прикладные методы оценивания распределения многомерных данных малой выборки. М.: ИПИ РАН, 2011, 146 с.
42. Кузнецова Л.Д. Разработка проблемно-ориентированных компонентов системы управления образовательными процессами высшего профессионального образования в современных условиях. Диссертация на соискание степени канд. техн. наук. Воронеж: Воронежский государственный педагогический университет, 2009, с. 167.
43. Лавров В.В., Спиринов Н.А. Методы планирования и обработки результатов инженерного эксперимента. Екатеринбург: ГОУ ВПО УГТУ-УПИ, 2004, 257 с.
44. Лапшина И.В., Усенко О.А., Алексеева А.В. Повышение эффективности работы пожарных спасательных отрядов средствами статистической обработки оперативных данных малого объема. В: Актуальные проблемы гуманитарных и естественных наук, 2015, №2-1, с. 24-29.
45. Легалов А.И., Ледаев Д.Н., Анкудинов А.В. Поддержка многокритериального анализа в сложных иерархических системах. В: Вестник НГУ. Серия: Информационные технологии, 2010, т.8, №1, с. 35-44.
46. Лейфер Л. А., Голубев М. П. Оценка имущества на основе сравнения продаж. Проблема малой выборки и пути её решения. В: Имущественные отношения в РФ, 2005, №2, с. 47 – 52.
47. Лемешко, Б. Ю. Робастные методы оценивания и отбраковка аномальных измерений. В: Заводская лаборатория, 1997, Т. 63, №5, с. 43-49.
48. Лумельский Я.П. Статистические оценки результатов контроля качества, М.: Изд-во стандартов, 1979, 199 с.
49. Львовский Е.Н. Статистические методы построения эмпирических формул: учеб. пособие для вузов. М. : Высшая школа, 1988, 239 с.
50. Маликов Я.В. Коагуляция, потенцированная аргоном, при лапароскопической холецистэктомии В: Бюллетень сибирской медицины, 2011, № 1, с. 81–87.
51. Магнус Я.Р., Катышев П.К., Пересецкий А.А Эконометрика. Начальный курс: Учеб. – 6-изд., перераб. и доп. - М.: Дело, 2004, 576 с.
52. Митропольский, А.К. Техника статистических вычислений. М. : Наука, 1971, 576 с.

53. Мотричкин К.В., Степанов А.В. Прогнозирование развития экономики региона в условиях неполноты информации. В: Актуальні проблеми економіки, 2012, №11(137), с. 412 – 417.
54. Мышкис А.Д. Элементы теории математических моделей. Изд. 3-е, исправленное. М.: КомКнига, 2007, 192с.
55. Носовский А. М., Пихлак А. Э., Логачев В. А., Чурсинова И. И., Мутьева Н. А. Статистика малых выборок в медицинских исследованиях. В: Российский медицинский журнал, 2013, №6, с. 57-60.
56. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. Москва: КНОРУС, 2010, 192с.
57. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло. В: Научный журнал КубГАУ, 2015, №114(10), с.27-41.
58. Орлов А.И. Прикладная статистика. Учебник для вузов. Москва: Экзамен, 2006, 672с.
59. Орлов А.И. Теория принятия решений. Учебник для вузов. Москва: Экзамен, 2006, 576с.
60. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов. Ростов-на-Дону: Феникс, 2009, 572с.
61. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? В: Заводская лаборатория. Диагностика материалов, 1991, т.57, №7, с.64-66.
62. Пешес Л.Я. Основы теории ускоренных испытаний на надежность /Л.Я. Пешес, М.Д. Степанова, Минск.: Наука и техника, 1972, 165 с.
63. Покидько С.В., Кочура С.Г. Корректировка результатов ресурсных испытаний электромеханических устройств при малых выборках В: Известия ВУЗОВ. Приборостроение, 2011., т. 54, № 4, с. 38-42.
64. Попукайло В.С. Исследование критериев грубых ошибок применительно к выборкам малого объема В: Радіоелектронні і комп'ютерні системи, 2015, №3(73), с. 39-44.
65. 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
66. 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

67. 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.

68. 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

69. 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

70. 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.

71. 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>

72. Попукайло В.С. Построение таблицы исходных данных для многомерного моделирования по выборкам малого объема. В: Тезисы IX Международной конференции. Тирасполь, 8-10 октября 2015г. Тирасполь: Изд-во Приднестровского университета, 2015, 192с – с.176-177.

73. 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.

74. 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.

75. Поршнев А.Г., Румянцева З.П., Соломатина Н.А. Управление организацией: Учебник. Москва: ИНФРА-М, 1998, 669с.

76. Путивцева Н.В. и др. Система поддержки принятия решений при выполнении проектов. В: Научные вести Белгородского государственного университета. Серия: Экономика. Информатика, 2015, №7(204), с.170-174.

77. Рылёв С.С. Автоматизация контроля цвета хлебобулочных изделий при управлении процессами выпечки. Автореферат диссертации на соискание ученой степени к.т.н. Воронеж, ФГОУ ВПО «Воронежский государственный университет инженерных технологий», 2012, 20 с.

78. Самойленко А.П., Горбунова Е.Б. Полиномиальная интерполяция при синтезе моделей технологических объектов по выборкам данных критически ограниченного объема. В: Известия Южного федерального университета. Технические науки, 2013, №11(148) с. 24-31.

79. Самойленко А.П., Усенко О.А. Компьютерные технологии обработки статистических массивов эмпирических данных критически малых объемов. В: Известия ТРТУ – ДонНГУ. Материалы Пятого Международного научно-практического семинара «Практика и перспективы развития партнерства в сфере высшей школы», Таганрог: Изд-во ТРТУ. Кн. 1, 2006, №6, с. 159-163.

80. Самойленко А. П., Усенко О. А. Статистический анализатор для диагностики и прогноза экосистемы в реальном масштабе времени. В: Известия ЮФУ. Технические науки, 2001, №2, с. 36-39.

81. Сизова Т.М. Статистика: Учебное пособие, СПб.: СПб ГУИТМО, 2005. 80с.

82. Снежко В. Л. Статистическая обработка результатов гидравлического эксперимента / В. Л. Снежко, С. И. Хусни. В: Молодой ученый, 2011, №8. Т.1, с. 83-86.

83. 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.

84. Столяренко Ю. А. Метод точечных распределений. В: Радіоелектронні і комп'ютерні системи, 2012, № 6, с. 75-77.

85. 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.

86. Сухорученков Б.И. Анализ малой выборки. Прикладные статистические методы. М.: Вузовская книга, 2010, 384с.

87. Трахтенгерц Э.А. Компьютерная поддержка принятия решений. М.: Синтег, 1998, 376с.
88. Хамханова Д.Н. Метрологическое обеспечение экспертных измерений для контроля качества продукции пищевой промышленности (на примере хлебобулочных изделий). Автореферат диссертации на соискание ученой степени к.т.н. Томск, ФГБОУ ВПО «ВСГУТУ », 2012, 41 с.
89. Хоконова Т. М., Хоконова М. З. Роль непараметрических методов в медико-биологических исследованиях В: APRIORI. Серия: Естественные и технические науки . 2014, №4, с.12.
90. Чавчанидзе В.В., Кумсишвили В.А. Об определении законов распределения на основе малого числа наблюдений. В: Применение вычислительной техники для автоматизации производства (Труды совещания 1959г.). М., Машгиз, 1961, с. 71-75.
91. Чебышев П.Л. Полное собрание сочинений, т.5. М.: Изд-во АН СССР, 1951, 474с.
92. Шор Я.Б. Статистические методы анализа и контроля качества и надежности, М.: Советское радио, 1962, 552 с.
93. Arason, S., Ásgeirsson, E.I., Margeirsson, S., Olsen, P., Stefánsson, H., Decision Support Systems for the Food Industry. In: Book Chapter. Intelligent Systems Reference Library. 2010. Volume 4. II. P. 295-315. doi: 10.1007/978-3-642-13639-9\_12
94. Association Internationale de la Boulangerie Industrielle. AIBI Bread Market Report 2013, 2015. <http://www.aibi.eu/wp-content/uploads/draft-AIBI-Bread-Market-report-2013.pdf>
95. Gaidric C. Luarea decizilor. Metode si tehnologii. Chisinau: Stiinta, 1998, 162p.
96. Bablekos G.D., Michaelides S.A., Konstantinos A.A. Effects of laparoscopic cholecystectomy on lung function: A systematic review In: World J. Gastroenterol, 2014, December, 20(46), p. 603–617.
97. Barnett V., Lewis T. Outliers in statistical data. Wiley, 1994, 584 p.
98. Budaev S. Multivariate methods and small sample size: combining with small effect size, 2011 [Preprint]. <http://cogprints.org/8672/1/dochtermann-jenkins-response.pdf>
99. Castaneda et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. In:Journal of Clinical Bioinformatics, 2015, №5:4, doi: 10.1186/s13336-015-0019-3.
100. Catani M., De Milito R., Romagnoli F. The best timing of surgery in laparoscopic cholecystectomy for acute cholecystitis: when and how is to be performed In: Hepatogastroenterology, 2008, Nov-Dec, 55(88), p. 1993–1996.

- 101.Chalal M., Boucher X., Marques G. Decision support system for servitization of industrial SMEs: a modelling and simulation approach In: *Journal of Decision Systems*, 2015, vol. 24, № 4, p. 355-382. doi: 10.1080/12460125.2015.1074836
- 102.Cojocaru S., Gaidric C. Clinical Decision Support System Sonares. In: *International Journal Information Theories & applications*, vol.15, 2008, №2, p. 137-143.
- 103.Dochtermann N.A., Jenkins S.H. Multivariate methods and small sample sizes In: *Ethology*, 2011, №117(2), p.95-101.
- 104.Dolgov Y. A., Dolgov A. Y. Stochastic Check for Control of Electronic Wares Quality In: *Trans. Of 10-th Intern. Symp. Of Applied Stochastic Models and Data Analysis. – 12-15 june 2001: Vol. 1/2. Univer. Techologic de Compiene. France*, p. 387-390.
- 105.Douglas C. Montgomery. *Design and Analysis of Experiments*. Wiley: Arizona State Univerisity, 2013, 757p.
- 106.Dukić D., Mesarić J. Flour Inventory Decision Support System in Baking Industry In: *Proceedings of the 5th International Congress" Flour-Bread'09"*, 2009, p. 92-98.
- 107.Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, Pa.: SIAM, 1982, 135 p.
- 108.Everitt B. S., Skrondal A. *The Cambridge Dictionary of Statistics*, 4 edition. Cambridge University Press, 2010, 480p.
- 109.Fisher R.A. *Statistical methods and scientific inference*. NY: Macmillan Pub Co, 1973, 190 p.
- 110.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.
- 111.Gaidric C. Decision support systems in ultrasound diagnostics. In: *Computer Science Journal of Moldova*, vol.17, 2009, №3(51), p. 255-277.
- 112.Genc V., Sulaimanov M., Cipe G. What necessitates the conversion to open cholecystectomy? A retrospective analysis of 5164 consecutive laparoscopic operations In: *Clinics*, 2011, 66(3), p. 417–420.
- 113.González-Rodríguez F.J., Paredes-Cotoré J.P., Pontón C. Early or delayed laparoscopic cholecystectomy in acute cholecystitis? Conclusions of a controlled trial In: *Hepatogastroenterology*, 2009, Jan-Feb, 56(89), p. 11–16.
- 114.Goodwin L.D., Leech N.L. Understanding Correlation: Factors That Affect the Size of r In: *The Journal of Experimental Education*, 2006, 74(3), p. 251–266.

- 115.Gurov S. 0-event: Point and Interval Estimates. In: Mathematics and Statistics, 2013, №1(2): p. 53-58.
- 116.Harbuzz Y. Aviation Material and Construction Fatigue Damage Assessment on the Basis of Acoustic Emission Signals. Summary of Doctoral Thesis. R.: RTU Press, 2015. 33 p.
- 117.Hays W. Statistics. Boston: Cengage Learning, 1994, 1136 p.
- 118.Hurlbert S.H. On misinterpretations of pseudoreplication and related issues: a reply to Oksanen. In: Oikos, 2004, №104, p.591-597.
- 119.Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint et al. Power failure: why small sample size undermines the reliability of neuroscience. In: Nature Reviews Neuroscience, May 2013, №14(5), p.365-376.
- 120.Lantmännen Unibake. 2013 Bread Barometer Survey: The Future of Bakery is Freshly Baked, 2014. <http://www.lantmannen-unibake.com/en-GB/News/Bread-Barometer-Survey-Results/>
- 121.Lazic S.E. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? In: BMC Neuroscience, 2010, №11:5, p.17.
- 122.Lwanga S.K., Lemeshov S. Sample size determination in health studies. A Practical manual. Geneva: World Health Organization, 1991, 30 p.
- 123.Maksimova E. The Main Trends of the Bakery Market. В: Хлебопечение / кондитерская сфера, 2016, №1(63), p. 16-21.
- 124.Marušák R., Kašpar J., Vopěnka P. Decision Support Systems (DSS) Optimal – A Case Study from the Czech Republic. In: Forests, 2015, №6, p. 163-182.
- 125.McBane, George C. Programs to Compute Distribution Functions and Critical Values for Extreme Value Ratios for Outlier Detection. In: Statistical Software, 2006, №16(3), p. 1-9.
- 126.Mehtätalo L., Comas C., Pukkala T., Palahi M. Combining a predicted diameter distribution with an estimate based on a small sample of diameters. In: Canadian Journal of Forest Research, 2011, №41, p. 750-762.
- 127.Moinester M., Gottfried R. Sample size estimation for correlation with pre-specified confidence interval. In: The Quantitative Methods for Psychology, 2014, №10(2), p. 124-130.
- 128.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.

129. 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.
130. Popukaylo V. Small size sample mathematical modeling In: Meridian Ingineresc 2015, №4(595), c.25-30.
131. Power D. J., Sharda R. Model-driven decision support systems: concepts and research directions. In: Decision Support Systems, 2007, №43, p.1044-1061 doi:10.1016/j.dss.2005.05.030
132. Rajalakshmi K., Chandra Mohan S., Dhinesh Babu S. Decision Support System in Healthcare Industry. In: International Journal of Computer Applications, 2011, Volume 26, №9, p.42-44.
133. Rhiel G.S., Chaffin W.W. An Investigation of the Large-Sample/Small-Sample Approach to the One-Sample Test for a Mean (Sigma Unknown) In: Journal of Statistics Education, 1996, vol.4, №3, <http://www.amstat.org/publications/jse/v4n3/rhiel.html>
134. Rooh-ul-Muqimet, Jan Q., Zarin M. Complications of Laparoscopic Cholecystectomy In: World Journal of Laparoscopic Surgery, 2008, January-April, №1(1), p.1-5.
135. Satterthwaite F.E. Random balance experimentation. In: Technometrics, 1959, №1(2), p. 111-137.
136. Secrieru I. Structured knowledge management techniques for the development of interactive and adaptive decision support system. In: Computer Science Journal of Moldova, v.17, 2009, №1(49), p. 58-73.
137. Seltman H.J. Experimental Design and Analysis. Pittsburgh, PA: Carnegie Mellon University, 2012, 428p.
138. Seo S. A review and comparison of methods for detecting outliers in univariate data sets. University of Pittsburgh, 2006. <http://d-scholarship.pitt.edu/7948/1/Seo.pdf>
139. Shestakova T. Mathematical Modeling in Integrated Circuit Technologies In: Intern. Conf. on Automation, Quality and Testing, Robotics A&QT-R 2002 (THETA). Cluj-Napoca, Romania, 2002, T. I, p. 274.
140. Silverman B.W. Density estimation for statistics and data analysis. Chapman and Hall: 1986, 175 p.
141. Spiegelhalter D.J., Knill-Jones R.P. Statistical and Knowledge-Based Approaches to Clinical Decision-Support Systems, with an Application in Gastroenterology. In: Journal of the Royal Statistical Society. Series A (General), 1984, Vol. 147, №1, p. 35-77.

142. Spiegelhalter D.J., Best N.G., Carlin B.P., Van der Linde, A. The deviance information criterion: 12 years on. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 2014, №76, p. 485-493.
143. Student. The probable error of a mean. In: *Biometrika*, 1908, № 6(1), p. 1-25
144. Swartz et al. Small sample properties of rare variant analysis methods. In: *BMC Proceedings*, 2014, №8(Suppl 1): S13.
145. Taborsky M. Sample size in the study of behaviour. In: *Ethology*, 2010, №116, p. 185–202.
146. Teddlie C., Fen Yu Mixed Methods Sampling: A Typology With Examples. In: *Journal of Mixed Methods Research*, 2007, №1, p.77-100.
147. Thode H.C. *Testing for Normality*. Florida: CRC Press, 2002, 368p.
148. Vacik H., José G. Borges, Garcia-Gonzalo J., Ljusk-Ola Eriksson. Decision Support for the Provision of Ecosystem Services under Climate Change: An Editorial. In: *Forests*, 2015, №6, p. 3212-3217.
149. Wackerly D., Mendenhall W., Scheaffer R. *Mathematical Statistics with Applications*, Seventh Edition. Thomson Brooks/Cole, 2008, p. 944.
150. Wilcox R.R. *Introduction to robust estimation and hypothesis testing*. 3<sup>rd</sup> edition. NY: Academic Press, 2011, 608p.
151. de Winter, J.C.F.. Using the Student's t-test with extremely small sample sizes. In: *Practical Assessment, Research & Evaluation*, 2013, 18(10), Available online: <http://pareonline.net/getvn.asp?v=18&n=10>.
152. De Winter, J. C. F. Small sample sizes, overextraction, and unrealistic expectations: A commentary on M. Mattsson. In: *Accident Analysis and Prevention*, 2013, №50, p. 776–777.
153. Weirich P. *Decision Space: Multidimensional Utility Analysis*. Cambridge: Cambridge University Press, 2015, 288p.
154. Weirich P. *Models of Decision-Making: Simplifying Choices*. Cambridge: Cambridge University Press, 2015, 276p.
155. Wheeler D.J. *Normality and the Process Behavior Chart*. Knoxville: Statistical Process Controls, 2000, 149p.
156. Wooldridge J.M. *Introductory Econometrics, A modern approach*. South-Western, Cengage learning. Michigan State University, 2013, 878p.
157. Yadav R.P., Adhikary S., Agrawal C.S. A comparative study of early vs. delayed laparoscopic cholecystectomy in acute cholecystitis In: *Kathmandu Univ Med J (KUMJ)*, 2009, Jan- Mar, 7(25), p. 16–20.

158. Yu-Meng Ding, Yan Zhong Dang Problem-oriented Decision Support in Manufacturing Industry. In: International Conference on Management Science and Management Innovation, 2014, p. 568-574.

159. Zasada M. Evaluation of the double normal distribution for tree diameter distribution modeling. In: Silva Fennica, vol. 47, № 2, article id 956.

160. Zerbet A., Nikulin M. A new statistics for detecting outliers in exponential case In: Communications in Statistics: Theory and Methods, 2003, v.32, p. 573–584.

**Приложение А1. Таблицы расчета построения исходной таблицы  
многомерных данных**

Таблица А1.1 – Таблица ненормированных плотностей вероятности для фактора 2

$X_{2j}$	$X_{2f}$							
	<b>89,65</b>	<b>90,6</b>	<b>92,5</b>	<b>95,8</b>	<b>99,4</b>	<b>100,9</b>	<b>102,6</b>	<b>108,5</b>
82,84657	0,2255	0,1445	0,0499					
83,8928	0,3442	0,2352	0,0922					
84,93903	0,4896	0,3566	0,1589	0,0225				
85,98525	0,6491	0,504	0,2552	0,0451				
87,03148	0,802	0,6638	0,382	0,0842				
88,07771	0,9235	0,8149	0,533	0,1468	0,0162			
89,12393	0,9911	0,9323	0,693	0,2383	0,0334	0,0115		
90,17016	0,9913	0,9941	0,8397	0,3606	0,0645	0,0246		
91,21639	0,9241	0,9878	0,9484	0,5086	0,1159	0,0489	0,0155	
92,26262	0,8028	0,9149	0,9982	0,6686	0,1941	0,0907	0,0321	
93,30884	0,65	0,7897	0,9792	0,819	0,3031	0,1566	0,0622	
94,35507	0,4905	0,6353	0,8952	0,935	0,4409	0,252	0,1122	
95,4013	0,345	0,4763	0,7627	0,9949	0,5978	0,378	0,1887	
96,44752	0,2261	0,3328	0,6057	0,9866	0,7554	0,5284	0,2958	
97,49375	0,1381	0,2167	0,4482	0,9118	0,8897	0,6884	0,4322	0,0203
98,53998	0,0786	0,1315	0,3092	0,7854	0,9765	0,8359	0,5884	0,0411
99,5862	0,0417	0,0744	0,1987	0,6305	0,9989	0,946	0,7466	0,0776
100,63243	0,0206	0,0392	0,1191	0,4717	0,9523	0,9977	0,8829	0,1365
101,67866		0,0193	0,0665	0,3289	0,8461	0,9807	0,9731	0,2238
102,72488			0,0346	0,2137	0,7007	0,8984	0,9995	0,3419
103,77111			0,0168	0,1294	0,5408	0,767	0,9568	0,487
104,81734				0,0731	0,3889	0,6103	0,8537	0,6464
105,86357				0,0384	0,2607	0,4526	0,7098	0,7996
106,90979				0,0188	0,1629	0,3128	0,5501	0,9219
107,95602					0,0948	0,2015	0,3973	0,9905
109,00225					0,0515	0,121	0,2674	0,9919
110,04847					0,026	0,0677	0,1678	0,9257
111,0947					0,0123	0,0353	0,0981	0,8052
112,14093						0,0171	0,0534	0,6528
113,18715							0,0271	0,4932

Таблица А1.2 – Таблица ненормированных плотностей вероятности для фактора 3

$X_{3j}$	$X_{3f}$							
	<b>64,97</b>	<b>65,78</b>	<b>66,11</b>	<b>66,71</b>	<b>66,85</b>	<b>67,71</b>	<b>68,13</b>	<b>68,58</b>
64,0963	0,5004	0,0764	0,0253					
64,29335	0,6601	0,1347	0,0501					

64,4904	0,8117	0,2212	0,0926	0,0115				
64,68745	0,9301	0,3387	0,1595	0,0245	0,0144			
64,8845	0,9934	0,4832	0,2561	0,0487	0,0301			
65,08155	0,9888	0,6424	0,3831	0,0902	0,0586			
65,2786	0,9172	0,7961	0,5342	0,1559	0,1065			
65,47565	0,793	0,9194	0,6942	0,2511	0,1803			
65,6727	0,639	0,9896	0,8407	0,3768	0,2844	0,0232		
65,86975	0,4798	0,9927	0,949	0,5271	0,4183	0,0463		
66,0668	0,3358	0,9281	0,9983	0,6871	0,5733	0,0864	0,021	
66,26385	0,219	0,8087	0,9788	0,8348	0,7322	0,15	0,0425	
66,4609	0,1332	0,6567	0,8943	0,9453	0,8717	0,2429	0,0799	0,017
66,65795	0,0754	0,497	0,7616	0,9975	0,9671	0,3664	0,1401	0,035
66,855	0,0398	0,3506	0,6044	0,9811	1	0,5153	0,2289	0,0673
67,05205	0,0196	0,2304	0,4471	0,8993	0,9636	0,6752	0,3485	0,1203
67,2491		0,1412	0,3082	0,7683	0,8655	0,8247	0,4947	0,2005
67,44615		0,0806	0,198	0,6117	0,7244	0,9388	0,6543	0,3116
67,6432		0,0429	0,1186	0,4539	0,5651	0,996	0,8066	0,4511
67,84025		0,0213	0,0662	0,3139	0,4109	0,9847	0,9267	0,6087
68,0373			0,0344	0,2023	0,2784	0,9074	0,9922	0,7656
68,23435			0,0167	0,1215	0,1758	0,7793	0,9902	0,8973
68,4314				0,068	0,1035	0,6237	0,9209	0,9802
68,62845				0,0355	0,0568	0,4653	0,7982	0,9979
68,8255				0,0173	0,029	0,3234	0,6448	0,9468
69,02255					0,0138	0,2096	0,4855	0,8372
69,2196						0,1265	0,3406	0,69
69,41665						0,0712	0,2228	0,53
69,6137						0,0374	0,1358	0,3794
69,81075						0,0183	0,0771	0,2531

Таблица А1.3 – Таблица ненормированных плотностей вероятности для фактора 4

$X_{4j}$	$X_{4f}$							
	<b>-36,08</b>	<b>-32,09</b>	<b>-31,83</b>	<b>-30,21</b>	<b>-28,9</b>	<b>-28,32</b>	<b>-27,45</b>	<b>-27,29</b>
-36,93866	0,8918	0,026	0,0174					
-36,46242	0,9775	0,0514	0,0357					
-35,98618	0,9986	0,0947	0,0684					
-35,50993	0,9508	0,1626	0,1221	0,0128				
-35,03369	0,8437	0,2604	0,2031	0,027				
-34,55745	0,6977	0,3885	0,315	0,0531				
-34,0812	0,5377	0,5403	0,4552	0,0976	0,0155			
-33,60496	0,3863	0,7002	0,6131	0,167	0,0321	0,0131		
-33,12871	0,2586	0,8457	0,7696	0,2664	0,0622	0,0276		
-32,65247	0,1613	0,9521	0,9003	0,396	0,1123	0,0542	0,015	0,0115
-32,17623	0,0938	0,9988	0,9816	0,5486	0,1888	0,0993	0,0312	0,0245
-31,69998	0,0508	0,9767	0,9974	0,7084	0,296	0,1696	0,0605	0,0488

-31,22374	0,0257	0,89	0,9445	0,8525	0,4324	0,27	0,1095	0,0904
-30,74749	0,0121	0,7559	0,8336	0,9561	0,5886	0,4005	0,1848	0,1562
-30,27125		0,5983	0,6857	0,9994	0,7468	0,5536	0,2905	0,2515
-29,79501		0,4414	0,5257	0,9736	0,883	0,7133	0,4257	0,3774
-29,31876		0,3034	0,3756	0,884	0,9731	0,8565	0,5814	0,5277
-28,84252		0,1944	0,2501	0,748	0,9995	0,9585	0,74	0,6878
-28,36627		0,1161	0,1552	0,5899	0,9567	0,9997	0,8778	0,8354
-27,89003		0,0646	0,0898	0,4335	0,8535	0,9717	0,9704	0,9456
-27,41379		0,0335	0,0484	0,297	0,7096	0,8803	0,9998	0,9976
-26,93754		0,0162	0,0243	0,1896	0,5499	0,7432	0,96	0,9809
-26,4613			0,0114	0,1128	0,3971	0,5848	0,8592	0,8988
-25,98505				0,0625	0,2673	0,4289	0,7166	0,7676
-25,50881				0,0323	0,1677	0,2931	0,557	0,611
-25,03257				0,0156	0,098	0,1867	0,4035	0,4532
-24,55632					0,0534	0,1108	0,2725	0,3133
-24,08008					0,0271	0,0613	0,1714	0,2019
-23,60384					0,0128	0,0316	0,1005	0,1212
-23,12759						0,0152	0,055	0,0678

Таблица А1.4 – Таблица ненормированных плотностей вероятности для выходной величины Y

$Y_j$	$Y_f$							
	55,94	57,18	72,03	72,22	75,48	76,34	79,12	81,9
49,64024	0,557	0,4324						
51,18567	0,7165	0,5887						
52,73109	0,8591	0,7469						
54,27652	0,96	0,8831						
55,82195	0,9998	0,9732	0,0208	0,019				
57,36738	0,9704	0,9995	0,042	0,0387				
58,91281	0,8778	0,9567	0,0791	0,0734	0,0175	0,0113		
60,45824	0,74	0,8534	0,1388	0,13	0,0359	0,0242		
62,00367	0,5815	0,7095	0,2271	0,2146	0,0687	0,0483	0,0133	
63,5491	0,4258	0,5498	0,3462	0,33	0,1226	0,0896	0,028	
65,09453	0,2906	0,397	0,492	0,473	0,2038	0,1549	0,055	0,0155
66,63996	0,1848	0,2672	0,6515	0,6318	0,3159	0,2497	0,1006	0,0323
68,18539	0,1096	0,1676	0,8041	0,7866	0,4563	0,3751	0,1715	0,0624
69,73082	0,0605	0,098	0,925	0,9127	0,6142	0,5251	0,2725	0,1126
71,27625	0,0312	0,0534	0,9917	0,987	0,7706	0,6851	0,4036	0,1893
72,82168	0,015	0,0271	0,9908	0,9947	0,901	0,8331	0,5571	0,2966
74,36711		0,0128	0,9226	0,9343	0,9819	0,9442	0,7167	0,4331
75,91254			0,8007	0,8179	0,9972	0,9973	0,8592	0,5894
77,45797			0,6476	0,6672	0,9439	0,9817	0,9601	0,7475
79,0034			0,4882	0,5073	0,8327	0,9007	0,9998	0,8836
80,54883			0,3429	0,3595	0,6846	0,7701	0,9703	0,9734
82,09426			0,2245	0,2374	0,5246	0,6137	0,8777	0,9994

83,63969			0,137	0,1462	0,3746	0,4558	0,7399	0,9563
85,18512			0,0779	0,0838	0,2493	0,3155	0,5813	0,8529
86,73055			0,0413	0,0448	0,1547	0,2035	0,4256	0,7089
88,27598			0,0204	0,0223	0,0894	0,1223	0,2905	0,5491
89,82141					0,0482	0,0685	0,1847	0,3964
91,36683					0,0242	0,0358	0,1095	0,2667
92,91226					0,0113	0,0174	0,0605	0,1672
94,45769							0,0311	0,0977

Таблица А1.5 – Состыковка виртуальных значений для 2 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
		88,07771	0,0162	66,4609	0,017			58,91281	0,0175
		89,12393	0,0334	66,65795	0,035	-36,93866	0,026	60,45824	0,0359
		90,17016	0,0645	66,855	0,0673	-36,46242	0,0514	62,00367	0,0687
		91,21639	0,1159	67,05205	0,1203	-35,98618	0,0947	63,5491	0,1226
		92,26262	0,1941	67,2491	0,2005	-35,50993	0,1626	65,09453	0,2038
		93,30884	0,3031	67,44615	0,3116	-35,03369	0,2604	66,63996	0,3159
		94,35507	0,4409	67,6432	0,4511	-34,55745	0,3885	68,18539	0,4563
		95,4013	0,5978	67,84025	0,6087	-34,0812	0,5403	69,73082	0,6142
		96,44752	0,7554	68,0373	0,7656	-33,60496	0,7002	71,27625	0,7706
		97,49375	0,8897	68,23435	0,8973	-33,12871	0,8457	72,82168	0,901
0,64143	0,9779	98,53998	0,9765	68,4314	0,9802	-32,65247	0,9521	74,36711	0,9819
0,64466	0,9985	99,5862	0,9989	68,62845	0,9979	-32,17623	0,9988	75,91254	0,9972
0,64788	0,9504	100,63243	0,9523	68,8255	0,9468	-31,69998	0,9767	77,45797	0,9439
0,65111	0,843	101,67866	0,8461	69,02255	0,8372	-31,22374	0,89	79,0034	0,8327
0,65434	0,6968	102,72488	0,7007	69,2196	0,69	-30,74749	0,7559	80,54883	0,6846
0,65757	0,5367	103,77111	0,5408	69,41665	0,53	-30,27125	0,5983	82,09426	0,5246
0,6608	0,3853	104,81734	0,3889	69,6137	0,3794	-29,79501	0,4414	83,63969	0,3746
0,66403	0,2577	105,86357	0,2607	69,81075	0,2531	-29,31876	0,3034	85,18512	0,2493
0,66725	0,1609	106,90979	0,1629			-28,84252	0,1944	86,73055	0,1547
0,67048	0,0935	107,95602	0,0948			-28,36627	0,1161	88,27598	0,0894
0,67371	0,0506	109,00225	0,0515			-27,89003	0,0646	89,82141	0,0482
0,67694	0,0256	110,04847	0,026			-27,41379	0,0335	91,36683	0,0242
0,68017	0,012	111,0947	0,0123			-26,93754	0,0162	92,91226	0,0113

Таблица А1.6 – Состыковка виртуальных значений для 3 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
		97,49375	0,0203					62,00367	0,0133
0,64143	0,0277	98,53998	0,0411					63,5491	0,028
0,64466	0,0545	99,5862	0,0776					65,09453	0,055
0,64788	0,0997	100,63243	0,1365					66,63996	0,1006
0,65111	0,1702	101,67866	0,2238					68,18539	0,1715
0,65434	0,2709	102,72488	0,3419					69,73082	0,2725
0,65757	0,4016	103,77111	0,487					71,27625	0,4036
0,6608	0,555	104,81734	0,6464	64,0963	0,5004			72,82168	0,5571
0,66403	0,7147	105,86357	0,7996	64,29335	0,6601			74,36711	0,7167
0,66725	0,8573	106,90979	0,9219	64,4904	0,8117	-36,93866	0,8918	75,91254	0,8592

0,67048	0,959	107,95602	0,9905	64,68745	0,9301	-36,46242	0,9775	77,45797	0,9601
0,67371	0,9997	109,00225	0,9919	64,8845	0,9934	-35,98618	0,9986	79,0034	0,9998
0,67694	0,9712	110,04847	0,9257	65,08155	0,9888	-35,50993	0,9508	80,54883	0,9703
0,68017	0,8793	111,0947	0,8052	65,2786	0,9172	-35,03369	0,8437	82,09426	0,8777
0,6834	0,7419	112,14093	0,6528	65,47565	0,793	-34,55745	0,6977	83,63969	0,7399
0,68662	0,5838	113,18715	0,4932	65,6727	0,639	-34,0812	0,5377	85,18512	0,5813
0,68985	0,4279			65,86975	0,4798	-33,60496	0,3863	86,73055	0,4256
0,69308	0,2922			66,0668	0,3358	-33,12871	0,2586	88,27598	0,2905
0,69631	0,186			66,26385	0,219	-32,65247	0,1613	89,82141	0,1847
0,69954	0,1103			66,4609	0,1332	-32,17623	0,0938	91,36683	0,1095
0,70277	0,061			66,65795	0,0754	-31,69998	0,0508	92,91226	0,0605
0,706	0,0314			66,855	0,0398	-31,22374	0,0257	94,45769	0,0311
0,70922	0,0151			67,05205	0,0196	-30,74749	0,0121		

Таблица А1.7 – Состыковка виртуальных значений для 4 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
0,6608	0,0192					-33,60496	0,0131		
0,66403	0,0391			65,6727	0,0232	-33,12871	0,0276	55,82195	0,0208
0,66725	0,0741	82,84657	0,0499	65,86975	0,0463	-32,65247	0,0542	57,36738	0,042
0,67048	0,1311	83,8928	0,0922	66,0668	0,0864	-32,17623	0,0993	58,91281	0,0791
0,67371	0,2162	84,93903	0,1589	66,26385	0,15	-31,69998	0,1696	60,45824	0,1388
0,67694	0,3321	85,98525	0,2552	66,4609	0,2429	-31,22374	0,27	62,00367	0,2271
0,68017	0,4756	87,03148	0,382	66,65795	0,3664	-30,74749	0,4005	63,5491	0,3462
0,6834	0,6346	88,07771	0,533	66,855	0,5153	-30,27125	0,5536	65,09453	0,492
0,68662	0,7887	89,12393	0,693	67,05205	0,6752	-29,79501	0,7133	66,63996	0,6515
0,68985	0,9143	90,17016	0,8397	67,2491	0,8247	-29,31876	0,8565	68,18539	0,8041
0,69308	0,9876	91,21639	0,9484	67,44615	0,9388	-28,84252	0,9585	69,73082	0,925
0,69631	0,9942	92,26262	0,9982	67,6432	0,996	-28,36627	0,9997	71,27625	0,9917
0,69954	0,9327	93,30884	0,9792	67,84025	0,9847	-27,89003	0,9717	72,82168	0,9908
0,70277	0,8154	94,35507	0,8952	68,0373	0,9074	-27,41379	0,8803	74,36711	0,9226
0,706	0,6644	95,4013	0,7627	68,23435	0,7793	-26,93754	0,7432	75,91254	0,8007
0,70922	0,5049	96,44752	0,6057	68,4314	0,6237	-26,4613	0,5848	77,45797	0,6476
0,71245	0,3574	97,49375	0,4482	68,62845	0,4653	-25,98505	0,4289	79,0034	0,4882
0,71568	0,2357	98,53998	0,3092	68,8255	0,3234	-25,50881	0,2931	80,54883	0,3429
0,71891	0,1449	99,5862	0,1987	69,02255	0,2096	-25,03257	0,1867	82,09426	0,2245
0,72214	0,083	100,63243	0,1191	69,2196	0,1265	-24,55632	0,1108	83,63969	0,137
0,72537	0,0443	101,67866	0,0665	69,41665	0,0712	-24,08008	0,0613	85,18512	0,0779
0,72859	0,0221	102,72488	0,0346	69,6137	0,0374	-23,60384	0,0316	86,73055	0,0413
		103,77111	0,0168	69,81075	0,0183	-23,12759	0,0152	88,27598	0,0204

Таблица А1.9 – Состыковка виртуальных значений для 5 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
0,67694	0,0198					-34,0812	0,0155	58,91281	0,0113
0,68017	0,0403	84,93903	0,0225	64,0963	0,0253	-33,60496	0,0321	60,45824	0,0242
0,6834	0,0762	85,98525	0,0451	64,29335	0,0501	-33,12871	0,0622	62,00367	0,0483
0,68662	0,1342	87,03148	0,0842	64,4904	0,0926	-32,65247	0,1123	63,5491	0,0896
0,68985	0,2205	88,07771	0,1468	64,68745	0,1595	-32,17623	0,1888	65,09453	0,1549
0,69308	0,3378	89,12393	0,2383	64,8845	0,2561	-31,69998	0,296	66,63996	0,2497

0,69631	0,4823	90,17016	0,3606	65,08155	0,3831	-31,22374	0,4324	68,18539	0,3751
0,69954	0,6416	91,21639	0,5086	65,2786	0,5342	-30,74749	0,5886	69,73082	0,5251
0,70277	0,7954	92,26262	0,6686	65,47565	0,6942	-30,27125	0,7468	71,27625	0,6851
0,706	0,919	93,30884	0,819	65,6727	0,8407	-29,79501	0,883	72,82168	0,8331
0,70922	0,9894	94,35507	0,935	65,86975	0,949	-29,31876	0,9731	74,36711	0,9442
0,71245	0,9929	95,4013	0,9949	66,0668	0,9983	-28,84252	0,9995	75,91254	0,9973
0,71568	0,9287	96,44752	0,9866	66,26385	0,9788	-28,36627	0,9567	77,45797	0,9817
0,71891	0,8094	97,49375	0,9118	66,4609	0,8943	-27,89003	0,8535	79,0034	0,9007
0,72214	0,6575	98,53998	0,7854	66,65795	0,7616	-27,41379	0,7096	80,54883	0,7701
0,72537	0,4977	99,5862	0,6305	66,855	0,6044	-26,93754	0,5499	82,09426	0,6137
0,72859	0,3515	100,63243	0,4717	67,05205	0,4471	-26,4613	0,3971	83,63969	0,4558
0,73182	0,2311	101,67866	0,3289	67,2491	0,3082	-25,98505	0,2673	85,18512	0,3155
0,73505	0,1416	102,72488	0,2137	67,44615	0,198	-25,50881	0,1677	86,73055	0,2035
		103,77111	0,1294	67,6432	0,1186	-25,03257	0,098	88,27598	0,1223
		104,81734	0,0731	67,84025	0,0662	-24,55632	0,0534	89,82141	0,0685
		105,86357	0,0384	68,0373	0,0344	-24,08008	0,0271	91,36683	0,0358
		106,90979	0,0188	68,23435	0,0167	-23,60384	0,0128	92,91226	0,0174

Таблица А1.10 – Состыковка виртуальных значений для 6 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
		89,12393	0,0115			-32,65247	0,015	55,82195	0,019
0,65111	0,0206	90,17016	0,0246	66,0668	0,021	-32,17623	0,0312	57,36738	0,0387
0,65434	0,0417	91,21639	0,0489	66,26385	0,0425	-31,69998	0,0605	58,91281	0,0734
0,65757	0,0787	92,26262	0,0907	66,4609	0,0799	-31,22374	0,1095	60,45824	0,13
0,6608	0,1382	93,30884	0,1566	66,65795	0,1401	-30,74749	0,1848	62,00367	0,2146
0,66403	0,2263	94,35507	0,252	66,855	0,2289	-30,27125	0,2905	63,5491	0,33
0,66725	0,3448	95,4013	0,378	67,05205	0,3485	-29,79501	0,4257	65,09453	0,473
0,67048	0,4904	96,44752	0,5284	67,2491	0,4947	-29,31876	0,5814	66,63996	0,6318
0,67371	0,65	97,49375	0,6884	67,44615	0,6543	-28,84252	0,74	68,18539	0,7866
0,67694	0,8029	98,53998	0,8359	67,6432	0,8066	-28,36627	0,8778	69,73082	0,9127
0,68017	0,9242	99,5862	0,946	67,84025	0,9267	-27,89003	0,9704	71,27625	0,987
0,6834	0,9914	100,63243	0,9977	68,0373	0,9922	-27,41379	0,9998	72,82168	0,9947
0,68662	0,9912	101,67866	0,9807	68,23435	0,9902	-26,93754	0,96	74,36711	0,9343
0,68985	0,9236	102,72488	0,8984	68,4314	0,9209	-26,4613	0,8592	75,91254	0,8179
0,69308	0,802	103,77111	0,767	68,62845	0,7982	-25,98505	0,7166	77,45797	0,6672
0,69631	0,649	104,81734	0,6103	68,8255	0,6448	-25,50881	0,557	79,0034	0,5073
0,69954	0,4895	105,86357	0,4526	69,02255	0,4855	-25,03257	0,4035	80,54883	0,3595
0,70277	0,344	106,90979	0,3128	69,2196	0,3406	-24,55632	0,2725	82,09426	0,2374
0,706	0,2253	107,95602	0,2015	69,41665	0,2228	-24,08008	0,1714	83,63969	0,1462
0,70922	0,1377	109,00225	0,121	69,6137	0,1358	-23,60384	0,1005	85,18512	0,0838
0,71245	0,0784	110,04847	0,0677	69,81075	0,0771	-23,12759	0,055	86,73055	0,0448
0,71568	0,0416	111,0947	0,0353					88,27598	0,0223
0,71891	0,0205	112,14093	0,0171						

Таблица А1.11 – Состыковка виртуальных значений для 7 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
0,65757	0,0182	91,21639	0,0155			-35,50993	0,0128	65,09453	0,0155
0,6608	0,0373	92,26262	0,0321			-35,03369	0,027	66,63996	0,0323

0,66403	0,0711	93,30884	0,0622	64,0963	0,0764	-34,55745	0,0531	68,18539	0,0624
0,66725	0,1262	94,35507	0,1122	64,29335	0,1347	-34,0812	0,0976	69,73082	0,1126
0,67048	0,2091	95,4013	0,1887	64,4904	0,2212	-33,60496	0,167	71,27625	0,1893
0,67371	0,3229	96,44752	0,2958	64,68745	0,3387	-33,12871	0,2664	72,82168	0,2966
0,67694	0,4647	97,49375	0,4322	64,8845	0,4832	-32,65247	0,396	74,36711	0,4331
0,68017	0,6232	98,53998	0,5884	65,08155	0,6424	-32,17623	0,5486	75,91254	0,5894
0,6834	0,7789	99,5862	0,7466	65,2786	0,7961	-31,69998	0,7084	77,45797	0,7475
0,68662	0,9068	100,63243	0,8829	65,47565	0,9194	-31,22374	0,8525	79,0034	0,8836
0,68985	0,9845	101,67866	0,9731	65,6727	0,9896	-30,74749	0,9561	80,54883	0,9734
0,69308	0,9961	102,72488	0,9995	65,86975	0,9927	-30,27125	0,9994	82,09426	0,9994
0,69631	0,9392	103,77111	0,9568	66,0668	0,9281	-29,79501	0,9736	83,63969	0,9563
0,69954	0,8252	104,81734	0,8537	66,26385	0,8087	-29,31876	0,884	85,18512	0,8529
0,70277	0,6757	105,86357	0,7098	66,4609	0,6567	-28,84252	0,748	86,73055	0,7089
0,706	0,5156	106,90979	0,5501	66,65795	0,497	-28,36627	0,5899	88,27598	0,5491
0,70922	0,3671	107,95602	0,3973	66,855	0,3506	-27,89003	0,4335	89,82141	0,3964
0,71245	0,2434	109,00225	0,2674	67,05205	0,2304	-27,41379	0,297	91,36683	0,2667
0,71568	0,1503	110,04847	0,1678	67,2491	0,1412	-26,93754	0,1896	92,91226	0,1672
0,71891	0,0865	111,0947	0,0981	67,44615	0,0806	-26,4613	0,1128	94,45769	0,0977
0,72214	0,0464	112,14093	0,0534	67,6432	0,0429	-25,98505	0,0625		
0,72537	0,0232	113,18715	0,0271	67,84025	0,0213	-25,50881	0,0323		
						-25,03257	0,0156		

Таблица А1.12 – Состыковка виртуальных значений для 8 строки

$X_1$	$P_{X1}$	$X_2$	$P_{X2}$	$X_3$	$P_{X3}$	$X_4$	$P_{X4}$	$Y$	$P_Y$
0,6608	0,0119			64,68745	0,0144	-36,93866	0,0174		
0,66403	0,0254			64,8845	0,0301	-36,46242	0,0357		
0,66725	0,0502			65,08155	0,0586	-35,98618	0,0684		
0,67048	0,0929			65,2786	0,1065	-35,50993	0,1221		
0,67371	0,1599	82,84657	0,1445	65,47565	0,1803	-35,03369	0,2031		
0,67694	0,2567	83,8928	0,2352	65,6727	0,2844	-34,55745	0,315		
0,68017	0,384	84,93903	0,3566	65,86975	0,4183	-34,0812	0,4552		
0,6834	0,5352	85,98525	0,504	66,0668	0,5733	-33,60496	0,6131	49,64024	0,557
0,68662	0,6948	87,03148	0,6638	66,26385	0,7322	-33,12871	0,7696	51,18567	0,7165
0,68985	0,8413	88,07771	0,8149	66,4609	0,8717	-32,65247	0,9003	52,73109	0,8591
0,69308	0,9494	89,12393	0,9323	66,65795	0,9671	-32,17623	0,9816	54,27652	0,96
0,69631	0,9984	90,17016	0,9941	66,855	1	-31,69998	0,9974	55,82195	0,9998
0,69954	0,9784	91,21639	0,9878	67,05205	0,9636	-31,22374	0,9445	57,36738	0,9704
0,70277	0,8936	92,26262	0,9149	67,2491	0,8655	-30,74749	0,8336	58,91281	0,8778
0,706	0,7605	93,30884	0,7897	67,44615	0,7244	-30,27125	0,6857	60,45824	0,74
0,70922	0,6037	94,35507	0,6353	67,6432	0,5651	-29,79501	0,5257	62,00367	0,5815
0,71245	0,4464	95,4013	0,4763	67,84025	0,4109	-29,31876	0,3756	63,5491	0,4258
0,71568	0,3075	96,44752	0,3328	68,0373	0,2784	-28,84252	0,2501	65,09453	0,2906
0,71891	0,1975	97,49375	0,2167	68,23435	0,1758	-28,36627	0,1552	66,63996	0,1848
0,72214	0,1182	98,53998	0,1315	68,4314	0,1035	-27,89003	0,0898	68,18539	0,1096
0,72537	0,0659	99,5862	0,0744	68,62845	0,0568	-27,41379	0,0484	69,73082	0,0605
0,72859	0,0343	100,63243	0,0392	68,8255	0,029	-26,93754	0,0243	71,27625	0,0312
0,73182	0,0166	101,67866	0,0193	69,02255	0,0138	-26,4613	0,0114	72,82168	0,015

Таблица А1.13 – Итоговая таблица виртуальной выборки

№	Факторы $X_i$				Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	
1	0,680	84,939	65,673	-29,795	49,640
2	0,683	85,985	65,870	-29,319	51,186
3	0,687	87,031	66,067	-28,843	52,731
4	0,690	88,078	66,264	-28,366	54,277
5	0,693	89,124	66,461	-27,890	55,822
6	0,696	90,170	66,658	-27,414	57,367
7	0,700	91,216	66,855	-26,938	58,913
8	0,703	92,263	67,052	-26,461	60,458
9	0,706	93,309	67,249	-25,985	62,004
10	0,709	94,355	67,446	-25,509	63,549
11	0,712	95,401	67,643	-25,033	65,095
12	0,716	96,448	67,840	-24,556	66,640
13	0,719	97,494	68,037	-24,080	68,185
14	0,722	98,540	68,234	-23,604	69,731
15	0,725	99,586	68,431	-23,128	71,276
16	0,641	98,540	68,431	-32,652	74,367
17	0,645	99,586	68,628	-32,176	75,913
18	0,648	100,632	68,826	-31,700	77,458
19	0,651	101,679	69,023	-31,224	79,003
20	0,654	102,725	69,220	-30,747	80,549
21	0,658	103,771	69,417	-30,271	82,094
22	0,661	104,817	69,614	-29,795	83,640
23	0,664	105,864	69,811	-29,319	85,185
24	0,667	106,910	64,490	-36,939	75,913
25	0,670	107,956	64,687	-36,462	77,458
26	0,674	109,002	64,885	-35,986	79,003
27	0,677	110,048	65,082	-35,510	80,549
28	0,680	111,095	65,279	-35,034	82,094
29	0,683	112,141	65,476	-34,557	83,640
30	0,687	113,187	65,673	-34,081	85,185
31	0,667	82,847	65,870	-32,652	57,367
32	0,670	83,893	66,067	-32,176	58,913
33	0,674	84,939	66,264	-31,700	60,458
34	0,677	85,985	66,461	-31,224	62,004
35	0,680	87,031	66,658	-30,747	63,549
36	0,683	88,078	66,855	-30,271	65,095
37	0,687	89,124	67,052	-29,795	66,640
38	0,690	90,170	67,249	-29,319	68,185
39	0,693	91,216	67,446	-28,843	69,731
40	0,696	92,263	67,643	-28,366	71,276
41	0,700	93,309	67,840	-27,890	72,822
42	0,703	94,355	68,037	-27,414	74,367

№	Факторы $X_i$				Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	
43	0,706	95,401	68,234	-26,938	75,913
44	0,709	96,448	68,431	-26,461	77,458
45	0,712	97,494	68,628	-25,985	79,003
46	0,716	98,540	68,826	-25,509	80,549
47	0,719	99,586	69,023	-25,033	82,094
48	0,722	100,632	69,220	-24,556	83,640
49	0,725	101,679	69,417	-24,080	85,185
50	0,729	102,725	69,614	-23,604	86,731
51	0,680	84,939	64,096	-33,605	60,458
52	0,683	85,985	64,293	-33,129	62,004
53	0,687	87,031	64,490	-32,652	63,549
54	0,690	88,078	64,687	-32,176	65,095
55	0,693	89,124	64,885	-31,700	66,640
56	0,696	90,170	65,082	-31,224	68,185
57	0,700	91,216	65,279	-30,747	69,731
58	0,703	92,263	65,476	-30,271	71,276
59	0,706	93,309	65,673	-29,795	72,822
60	0,709	94,355	65,870	-29,319	74,367
61	0,712	95,401	66,067	-28,843	75,913
62	0,716	96,448	66,264	-28,366	77,458
63	0,719	97,494	66,461	-27,890	79,003
64	0,722	98,540	66,658	-27,414	80,549
65	0,725	99,586	66,855	-26,938	82,094
66	0,729	100,632	67,052	-26,461	83,640
67	0,732	101,679	67,249	-25,985	85,185
68	0,735	102,725	67,446	-25,509	86,731
69	0,651	90,170	66,067	-32,176	57,367
70	0,654	91,216	66,264	-31,700	58,913
71	0,658	92,263	66,461	-31,224	60,458
72	0,661	93,309	66,658	-30,747	62,004
73	0,664	94,355	66,855	-30,271	63,549
74	0,667	95,401	67,052	-29,795	65,095
75	0,670	96,448	67,249	-29,319	66,640
76	0,674	97,494	67,446	-28,843	68,185
77	0,677	98,540	67,643	-28,366	69,731
78	0,680	99,586	67,840	-27,890	71,276
79	0,683	100,632	68,037	-27,414	72,822
80	0,687	101,679	68,234	-26,938	74,367
81	0,690	102,725	68,431	-26,461	75,913
82	0,693	103,771	68,628	-25,985	77,458
83	0,696	104,817	68,826	-25,509	79,003
84	0,700	105,864	69,023	-25,033	80,549
85	0,703	106,910	69,220	-24,556	82,094

№	Факторы $X_i$				Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	
86	0,706	107,956	69,417	-24,080	83,640
87	0,709	109,002	69,614	-23,604	85,185
88	0,712	110,048	69,811	-23,128	86,731
89	0,664	93,309	64,096	-34,557	68,185
90	0,667	94,355	64,293	-34,081	69,731
91	0,670	95,401	64,490	-33,605	71,276
92	0,674	96,448	64,687	-33,129	72,822
93	0,677	97,494	64,885	-32,652	74,367
94	0,680	98,540	65,082	-32,176	75,913
95	0,683	99,586	65,279	-31,700	77,458
96	0,687	100,632	65,476	-31,224	79,003
97	0,690	101,679	65,673	-30,747	80,549
98	0,693	102,725	65,870	-30,271	82,094
99	0,696	103,771	66,067	-29,795	83,640
100	0,700	104,817	66,264	-29,319	85,185
101	0,703	105,864	66,461	-28,843	86,731
102	0,706	106,910	66,658	-28,366	88,276
103	0,709	107,956	66,855	-27,890	89,821
104	0,712	109,002	67,052	-27,414	91,367
105	0,716	110,048	67,249	-26,938	92,912
106	0,719	111,095	67,446	-26,461	94,458
107	0,683	85,985	66,067	-33,605	49,640
108	0,687	87,031	66,264	-33,129	51,186
109	0,690	88,078	66,461	-32,652	52,731
110	0,693	89,124	66,658	-32,176	54,277
111	0,696	90,170	66,855	-31,700	55,822
112	0,700	91,216	67,052	-31,224	57,367
113	0,703	92,263	67,249	-30,747	58,913
114	0,706	93,309	67,446	-30,271	60,458
115	0,709	94,355	67,643	-29,795	62,004
116	0,712	95,401	67,840	-29,319	63,549
117	0,716	96,448	68,037	-28,843	65,095
118	0,719	97,494	68,234	-28,366	66,640
119	0,722	98,540	68,431	-27,890	68,185
120	0,725	99,586	68,628	-27,414	69,731
121	0,729	100,632	68,826	-26,938	71,276
122	0,732	101,679	69,023	-26,461	72,822

## Приложение А2. Описание программного обеспечения

Для наглядности общий алгоритм работы программы представлен на рисунке А2.1.



Рисунок А2.1 – Общий алгоритм работы программы

Формирование корреляционной матрицы происходит по алгоритму, показанному на рисунке А2.2.

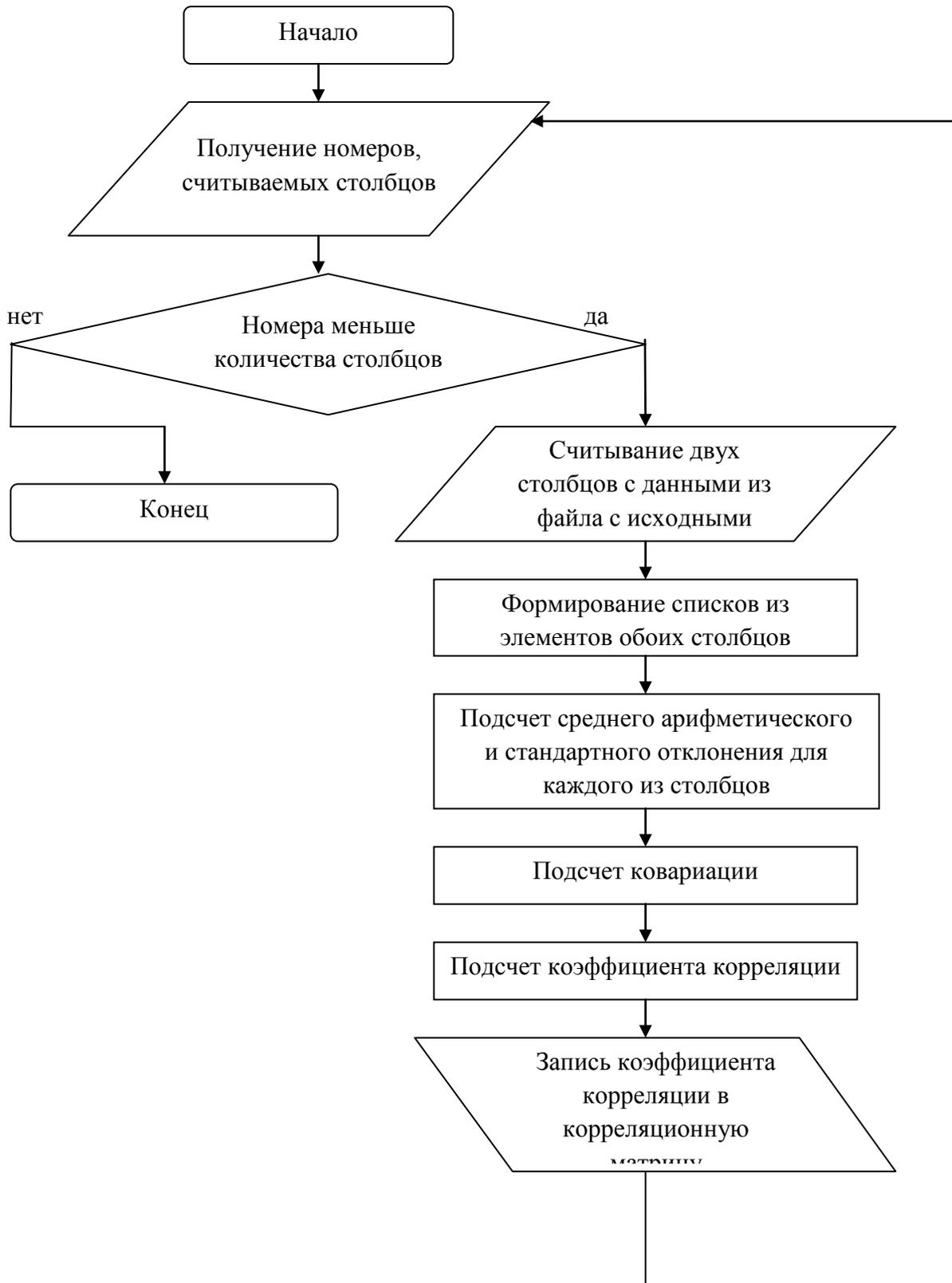


Рисунок А2.2 – Алгоритм формирования корреляционной матрицы

Построение графа корреляционных плеяд (в виде матрицы смежности) происходит по алгоритму, показанному на рисунке А2.3.

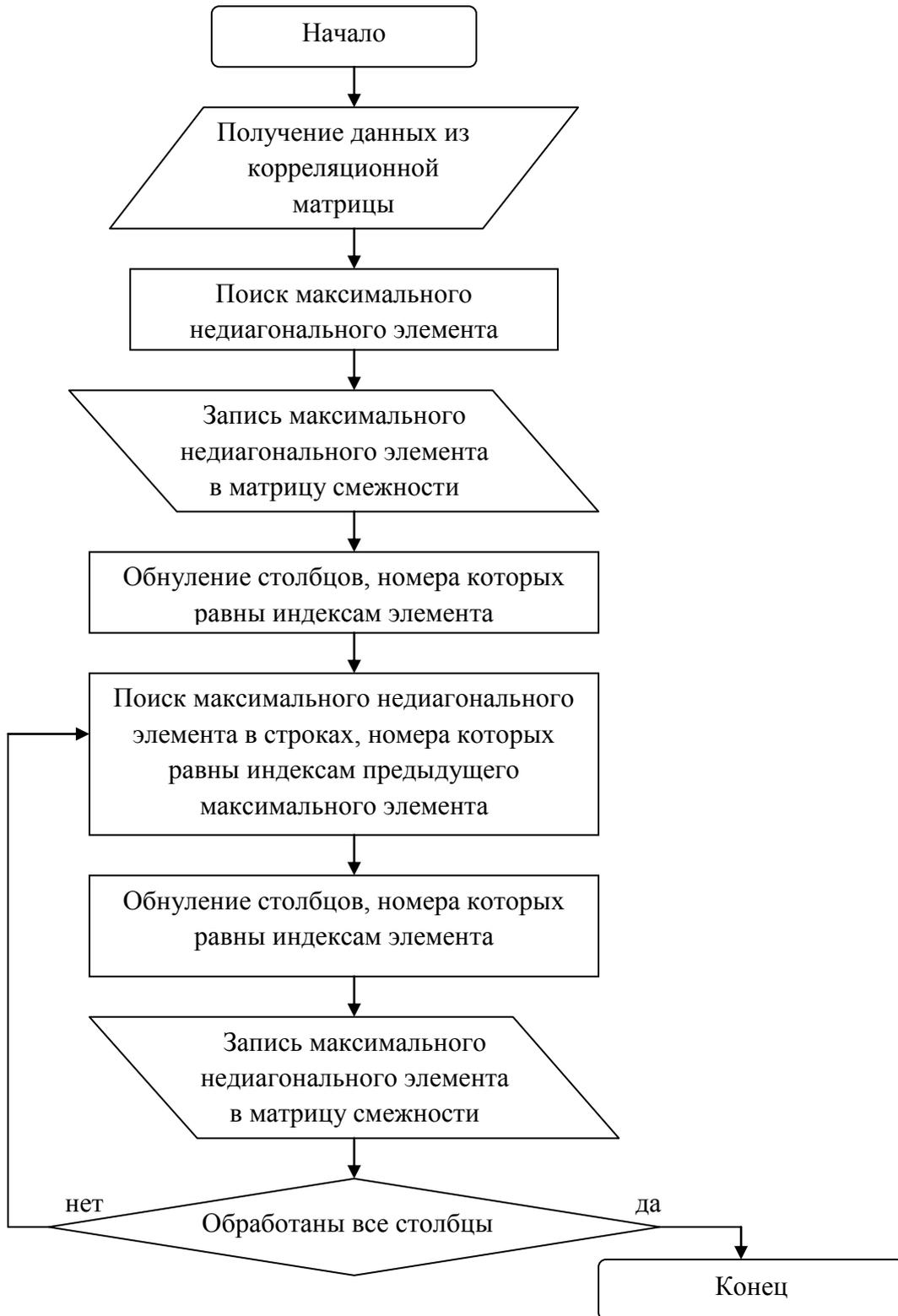


Рисунок А2.3 – Алгоритм построения графа корреляционных плеяд

Выделение характеризующих факторов в корреляционных плеядах, обнуление остальных графа происходит по алгоритму, показанному на рисунке А2.4.

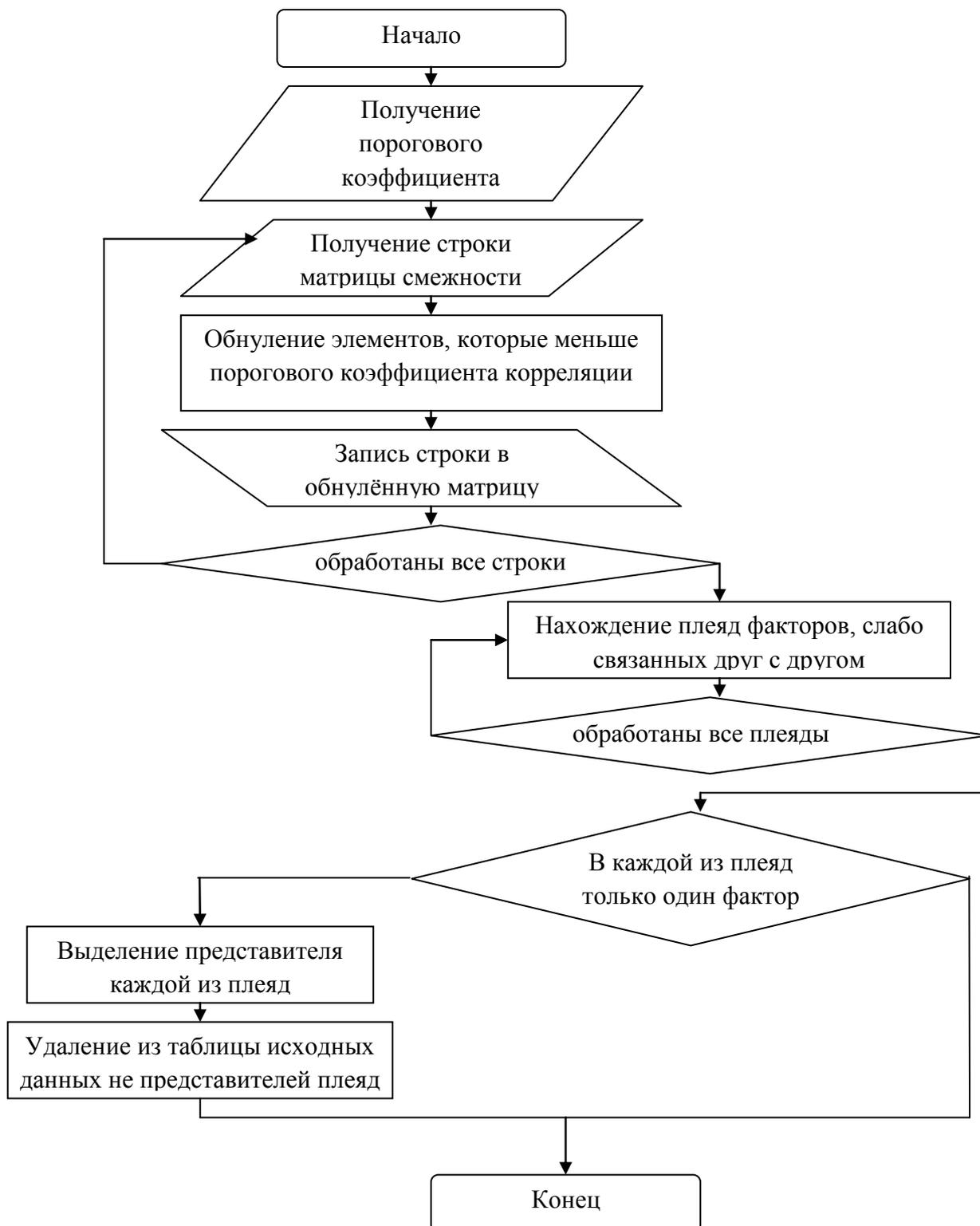


Рисунок А2.4 – Алгоритм выделения характеризующих факторов

1. Программа *Correlation* состоит из одного класса *Main.cs*, реализующего всю логику программы. Класс *Main.cs* содержит следующие функции:

**public static void readColumnFromCSV(string fileName, string FileNameOut, int columnNumber1, int columnNumber2, out List<Double> resultList1, out List<Double> resultList2)** - эта функция считывает два столбца из файла с исходными данными, и формирует из них списки.

```
string fileName           //имя исходного файла
string FfileNameOut       //имя файла для записи результатов
int columnNumber1, columnNumber2 //номера обрабатываемых столбцов
List<Double> resultList1, resultList2 //полученные списки
```

**public static void Correlation(string FileNameOut, ref List<Double> resultList1, ref List<Double> resultList2)** - данная функция подсчитывает коэффициент корреляции Пирсона между элементами списков *resultList1* и *resultList2* и записывает его в корреляционную матрицу.

```
double sumTempDifferense //ковариация между элементами списков.
```

**public static void standotklon(double[] x1, double[] x2, out double Sx1, out double Sx2, out double averx1, out double averx2)** - эта функция получает два массива данных с экспериментальными данными и подсчитывает их средние значения и стандартные отклонения.

```
double[] x1, x2           //массивы исходных данных
double Sx1, Sx2           //стандартные отклонения массивов
double averx1, averx2     //средние арифметические массивов
```

**public static void correlToPleyad(float predel, string FileNameOut, string FileNameIn, ref bool end, ref int Number)** - данная функция формирует из корреляционной матрицы матрицу смежности графа корреляционных плеяд и записывает её в файл.

```
float predel              //пороговый коэффициент корреляции
ref int Number            //номер графа корреляционных плеяд
string[] tempStringMass  //строка из корреляционной матрицы
float[,] tempCorrelation //корреляционная матрица
float[,] grafPleyadTable //матрица смежности графа плеяд
```

```
float max //максимальный элемент матрицы
int indexStr, indexStl //индексы элемента матрицы
```

**public static void searchWithTwoLines(int k, ref int indexStr, ref int indexStl, ref float max, float[,] tempCorrelation)** - эта функция производит поиск максимального элемента в двух заданных строках корреляционной матрицы.

```
int tempIndexSt, tempIndexStr //временные индексы найденного
максимального элемента
```

**public static void pleyadToNull(float predel, string FileNameOut, string FileNameIn, ref bool end, ref int Number)** - данная функция реализует удаление из матрицы смежности графа корреляционных плеяд элементы, которые меньше порогового коэффициента корреляции, таким образом разделяя плеяды между собой.

```
string[] tempStringMass //строка из матрицы смежности
float[,] tempCorrelation //матрица смежности
float predel //пороговый коэффициент корреляции
List<List<Int32>> listOfListLinked //список списков связанных между собой
факторов в матрице смежности графа корреляционных плеяд
```

**verifyCtr(tempCorrelation, i, ref verifyList, ref listLinkedWithStr,k)** - эта функция формирует список факторов, связанных между собой для каждого из факторов матрицы смежности графа корреляционных плеяд.

```
List<Int32> verifyList //список обработанных факторов
List<Int32> listLinkedWithStr //список факторов, связанных со строкой
```

**public static void delColumnFromData(List<List<Int32>> listOfListLinked, string FileNameOut, ref int Number)** - данная функция удаляет столбец из матрицы смежности корреляционных плеяд.

```
List<Int32> Podspisok //список факторов одной из плеяд
Int32 el //фактор из списка факторов одной из плеяд
List<int> toDelete //список факторов, помеченных на удаление
string tempString //строка исходной корреляционной матрицы
string[] tempArray //массив элементов строки исходной корреляционной
матрицы
```

```
string[] temp  
матрицы
```

```
//массив элементов строки полученной корреляционной
```

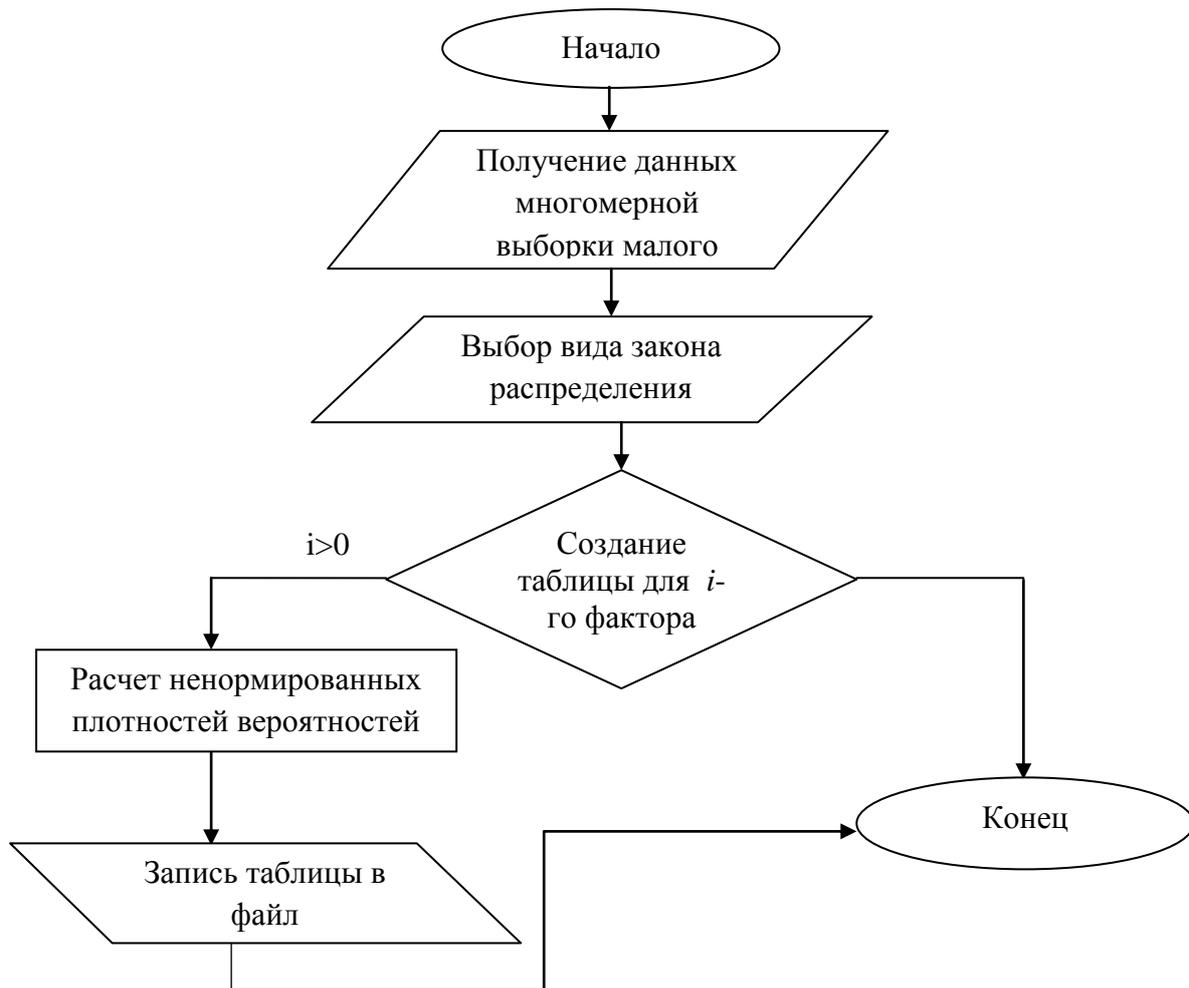


Рисунок А2.5 – Общий алгоритм работы программы *Mtr*

1. Программа *Mtr* состоит из класса *Program.cs*, в котором происходит первоначальная обработка входных данных, выбирается вид закона распределения и записываются в файлы результирующие таблицы для каждого из факторов. Основная логика работы программы находится в родительском классе *MTRNormal*, который реализует метод точечных распределений для случая нормального закона распределения случайной величины. Остальные классы являются дочерними и наследуют логику этого класса, переопределяя необходимые функции. Опишем основные функции, входящие в класс *MTRNormal*:

**private Double getXAverage()** – возвращает математическое ожидание малой выборки.

**private Double getDisp()** – возвращает дисперсию малой выборки.

**protected virtual Double getA()** – возвращает нижнюю границу существования интервала контролируемой величины  $X$ .

**protected virtual Double getB()** – возвращает верхнюю границу существования интервала контролируемой величины  $X$ .

**protected virtual Double getRoShtrih()** – возвращает вспомогательный коэффициент, в зависимости от вида закона распределения и объема выборки.

**public void calc()** – функция, рассчитывающая значения ненормированной оценки плотности в центре  $j$ -го интервала дискретности.

Double  $xi\_n$  // Нижняя граница интервала определений  $i$ -ой дельтаобразной функции

Double  $xi\_v$  // Верхняя граница интервала определений  $i$ -ой дельтаобразной функции

Dictionary<int, Dictionary<int, double>> // Результирующий словарь

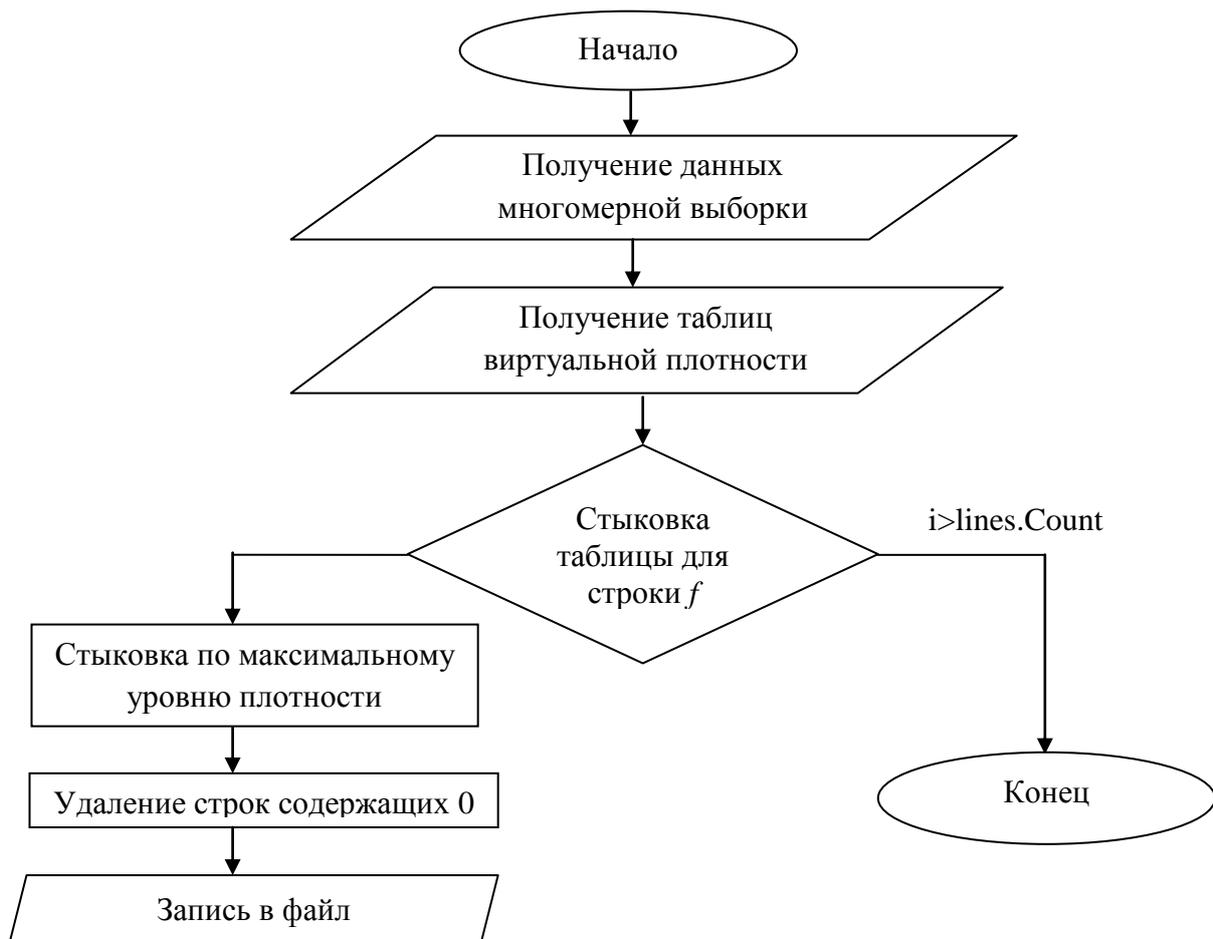


Рисунок A2.6 – Общий алгоритм работы программы *Comb*

Программа *Comb* состоит из 6 классов: *Program*, *Worker*, *LinkTable*, *DataTable*, *ResColumn*, *ResTable*.

Класс *LinkTable* содержит таблицу исходных данных и методы для работы с ней. Выдаёт по номеру эксперимента значения его факторов.

Класс *DataTable* содержит таблицы виртуального распределения плотностей вероятности и операции для работы с ними.

Класс *ResTable* отвечает за хранение итоговых таблиц данных для каждого из экспериментов. В классе реализованы следующие методы:

**public void calcShifts()** – данный метод рассчитывает сдвиг для состыковки относительно первого столбца.

**protected int getMaxShift(),protected int getMinShift()** – получают, соответственно максимальный и минимальные сдвиги в таблице.

**public List<List<double>> calcRes(bool returnProbs)** – возвращает результирующую таблицу для соответствующей строки исходных данных.

Класс *ResColumn* отвечает за формирование данных в процессе состыковки. Его методы возвращают значения максимальной плотности вероятности, а также значения экспериментальных данных и их вероятности по индексу в таблице данных.

Класс *Program* запускает на выполнение класс *Worker*, в который передаёт имена файлов входных и выходных данных.

Вся логика программы хранится в классе *Worker*, который загружает таблицы данных и связей, управляет расчетами, а также выводит результаты преобразований. Рассмотрим основные функции, входящие в данный класс:

**public void loadLinkTable(String fileName, char delimiter)** – данная функция загружает из *.csv* файла таблицу исходных данных.

```
String filename          //имя считываемого файла
Char delimiter          //знак-разделитель
```

**public void loadDataTable(int factorNumber, String fileName, char delimiter)** – загружает из *.csv* файла таблицы с данными.

```
int factorNumber        //номер считываемого фактора
String filename         //имя считываемого файла
Char delimiter          //знак-разделитель
```

**public void doWork(String outFileName, bool outProbs)** – метод *doWork* формирует результирующую таблицу, вызывая классы *ResColumn*, *ResTable* и записывает её в файл.

```
String outFileName     //имя сохраняемого файла
bool outProbs          //переменная, указывающая, записывать ли значения
плотностей вероятности
```

### Приложение А3. Расчет модели сложности операции

Таблица А3.1 – Таблица ненормированных плотностей вероятности для фактора 1

$X_{1j}$	$X_{1f}$							
	33	44	43	30	65	66	70	52
14,6712	0,1396			0,2523				
17,1335	0,2287	0,0146	0,0198	0,3790				
19,5958	0,3489	0,0305	0,0404	0,5303				
22,0582	0,4958	0,0595	0,0765	0,6910				
24,5205	0,6562	0,1082	0,1352	0,8387				0,0120
26,9828	0,8088	0,1832	0,2224	0,9481				0,0255
29,4452	0,9286	0,2890	0,3407	0,9982				0,0507
31,9075	0,9930	0,4245	0,4862	0,9789				0,0939
34,3698	0,9891	0,5807	0,6463	0,8941				0,1618
36,8322	0,9175	0,7400	0,8002	0,7607				0,2597
39,2945	0,7928	0,8783	0,9227	0,6028	0,0208	0,0153		0,3883
41,7568	0,6380	0,9709	0,9910	0,4449	0,0422	0,0319		0,5407
44,2192	0,4783	0,9997	0,9913	0,3058	0,0796	0,0620	0,0203	0,7013
46,6815	0,3339	0,9587	0,9236	0,1958	0,1400	0,1123	0,0413	0,8472
49,1438	0,2171	0,8564	0,8016	0,1168	0,2292	0,1892	0,0782	0,9533
51,6062	0,1315	0,7125	0,6479	0,0649	0,3495	0,2970	0,1377	0,9991
54,0685	0,0742	0,5521	0,4878	0,0336	0,4965	0,4342	0,2260	0,9752
56,5308	0,0390	0,3985	0,3420	0,0162	0,6568	0,5913	0,3454	0,8867
58,9932	0,0191	0,2679	0,2234		0,8094	0,7500	0,4917	0,7508
61,4555		0,1677	0,1359		0,9290	0,8860	0,6519	0,5922
63,9178		0,0978	0,0770		0,9932	0,9749	0,8051	0,4350
66,3802		0,0531	0,0406		0,9889	0,9992	0,9261	0,2977
68,8425		0,0269	0,0200		0,9171	0,9538	0,9922	0,1897
71,3048		0,0127			0,7922	0,8480	0,9901	0,1126
73,7672					0,6374	0,7022	0,9202	0,0623
76,2295					0,4776	0,5416	0,7966	0,0321
78,6918					0,3333	0,3891	0,6423	0,0154
81,1542					0,2167	0,2603	0,4823	
83,6165					0,1312	0,1622	0,3374	
86,0788					0,0740	0,0942	0,2198	

Таблица А3.2 – Таблица ненормированных плотностей вероятности для фактора 2

$X_{2j}$	$X_{2f}$							
	1	1	1	7	15	10	7	1
0,6672	0,9946	0,9946	0,9946	0,1392		0,0138	0,1392	0,9946
1,5136	0,9871	0,9871	0,9871	0,2276	0,0001	0,0290	0,2276	0,9871
2,3599	0,9131	0,9131	0,9131	0,3469	0,0004	0,0567	0,3469	0,9131

3,2063	0,7871	0,7871	0,7871	0,4928	0,0011	0,1034	0,4928	0,7871
4,0527	0,6324	0,6324	0,6324	0,6524	0,0028	0,1757	0,6524	0,6324
4,8991	0,4736	0,4736	0,4736	0,8049	0,0066	0,2782	0,8049	0,4736
5,7454	0,3305	0,3305	0,3305	0,9255	0,0148	0,4107	0,9255	0,3305
6,5918	0,2149	0,2149	0,2149	0,9918	0,0309	0,5649	0,9918	0,2149
7,4382	0,1303	0,1303	0,1303	0,9906	0,0601	0,7242	0,9906	0,1303
8,2845	0,0736	0,0736	0,0736	0,9221	0,1089	0,8653	0,9221	0,0736
9,1309	0,0388	0,0388	0,0388	0,7999	0,1839	0,9635	0,7999	0,0388
9,9773	0,0190	0,0190	0,0190	0,6467	0,2893	1,0000	0,6467	0,0190
10,8236	0,0087	0,0087	0,0087	0,4873	0,4242	0,9672	0,4873	0,0087
11,6700	0,0037	0,0037	0,0037	0,3422	0,5797	0,8719	0,3422	0,0037
12,5164	0,0015	0,0015	0,0015	0,2240	0,7384	0,7325	0,2240	0,0015
13,3628	0,0005	0,0005	0,0005	0,1366	0,8765	0,5735	0,1366	0,0005
14,2091	0,0002	0,0002	0,0002	0,0777	0,9697	0,4185	0,0777	0,0002
15,0555	0,0001	0,0001	0,0001	0,0412	0,9998	0,2846	0,0412	0,0001
15,9019				0,0203	0,9608	0,1804	0,0203	
16,7482				0,0094	0,8605	0,1066	0,0094	
17,5946				0,0040	0,7182	0,0587	0,0040	
18,4410				0,0016	0,5587	0,0301	0,0016	
19,2874				0,0006	0,4050	0,0144	0,0006	
20,1337				0,0002	0,2737	0,0064	0,0002	
20,9801				0,0001	0,1723	0,0027	0,0001	
21,8265					0,1011	0,0010		
22,6728					0,0553	0,0004		
23,5192					0,0282	0,0001		
24,3656					0,0134	0,0000		
25,2119					0,0059	0,0000		

Таблица А3.3 – Таблица ненормированных плотностей вероятности для фактора 3

$X_{3j}$	$X_{3f}$							
	98	90	110	110	120	110	110	90
79,5865	0,0183	0,2782						0,2782
81,3219	0,0376	0,4113						0,4113
83,0573	0,0718	0,5663						0,5663
84,7928	0,1277	0,7262						0,7262
86,5282	0,2117	0,8674						0,8674
88,2636	0,3268	0,9651						0,9651
89,9990	0,4699	1,0000						1,0000
91,7344	0,6293	0,9651	0,0195	0,0195		0,0195	0,0195	0,9651
93,4698	0,7850	0,8676	0,0398	0,0398		0,0398	0,0398	0,8676
95,2052	0,9120	0,7264	0,0756	0,0756		0,0756	0,0756	0,7264
96,9406	0,9868	0,5665	0,1337	0,1337		0,1337	0,1337	0,5665
98,6761	0,9946	0,4115	0,2203	0,2203		0,2203	0,2203	0,4115

100,4115	0,9337	0,2784	0,3380	0,3380		0,3380	0,3380	0,2784
102,1469	0,8164	0,1754	0,4831	0,4831	0,0233	0,4831	0,4831	0,1754
103,8823	0,6648	0,1029	0,6430	0,6430	0,0467	0,6430	0,6430	0,1029
105,6177	0,5043	0,0563	0,7973	0,7973	0,0871	0,7973	0,7973	0,0563
107,3531	0,3563	0,0286	0,9207	0,9207	0,1515	0,9207	0,9207	0,0286
109,0885	0,2344	0,0136	0,9902	0,9902	0,2455	0,9902	0,9902	0,0136
110,8239	0,1437		0,9920	0,9920	0,3703	0,9920	0,9920	
112,5594	0,0820		0,9256	0,9256	0,5204	0,9256	0,9256	
114,2948	0,0436		0,8044	0,8044	0,6811	0,8044	0,8044	
116,0302	0,0216		0,6512	0,6512	0,8303	0,6512	0,6512	
117,7656			0,4909	0,4909	0,9428	0,4909	0,4909	
119,5010			0,3447	0,3447	0,9971	0,3447	0,3447	
121,2364			0,2255	0,2255	0,9821	0,2255	0,2255	
122,9718			0,1374	0,1374	0,9011	0,1374	0,1374	
124,7072			0,0779	0,0779	0,7700	0,0779	0,0779	
126,4427			0,0412	0,0412	0,6128	0,0412	0,0412	
128,1781			0,0203	0,0203	0,4543	0,0203	0,0203	
129,9135					0,3137			

Таблица А3.4 – Таблица ненормированных плотностей вероятности для фактора 4

$X_{4j}$	$X_{4f}$							
	38	38	40	37	45	55	43	40
28,3027	0,0237	0,0237		0,0492				
29,2474	0,0473	0,0473		0,0913				
30,1920	0,0883	0,0883	0,0217	0,1580				0,0217
31,1367	0,1533	0,1533	0,0438	0,2544				0,0438
32,0813	0,2479	0,2479	0,0824	0,3816				0,0824
33,0259	0,3734	0,3734	0,1442	0,5332			0,0190	0,1442
33,9706	0,5239	0,5239	0,2352	0,6939			0,0389	0,2352
34,9152	0,6846	0,6846	0,3572	0,8411	0,0174		0,0741	0,3572
35,8598	0,8333	0,8333	0,5054	0,9496	0,0359		0,1313	0,5054
36,8045	0,9447	0,9447	0,6659	0,9985	0,0690		0,2169	0,6659
37,7491	0,9975	0,9975	0,8173	0,9779	0,1233		0,3336	0,8173
38,6938	0,9810	0,9810	0,9343	0,8921	0,2053		0,4779	0,9343
39,6384	0,8986	0,8986	0,9948	0,7579	0,3184		0,6377	0,9948
40,5830	0,7667	0,7667	0,9866	0,5998	0,4599		0,7925	0,9866
41,5277	0,6093	0,6093	0,9113	0,4421	0,6187		0,9173	0,9113
42,4723	0,4509	0,4509	0,7840	0,3035	0,7754		0,9890	0,7840
43,4170	0,3109	0,3109	0,6282	0,1941	0,9050		0,9931	0,6282
44,3616	0,1996	0,1996	0,4689	0,1156	0,9839		0,9288	0,4689
45,3062	0,1194	0,1194	0,3259	0,0641	0,9963	0,0237	0,8091	0,3259
46,2509	0,0665	0,0665	0,2110	0,0331	0,9396	0,0475	0,6565	0,2110
47,1955	0,0345	0,0345	0,1273	0,0159	0,8254	0,0885	0,4962	0,1273

48,1402	0,0167	0,0167	0,0715		0,6753	0,1536	0,3492	0,0715
49,0848			0,0374		0,5146	0,2483	0,2290	0,0374
50,0294			0,0182		0,3653	0,3739	0,1398	0,0182
50,9741					0,2415	0,5245	0,0795	
51,9187					0,1487	0,6852	0,0421	
52,8633					0,0853	0,8338	0,0208	
53,8080					0,0455	0,9450		
54,7526					0,0227	0,9976		
55,6973						0,9808		

Таблица А3.5 – Таблица ненормированных плотностей вероятности для фактора 5

$X_{sj}$	$X_{sf}$							
	3	4	5	2	5	5	5	4
1,5127	0,0888			0,7711				
1,6929	0,1541			0,9019				
1,8730	0,2490			0,9825				
2,0532	0,3748	0,0158		0,9969				0,0158
2,2333	0,5255	0,0328		0,9421				0,0328
2,4135	0,6862	0,0636		0,8293				0,0636
2,5937	0,8346	0,1147		0,6799				0,1147
2,7738	0,9455	0,1928		0,5192				0,1928
2,9540	0,9977	0,3019		0,3693				0,3019
3,1341	0,9805	0,4401	0,0221	0,2446	0,0221	0,0221	0,0221	0,4401
3,3143	0,8975	0,5977	0,0446	0,1509	0,0446	0,0446	0,0446	0,5977
3,4944	0,7652	0,7559	0,0836	0,0867	0,0836	0,0836	0,0836	0,7559
3,6746	0,6076	0,8906	0,1462	0,0464	0,1462	0,1462	0,1462	0,8906
3,8548	0,4494	0,9772	0,2379	0,0231	0,2379	0,2379	0,2379	0,9772
4,0349	0,3096	0,9987	0,3608		0,3608	0,3608	0,3608	0,9987
4,2151	0,1987	0,9506	0,5094		0,5094	0,5094	0,5094	0,9506
4,3952	0,1187	0,8428	0,6701		0,6701	0,6701	0,6701	0,8428
4,5754	0,0661	0,6960	0,8209		0,8209	0,8209	0,8209	0,6960
4,7556	0,0343	0,5353	0,9367		0,9367	0,9367	0,9367	0,5353
4,9357	0,0165	0,3835	0,9955		0,9955	0,9955	0,9955	0,3835
5,1159		0,2559	0,9854		0,9854	0,9854	0,9854	0,2559
5,2960		0,1590	0,9085		0,9085	0,9085	0,9085	0,1590
5,4762		0,0920	0,7802		0,7802	0,7802	0,7802	0,0920
5,6563		0,0496	0,6240		0,6240	0,6240	0,6240	0,0496
5,8365		0,0249	0,4649		0,4649	0,4649	0,4649	0,0249
6,0167		0,0117	0,3226		0,3226	0,3226	0,3226	0,0117
6,1968			0,2085		0,2085	0,2085	0,2085	
6,3770			0,1255		0,1255	0,1255	0,1255	
6,5571			0,0704		0,0704	0,0704	0,0704	
6,7373			0,0367		0,0367	0,0367	0,0367	

Таблица А3.6 – Таблица ненормированных плотностей вероятности для фактора 6

$X_{6j}$	$X_{6f}$							
	7	12	20	15	20	10	15	22
2,8213	0,4225	0,0156				0,0786		
3,6699	0,5785	0,0326				0,1384		
4,5184	0,7379	0,0632				0,2270		
5,3669	0,8767	0,1140				0,3467		
6,2155	0,9701	0,1918		0,0222		0,4932	0,0222	
7,0640	0,9998	0,3005		0,0447		0,6535	0,0447	
7,9125	0,9597	0,4385		0,0838		0,8065	0,0838	
8,7610	0,8581	0,5959		0,1465		0,9270	0,1465	
9,6096	0,7146	0,7543		0,2384		0,9925	0,2384	
10,4581	0,5543	0,8893	0,0112	0,3613	0,0112	0,9897	0,3613	
11,3066	0,4004	0,9766	0,0240	0,5101	0,0240	0,9192	0,5101	
12,1552	0,2694	0,9988	0,0480	0,6707	0,0480	0,7952	0,6707	
13,0037	0,1689	0,9515	0,0893	0,8215	0,0893	0,6407	0,8215	0,0184
13,8522	0,0986	0,8443	0,1549	0,9371	0,1549	0,4808	0,9371	0,0378
14,7007	0,0536	0,6977	0,2501	0,9956	0,2501	0,3361	0,9956	0,0721
15,5493	0,0271	0,5371	0,3762	0,9852	0,3762	0,2188	0,9852	0,1283
16,3978	0,0128	0,3850	0,5271	0,9081	0,5271	0,1327	0,9081	0,2125
17,2463		0,2571	0,6878	0,7796	0,6878	0,0749	0,7796	0,3279
18,0948		0,1599	0,8360	0,6233	0,8360	0,0394	0,6233	0,4712
18,9434		0,0926	0,9464	0,4642	0,9464	0,0193	0,4642	0,6306
19,7919		0,0500	0,9979	0,3220	0,9979		0,3220	0,7862
20,6404		0,0251	0,9800	0,2081	0,9800		0,2081	0,9128
21,4890		0,0118	0,8964	0,1252	0,8964		0,1252	0,9872
22,3375			0,7637	0,0702	0,7637		0,0702	0,9944
23,1860			0,6060	0,0366	0,6060		0,0366	0,9329
24,0345			0,4479	0,0178	0,4479		0,0178	0,8152
24,8831			0,3083		0,3083			0,6635
25,7316			0,1977		0,1977			0,5030
26,5801			0,1180		0,1180			0,3552
27,4287			0,0657		0,0657			0,2336

Таблица А3.7 – Таблица ненормированных плотностей вероятности для фактора 7

$X_{7j}$	$X_{7f}$							
	7	12	10	0	15	10	15	15
0,0012	0,0238			0,9948				
0,1521	0,0476			0,9792				
0,1905	0,0887			0,9981				
1,0153	0,1540		0,0147	0,9476		0,0147		
1,8400	0,2489		0,0309	0,8379		0,0309		
2,6648	0,3747		0,0602	0,6901		0,0602		
3,4895	0,5254	0,0228	0,1093	0,5294		0,1093		

4,3143	0,6861	0,0457	0,1848	0,3782		0,1848		
5,1391	0,8345	0,0855	0,2911	0,2517		0,2911		
5,9638	0,9455	0,1491	0,4270	0,1560	0,0141	0,4270	0,0141	0,0141
6,7886	0,9977	0,2421	0,5835	0,0901	0,0295	0,5835	0,0295	0,0295
7,6133	0,9805	0,3660	0,7427	0,0484	0,0578	0,7427	0,0578	0,0578
8,4381	0,8976	0,5155	0,8804	0,0243	0,1055	0,8804	0,1055	0,1055
9,2629	0,7653	0,6762	0,9720	0,0113	0,1792	0,9720	0,1792	0,1792
10,0876	0,6078	0,8261	0,9996		0,2835	0,9996	0,2835	0,2835
10,9124	0,4495	0,9401	0,9575		0,4178	0,9575	0,4178	0,4178
11,7371	0,3097	0,9964	0,8542		0,5735	0,8542	0,5735	0,5735
12,5619	0,1987	0,9836	0,7098		0,7331	0,7098	0,7331	0,7331
13,3867	0,1188	0,9044	0,5493		0,8729	0,5493	0,8729	0,8729
14,2114	0,0661	0,7746	0,3960		0,9680	0,3960	0,9680	0,9680
15,0362	0,0343	0,6179	0,2659		0,9999	0,2659	0,9999	0,9999
15,8609	0,0166	0,4590	0,1663		0,9620	0,1663	0,9620	0,9620
16,6857		0,3176	0,0968		0,8621	0,0968	0,8621	0,8621
17,5105		0,2047	0,0525		0,7195	0,0525	0,7195	0,7195
18,3352		0,1229	0,0265		0,5593	0,0265	0,5593	0,5593
19,1600		0,0687	0,0125		0,4050	0,0125	0,4050	0,4050
19,9847		0,0358			0,2731		0,2731	0,2731
20,8095		0,0174			0,1716		0,1716	0,1716
21,6343					0,1004		0,1004	0,1004
22,4590					0,0547		0,0547	0,0547

Таблица А3.8 – Таблица ненормированных плотностей вероятности для фактора 8

$X_{8j}$	$X_{8f}$							
	7	5	5	6	8	5	14	3
-1,1132								0,1212
-0,5795		0,0206	0,0206			0,0206		0,2022
-0,0458		0,0417	0,0417			0,0417		0,3143
0,4878		0,0789	0,0789	0,0226		0,0789		0,4551
1,0215	0,0116	0,1388	0,1388	0,0454		0,1388		0,6136
1,5552	0,0248	0,2275	0,2275	0,0850		0,2275		0,7707
2,0888	0,0493	0,3474	0,3474	0,1483	0,0128	0,3474		0,9016
2,6225	0,0916	0,4940	0,4940	0,2410	0,0271	0,4940		0,9824
3,1562	0,1583	0,6543	0,6543	0,3646	0,0536	0,6543		0,9970
3,6898	0,2549	0,8072	0,8072	0,5139	0,0985	0,8072		0,9424
4,2235	0,3822	0,9275	0,9275	0,6745	0,1688	0,9275		0,8296
4,7572	0,5339	0,9927	0,9927	0,8247	0,2693	0,9927		0,6803
5,2908	0,6946	0,9895	0,9895	0,9392	0,4003	0,9895		0,5196
5,8245	0,8417	0,9187	0,9187	0,9962	0,5541	0,9187		0,3696
6,3582	0,9499	0,7944	0,7944	0,9841	0,7144	0,7944		0,2449
6,8918	0,9985	0,6399	0,6399	0,9055	0,8580	0,6399		0,1511
7,4255	0,9777	0,4800	0,4800	0,7761	0,9597	0,4800		0,0869

7,9592	0,8916	0,3354	0,3354	0,6195	0,9998	0,3354		0,0465
8,4928	0,7573	0,2183	0,2183	0,4606	0,9702	0,2183	0,0227	0,0232
9,0265	0,5991	0,1323	0,1323	0,3190	0,8768	0,1323	0,0457	
9,5602	0,4414	0,0747	0,0747	0,2057	0,7381	0,0747	0,0855	
10,0938	0,3030	0,0393	0,0393	0,1236	0,5787	0,0393	0,1490	
10,6275	0,1937	0,0192	0,0192	0,0692	0,4226	0,0192	0,2420	
11,1612	0,1153			0,0360	0,2875		0,3659	
11,6948	0,0639			0,0175	0,1821		0,5153	
12,2285	0,0330				0,1075		0,6760	
12,7622	0,0159				0,0591		0,8260	
13,2958					0,0302		0,9400	
13,8295					0,0144		0,9964	
14,3632							0,9837	

Таблица А3.9 – Таблица ненормированных плотностей вероятности для фактора 9

$X_{9j}$	$X_{9f}$							
	9,8	8,9	9,5	8,9	9	12,6	9,7	12,5
6,5281		0,0380		0,0380	0,0286			
6,7753		0,0725	0,0133	0,0725	0,0563			
7,0225	0,0113	0,1288	0,0282	0,1288	0,1029		0,0155	
7,2697	0,0242	0,2132	0,0555	0,2132	0,1754		0,0323	
7,5169	0,0483	0,3288	0,1016	0,3288	0,2783		0,0626	
7,7641	0,0898	0,4723	0,1734	0,4723	0,4114		0,1132	
8,0113	0,1556	0,6318	0,2757	0,6318	0,5665		0,1905	
8,2585	0,2512	0,7872	0,4081	0,7872	0,7264		0,2987	
8,5057	0,3776	0,9136	0,5628	0,9136	0,8676		0,4363	
8,7529	0,5286	0,9875	0,7229	0,9875	0,9651		0,5936	
9,0001	0,6893	0,9942	0,8648	0,9942	1,0000		0,7522	
9,2473	0,8373	0,9323	0,9636	0,9323	0,9651		0,8877	
9,4945	0,9472	0,8142	1,0000	0,8142	0,8675		0,9757	
9,7417	0,9980	0,6624	0,9666	0,6624	0,7262		0,9990	0,0120
9,9889	0,9795	0,5019	0,8702	0,5019	0,5663	0,0190	0,9526	0,0256
10,2361	0,8953	0,3542	0,7298	0,3542	0,4113	0,0388	0,8461	0,0508
10,4833	0,7623	0,2328	0,5700	0,2328	0,2782	0,0739	0,6999	0,0940
10,7305	0,6045	0,1425	0,4146	0,1425	0,1753	0,1310	0,5393	0,1619
10,9777	0,4464	0,0813	0,2809	0,0813	0,1029	0,2165	0,3870	0,2599
11,2249	0,3071	0,0432	0,1773	0,0432	0,0562	0,3330	0,2587	0,3885
11,4721	0,1968	0,0214	0,1042	0,0214	0,0286	0,4772	0,1611	0,5410
11,7193	0,1174		0,0571		0,0136	0,6370	0,0934	0,7016
11,9665	0,0653		0,0291			0,7919	0,0504	0,8475
12,2137	0,0338		0,0138			0,9169	0,0254	0,9534
12,4609	0,0163					0,9888	0,0119	0,9991
12,7081						0,9932		0,9751
12,9553						0,9292		0,8865

13,2025						0,8097		0,7506
13,4497						0,6572		0,5919
13,6969						0,4968		0,4348

Таблица А3.10 – Таблица ненормированных плотностей вероятности для выходной величины Y

$Y_i$	$Y_f$							
	2	3	3	2	3	3	3	3
1,6760	0,5068			0,5068				
1,7501	0,6674			0,6674				
1,8242	0,8186			0,8186				
1,8982	0,9351			0,9351				
1,9723	0,9950			0,9950				
2,0464	0,9862			0,9862				
2,1204	0,9103			0,9103				
2,1945	0,7827	0,0150	0,0150	0,7827	0,0150	0,0150	0,0150	0,0150
2,2686	0,6268	0,0313	0,0313	0,6268	0,0313	0,0313	0,0313	0,0313
2,3426	0,4675	0,0609	0,0609	0,4675	0,0609	0,0609	0,0609	0,0609
2,4167	0,3248	0,1104	0,1104	0,3248	0,1104	0,1104	0,1104	0,1104
2,4908	0,2101	0,1865	0,1865	0,2101	0,1865	0,1865	0,1865	0,1865
2,5648	0,1266	0,2933	0,2933	0,1266	0,2933	0,2933	0,2933	0,2933
2,6389	0,0711	0,4298	0,4298	0,0711	0,4298	0,4298	0,4298	0,4298
2,7130	0,0372	0,5865	0,5865	0,0372	0,5865	0,5865	0,5865	0,5865
2,7870	0,0181	0,7455	0,7455	0,0181	0,7455	0,7455	0,7455	0,7455
2,8611		0,8825	0,8825		0,8825	0,8825	0,8825	0,8825
2,9352		0,9731	0,9731		0,9731	0,9731	0,9731	0,9731
3,0092		0,9994	0,9994		0,9994	0,9994	0,9994	0,9994
3,0833		0,9561	0,9561		0,9561	0,9561	0,9561	0,9561
3,1574		0,8518	0,8518		0,8518	0,8518	0,8518	0,8518
3,2314		0,7069	0,7069		0,7069	0,7069	0,7069	0,7069
3,3055		0,5464	0,5464		0,5464	0,5464	0,5464	0,5464
3,3796		0,3933	0,3933		0,3933	0,3933	0,3933	0,3933
3,4536		0,2637	0,2637		0,2637	0,2637	0,2637	0,2637
3,5277		0,1647	0,1647		0,1647	0,1647	0,1647	0,1647
3,6018		0,0958	0,0958		0,0958	0,0958	0,0958	0,0958
3,6758		0,0519	0,0519		0,0519	0,0519	0,0519	0,0519
3,7499		0,0262	0,0262		0,0262	0,0262	0,0262	0,0262
3,8240		0,0123	0,0123		0,0123	0,0123	0,0123	0,0123

Таблица А3.11 – Таблица виртуальной выборки

№	Факторы $X_i$									Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	$X_{5f}$	$X_{6f}$	$X_{7f}$	$X_{8f}$	$X_{9f}$	
1	31,9075	0,6672	98,6761	37,7491	2,9540	7,0640	6,7886	6,8918	9,7417	1,9723
2	34,3698	1,5136	100,4115	38,6938	3,1341	7,9125	7,6133	7,4255	9,9889	2,0464
3	36,8322	2,3599	102,1469	39,6384	3,3143	8,7610	8,4381	7,9592	10,2361	2,1204
4	39,2945	3,2063	103,8823	40,5830	3,4944	9,6096	9,2629	8,4928	10,4833	2,1945
5	41,7568	4,0527	105,6177	41,5277	3,6746	10,4581	10,0876	9,0265	10,7305	2,2686
6	44,2192	4,8991	107,3531	42,4723	3,8548	11,3066	10,9124	9,5602	10,9777	2,3426
7	46,6815	5,7454	109,0885	43,4170	4,0349	12,1552	11,7371	10,0938	11,2249	2,4167
8	49,1438	6,5918	110,8239	44,3616	4,2151	13,0037	12,5619	10,6275	11,4721	2,4908
9	51,6062	7,4382	112,5594	45,3062	4,3952	13,8522	13,3867	11,1612	11,7193	2,5648
10	54,0685	8,2845	114,2948	46,2509	4,5754	14,7007	14,2114	11,6948	11,9665	2,6389
11	56,5308	9,1309	116,0302	47,1955	4,7556	15,5493	15,0362	12,2285	12,2137	2,7130
12	44,2192	0,6672	89,9990	37,7491	4,0349	12,1552	11,7371	4,7572	9,0001	3,0092
13	46,6815	1,5136	91,7344	38,6938	4,2151	13,0037	12,5619	5,2908	9,2473	3,0833
14	49,1438	2,3599	93,4698	39,6384	4,3952	13,8522	13,3867	5,8245	9,4945	3,1574
15	51,6062	3,2063	95,2052	40,5830	4,5754	14,7007	14,2114	6,3582	9,7417	3,2314
16	54,0685	4,0527	96,9406	41,5277	4,7556	15,5493	15,0362	6,8918	9,9889	3,3055
17	56,5308	4,8991	98,6761	42,4723	4,9357	16,3978	15,8609	7,4255	10,2361	3,3796
18	58,9932	5,7454	100,4115	43,4170	5,1159	17,2463	16,6857	7,9592	10,4833	3,4536
19	61,4555	6,5918	102,1469	44,3616	5,2960	18,0948	17,5105	8,4928	10,7305	3,5277
20	63,9178	7,4382	103,8823	45,3062	5,4762	18,9434	18,3352	9,0265	10,9777	3,6018
21	66,3802	8,2845	105,6177	46,2509	5,6563	19,7919	19,1600	9,5602	11,2249	3,6758
22	68,8425	9,1309	107,3531	47,1955	5,8365	20,6404	19,9847	10,0938	11,4721	3,7499
23	44,2192	0,6672	110,8239	39,6384	4,9357	19,7919	10,0876	4,7572	9,4945	3,0092
24	46,6815	1,5136	112,5594	40,5830	5,1159	20,6404	10,9124	5,2908	9,7417	3,0833
25	49,1438	2,3599	114,2948	41,5277	5,2960	21,4890	11,7371	5,8245	9,9889	3,1574
26	51,6062	3,2063	116,0302	42,4723	5,4762	22,3375	12,5619	6,3582	10,2361	3,2314
27	54,0685	4,0527	117,7656	43,4170	5,6563	23,1860	13,3867	6,8918	10,4833	3,3055
28	56,5308	4,8991	119,5010	44,3616	5,8365	24,0345	14,2114	7,4255	10,7305	3,3796
29	58,9932	5,7454	121,2364	45,3062	6,0167	24,8831	15,0362	7,9592	10,9777	3,4536
30	61,4555	6,5918	122,9718	46,2509	6,1968	25,7316	15,8609	8,4928	11,2249	3,5277
31	63,9178	7,4382	124,7072	47,1955	6,3770	26,5801	16,6857	9,0265	11,4721	3,6018
32	66,3802	8,2845	126,4427	48,1402	6,5571	27,4287	17,5105	9,5602	11,7193	3,6758
33	22,0582	4,0527	105,6177	33,9706	1,5127	12,1552	10,0876	4,2235	8,2585	1,7501
34	24,5205	4,8991	107,3531	34,9152	1,6929	13,0037	10,9124	4,7572	8,5057	1,8242
35	26,9828	5,7454	109,0885	35,8598	1,8730	13,8522	11,7371	5,2908	8,7529	1,8982
36	29,4452	6,5918	110,8239	36,8045	2,0532	14,7007	12,5619	5,8245	9,0001	1,9723
37	31,9075	7,4382	112,5594	37,7491	2,2333	15,5493	13,3867	6,3582	9,2473	2,0464
38	34,3698	8,2845	114,2948	38,6938	2,4135	16,3978	14,2114	6,8918	9,4945	2,1204
39	36,8322	9,1309	116,0302	39,6384	2,5937	17,2463	15,0362	7,4255	9,7417	2,1945

№	Факторы $X_i$									Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	$X_{5f}$	$X_{6f}$	$X_{7f}$	$X_{8f}$	$X_{9f}$	
40	39,2945	9,9773	117,7656	40,5830	2,7738	18,0948	15,8609	7,9592	9,9889	2,2686
41	41,7568	10,8236	119,5010	41,5277	2,9540	18,9434	16,6857	8,4928	10,2361	2,3426
42	44,2192	11,6700	121,2364	42,4723	3,1341	19,7919	17,5105	9,0265	10,4833	2,4167
43	46,6815	12,5164	122,9718	43,4170	3,3143	20,6404	18,3352	9,5602	10,7305	2,4908
44	49,1438	13,3628	124,7072	44,3616	3,4944	21,4890	19,1600	10,0938	10,9777	2,5648
45	51,6062	14,2091	126,4427	45,3062	3,6746	22,3375	19,9847	10,6275	11,2249	2,6389
46	54,0685	15,0555	128,1781	46,2509	3,8548	23,1860	20,8095	11,1612	11,4721	2,7130
47	39,2945	6,5918	102,1469	35,8598	3,1341	11,3066	6,7886	2,6225	6,5281	2,2686
48	41,7568	7,4382	103,8823	36,8045	3,3143	12,1552	7,6133	3,1562	6,7753	2,3426
49	44,2192	8,2845	105,6177	37,7491	3,4944	13,0037	8,4381	3,6898	7,0225	2,4167
50	46,6815	9,1309	107,3531	38,6938	3,6746	13,8522	9,2629	4,2235	7,2697	2,4908
51	49,1438	9,9773	109,0885	39,6384	3,8548	14,7007	10,0876	4,7572	7,5169	2,5648
52	51,6062	10,8236	110,8239	40,5830	4,0349	15,5493	10,9124	5,2908	7,7641	2,6389
53	54,0685	11,6700	112,5594	41,5277	4,2151	16,3978	11,7371	5,8245	8,0113	2,7130
54	56,5308	12,5164	114,2948	42,4723	4,3952	17,2463	12,5619	6,3582	8,2585	2,7870
55	58,9932	13,3628	116,0302	43,4170	4,5754	18,0948	13,3867	6,8918	8,5057	2,8611
56	61,4555	14,2091	117,7656	44,3616	4,7556	18,9434	14,2114	7,4255	8,7529	2,9352
57	63,9178	15,0555	119,5010	45,3062	4,9357	19,7919	15,0362	7,9592	9,0001	3,0092
58	66,3802	15,9019	121,2364	46,2509	5,1159	20,6404	15,8609	8,4928	9,2473	3,0833
59	68,8425	16,7482	122,9718	47,1955	5,2960	21,4890	16,6857	9,0265	9,4945	3,1574
60	71,3048	17,5946	124,7072	48,1402	5,4762	22,3375	17,5105	9,5602	9,7417	3,2314
61	73,7672	18,4410	126,4427	49,0848	5,6563	23,1860	18,3352	10,0938	9,9889	3,3055
62	76,2295	19,2874	128,1781	50,0294	5,8365	24,0345	19,1600	10,6275	10,2361	3,3796
63	78,6918	20,1337	129,9135	50,9741	6,0167	24,8831	19,9847	11,1612	10,4833	3,4536
64	46,6815	3,2063	96,9406	47,1955	3,4944	2,8213	3,4895	0,4878	10,7305	2,4167
65	49,1438	4,0527	98,6761	48,1402	3,6746	3,6699	4,3143	1,0215	10,9777	2,4908
66	51,6062	4,8991	100,4115	49,0848	3,8548	4,5184	5,1391	1,5552	11,2249	2,5648
67	54,0685	5,7454	102,1469	50,0294	4,0349	5,3669	5,9638	2,0888	11,4721	2,6389
68	56,5308	6,5918	103,8823	50,9741	4,2151	6,2155	6,7886	2,6225	11,7193	2,7130
69	58,9932	7,4382	105,6177	51,9187	4,3952	7,0640	7,6133	3,1562	11,9665	2,7870
70	61,4555	8,2845	107,3531	52,8633	4,5754	7,9125	8,4381	3,6898	12,2137	2,8611
71	63,9178	9,1309	109,0885	53,8080	4,7556	8,7610	9,2629	4,2235	12,4609	2,9352
72	66,3802	9,9773	110,8239	54,7526	4,9357	9,6096	10,0876	4,7572	12,7081	3,0092
73	68,8425	10,8236	112,5594	55,6973	5,1159	10,4581	10,9124	5,2908	12,9553	3,0833
74	51,6062	0,6672	98,6761	36,8045	3,6746	8,7610	9,2629	10,0938	8,0113	2,4908
75	54,0685	1,5136	100,4115	37,7491	3,8548	9,6096	10,0876	10,6275	8,2585	2,5648
76	56,5308	2,3599	102,1469	38,6938	4,0349	10,4581	10,9124	11,1612	8,5057	2,6389
77	58,9932	3,2063	103,8823	39,6384	4,2151	11,3066	11,7371	11,6948	8,7529	2,7130
78	61,4555	4,0527	105,6177	40,5830	4,3952	12,1552	12,5619	12,2285	9,0001	2,7870
79	63,9178	4,8991	107,3531	41,5277	4,5754	13,0037	13,3867	12,7622	9,2473	2,8611

№	Факторы $X_i$									Y
	$X_{1f}$	$X_{2f}$	$X_{3f}$	$X_{4f}$	$X_{5f}$	$X_{6f}$	$X_{7f}$	$X_{8f}$	$X_{9f}$	
80	66,3802	5,7454	109,0885	42,4723	4,7556	13,8522	14,2114	13,2958	9,4945	2,9352
81	68,8425	6,5918	110,8239	43,4170	4,9357	14,7007	15,0362	13,8295	9,7417	3,0092
82	71,3048	7,4382	112,5594	44,3616	5,1159	15,5493	15,8609	14,3632	9,9889	3,0833
83	51,6062	0,6672	89,9990	39,6384	4,0349	22,3375	15,0362	3,1562	12,4609	3,0092
84	54,0685	1,5136	91,7344	40,5830	4,2151	23,1860	15,8609	3,6898	12,7081	3,0833
85	56,5308	2,3599	93,4698	41,5277	4,3952	24,0345	16,6857	4,2235	12,9553	3,1574
86	58,9932	3,2063	95,2052	42,4723	4,5754	24,8831	17,5105	4,7572	13,2025	3,2314
87	61,4555	4,0527	96,9406	43,4170	4,7556	25,7316	18,3352	5,2908	13,4497	3,3055
88	63,9178	4,8991	98,6761	44,3616	4,9357	26,5801	19,1600	5,8245	13,6969	3,3796

**Приложение А4. Балльная оценка органолептических показателей качества хлебобулочных изделий**

Таблица А4.1 – Балльная оценка органолептических показателей качества хлебобулочных изделий

Показатели	Баллы	Хлеб формовой и подовый из пшеничной муки	Булочные изделия	Сдобные изделия
		<i>Соответствующие каждому баллу шкалы словесные характеристики отдельных показателей качества изделий</i>		
<b>Форма</b>	5	Симметричная с заметно выпуклой верхней коркой для формового, четкая симметричная форма для подового	Правильная, симметричная, с хорошо закатанными концами у городских булок и батонов	Правильная с четко выраженным рисунком
	4	Достаточно симметричная относительно верхней корки	Достаточно симметричная относительно верхней корки и по периметру или по обоим краям	Достаточно точно выраженный рисунок
	3	Слишком округлая или недостаточно выраженная выпуклость верхней корки, немного несимметричная относительно верхней корки	Слишком округлая или немного расплывчатая, немного несимметричная относительно верхней корки и по периметру или по обоим краям	Немного расплывчатый рисунок
	2	Плоская верхняя корка у формового, несимметричная или заметно расплывчатая у подового	Несимметричная или заметно расплывчатая	Расплывчатая, недостаточно выраженный рисунок
	1	Вогнутая верхняя корка у формового, сильно расплывчатая у подового, искаженная, деформированная	Сильно расплывчатая, искаженная или деформированная	Расплывчатая, невыраженный рисунок, деформированная
<b>Цвет корки</b>	5	От золотистого до светло-коричневого, равномерный		
	4	Светло-золотистый или коричневый, достаточно равномерный		
	3	Желтый или интенсивно темно-коричневый, недостаточно равномерный		
	2	Светло-желтый или от желтого до коричневого, но сильно неравномерный		

Показатели	Баллы	Хлеб формовой и подовый из пшеничной муки	Булочные изделия	Сдобные изделия
		<i>Соответствующие каждому баллу шкалы словесные характеристики отдельных показателей качества изделий</i>		
	1	Бледный или горелый		
Поверхность	5	Гладкая (за исключением пшеничного обойного), без трещин и подрывов, глянцевая	Гладкая, четко выраженные надрезы или гребешок, глянцевая	Гладкая, глянцевая, отличная отделка поверхности
	4	Достаточно гладкая, едва заметные трещины и подрывы, глянцевая	Достаточно гладкая, достаточно выраженные надрезы или гребешок, глянцевая	Достаточно гладкая, глянцевая или хорошая отделка поверхности
	3	Слегка пузырчатая, шероховатая; заметные, но не крупные трещины и подрывы, глянец слабый	Едва заметные трещины, недостаточно выраженные надрезы или гребешок, глянец слабый	Мелкие пузыри, трещины, слабый глянец или удовлетворительная отделка поверхности
	2	Пузырчатая, бугристая, крупные трещины и подрывы, матовая поверхность	Небольшие трещины и подрывы, невыраженные надрезы или гребешок, матовая поверхность	Небольшие трещины и подрывы, матовая или недостаточно удовлетворительная отделка поверхности
	1	Разорванная корка, боковые при-тиски	Крупные трещины и подрывы, пузыри, боковые притиски	Крупные трещины и пузыри, боковые притиски, плохая отделка поверхности
Состояние мякиша	5	Очень мягкий, нежный, эластичный		
	4	Мягкий, эластичный		
	3	Удовлетворительно мягкий (немного уплотненный), эластичный		
	2	Заметно уплотненный, крошащийся, заметно заминающийся		
	1	Сильно заминающийся, влажный на ощупь, липкий		
Пористость	5	Хорошо развитая и равномерная, поры мелкие и тонкостенные		
	4	Достаточно развитая и достаточно равномерная, поры мелкие и средние, тонкостенные		
	3	Удовлетворительно развитая пористость, поры различной величины, средней толщины, распределены неравномерно на срезе мякиша		
	2	Поры мелкие, недоразвитые или крупные толстостенные, незначительное количество плотных (беспористых) участков мякиша, разрыв мякиша, заметное отслоение мякиша от корки		
	1	Оторванный от верхней корки и осевший мякиш, закал, плотный (неразрыхленный) мякиш, следы непромеса		

Показатели	Баллы	Хлеб формовой и подовый из пшеничной муки	Булочные изделия	Сдобные изделия
		<i>Соответствующие каждому баллу шкалы словесные характеристики отдельных показателей качества изделий</i>		
Аромат (запах)	5	Интенсивно выраженный, характерный для данного вида изделий		
	4	Выраженный, характерный для данного вида изделий		
	3	Слабовыраженный, характерный для данного вида изделий		
	2	Невыраженный, слегка посторонний		
	1	Сильно кислый, посторонний, неприятный		
Вкус	5	Интенсивно выраженный, характерный для данного вида изделий		
	4	Выраженный, характерный для данного вида изделий		
	3	Слабовыраженный, характерный для данного вида изделий		
	2	Пресноватый, слегка кислый, слегка тестовый		
	1	Совершенно пресный, резко кислый, пересоленный, посторонний, неприятный, хруст на зубах		

## **ДЕКЛАРАЦИЯ ОБ ОТВЕТСТВЕННОСТИ**

Нижеподписавшийся, заявляю под личную ответственность, что материалы, представленные в докторской диссертации, являются результатом личных научных исследований и разработок. Осознаю, что в противном случае, буду нести ответственность в соответствии с действующим законодательством.

Попукайло Владимир

Подпись

Число

## РЕЗЮМЕ АВТОРА

**Персональные данные:** Попукайло Владимир, 09.12.1988, Бендеры, Молдова.

**Образование:**

**2004-2008 Техник.** Технический колледж им Ю.А. Гагарина.  
Специальность: «Вычислительные машины, комплексы, системы и сети»

**2008-2012 Инженер.** Университет им. Т.Г. Шевченко (г. Тирасполь), Инженерно-технический институт. Специальность: «Вычислительные машины, комплексы, системы и сети».

**2012-2015 Аспирант** при Университете им. Т.Г. Шевченко (г. Тирасполь).  
Специальность: «Математическое моделирование, численные методы и комплексы программ»

**2016-2017** Профессиональная переподготовка по специальности «Анализ данных», Санкт-Петербургский Академический Университет Российской Академии Наук.

**Профессиональная деятельность:**

**2008-2010** Системный администратор ООО «ВлаСаНа», г. Бендеры

**2010-2012.** Технический директор ООО «ВлаСаНа», г. Бендеры

**2012** – настоящее время. Преподаватель кафедры ИТиАУПП, Инженерно-технический Институт, Университет им. Т.Г. Шевченко (г. Тирасполь)

**2015** – настоящее время. Ассоциированный сотрудник Института Математики и Информатики Академии Наук Молдовы.

**2016** – настоящее время. Младший научный сотрудник НИЛ «Математическое моделирование», Инженерно-технический Институт, Университет им. Т.Г. Шевченко (г. Тирасполь)

**Область исследования:** математическое моделирование, математическая статистика, теория принятия решений.

**Публикации:** Опубликовано 16 научных работ, в том числе 5 в рецензируемых научных журналах из списка ВАК Молдовы, Украины и Российской Федерации, 2 учебных пособия. Участвовал в работе 6 международных научных конференций.

**Знание языков:** Русский (родной), Английский (хорошо).

**Контактные данные:**

Тел. +(373) 77844001

e-mail: vsp.science@gmail.com

